Collected Works

William Stein

March 2012

Contents

1 Empirical Evidence for the Birch and Swinnerton-Dyer Conjecture for Modular Jacobians of Genus 2 Curves, with E.V. Flynn, F. Leprévost, E.F. Schafer, M. Stoll, and J.L. Wetherell

MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

EMPIRICAL EVIDENCE FOR THE BIRCH AND SWINNERTON-DYER CONJECTURES FOR MODULAR JACOBIANS OF GENUS 2 CURVES

E. VICTOR FLYNN, FRANCK LEPRÉVOST, EDWARD F. SCHAEFER, WILLIAM A. STEIN, MICHAEL STOLL, AND JOSEPH L. WETHERELL

Abstract. This paper provides empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves. The second of these conjectures relates six quantities associated to a Jacobian over the rational numbers. One of these six quantities is the size of the Shafarevich-Tate group. Unable to compute that, we computed the five other quantities and solved for the last one. In all 32 cases, the result is very close to an integer that is a power of 2. In addition, this power of 2 agrees with the size of the 2-torsion of the Shafarevich-Tate group, which we could compute.

1. Introduction

The conjectures of Birch and Swinnerton-Dyer, originally stated for elliptic curves over $\mathbf{Q}$, have been a constant source of motivation for the study of elliptic curves, with the ultimate goal being to find a proof. This has resulted not only in a better theoretical understanding, but also in the development of better algorithms for computing the analytic and arithmetic invariants that are so intriguingly related by them. We now know that the first and, up to a non-zero rational factor, the second conjecture hold for modular elliptic curves over $\mathbf{Q}$ [1] in the analytic rank 0 and 1 cases (see [GZ, Ko, Wal1, Wal2]). Furthermore, a number of people have provided numerical evidence for the conjectures for a large number of elliptic curves; see for example [BGZ, BSD, Ca, Cr2].

Received by the editor June 5, 2000.

1991 *Mathematics Subject Classification.* Primary 11G40; Secondary 11G10, 11G30, 14H25, 14H40, 14H45.

Key words and phrases. Birch and Swinnerton-Dyer conjecture, genus 2, Jacobian, modular abelian variety.

The first author thanks the Nuffield Foundation (Grant SCI/180/96/71/G) for financial support. The second author did some of the research at the Max-Planck Institut für Mathematik and the Technische Universität Berlin. The third author thanks the National Security Agency (Grant MDA904-99-1-0013). The fourth author was supported by a Sarah M. Hallam fellowship. The fifth author did some of the research at the Max-Planck-Institut für Mathematik. The sixth author thanks the National Science Foundation (Grant DMS-9705959). The authors had useful conversations with John Cremona, Qing Liu, Karl Rubin and Peter Swinnerton-Dyer and are grateful to Xiangdong Wang and Michael Müller for making data available to them.

[1] It has recently been announced by Breuil, Conrad, Diamond and Taylor that they have extended Wiles' results and shown that all elliptic curves over $\mathbf{Q}$ are modular (see [BCDT]).

©1997 American Mathematical Society

By now, our theoretical and algorithmic knowledge of curves of genus 2 and their Jacobians has reached a state that makes it possible to conduct similar investigations. The Birch and Swinnerton-Dyer conjectures have been generalized to arbitrary abelian varieties over number fields by Tate [Ta]. If J is the Jacobian of a genus 2 curve over $\mathbf{Q}$, then the first conjecture states that the order of vanishing of the L-series of the Jacobian at $s = 1$ (the *analytic rank*) is equal to the Mordell-Weil rank of the Jacobian. The second conjecture is that

$$\lim_{s\to 1}(s-1)^{-r}L(J,s) = \Omega \cdot \mathrm{Reg} \cdot \prod_p c_p \cdot \#\text{Ш}(J,\mathbf{Q}) \cdot (\#J(\mathbf{Q})_{\mathrm{tors}})^{-2}\,. \tag{1.1}$$

In this equation, $L(J,s)$ is the L-series of the Jacobian J, and r is its analytic rank. We use Ω to denote the integral over $J(\mathbf{R})$ of a particular differential 2-form; the precise choice of this differential is described in Section 3.5. Reg is the regulator of $J(\mathbf{Q})$. For primes p, we use c_p to denote the size of $J(\mathbf{Q}_p)/J^0(\mathbf{Q}_p)$, where $J^0(\mathbf{Q}_p)$ is defined in Section 3.4. We let $\text{Ш}(J,\mathbf{Q})$ be the Shafarevich-Tate group of J over $\mathbf{Q}$, and we let $J(\mathbf{Q})_{\mathrm{tors}}$ denote the torsion subgroup of $J(\mathbf{Q})$.

As in the case of elliptic curves, the first conjecture assumes that the L-series can be analytically continued to $s = 1$, and the second conjecture additionally assumes that the Shafarevich-Tate group is finite. Neither of these assumptions is known to hold for arbitrary genus 2 curves. The analytic continuation of the L-series, however, is known to exist for modular abelian varieties over $\mathbf{Q}$, where an abelian variety is called *modular* if it is a quotient of the Jacobian $J_0(N)$ of the modular curve $X_0(N)$ for some level N. For simplicity, we will also call a genus 2 curve *modular* when its Jacobian is modular in this sense. So it is certainly a good idea to look at modular genus 2 curves over $\mathbf{Q}$, since we then at least know that the statement of the first conjecture makes sense. Moreover, for many modular abelian varieties it is also known that the Shafarevich-Tate group is finite, therefore the statement of the second conjecture also makes sense. As it turns out, all of our examples belong to this class. An additional benefit of choosing modular genus 2 curves is that one can find lists of such curves in the literature.

In this article, we provide empirical evidence for the Birch and Swinnerton-Dyer conjectures for such modular genus 2 curves. Since there is no known effective way of computing the size of the Shafarevich-Tate group, we computed the other five terms in equation (1.1) (in two different ways, if possible). This required several different algorithms, some of which were developed or improved while we were working on this paper. If one of these algorithms is already well described in the literature, then we simply cite it. Otherwise, we describe it here in some detail (in particular, algorithms for computing Ω and c_p).

For modular abelian varieties associated to newforms whose L-series have analytic rank 0 or 1, the first Birch and Swinnerton-Dyer conjecture has been proven. In such cases, the Shafarevich-Tate group is also known to be finite and the second conjecture has been proven, up to a non-zero rational factor. This all follows from results in [GZ, KL, Wal1, Wal2]. In our examples, all of the analytic ranks are either 0 or 1. Thus we already know that the first conjecture holds. Since the Jacobians we consider are associated to a quadratic conjugate pair of newforms, the analytic rank of the Jacobian is twice the analytic rank of either newform (see [GZ]).

The second Birch and Swinnerton-Dyer conjecture has not been proven for the cases we consider. In order to verify equation (1.1), we computed the five terms other than $\#Ш(J, \mathbf{Q})$ and solved for $\#Ш(J, \mathbf{Q})$. In each case, the value is an integer to within the accuracy of our calculations. This number is a power of 2, which coincides with the independently computed size of the 2-torsion subgroup of $Ш(J, \mathbf{Q})$. Hence, we have verified the second Birch and Swinnerton-Dyer conjecture for our curves at least numerically, if we assume that the Shafarevich-Tate group consists of 2-torsion only. (This is an ad hoc assumption based only on the fact that we do not know better.) See Section 6 for circumstances under which the verification is exact.

The curves are listed in Table 1, and the numerical results can be found in Table 2.

2. The Curves

Each of the genus 2 curves we consider is related to the Jacobian $J_0(N)$ of the modular curve $X_0(N)$ for some level N. When only one of these genus 2 curves arises from a given level N, then we denote this curve by C_N; when there are two curves coming from level N we use the notation $C_{N,A}$, $C_{N,B}$. The relationship of, say, C_N to $J_0(N)$ depends on the source. Briefly, from Hasegawa [Hs] we obtain quotients of $X_0(N)$ and from Wang [Wan] we obtain curves whose Jacobians are quotients of $J_0(N)$. In both cases the Jacobian J_N of C_N is isogenous to a 2-dimensional factor of $J_0(N)$. (When not referring to a specific curve, we will typically drop the subscript N from J.) In this way we can also associate C_N with a 2-dimensional subspace of $S_2(N)$, the space of cusp forms of weight 2 for $\Gamma_0(N)$.

We now discuss the precise source of the genus 2 curves we will consider. Hasegawa [Hs] has provided exact equations for all genus 2 curves which are quotients of $X_0(N)$ by a subgroup of the Atkin-Lehner involutions. There are 142 such curves. We are particularly interested in those where the Jacobian corresponds to a subspace of $S_2(N)$ spanned by a quadratic conjugate pair of newforms. There are 21 of these with level $N \leq 200$. For these curves we will provide evidence for the second conjecture. There are seven more such curves with $N > 200$. We can classify the other 2-dimensional subspaces into four types. There are 2-dimensional subspaces of oldforms that are irreducible under the action of the Hecke algebra. There are also 2-dimensional subspaces that are reducible under the action of the Hecke algebra and are spanned by two oldforms, two newforms or one of each. The Jacobians corresponding to the latter three kinds are always isogenous, over $\mathbf{Q}$, to the product of two elliptic curves. Given the small levels, these are elliptic curves for which Cremona [Cr2] has already provided evidence for the Birch and Swinnerton-Dyer conjectures. In Table 5, we describe the kind of cusp forms spanning the 2-dimensional subspace and the signs of their functional equations from the level at which they are newforms. The analytic and Mordell-Weil ranks were always the smallest possible given those signs.

The second set of curves was created by Wang [Wan] and is further discussed in [FM]. This set consists of 28 curves that were constructed by considering the spaces $S_2(N)$ with $N \leq 200$. Whenever a subspace spanned by a pair of quadratic conjugate newforms was found, these newforms were integrated to produce a quotient abelian variety A of $J_0(N)$. These quotients are *optimal* in the sense of [Ma], in that the kernel of the quotient map is connected.

The period matrix for A was created using certain intersection numbers. When all of the intersection numbers have the same value, then the polarization on A

induced from the canonical polarization of $J_0(N)$ is equivalent to a principal polarization. (Two polarizations are *equivalent* if they differ by an integer multiple.) Conversely, every 2-dimensional optimal quotient of $J_0(N)$ in which the induced polarization is equivalent to a principal polarization is found in this way.

Using theta functions, numerical approximations were found for the Igusa invariants of the abelian surfaces. These numbers coincide with rational numbers of fairly small height within the limits of the precision used for the computations. Wang then constructed curves defined over $\mathbf{Q}$ whose Igusa invariants are the rational numbers found. (There is one abelian surface at level $N = 177$ for which Wang was not able to find a curve.) If we assume that these rational numbers are the true Igusa invariants of the abelian surfaces, then it follows that Wang's curves have Jacobians isomorphic, over $\overline{\mathbf{Q}}$, to the principally polarized abelian surfaces in his list. Since the classification given by these invariants is only up to isomorphism over $\overline{\mathbf{Q}}$, the Jacobians of Wang's curves are not necessarily isomorphic to, but can be twists of, the optimal quotients of $J_0(N)$ over $\mathbf{Q}$ (see below).

There are four curves in Hasegawa's list which do not show up in Wang's list (they are listed in Table 1 with an H in the last column). Their Jacobians are quotients of $J_0(N)$, but are not optimal quotients. It is likely that there are modular genus 2 curves which neither are Atkin-Lehner quotients of $X_0(N)$ (in Hasegawa's sense) nor have Jacobians that are optimal quotients. These curves could be found by looking at the optimal quotient abelian surfaces and checking whether they are isogenous to a principally polarized abelian surface over $\mathbf{Q}$.

For 17 of the curves in Wang's list, the 2-dimensional subspace spanned by the newforms is the same as that giving one of Hasegawa's curves. In all of those cases, the curve given by Wang's equation is isomorphic, over $\mathbf{Q}$, to that given by Hasegawa. This verifies Wang's equations for these 17 curves. They are listed in Table 1 with HW in the last column.

The remaining eleven curves (listed in Table 1 with a W in the last column) derive from the other eleven optimal quotients in Wang's list. These are described in more detail in Section 2.1 below.

With the exception of curves C_{63}, $C_{117,A}$ and C_{189}, the Jacobians of all of our curves are absolutely simple, and the canonically polarized Jacobians have automorphism groups of size two. We showed that these Jacobians are absolutely simple using an argument like those in [Le, Sto1]. The automorphism group of the canonically polarized Jacobian of a hyperelliptic curve is isomorphic to the automorphism group of the curve (see [Mi2, Thm. 12.1]). Each automorphism of a hyperelliptic curve induces a linear fractional transformation on x-coordinates (see [CF, p. 1]). Each automorphism also permutes the six Weierstrass points. Once we believed we had found all of the automorphisms, we were able to show that there are no more by considering all linear fractional transformations sending three fixed Weierstrass points to any three Weierstrass points. In each case, we worked with sufficient accuracy to show that other linear fractional transformations did not permute the Weierstrass points.

Let ζ_3 denote a primitive third root of unity. The Jacobians of curves C_{63}, $C_{117,A}$ and C_{189} are each isogenous to the product of two elliptic curves over $\mathbf{Q}(\zeta_3)$, though not over $\mathbf{Q}$, where they are simple. These genus 2 curves have automorphism groups of size 12. In the following table we list the curve at the left. On the right we give one of the elliptic curves which is a factor of its Jacobian. The second factor is the conjugate.

$$\begin{array}{ll} C_{63}: & y^2 = x(x^2 + (9 - 12\zeta_3)x - 48\zeta_3) \\ C_{117,A}: & y^2 = x(x^2 - (12 + 27\zeta_3)x - (48 + 48\zeta_3)) \\ C_{189}: & y^2 = x^3 + (66 - 3\zeta_3)x^2 + (342 + 81\zeta_3)x + 105 + 21\zeta_3 \end{array}$$

Note that these three Jacobians are examples of abelian varieties 'with extra twist' as discussed in [Cr1], where they can be found in the tables on page 409.

2.1. **Models for the Wang-only curves.** As we have already noted, a modular genus 2 curve may be found by either, both, or neither of Wang's and Hasegawa's techniques. Hasegawa's method allows for the exact determination, over $\mathbf{Q}$, of the equation of any modular genus 2 curve it has found. On the other hand, if Wang's technique detects a modular genus 2 curve C_N, his method produces real approximations to a curve C'_N which is defined over $\mathbf{Q}$ and is isomorphic to C_N over $\overline{\mathbf{Q}}$. We will call C'_N a *twisted modular genus 2 curve.*

In this section we attempt to determine equations for the eleven modular genus 2 curves detected by Wang but not by Hasegawa. If we assume that Wang's equations for the twisted modular genus 2 curves are correct, we find that we are able to determine the twists. In turn, this gives us strong evidence that Wang's equations for the twisted curves were correct. Undoing the twist, we determine probable equations for the modular genus 2 curves. We end by providing further evidence for the correctness of these equations.

In what follows, we will use the notation of [Cr2] and recommend it as a reference on the general results that we assume here and in Section 4 and the appendix. Fix a level N and let $f(z) \in S_2(N)$. Then f has a Fourier expansion

$$f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z} .$$

For a newform f, we have $a_1 \neq 0$; so we can normalize it by setting $a_1 = 1$. In our cases, the a_n's are integers in a real quadratic field. For each prime p not dividing N, the corresponding Euler factor of the L-series $L(f,s)$ is $1 - a_p p^{-s} + p^{1-2s}$. Let $N(a_p)$ and $Tr(a_p)$ denote the norm and trace of a_p. The product of this Euler factor and its conjugate is $1 - Tr(a_p)\,p^{-s} + (N(a_p) + 2p)\,p^{-2s} - p\,Tr(a_p)\,p^{-3s} + p^2\,p^{-4s}$. Therefore, the characteristic polynomial of the p-Frobenius on the corresponding abelian variety over $\mathbf{F}_p$ is $x^4 - Tr(a_p)\,x^3 + (N(a_p) + 2p)\,x^2 - p\,Tr(a_p)\,x + p^2$. Let C be a curve, over $\mathbf{Q}$, whose Jacobian, over $\mathbf{Q}$, comes from the space spanned by f and its conjugate. Then we know that $p + 1 - \#C(\mathbf{F}_p) = Tr(a_p)$ and $\frac{1}{2}(\#C(\mathbf{F}_p)^2 + \#C(\mathbf{F}_{p^2})) - (p+1)\#C(\mathbf{F}_p) - p = N(a_p)$ (see [MS, Lemma 3]). For the odd primes less than 200, not dividing N, we computed $\#C(\mathbf{F}_p)$ and $\#C(\mathbf{F}_{p^2})$ for each curve given by one of Wang's equations. From these we could compute the characteristic polynomials of Frobenius and see if they agreed with those predicted by the a_p's of the newforms.

Of the eleven curves, the characteristic polynomials agreed for only four. In each of the remaining seven cases we found a twist of Wang's curve whose characteristic polynomials agreed with those predicted by the newform for all odd primes less than 200 not dividing N. Four of these twists were quadratic and three were of higher degree. It is these twists that appear in Table 1.

We can provide further evidence that these equations are correct. For each curve given in Table 1, it is easy to determine the primes of singular reduction. In

Section 3.4 we will provide techniques for determining which of those primes divides the conductor of its Jacobian. In each case, the primes dividing the conductor of the Jacobian of the curve are exactly the primes dividing the level N; this is necessary. With the exception of curve C_{188}, all the curves come from odd levels. We used Liu's `genus2reduction` program (`ftp://megrez.math.u-bordeaux.fr/pub/liu`) to compute the conductor of the curve. In each case (other than curve C_{188}), the conductor is the square of the level; this is also necessary. For curve C_{188}, the odd part of the conductor of the curve is the square of the odd part of the level.

In addition, since the Jacobians of the Wang curves are optimal quotients, we can compute $k \cdot \Omega$ (where k is the Manin constant, conjectured to be 1) using the newforms. In each case, these agree (to within the accuracy of our computations) with the Ω's computed using the equations for the curves. We can also compute the value of c_p for optimal quotients from the newforms, when p exactly divides N and the eigenvalue of the pth Atkin-Lehner involution is -1. When p exactly divides N and the eigenvalue of the pth Atkin-Lehner involution is $+1$, the component group is either 0, $\mathbf{Z}/2\mathbf{Z}$, or $(\mathbf{Z}/2\mathbf{Z})^2$. These results are always in agreement with the values computed using the equations for the curves. The algorithms based on the newforms are described in Section 4, those based on the equations of the curves are described in Section 3.

Lastly, we were able to compute the Mordell-Weil ranks of the Jacobians of the curves given by ten of these eleven equations. In each case it agrees with the analytic rank of the Jacobian, as deduced from the newforms.

It should be noted that curve $C_{125,B}$ is the $\sqrt{5}$-twist of curve $C_{125,A}$; the corresponding statement holds for the associated 2-dimensional subspaces of $S_2(125)$. Since curve $C_{125,A}$ is a Hasegawa curve, this proves that the equation given in Table 1 for curve $C_{125,B}$ is correct.

The a_p's and other information concerning Wang's curves are currently kept in a database at the Institut für experimentelle Mathematik in Essen, Germany. Most recently, this database was under the care of Michael Müller. William Stein also keeps a database of a_p's for newforms.

Remark 2.1. For the remainder of this paper we will assume that the equations for the curves given in Table 1 are correct; that is, that they are equations for the curves whose Jacobians are isogenous to a factor of $J_0(N)$ in the way described above. Some of the quantities can be computed either from the newform or from the equation for the curve. We performed both computations whenever possible, and view this duplicate effort as an attempt to verify our implementation of the algorithms rather than an attempt to verify the equations in Table 1. For most quantities, one method or the other is not guaranteed to produce a value; in this case, we simply quote the value from whichever method did succeed. The reader who is disturbed by this philosophy should ignore the Wang-only curves, since the equations for the Hasegawa curves can be proven to be correct.

3. Algorithms for genus 2 curves

In this section, we describe the algorithms that are based on the given models for the curves. We give algorithms that compute all terms on the right hand side of equation (1.1), with the exception of the size of the Shafarevich-Tate group. We describe, however, how to find the size of its 2-torsion subgroup. Note that these algorithms are for general genus 2 curves and do not depend on modularity.

3.1. **Torsion Subgroup.** The computation of the torsion subgroup of $J(\mathbf{Q})$ is straightforward. We used the technique described in [CF, pp. 78–82]. This technique is not always effective, however. For an algorithm working in all cases see [Sto3].

3.2. **Mordell-Weil rank and $Ш(J,\mathbf{Q})[2]$.** The group $J(\mathbf{Q})$ is a finitely generated abelian group and so is isomorphic to $\mathbf{Z}^r \oplus J(\mathbf{Q})_{\text{tors}}$ for some r called the Mordell-Weil rank. As noted above (see Section 1), we justifiably use r to denote both the analytic and Mordell-Weil ranks since they agree for all curves in Table 1.

We used the algorithm described in [FPS] to compute $\mathrm{Sel}^2_{\text{fake}}(J,\mathbf{Q})$ (notation from [PSc]), which is a quotient of the 2-Selmer group $\mathrm{Sel}^2(J,\mathbf{Q})$. More details on this algorithm can be found in [Sto2]. Theorem 13.2 of [PSc] explains how to get $\mathrm{Sel}^2(J,\mathbf{Q})$ from $\mathrm{Sel}^2_{\text{fake}}(J,\mathbf{Q})$. Let $M[2]$ denote the 2-torsion of an abelian group M and let $\dim V$ denote the dimension of an $\mathbf{F}_2$ vector space V. We have $\dim \mathrm{Sel}^2(J,\mathbf{Q}) = r + \dim J(\mathbf{Q})[2] + \dim Ш(J,\mathbf{Q})[2]$. In other words,

$$\dim\, Ш(J,\mathbf{Q})[2] = \dim \mathrm{Sel}^2(J,\mathbf{Q}) - r - \dim J(\mathbf{Q})[2].$$

It is interesting to note that in all 30 cases where $\dim Ш(J,\mathbf{Q})[2] \leq 1$, we were able to compute the Mordell-Weil rank independently from the analytic rank. The cases where $\dim Ш(J,\mathbf{Q})[2] = 1$ are discussed in more detail in Section 6. For both of the remaining cases we have $\dim Ш(J,\mathbf{Q})[2] = 2$. One of these cases is $C_{125,B}$. For this curve we computed $\mathrm{Sel}^{\sqrt{5}}(J_{125,B},\mathbf{Q})$ using the technique described in [Sc]. From this, we were able to determine that the Mordell-Weil rank is 0 independently from the analytic rank. For the other case, $C_{133,A}$, we could show that r had to be either 0 or 2 from the equation, but we needed the analytic computation to show that $r = 0$.

3.3. **Regulator.** When the Mordell-Weil rank is 0, then the regulator is 1. When the Mordell-Weil rank is positive, then to compute the regulator, we first need to find generators for $J(\mathbf{Q})/J(\mathbf{Q})_{\text{tors}}$. The regulator is the determinant of the canonical height pairing matrix on this set of generators. An algorithm for computing the generators and canonical heights is given in [FS]; it was used to find generators for $J(\mathbf{Q})/J(\mathbf{Q})_{\text{tors}}$ and to compute the regulators. In that article, the algorithm for computing height constants at the infinite prime is not clearly explained and there are some errors in the examples. A clear algorithm for computing infinite height constants is given in [Sto3]. In [Sto4], some improvements of the results and algorithms in [FS] and [Sto3] are discussed. The regulators in Table 2 have been double-checked using these improved algorithms.

3.4. **Tamagawa Numbers.** Let $\mathcal{O}$ be the integer ring in K which will be $\mathbf{Q}_p$ or $\mathbf{Q}_p^{\text{unr}}$ (the maximal unramified extension of $\mathbf{Q}_p$). Let $\mathcal{J}$ be the Néron model of J over $\mathcal{O}$. Define $\mathcal{J}^0$ to be the open subgroup scheme of $\mathcal{J}$ whose generic fiber is isomorphic to J over K and whose special fiber is the identity component of the closed fiber of $\mathcal{J}$. The group $\mathcal{J}^0(\mathcal{O})$ is isomorphic to a subgroup of $J(K)$ which we denote $J^0(K)$. The group $J(\mathbf{Q}_p^{\text{unr}})/J^0(\mathbf{Q}_p^{\text{unr}})$ is the component group of $\mathcal{J}$ over $\mathcal{O}_{\mathbf{Q}_p^{\text{unr}}}$. We are interested in computing $c_p = \#J(\mathbf{Q}_p)/J^0(\mathbf{Q}_p)$, which is sometimes called the Tamagawa number. Since Néron models are stable under unramified base extension, the $\mathrm{Gal}(\mathbf{Q}_p^{\text{unr}}/\mathbf{Q}_p)$-invariant subgroup of $J^0(\mathbf{Q}_p^{\text{unr}})$

is $J^0(\mathbf{Q}_p)$. Since $H^1(\mathrm{Gal}(\mathbf{Q}_p^{\mathrm{unr}}/\mathbf{Q}_p), J^0(\mathbf{Q}_p^{\mathrm{unr}}))$ is trivial (see [Mi1, p. 58]) we see that the $\mathrm{Gal}(\mathbf{Q}_p^{\mathrm{unr}}/\mathbf{Q}_p)$-invariant subgroup of $J(\mathbf{Q}_p^{\mathrm{unr}})/J^0(\mathbf{Q}_p^{\mathrm{unr}})$ is $J(\mathbf{Q}_p)/J^0(\mathbf{Q}_p)$.

There exist several discussions in the literature on constructing the group $J(\mathbf{Q}_p^{\mathrm{unr}})/J^0(\mathbf{Q}_p^{\mathrm{unr}})$ starting with an integral model of the underlying curve. For our purposes, we especially recommend Silverman's book [Si], Chapter IV, Sections 4 and 7. For a more detailed treatment, see [BLR, chap. 9] and [Ed2, §2]. One can find justifications for what we will do in these sources. While constructing such groups, we ran into a number of difficulties that we did not find described anywhere. For that reason, we will present examples of such difficulties that arose, as well as our methods of resolution. We do not claim that we will describe all situations that could arise.

When computing c_p we need a proper, regular model $\mathcal{C}$ for C over $\mathbf{Z}_p$. Let $\mathbf{Z}_p^{\mathrm{unr}}$ denote the ring of integers of $\mathbf{Q}_p^{\mathrm{unr}}$ and note that $\mathbf{Z}_p^{\mathrm{unr}}$ is a pro-étale Galois extension of $\mathbf{Z}_p$ with Galois group $\mathrm{Gal}(\mathbf{Z}_p^{\mathrm{unr}}/\mathbf{Z}_p) = \mathrm{Gal}(\mathbf{Q}_p^{\mathrm{unr}}/\mathbf{Q}_p)$. It follows that giving a model for C over $\mathbf{Z}_p$ is equivalent to giving a model for C over $\mathbf{Z}_p^{\mathrm{unr}}$ that is equipped with a Galois action. We have found it convenient to always work with the latter description. Thus for us, giving a model over $\mathbf{Z}_p$ will always mean giving a model over $\mathbf{Z}_p^{\mathrm{unr}}$ together with a Galois action.

In order to find a proper, regular model for C over $\mathbf{Z}_p$, we start with the models in Table 1. Technically, we consider the curves to be the two affine pieces $y^2+g(x)y = f(x)$ and $v^2 + u^3g(1/u)v = u^6f(1/u)$, glued together by $ux = 1$, $v = u^3y$. We blow them up at all points that are not regular until we have a regular model. (A point is *regular* if the cotangent space there has two generators.) These curves are all proper, and this is not affected by blowing up.

Let $\mathcal{C}_p$ denote the special fiber of $\mathcal{C}$ over $\mathbf{Z}_p^{\mathrm{unr}}$. The group $J(\mathbf{Q}_p^{\mathrm{unr}})/J^0(\mathbf{Q}_p^{\mathrm{unr}})$ is isomorphic to a quotient of the degree 0 part of the free group on the irreducible components of $\mathcal{C}_p$. Let the irreducible components be denoted $\mathcal{D}_i$ for $1 \le i \le n$, and let the multiplicity of $\mathcal{D}_i$ in $\mathcal{C}_p$ be d_i. Then the degree 0 part of the free group has the form

$$L = \{\sum_{i=1}^n \alpha_i \mathcal{D}_i \mid \sum_{i=1}^n d_i\alpha_i = 0\}\,.$$

In order to describe the group that we quotient out by, we must discuss the intersection pairing. For components $\mathcal{D}_i$ and $\mathcal{D}_j$ of the special fiber, let $\mathcal{D}_i{\cdot}\mathcal{D}_j$ denote their intersection pairing. In all of the special fibers that arise in our examples, distinct components intersect transversally. Thus, if $i \neq j$, then $\mathcal{D}_i \cdot \mathcal{D}_j$ equals the number of points at which $\mathcal{D}_i$ and $\mathcal{D}_j$ intersect. The case of self-intersection ($i = j$) is discussed below.

The kernel of the map from L to $J(\mathbf{Q}_p^{\mathrm{unr}})/J^0(\mathbf{Q}_p^{\mathrm{unr}})$ is generated by divisors of the form

$$[\mathcal{D}_j] = \sum_{i=1}^n (\mathcal{D}_j \cdot \mathcal{D}_i)\mathcal{D}_i$$

for each component $\mathcal{D}_j$. We can deduce $\mathcal{D}_j \cdot \mathcal{D}_j$ by noting that $[\mathcal{D}_j]$ must be contained in the group L. This follows from the fact that the intersection pairing of $\mathcal{C}_p = \sum d_i\mathcal{D}_i$ with any irreducible component is 0.

Example 1. Curve $C_{65,B}$ over $\mathbf{Z}_2$.

The Jacobian of $C_{65,B}$ is a quotient of the Jacobian of $X_0(65)$. Since 65 is odd, $J_0(65)$ has good reducation at 2; however, $C_{65,B}$ has singular reduction at 2. Since the equation for this curve is conjectural (it is a Wang-only curve), it will be nice to verify that 2 does not divide the conductor of its Jacobian, i.e. that the Jacobian has good reduction at 2. In addition, we will need a proper, regular model for this curve in order to find Ω.

We start with the arithmetic surface over $\mathbf{Z}_2^{\text{unr}}$ given by the two pieces $y^2 = f(x) = -x^6 + 10x^5 - 32x^4 + 20x^3 + 40x^2 + 6x - 1$ and $v^2 = u^6 f(1/u)$. (Here and in the following we will not specify the gluing maps.) This arithmetic surface is regular at $u = 0$ so we focus our attention on the first affine piece. The special fiber of $y^2 = f(x)$ over $\mathbf{Z}_2^{\text{unr}}$ is given by $(y + x^3 + 1)^2 = 0 \pmod 2$; this is a genus 0 curve of multiplicity 2 that we denote A. This model is not regular at the two points $(x - \alpha, y, 2)$, where α is a root of $x^2 - 3x - 1$. The current special fiber is in Figure 1 and is labelled *Fiber 1*.

We fix α and move $(x - \alpha, y, 2)$ to the origin using the substitution $x_0 = x - \alpha$. We get

$$y^2 = -x_0^6 + (-6\alpha+10)x_0^5 + (5\alpha-47)x_0^4 + (-28\alpha+60)x_0^3 + (-11\alpha-2)x_0^2 + (-24\alpha-16)x_0$$

which we rewrite as the pair of equations

$$\begin{aligned} g_1(x_0, y, p) &= -x_0^6 + (-3\alpha + 5)px_0^5 + (5\alpha - 47)x_0^4 + (-7\alpha + 15)p^2x_0^3 \\ &\qquad + (-11\alpha - 2)x_0^2 + (-3\alpha - 2)p^3x_0 - y^2 \\ &= 0, \\ p &= 2. \end{aligned}$$

To blow up at (x_0, y, p), we introduce projective coordinates (x_1, y_1, p_1) with $x_0y_1 = x_1y$, $x_0p_1 = x_1p$, and $yp_1 = y_1p$. We look in three affine pieces that cover the blow-up of $g_1(x_0, y, p) = 0$, $p = 2$ and check for regularity.

$x_1 = 1$**:** We have $y = x_0y_1$, $p = x_0p_1$. We get $g_2(x_0, y_1, p_1) = 0$, $x_0p_1 = 2$, where

$$\begin{aligned} g_2(x_0, y_1, p_1) &= x_0^{-2}g_1(x_0, x_0y_1, x_0p_1) \\ &= -x_0^4 + (-3\alpha + 5)p_1x_0^4 + (5\alpha - 47)x_0^2 + (-7\alpha + 15)p_1^2x_0^3 \\ &\qquad + (-11\alpha - 2) + (-3\alpha - 2)p_1^3x_0^2 - y_1^2\,. \end{aligned}$$

In the reduction we have either $x_0 = 0$ or $p_1 = 0$.

$x_0 = 0$**:** $(y_1 + \alpha + 1)^2 = 0$. This is a new component which we denote B. It has genus 0 and multiplicity 2. We check regularity along B at $(x_0, y_1 + \alpha + 1, p_1 - t, 2)$, with t in $\mathbf{Z}_2^{\text{unr}}$, and find that B is nowhere regular.

$p_1 = 0$**:** $(y_1 + x_0^2 + \alpha x_0 + (\alpha + 1))^2 = 0$. Using the gluing maps, we see that this is A.

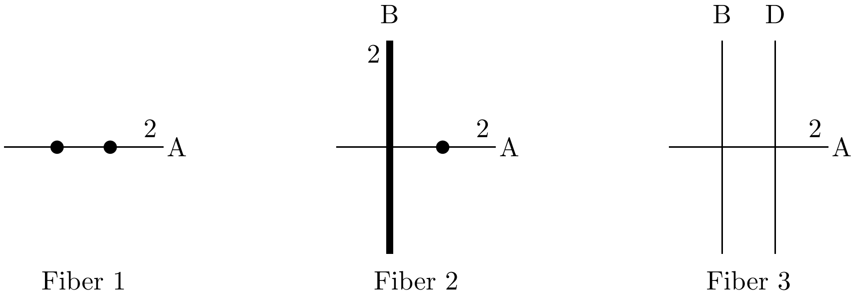
$y_1 = 1$**:** We get no new information from this affine piece.

$p_1 = 1$**:** We have $x_0 = x_1p$, $y = y_1p$. We get $g_3(x_1, y_1, p) = p^{-2}g_1(x_1p, y_1p, p) = 0$, $p = 2$. In the reduction we have

$p = 0$**:** $(y_1 + (\alpha + 1)x_1)^2 = 0$. Using the gluing maps, we see that this is B. It is nowhere regular.

The current special fiber is in Figure 1 and is labelled *Fiber 2*. It is not regular along B and at the other point on A which we have not yet blown up. The component B does not lie entirely in any one affine piece so we will blow up the affine pieces $x_1 = 1$ and $p_1 = 1$ along B.

FIGURE 1. Special fibers of curve $C_{65,B}$ over $\mathbf{Z}_2$; points not regular are thick



To blow up $x_1 = 1$ along B we make the substitution $y_2 = y_1 + \alpha + 1$ and replace each factor of 2 in a coefficient by $x_0 p_1$. We have $g_4(x_0, y_2, p_1) = 0$ and $x_0 p_1 = 2$, and we want to blow up along the line $(x_0, y_2, 2)$. Blowing up along a line is similar to blowing up at a point: since we are blowing up at $(x_0, y_2, 2) = (x_0, y_2)$, we introduce projective coordinates x_3, y_3 together with the relation $x_0 y_3 = x_3 y_2$. We consider two affine pieces that cover the blow-up of $x_1 = 1$.

$x_3 = 1$**:** We have $y_2 = y_3 x_0$. We get $g_5(x_0, y_3, p_1) = x_0^{-2} g_4(x_0, y_3 x_0, p_1) = 0$ and $x_0 p_1 = 2$. In the reduction we have

$x_0 = 0$**:** $y_3^2 + (\alpha + 1) y_3 p_1 + \alpha p_1^3 + p_1^2 + \alpha + 1 = 0$. This is B. It is now a non-singular genus 1 curve.

$p_1 = 0$**:** $(x_0 + y_3 + \alpha)^2 = 0$. This is A. The point where B meets A transversally is regular.

$y_3 = 1$**:** We get no new information from this affine piece.

When we blow up $p_1 = 1$ along B we get essentially the same thing and all points are again regular.

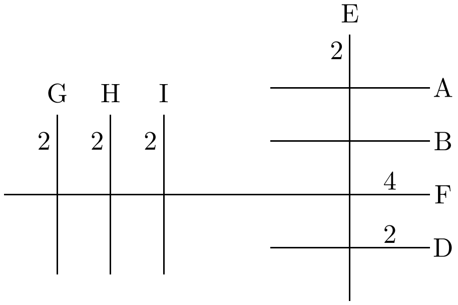
The other non-regular point on A is the conjugate of the one we blew up. Therefore, after performing the conjugate blow ups, it too will be a genus 1 component crossing A transversally. We denote this component D; it is conjugate to B.

We now have a proper, regular model $\mathcal{C}$ of C over $\mathbf{Z}_2$. Let $\mathcal{C}_2$ be the special fiber of this model; a diagram of $\mathcal{C}_2$ is in Figure 1 and is labelled *Fiber 3*. We can use $\mathcal{C}$ to show that the Néron model $\mathcal{J}$ of the Jacobian $J = J_{65,B}$ has good reduction at 2.

We know that the reduction of $\mathcal{J}^0$ is the extension of an abelian variety by a connected linear group. Since $\mathcal{C}$ is regular and proper, the abelian variety part of the reduction is the product of the Jacobians of the normalizations of the components of $\mathcal{C}_2$ (see [BLR, 9.3/11 and 9.5/4]). Thus, the abelian variety part is the product of the Jacobians of B and D. Since this is 2-dimensional, the reduction of $\mathcal{J}^0$ is an abelian variety. In other words, since the sum of the genera of the components of the special fiber is equal to the dimension of J, the reduction is an abelian variety. It follows that $\mathcal{J}$ has good reduction at 2, that the conductor of J is odd, and that $c_2 = 1$. As noted above, this gives further evidence that the equation given in Table 1 is correct.

Example 2. Curve C_{63} over $\mathbf{Z}_3$.

FIGURE 2. Special fiber of curve C_{63} over $\mathbf{Z}_3$



The Tamagawa number is often found using the intersection matrix and subdeterminants. This is not entirely satisfactory for cases where the special fiber has several components and a non-trivial Galois action. Here is an example of how to resolve this (see also [BL]).

When we blow up curve C_{63} over $\mathbf{Z}_3^{\mathrm{unr}}$, we get the special fiber shown in Figure 2. Elements of $\mathrm{Gal}(\mathbf{Q}_3^{\mathrm{unr}}/\mathbf{Q}_3)$ that do not fix the quadratic unramified extension of $\mathbf{Q}_3$ switch H and I. The other components are defined over $\mathbf{Q}_3$. All components have genus 0. The group $J(\mathbf{Q}_3^{\mathrm{unr}})/J^0(\mathbf{Q}_3^{\mathrm{unr}})$ is isomorphic to a quotient of

$$L = \{\alpha A+\beta B+\delta D+\epsilon E+\phi F+\gamma G+\eta H+\iota I \mid \alpha+\beta+2\delta+2\epsilon+4\phi+2\gamma+2\eta+2\iota = 0\}\,.$$

The kernel is generated by the following divisors.

$$\begin{array}{llll}
[A] = & -2A+E & [B] = & -2B+E \\
[D] = & -D+E & [E] = & A+B+D-4E+F \\
[F] = & E-2F+G+H+I & [G] = & F-2G \\
[H] = & F-2H & [I] = & F-2I
\end{array}$$

When we project away from A, we find that $J(\mathbf{Q}_3^{\mathrm{unr}})/J^0(\mathbf{Q}_3^{\mathrm{unr}})$ is isomorphic to

$$\langle B, D, E, F, G, H, I \mid E = 0, E = 2B, D = E, 4E = B + D + F,$$
$$2F = E + G + H + I, F = 2G = 2H = 2I\rangle.$$

At this point, it is straightforward to simplify the representation by elimination. Note that we projected away from A, which is Galois-invariant. It is best to continue eliminating Galois-invariant elements first. We find that this group is isomorphic to $\langle H, I \mid 2H = 2I = 0\rangle$ and elements of $\mathrm{Gal}(\mathbf{Q}_3^{\mathrm{unr}}/\mathbf{Q}_3)$ that do not fix the quadratic unramified extension of $\mathbf{Q}_3$ switch H and I. Therefore $J(\mathbf{Q}_3^{\mathrm{unr}})/J^0(\mathbf{Q}_3^{\mathrm{unr}}) \cong \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$ and $c_3 = \#J(\mathbf{Q}_3)/J^0(\mathbf{Q}_3) = 2$.

3.5. **Computing Ω.** By an *integral differential* (or *integral form*) on J we mean the pullback to J of a global relative differential form on the Néron model of J over $\mathbf{Z}$. The set of integral n-forms on J is a full-rank lattice in the $\mathbf{Q}$-vector space of global holomorphic n-forms on J. Since J is an abelian variety of dimension 2, the integral 1-forms are a free $\mathbf{Z}$-module of rank 2 and the integral 2-forms are a free $\mathbf{Z}$-module of rank 1. Moreover, the wedge of a basis for the integral 1-forms is a generator for the integral 2-forms. The quantity Ω is the integral, over the real points of J, of a generator for the integral 2-forms. (We choose the generator that leads to a positive integral.)

We now translate this into a computation on the curve C. Let $\{\omega_1, \omega_2\}$ be a $\mathbf{Q}$-basis for the holomorphic differentials on C and let $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ be a $\mathbf{Z}$-basis for the homology of $C(\mathbf{C})$. Create a 2×4 complex matrix $M_{\mathbf{C}} = [\int_{\gamma_j} \omega_i]$ by integrating the differentials over the homology and let $M_{\mathbf{R}} = \mathrm{Tr}_{\mathbf{C}/\mathbf{R}}(M_{\mathbf{C}})$ be the 2×4 real matrix whose entries are traces from the complex matrix. The columns of $M_{\mathbf{R}}$ generate a lattice Λ in $\mathbf{R}^2$. If we make the standard identification between the holomorphic 1-forms on J and the holomorphic differentials on C (see [Mi2]), then the notation $\int_{J(\mathbf{R})} |\omega_1 \wedge \omega_2|$ makes sense and its value can be computed as the area of a fundamental domain for Λ.

If $\{\omega_1, \omega_2\}$ is a basis for the integral 1-forms on J, then $\int_{J(\mathbf{R})} |\omega_1 \wedge \omega_2| = \Omega$. On the other hand, the computation of $M_{\mathbf{C}}$ is simplest if we choose $\omega_1 = dX/Y$, and $\omega_2 = X\,dX/Y$ with respect to a model for C of the form $Y^2 = F(X)$; in this case we obtain Ω by a simple change-of-basis calculation. This assumes, of course, that we know how to express a basis for the integral 1-forms in terms of the basis $\{\omega_1, \omega_2\}$; this is addressed in more detail below.

It is worth mentioning an alternate strategy. Instead of finding a $\mathbf{Z}$-basis for the homology of $C(\mathbf{C})$ one could find a $\mathbf{Z}$-basis $\{\gamma_1', \gamma_2'\}$ for the subgroup of the homology that is fixed by complex conjugation (call this the real homology). Integrating would give us a 2×2 real matrix $M_{\mathbf{R}}'$ and the determinant of $M_{\mathbf{R}}'$ would equal the integral of $\omega_1 \wedge \omega_2$ over the connected component of $J(\mathbf{R})$. In other words, the number of real connected components of J is equal to the index of the $\mathbf{C}/\mathbf{R}$-traces in the real homology.

We now come to the question of determining the differentials on C which correspond to the integral 1-forms on J. Call these the integral differentials on C. This computation can be done one prime at a time. At each prime p this is equivalent to determining a $\mathbf{Z}_p^{\mathrm{unr}}$-basis for the global relative differentials on any proper, regular model for C over $\mathbf{Z}_p^{\mathrm{unr}}$. In fact a more general class of models can be used; see the discussion of models with rational singularities in [BLR, §6.7] and [Li, §4.1].

We start with the model $y^2 + g(x)y = f(x)$ given in Table 1. Note that the substitution $X = x$ and $Y = 2y + g(x)$ gives us a model of the form $Y^2 = F(X)$. For integration purposes, our preferred differentials are $dX/Y = dx/(2y + g(x))$ and $X\,dX/Y = x\,dx/(2y + g(x))$. It is not hard to show that at primes of nonsingular reduction for the $y^2 + g(x)y = f(x)$ model, these differentials will generate the integral 1-forms. For each prime p of singular reduction we give the following algorithm. All steps take place over $\mathbf{Z}_p^{\mathrm{unr}}$.

Step 1: Compute explicit equations for a proper, regular model $\mathcal{C}$.

Step 2: Diagram the configuration of the special fiber of $\mathcal{C}$.

Step 3: (Optional) Identify exceptional components and blow them down in the configuration diagram. Repeat step 3 as necessary.

Step 4: (Optional) Remove components with genus 0 and self-intersection -2. Since C has genus greater than 1, there will be a component that is not of this kind.

(This step corresponds to contracting the given components. The model obtained would no longer be regular; it would, however, be a proper model with rational singularities. We will not need a diagram of the resulting configuration.)

Step 5: Determine a $\mathbf{Z}_p^{\mathrm{unr}}$-basis for the integral differentials. It suffices to check this on a dense open subset of each surviving component. Note that we have explicit equations for a dense open subset of each of these components from the model $\mathcal{C}$ in step 1. A pair of differentials $\{\eta_1, \eta_2\}$ will be a basis for the integral differentials (at p) if the following three statements are true.

a: The pair $\{\eta_1, \eta_2\}$ is a basis for the holomorphic differentials on C.

b: The reductions of η_1 and η_2 produce well-defined differentials mod p on an open subset of each surviving component.

c: If $a_1\eta_1 + a_2\eta_2 = 0 \pmod p$ on all surviving components, then $p|a_1$ and $p|a_2$.

Techniques for explicitly computing a proper, regular model are discussed in Section 3.4. A configuration diagram should include the genus, multiplicity and self-intersection number of each component and the number and type of intersections between components. Note that when an exceptional component is blown down, all of the self-intersection numbers of the components intersecting it will go up (towards 0). In particular, components which were not exceptional before may become exceptional in the new configuration.

Steps 3 and 4 are intended to make this algorithm more efficient for a human. They are entirely optional. For a computer implementation it may be easier to simply check every component than to worry about manipulating configurations.

The curves in Table 1 are given as $y^2 + g(x)y = f(x)$. We assumed, at first, that $dx/(2y + g(x))$ and $x\,dx/(2y + g(x))$ generate the integral differentials. We integrated these differentials around each of the four paths generating the complex homology and found a provisional Ω. Then we checked the proper, regular models to determine if these differentials really do generate the integral differentials and adjusted Ω when necessary. There were three curves where we needed to adjust Ω. We describe the adjustment for curve $C_{65,B}$ in the following example. For curve C_{63}, we used the differentials $3\,dx/(2y + g(x))$ and $x\,dx/(2y + g(x))$. For curve $C_{65,A}$, we used the differentials $3\,dx/(2y + g(x))$ and $3x\,dx/(2y + g(x))$.

Example 3. Curve $C_{65,B}$.

The primes of singular reduction for curve $C_{65,B}$ are 2, 5 and 13. In Example 1 of Section 3.4, we found a proper, regular model $\mathcal{C}$ for C over $\mathbf{Z}_2^{\mathrm{unr}}$. The configuration for the special fiber of $\mathcal{C}$ is sketched in Figure 1 under the label *Fiber 3.* Component A is exceptional and can be blown down to produce a model in which B and D cross transversally. Since B and D both have genus 1, we cannot eliminate either of these components. Furthermore, it suffices to check B, since D is its Galois conjugate.

To get from the equation of the curve listed in Table 1 to an affine containing an open subset of B we need to make the substitutions $x = x_0 - \alpha$ and $y = x_0(y_3x_0 - \alpha - 1)$. We also have $x_0p_1 = 2$. Using the substitutions and the relation $dx_0/x_0 = -dp_1/p_1$, we get

$$\frac{dx}{2y} = \frac{-dp_1}{2p_1(y_3x_0 - \alpha - 1)} \quad \text{and} \quad \frac{x\,dx}{2y} = \frac{-(x_0 + \alpha)\,dp_1}{2p_1(y_3x_0 - \alpha - 1)}\,.$$

Note that $p_1 - t$ is a uniformizer at $p_1 = t$ almost everywhere on B. When we multiply each differential by 2, then the denominator of each is almost everywhere non-zero; thus, dx/y and $x\,dx/y$ are integral at 2. Moreover, although the linear combination $(x-\alpha)\,dx/y$ is identically zero on B, it is not identically zero on D (its Galois conjugate is not identically zero on B). Thus, our new basis is correct at 2. We multiply the provisional Ω by 4 to get a new provisional Ω which is correct at 2.

Similar (but somewhat simpler) computations at the primes 5 and 13 show that no adjustment is needed at these primes. Thus, dx/y and $x\,dx/y$ form a basis for the integral differentials of curve $C_{65,B}$, and the correct value of Ω is 4 times our original guess.

4. Modular algorithms

In this section, we describe the algorithms that were used to compute some of the data from the newforms. This includes the analytic rank and leading coefficient of the L-series. For optimal quotients, the value of $k \cdot \Omega$ can also be found (k is the Manin constant), as well as partial information on the Tamagawa numbers c_p and the size of the torsion subgroup.

4.1. **Analytic rank of $L(J,s)$ and leading coefficient at $s=1$.**

Fix a Jacobian J corresponding to the 2-dimensional subspace of $S_2(N)$ spanned by quadratic conjugate, normalized newforms f and $\overline{f}$. Let W_N be the Fricke involution. The newforms f and $\overline{f}$ have the same eigenvalue ϵ_N with respect to W_N, namely $+1$ or -1. In the notation of Section 2, let

$$L(f,s)=\sum_{n=1}^{\infty}\frac{a_n}{n^s}$$

be the L-series of f; then $L(\overline{f},s)$ is the Dirichlet series whose coefficients are the conjugates of the coefficients of $L(f,s)$. (Recall that the a_n are integers in some real quadratic field.) The order of $L(f,s)$ at $s=1$ is even when $\epsilon_N=-1$ and odd when $\epsilon_N=+1$. We have $L(J,s)=L(f,s)L(\overline{f},s)$. Thus the analytic rank of J is 0 modulo 4 when $\epsilon_N=-1$ and 2 modulo 4 when $\epsilon_N=+1$. We found that the ranks were all 0 or 2. To prove that the analytic rank of J is 0, we need to show $L(f,1)\neq 0$ and $L(\overline{f},1)\neq 0$. In the case that $\epsilon_N=+1$, to prove that the analytic rank is 2, we need to show that $L'(f,1)\neq 0$ and $L'(\overline{f},1)\neq 0$. When $\epsilon_N=-1$, we can evaluate $L(f,1)$ as in [Cr2, §2.11]. When $\epsilon_N=+1$, we can evaluate $L'(f,1)$ as in [Cr2, §2.13]. Each appropriate $L(f,1)$ or $L'(f,1)$ was at least 0.1 and the errors in our approximations were all less than 10^{-67}. In this way we determined the analytic ranks, which we denote r. As noted in the introduction, the analytic rank equals the Mordell-Weil rank if $r=0$ or $r=2$. Thus, we can simply call r the rank, without fear of ambiguity.

To compute the leading coefficient of $L(J,s)$ at $s=1$, we note that $\lim_{s\to 1}L(J,s)/(s-1)^r=L^{(r)}(J,1)/r!$. In the $r=0$ case, we simply have $L(J,1)=L(f,1)L(\overline{f},1)$. In the $r=2$ case, we have $L''(J,s)=L''(f,s)L(\overline{f},s)+2L'(f,s)L'(\overline{f},s)+L(f,s)L''(\overline{f},s)$. Evaluating both sides at $s=1$ we get $\frac{1}{2}L''(J,1)=L'(f,1)L'(\overline{f},1)$.

4.2. **Computing $k\cdot\Omega$.**

Let J, f and $\overline{f}$ be as in Section 4.1 and assume J is an optimal quotient. Let V be the 2-dimensional space spanned by f and $\overline{f}$. Choose a basis $\{\omega_1,\omega_2\}$ for the subgroup of V consisting of forms whose q-expansion coefficients lie in $\mathbf{Z}$. Let $k\cdot\Omega$ be the volume of the real points of the quotient of $\mathbf{C}\times\mathbf{C}$ by the lattice of period integrals $(\int_\gamma \omega_1, \int_\gamma \omega_2)$ with γ in the integral homology

$H_1(X_0(N), \mathbf{Z})$. The rational number k is called the *Manin constant*. In practice we compute $k \cdot \Omega$ using modular symbols and a generalization to dimension 2 of the algorithm for computing periods described in [Cr2, §2.10]. When $L(J, 1) \neq 0$ the method of [Cr2, §2.11] coupled with Sections 4.1 and 4.3 can also be used to compute $k \cdot \Omega$.

A slight generalization of the argument of Proposition 2 of [Ed1] proves that k is, in fact, an integer. This generalization can be found in [AS2], where one also finds a conjecture that k must equal 1 for all optimal quotients of Jacobians of modular curves, which generalizes the longstanding conjecture of Manin that k equals 1 for all optimal elliptic curves. In unpublished work, Edixhoven has partially proven Manin's conjecture.

The computations of the present paper verify that k equals 1 for the optimal quotients that we are considering. Specifically, we computed $k \cdot \Omega$ as above and Ω as described in Section 3.5. The quotient of the two values was always well within 0.5 of 1.

4.3. **Computing** $L(J, 1)/(k \cdot \Omega)$**.** We compute the rational number $L(J, 1)/(k \cdot \Omega)$, for optimal quotients, using the algorithm in [AS1]. This algorithm generalizes the algorithm described in [Cr2, §2.8] to dimension greater than 1.

4.4. **Tamagawa numbers.** In this section we assume that p is a prime which exactly divides the conductor N of J. Under these conditions, Grothendieck [Gr] gave a description of the component group of J in terms of a monodromy pairing on certain character groups. (For more details, see Ribet [Ri, §2].) If, in addition, J is a new optimal quotient of $J_0(N)$, one deduces the following. When the eigenvalue for f of the Atkin-Lehner involution W_p is $+1$, then the rational component group of J is a subgroup of $(\mathbf{Z}/2\mathbf{Z})^2$. Furthermore, when the eigenvalue of W_p is -1, the algorithm described in [Ste] can be used to compute the value of c_p.

4.5. **Torsion subgroup.** To compute an integer divisible by the order of the torsion subgroup of J we make use of the following two observations. First, it is a consequence of the Eichler-Shimura relation [Sh, §7.9] that if p is a prime not dividing the conductor N of J and $f(T)$ is the characteristic polynomial of the endomorphism T_p of J, then $\#J(\mathbf{F}_p) = f(p + 1)$ (see [Cr2, §2.4] for an algorithm to compute $f(T)$). Second, if p is an odd prime at which J has good reduction, then the natural map $J(\mathbf{Q})_{\text{tors}} \to J(\mathbf{F}_p)$ is injective (see [CF, p. 70]). This does not depend on whether J is an optimal quotient. To obtain a lower bound on the torsion subgroup for optimal quotients, we use modular symbols and the Abel-Jacobi theorem [La, IV.2] to compute the order of the image of the rational point $(0) - (\infty) \in J_0(N)$.

5. Tables

In Table 1, we list the 32 curves described in Section 2. We give the level N from which each curve arose, an integral model for the curve, and list the source(s) from which it came (H for Hasegawa [Ha], W for Wang [Wan]). Throughout the paper, the curves are denoted C_N (or $C_{N,A}$, $C_{N,B}$).

In Table 2, we list the curve C_N simply by N, the level from which it arose. Let r denote the rank. We list $\lim_{s \to 1}(s - 1)^{-r} L(J, s)$ where $L(J, s)$ is the L-series for the Jacobian J of C_N and round off the results to five digits. The symbol Ω was defined in Section 3.5 and is also rounded to five digits. Let Reg denote the

N	Equation	Source
23	$y^2 + (x^3 + x + 1)y = -2x^5 - 3x^2 + 2x - 2$	HW
29	$y^2 + (x^3 + 1)y = -x^5 - 3x^4 + 2x^2 + 2x - 2$	HW
31	$y^2 + (x^3 + x^2 + 1)y = -x^5 - 5x^4 - 5x^3 + 3x^2 + 2x - 3$	HW
35	$y^2 + (x^3 + x)y = -x^5 - 8x^3 - 7x^2 - 16x - 19$	H
39	$y^2 + (x^3 + 1)y = -5x^4 - 2x^3 + 16x^2 - 12x + 2$	H
63	$y^2 + (x^3 - 1)y = 14x^3 - 7$	W
65,A	$y^2 + (x^3 + 1)y = -4x^6 + 9x^4 + 7x^3 + 18x^2 - 10$	W
65,B	$y^2 = -x^6 + 10x^5 - 32x^4 + 20x^3 + 40x^2 + 6x - 1$	W
67	$y^2 + (x^3 + x + 1)y = x^5 - x$	HW
73	$y^2 + (x^3 + x^2 + 1)y = -x^5 - 2x^3 + x$	HW
85	$y^2 + (x^3 + x^2 + x)y = x^4 + x^3 + 3x^2 - 2x + 1$	H
87	$y^2 + (x^3 + x + 1)y = -x^4 + x^3 - 3x^2 + x - 1$	HW
93	$y^2 + (x^3 + x^2 + 1)y = -2x^5 + x^4 + x^3$	HW
103	$y^2 + (x^3 + x^2 + 1)y = x^5 + x^4$	HW
107	$y^2 + (x^3 + x^2 + 1)y = x^4 - x^2 - x - 1$	HW
115	$y^2 + (x^3 + x + 1)y = 2x^3 + x^2 + x$	HW
117,A	$y^2 + (x^3 - 1)y = 3x^3 - 7$	W
117,B	$y^2 + (x^3 + 1)y = -x^6 - 3x^4 - 5x^3 - 12x^2 - 9x - 7$	W
125,A	$y^2 + (x^3 + x + 1)y = x^5 + 2x^4 + 2x^3 + x^2 - x - 1$	HW
125,B	$y^2 + (x^3 + x + 1)y = x^6 + 5x^5 + 12x^4 + 12x^3 + 6x^2 - 3x - 4$	W
133,A	$y^2 + (x^3 + x + 1)y = -2x^6 + 7x^5 - 2x^4 - 19x^3 + 2x^2 + 18x + 7$	W
133,B	$y^2 + (x^3 + x^2 + 1)y = -x^5 + x^4 - 2x^3 + 2x^2 - 2x$	HW
135	$y^2 + (x^3 + x + 1)y = x^4 - 3x^3 + 2x^2 - 8x - 3$	W
147	$y^2 + (x^3 + x^2 + x)y = x^5 + 2x^4 + x^3 + x^2 + 1$	HW
161	$y^2 + (x^3 + x + 1)y = x^3 + 4x^2 + 4x + 1$	HW
165	$y^2 + (x^3 + x^2 + x)y = x^5 + 2x^4 + 3x^3 + x^2 - 3x$	H
167	$y^2 + (x^3 + x + 1)y = -x^5 - x^3 - x^2 - 1$	HW
175	$y^2 + (x^3 + x^2 + 1)y = -x^5 - x^4 - 2x^3 - 4x^2 - 2x - 1$	W
177	$y^2 + (x^3 + x^2 + 1)y = x^5 + x^4 + x^3$	HW
188	$y^2 = x^5 - x^4 + x^3 + x^2 - 2x + 1$	W
189	$y^2 + (x^3 - 1)y = x^3 - 7$	W
191	$y^2 + (x^3 + x + 1)y = -x^3 + x^2 + x$	HW

TABLE 1. Levels, integral models and sources for curves

regulator, also rounded to five digits. We list the c_p's by primes of increasing order dividing the level N. We denote $J(\mathbf{Q})_{\text{tors}} = \Phi$ and list its size. We use Ш? to denote the size of $(\lim_{s\to 1}(s-1)^{-r}L(J,s)) \cdot (\#J(\mathbf{Q})_{\text{tors}})^2/(\Omega \cdot \text{Reg} \cdot \prod c_p)$, rounded to the nearest integer. We will refer to this as the *conjectured size of* Ш$(J, \mathbf{Q})$. For rank 0 optimal quotients this integer equals the (a priori) rational number $(L(J,1)/(k \cdot \Omega)) \cdot ((\#J(\mathbf{Q})_{\text{tors}})^2/\prod c_p)$; of course there is no rounding error in this computation. For all other cases the last column gives a bound on the accuracy of the computations; all values of Ш? were at least this close to the nearest integer before rounding.

N	r	$\lim\limits_{s\to 1} \frac{L(J,s)}{(s-1)^r}$	Ω	Reg	c_p's	Φ	Ш?	error
23	0	0.24843	2.7328	1	11	11	1	
29	0	0.29152	2.0407	1	7	7	1	
31	0	0.44929	2.2464	1	5	5	1	
35	0	0.37275	2.9820	1	16,2	16	1	$< 10^{-25}$
39	0	0.38204	10.697	1	28,1	28	1	$< 10^{-25}$
63	0	0.75328	4.5197	1	2,3	6	1	
65,A	0	0.45207	6.3289	1	7,1	14	2	
65,B	0	0.91225	5.4735	1	1,3	6	2	
67	2	0.23410	20.465	0.011439	1	1	1	$< 10^{-50}$
73	2	0.25812	24.093	0.010713	1	1	1	$< 10^{-49}$
85	2	0.34334	9.1728	0.018715	4,2	2	1	$< 10^{-26}$
87	0	1.4323	7.1617	1	5,1	5	1	
93	2	0.33996	18.142	0.0046847	4,1	1	1	$< 10^{-49}$
103	2	0.37585	16.855	0.022299	1	1	1	$< 10^{-49}$
107	2	0.53438	11.883	0.044970	1	1	1	$< 10^{-49}$
115	2	0.41693	10.678	0.0097618	4,1	1	1	$< 10^{-50}$
117,A	0	1.0985	3.2954	1	4,3	6	1	
117,B	0	1.9510	1.9510	1	4,1	2	1	
125,A	2	0.62996	13.026	0.048361	1	1	1	$< 10^{-50}$
125,B	0	2.0842	2.6052	1	5	5	4	
133,A	0	2.2265	2.7832	1	5,1	5	4	
133,B	2	0.43884	15.318	0.028648	1,1	1	1	$< 10^{-49}$
135	0	1.5110	4.5331	1	3,1	3	1	
147	2	0.61816	13.616	0.045400	2,2	2	1	$< 10^{-50}$
161	2	0.82364	11.871	0.017345	4,1	1	1	$< 10^{-47}$
165	2	0.68650	9.5431	0.071936	4,2,2	4	1	$< 10^{-26}$
167	2	0.91530	7.3327	0.12482	1	1	1	$< 10^{-47}$
175	0	0.97209	4.8605	1	1,5	5	1	
177	2	0.90451	13.742	0.065821	1,1	1	1	$< 10^{-45}$
188	2	1.1708	11.519	0.011293	9,1	1	1	$< 10^{-44}$
189	0	1.2982	3.8946	1	1,3	3	1	
191	2	0.95958	17.357	0.055286	1	1	1	$< 10^{-44}$

Table 2. Conjectured sizes of Ш$(J, \mathbf{Q})$

In Table 3 are generators of $J(\mathbf{Q})/J(\mathbf{Q})_{\text{tors}}$ for the curves whose Jacobians have Mordell-Weil rank 2. The generators are given as divisor classes. Whenever possible, we have chosen generators of the form $[P - Q]$ where P and Q are rational points on the curve. Curve 167 is the only example where this is not the case, since the degree zero divisors supported on the (known) rational points on C_{167} generate a subgroup of index two in the full Mordell-Weil group. Affine points are given by their x and y coordinates in the model given in Table 1. There are two points at infinity in the normalization of the curves described by our equations, with the exception of curve C_{188}. These are denoted by ∞_a, where a is the value of the function y/x^3 on the point in question. The (only) point at infinity on curve C_{188} is simply denoted ∞.

N	Generators of $J(\mathbf{Q})/J(\mathbf{Q})_{\mathrm{tors}}$	
67	$[(0,0)-\infty_{-1}]$	$[(0,0)-(0,-1)]$
73	$[(0,-1)-\infty_{-1}]$	$[(0,0)-\infty_{-1}]$
85	$[(1,1)-\infty_{-1}]$	$[(-1,3)-\infty_0]$
93	$[(-1,1)-\infty_0]$	$[(1,-3)-(-1,-2)]$
103	$[(0,0)-\infty_{-1}]$	$[(0,-1)-(0,0)]$
107	$[\infty_{-1}-\infty_0]$	$[(-1,-1)-\infty_{-1}]$
115	$[(1,-4)-\infty_0]$	$[(1,1)-(-2,2)]$
125,A	$[\infty_{-1}-\infty_0]$	$[(-1,0)-\infty_{-1}]$
133,B	$[\infty_{-1}-\infty_0]$	$[(0,-1)-\infty_{-1}]$
147	$[\infty_{-1}-\infty_0]$	$[(-1,-1)-\infty_0]$
161	$[(1,2)-(-1,1)]$	$[(\frac{1}{2},-3)-(1,2)]$
165	$[(1,1)-\infty_{-1}]$	$[(0,0)-\infty_0]$
167	$[(-1,1)-\infty_0]$	$[(i,0)+(-i,0)-\infty_0-\infty_{-1}]$
177	$[(0,-1)-\infty_0]$	$[(0,0)-(0,-1)]$
188	$[(0,-1)-\infty]$	$[(0,1)-(1,-2)]$
191	$[\infty_{-1}-\infty_0]$	$[(0,-1)-\infty_0]$

TABLE 3. Generators of $J(\mathbf{Q})/J(\mathbf{Q})_{\mathrm{tors}}$ in rank 2 cases

N	Prime	Type	Prime	Type	N	Prime	Type	Prime	Type
23	23	I_{3-2-1}			117,A	3	$\mathrm{III}-\mathrm{III}^*-0$	13	I_{1-1-1}
29	29	I_{3-1-1}			117,B	3	I^*_{3-1-1}	13	I_{1-1-0}
31	31	I_{2-1-1}			125,A	5	$\mathrm{VIII}-1$		
35	5	I_{3-2-2}	7	I_{2-1-0}	125,B	5	$\mathrm{IX}-3$		
39	3	I_{6-2-2}	13	I_{1-1-0}	133,A	7	I_{2-1-1}	19	I_{1-1-0}
63	3	$2\mathrm{I}_0^*-0$	7	I_{1-1-1}	133,B	7	I_{1-1-0}	19	I_{1-1-0}
65,A	3	$\mathrm{I}_0-\mathrm{I}_0-1$	5	I_{3-1-1}	135	3	III	5	I_{3-1-0}
65,A	13	I_{1-1-0}			147	3	I_{2-1-0}	7	VII
65,B	2	$\mathrm{I}_0-\mathrm{I}_0-1$	5	I_{3-1-0}	161	7	I_{2-2-0}	23	I_{1-1-0}
65,B	13	I_{1-1-1}			165	3	I_{2-2-0}	5	I_{2-1-0}
67	67	I_{1-1-0}			165	11	I_{2-1-0}		
73	73	I_{1-1-0}			167	167	I_{1-1-0}		
85	5	I_{2-2-0}	17	I_{2-1-0}	175	5	$\mathrm{II}-\mathrm{II}-0$	7	I_{2-1-1}
87	3	I_{2-1-1}	29	I_{1-1-0}	177	3	I_{1-1-0}	59	I_{1-1-0}
93	3	I_{2-2-0}	31	I_{1-1-0}	188	2	$\mathrm{IV}-\mathrm{IV}-0$	47	I_{1-1-0}
103	103	I_{1-1-0}			189	3	$\mathrm{II}-\mathrm{IV}^*-0$	7	I_{1-1-1}
107	107	I_{1-1-0}			191	191	I_{1-1-0}		
115	5	I_{2-2-0}	23	I_{1-1-0}					

TABLE 4. Namikawa and Ueno classification of special fibers

In Table 4 are the reduction types, from the classification of [NU], of the special fibers of the minimal, proper, regular models of the curves for each of the primes of singular reduction for the curve. They are the same as the primes dividing the level except that curve $C_{65,A}$ has singular reduction at the prime 3 and curve $C_{65,B}$ has singular reduction at the prime 2.

6. Discussion of Shafarevich-Tate groups and evidence for the second conjecture

From Section 3.2 we have $\dim Ш(J,\mathbf{Q})[2] = \dim \mathrm{Sel}^2(J,\mathbf{Q}) - r - \dim J(\mathbf{Q})[2]$. With the exception of curves $C_{65,A}$, $C_{65,B}$, $C_{125,B}$, and $C_{133,A}$ we have $\dim Ш(J,\mathbf{Q})[2] = 0$. Thus we expect $\#Ш(J,\mathbf{Q})$ to be an odd square. In each case, the conjectured size of $Ш(J,\mathbf{Q})$ is 1. For curves $C_{65,A}$, $C_{65,B}$, $C_{125,B}$ and $C_{133,A}$ we have $\dim Ш(J,\mathbf{Q})[2] = 1, 1, 2$ and 2 and the conjectured size of $Ш(J,\mathbf{Q}) = 2, 2, 4$ and 4, respectively. We see that in each case, the (conjectured) size of the odd part of $Ш(J,\mathbf{Q})$ is 1 and the 2-part is accounted for by its 2-torsion.

Recall that for rank 0 optimal quotients we are able to exactly determine the value which the second Birch and Swinnerton-Dyer conjecture predicts for $Ш(J,\mathbf{Q})$. From the previous paragraph, we then see that equation (1.1) holds if and only if $Ш(J,\mathbf{Q})$ is killed by 2.

It is also interesting to consider deficient primes. A prime p is *deficient* with respect to a curve C of genus 2, if C has no degree 1 rational divisor over $\mathbf{Q}_p$. From [PSt], the number of deficient primes has the same parity as $\dim Ш(J,\mathbf{Q})[2]$. Curve $C_{65,A}$ has one deficient prime 3. Curve $C_{65,B}$ has one deficient prime 2. Curve $C_{117,B}$ has two deficient primes 3 and ∞. The rest of the curves have no deficient primes.

Since we have found r (analytic rank) independent points on each Jacobian, we have a direct proof that the Mordell-Weil rank must equal the analytic rank if $\dim Ш(J,\mathbf{Q})[2] = 0$. For curves $C_{65,A}$ and $C_{65,B}$, the presence of an odd number of deficient primes gives us a similar result. For $C_{125,B}$ we used a $\sqrt{5}$-Selmer group to get a similar result. Thus, we have an independent proof of equality between analytic and Mordell-Weil ranks for all curves except $C_{133,A}$.

The 2-Selmer groups have the same dimensions for the pairs $C_{125,A}$, $C_{125,B}$ and $C_{133,A}$, $C_{133,B}$. For each pair, the Mordell-Weil rank is 2 for one curve and the 2-torsion of the Shafarevich-Tate group has dimension 2 for the other. In addition, the two Jacobians, when canonically embedded into $J_0(N)$, intersect in their 2-torsion subgroups, and one can check that their 2-Selmer groups become equal under the identification of $H^1(\mathbf{Q}, J_{N,A}[2])$ with $H^1(\mathbf{Q}, J_{N,B}[2])$ induced by the identification of the 2-torsion subgroups. Thus these are examples of the principle of a 'visible part of a Shafarevich-Tate group' as discussed in [CM].

Appendix: Other Hasegawa curves

In Table 5 is data concerning all 142 of Hasegawa's curves in the order presented in his paper. Let us explain the entries. The first column in each set of three columns gives the level, N. The second column gives a classification of the cusp forms spanning the 2-dimensional subspace of $S_2(N)$ corresponding to the Jacobian. When that subspace is irreducible with respect to the action of the Hecke algebra and is spanned by two newforms or two oldforms, we write $2n$ or $2o$, respectively. When that subspace is reducible and is spanned by two oldforms, two newforms or one of each, we write oo, nn and on, respectively. The third column contains the sign of the functional equation at the level M at which the cusp form is a newform. This is the negative of ϵ_M (described in Section 4.1). The order of the two signs in the third column agrees with that of the forms listed in the second column. We

22	*oo*	++	58	*nn*	+−	87	2*o*	++	129	*on*	−−	198	2*o*	+−
23	2*n*	++	60	*oo*	++	88	*on*	+−	130	*on*	−+	204	2*o*	+−
26	*nn*	++	60	2*o*	++	90	*on*	++	132	*oo*	++	205	2*n*	−−
28	*oo*	++	60	2*o*	++	90	*oo*	++	133	2*n*	−−	206	2*o*	−−
29	2*n*	++	62	2*o*	++	90	*oo*	++	134	2*o*	−−	209	2*n*	−−
30	*on*	++	66	*nn*	++	90	*oo*	++	135	*on*	+−	210	*on*	+−
30	*oo*	++	66	2*o*	++	91	*nn*	−−	138	*nn*	+−	213	2*n*	−−
30	*on*	++	66	2*o*	++	93	2*n*	−−	138	*on*	+−	215	*on*	−−
31	2*n*	++	66	*on*	++	98	*oo*	++	140	*oo*	++	221	2*n*	−−
33	*on*	++	67	2*n*	−−	100	*oo*	++	142	*nn*	+−	230	2*o*	−−
35	2*n*	++	68	*oo*	++	102	*on*	+−	143	*on*	+−	255	2*o*	−−
37	*nn*	+−	69	2*o*	++	102	*on*	+−	146	2*o*	−−	266	2*o*	−−
38	*on*	++	70	*on*	++	103	2*n*	−−	147	2*n*	−−	276	2*o*	+−
39	2*n*	++	70	2*o*	++	104	2*o*	++	150	*on*	++	284	2*o*	+−
40	*on*	++	70	2*o*	++	106	*on*	−−	153	*on*	+−	285	*on*	−−
40	*oo*	++	70	2*o*	++	107	2*n*	−−	154	*on*	−−	286	*on*	−−
42	*on*	++	72	*on*	++	110	*on*	++	156	*oo*	++	287	2*n*	−−
42	*oo*	++	72	*oo*	++	111	*oo*	+−	158	*on*	−−	299	2*n*	−−
42	*on*	++	73	2*n*	−−	112	*on*	+−	161	2*n*	−−	330	2*o*	−−
42	*oo*	++	74	*oo*	+−	114	*oo*	+−	165	2*n*	−−	357	2*n*	−−
44	2*o*	++	77	*on*	+−	115	2*n*	−−	166	*on*	−−	380	2*o*	+−
46	2*o*	++	78	*oo*	++	116	2*o*	+−	167	2*n*	−−	390	*on*	−−
48	*on*	++	78	2*o*	++	117	2*o*	++	168	2*o*	++			
48	*oo*	++	80	*oo*	++	120	*oo*	++	170	2*o*	−−			
50	*nn*	++	84	*oo*	++	120	*on*	++	177	2*n*	−−			
52	*oo*	++	84	*oo*	++	121	*on*	+−	180	2*o*	++			
52	*oo*	++	84	*oo*	++	122	*on*	−−	184	*on*	+−			
54	*on*	++	84	*oo*	++	125	2*n*	−−	186	2*o*	−−			
57	*on*	+−	85	2*n*	−−	126	*oo*	++	190	*on*	+−			
57	*on*	+−	87	2*n*	++	126	*on*	++	191	2*n*	−−			

TABLE 5. Spaces of cusp forms associated to Hasegawa's curves

include this information for those who would like to further study these curves. The curves with $N < 200$ classified as $2n$ appeared already in Table 1.

The smallest possible Mordell-Weil ranks corresponding to $++$, $+-$, $-+$ and $--$, predicted by the first Birch and Swinnerton-Dyer conjecture, are 0, 1, 1 and 2 respectively. In all cases, those were, in fact, the Mordell-Weil ranks. This was determined by computing 2-Selmer groups with a computer program based on [Sto2]. Of course, these are cases where the first Birch and Swinnerton-Dyer conjecture is already known to hold. In the cases where the Mordell-Weil rank is positive, the Mordell-Weil group has a subgroup of finite index generated by degree zero divisors supported on rational points with x-coordinates with numerators bounded by 7 (in absolute value) and denominators by 12 with one exception. On the second curve with $N = 138$, the divisor class $[(3+2\sqrt{2}, 80+56\sqrt{2}) + (3-2\sqrt{2}, 80-56\sqrt{2}) - 2\infty]$ generates a subgroup of finite index in the Mordell-Weil group.

References

[AS1] A. Agashé and W.A. Stein, *Some abelian varieties with visible Shafarevich-Tate groups*, preprint, 2000.

[AS2] A. Agashé, and W.A. Stein, *The generalized Manin constant, congruence primes, and the modular degree*, in preparation, 2000.

[BSD] B. Birch and H.P.F. Swinnerton-Dyer, *Notes on elliptic curves. II*, J. reine angew. Math., **218** (1965), 79–108. MR 31 #3419

[BL] S. Bosch and Q. Liu, *Rational points of the group of components of a Néron model*, Manuscripta Math, **98** (1999), 275–293.

[BLR] S. Bosch, W. Lütkebohmert and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR **91i**:14034

[BCDT] C. Breuil, B. Conrad, F. Diamond and R. Taylor *On the modularity of elliptic curves over* **Q***: Wild 3-adic exercises.* `http://abel.math.harvard.edu/HTML/Individuals/Richard_Taylor.html` (2000).

[BGZ] J. Buhler, B.H. Gross and D.B. Zagier, *On the conjecture of Birch and Swinnerton-Dyer for an elliptic curve of rank* 3. Math. Comp., **44** (1985), 473–481. MR **86g**:11037

[Ca] J.W.S. Cassels, *Arithmetic on curves of genus 1. VIII. On conjectures of Birch and Swinnerton-Dyer.*, J. reine angew. Math., **217** (1965), 180–199. MR 31 #3420

[CF] J.W.S. Cassels and E.V. Flynn, *Prolegomena to a middlebrow arithmetic of curves of genus 2*, London Math. Soc., Lecture Note Series 230, Cambridge Univ. Press, Cambridge, 1996. MR **97i**:11071

[Cr1] J.E. Cremona, *Abelian varieties with extra twist, cusp forms, and elliptic curves over imaginary quadratic fields*, J. London Math. Soc. (2), **45** (1992), 404–416. MR **93h**:11056

[Cr2] J.E. Cremona, *Algorithms for modular elliptic curves. 2nd edition*, Cambridge Univ. Press, Cambridge, 1997. MR **93m**:11053

[CM] J.E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, to appear in *Experiment. Math.*

[Ed1] B. Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Progr. Math., 89, Birkhauser Boston, Boston, MA, 1991, pp. 25–39.

[Ed2] B. Edixhoven, *L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein"*, Astérisque, No. 196–197 (1992), 159–170. MR **92k**:11059

[FPS] E.V. Flynn, B. Poonen and E.F. Schaefer, *Cycles of quadratic polynomials and rational points on a genus-two curve*, Duke Math. J., **90** (1997), 435–463. MR **98j**:11048

[FS] E.V. Flynn and N.P. Smart, *Canonical heights on the Jacobians of curves of genus 2 and the infinite descent*, Acta Arith., **79** (1997), 333–352. MR **98f**:11066

[FM] G. Frey and M. Müller, *Arithmetic of modular curves and applications*, in *Algorithmic algebra and number theory*, Ed. Matzat et al., Springer-Verlag, Berlin, 1999, pp. 11–48. MR **00a**:11095

[GZ] B.H. Gross and D.B. Zagier, *Heegner points and derivatives of L-series*, Invent. Math., **84** (1986), 225–320. MR **87j**:11057

[Gr] A. Grothendieck, *Groupes de monodromie en géométrie algébrique*, SGA 7 I, Exposé IX, Lecture Notes in Math. vol. 288, Springer, Berlin–Heidelberg–New York, 1972, pp. 313–523. MR 50 #7134

[Ha] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. 52, Springer-Verlag, New York, 1977. MR 57 #3116

[Hs] Y. Hasegawa, *Table of quotient curves of modular curves $X_0(N)$ with genus 2*, Proc. Japan. Acad., **71** (1995), 235–239. MR **97e**:11071

[Ko] V.A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E,\mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat., **52** (1988), 522–540. MR **89m**:11056

[KL] V.A. Kolyvagin and D.Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Leningrad Math J., **1** (1990), 1229–1253. MR **91c**:11032

[La] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1976. MR 55 #2751

[Le] F. Leprévost, *Jacobiennes de certaines courbes de genre 2: torsion et simplicité*, J. Théor. Nombres Bordeaux, **7** (1995), 283–306. MR **98a**:11078

[Li] Q. Liu, *Conducteur et discriminant minimal de courbes de genre 2*, Compos. Math., **94** (1994), 51–79. MR **96b**:14038

[Ma] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math., **44** (1978), 129–162. MR **80h**:14022

[MS] J.R. Merriman and N.P. Smart, *Curves of genus 2 with good reduction away from 2 with a rational Weierstrass point*, Math. Proc. Cambridge Philos. Soc., **114** (1993), 203–214. MR **94h**:14031

[Mi1] J.S. Milne, *Arithmetic duality theorems*, Academic Press, Boston, 1986. MR **88e**:14028

[Mi2] J.S. Milne, *Jacobian varieties*, in: *Arithmetic geometry*, Ed. G. Cornell, G. and J.H. Silverman, Springer-Verlag, New York, 1986, pp. 167–212. MR **89b**:14029

[NU] Y. Namikawa and K. Ueno, *The complete classification of fibres in pencils of curves of genus two*, Manuscripta Math., **9** (1973), 143–186. MR 51 #5595

[PSc] B. Poonen and E.F. Schaefer, *Explicit descent for Jacobians of cyclic covers of the projective line*, J. reine angew. Math., **488** (1997), 141–188. MR **98k**:11087

[PSt] B. Poonen and M. Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2), **150** (1999), 1109–1149.

[Ri] K. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. math., **100** (1990), 431–476. MR **91g**:11066

[Sc] E.F. Schaefer, *Computing a Selmer group of a Jacobian using functions on the curve*, Math. Ann., **310** (1998), 447-471. MR **99h**:11063

[Sh] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, 1994. MR **95e**:11048

[Si] J.H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Grad. Texts in Math. 151, Springer-Verlag, New York, 1994. MR **96b**:11074

[Ste] W.A. Stein, *Component groups of optimal quotients of Jacobians*, preprint, 2000.

[Sto1] M. Stoll, *Two simple 2-dimensional abelian varieties defined over* $\mathbf{Q}$ *with Mordell-Weil rank at least* 19, C. R. Acad. Sci. Paris, Série I, **321** (1995), 1341–1344. MR **96j**:11084

[Sto2] M. Stoll, *Implementing 2-descent for Jacobians of hyperelliptic curves*, preprint, 2000.

[Sto3] M. Stoll, *On the height constant for curves of genus two*, Acta Arith, **90** (1999), 183–201.

[Sto4] M. Stoll, *On the height constant for curves of genus two, II*, in preparation.

[Ta] J. Tate, *On the conjectures of Birch and Swinneron-Dyer and a geometric analog.* Séminaire Bourbaki, **306** 1965/1966. MR 1 610977

[Wal1] J.-L. Waldspurger, *Correspondances de Shimura*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2, (Warsaw, 1983), 1984, pp. 525–531. MR **86m**:11036

[Wal2] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier*, J. Math. Pures Appl. (9), **60** (1981), 375–484. MR **83h**:10061

[Wan] X. Wang, *2-dimensional simple factors of* $J_0(N)$, Manuscripta Math., **87** (1995), 179–197. MR **96h**:11059

Department of Mathematical Sciences, University of Liverpool, P.O.Box 147, Liverpool L69 3BX, England
E-mail address: evflynn@liverpool.ac.uk

Université Grenoble I, Institut Fourier, BP 74, F-38402 Saint Martin d'Hères Cedex, France
E-mail address: leprevot@math.jussieu.fr

Department of Mathematics and Computer Science, Santa Clara University, Santa Clara, CA 95053, USA
E-mail address: eschaefe@math.scu.edu

Department of Mathematics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA
E-mail address: was@math.berkeley.edu

Mathematisches Institut der Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany
E-mail address: stoll@math.uni-duesseldorf.de

Department of Mathematics, University of Southern California, 1042 W. 36th Place, Los Angeles, CA 90089-1113, USA
E-mail address: jlwether@alum.mit.edu

2 Component Groups of Quotients of $J_0(N)$, with D. Kohel

Component Groups of Quotients of $J_0(N)$

Copyright Springer-Verlag

David Kohel[1] and William A. Stein[2]

[1] University of Sydney
kohel@maths.usyd.edu.au
http://www.maths.usyd.edu.au:8000/u/kohel/
[2] University of California at Berkeley,
was@math.berkeley.edu
http://shimura.math.berkeley.edu/~was

Abstract. Let f be a newform of weight 2 on $\Gamma_0(N)$, and let A_f be the corresponding optimal abelian variety quotient of $J_0(N)$. We describe an algorithm to compute the order of the component group of A_f at primes p that exactly divide N. We give a table of orders of component groups for all f of level $N \leq 127$ and five examples in which the component group is very large, as predicted by the Birch and Swinnerton-Dyer conjecture.

1 Introduction

Let $X_0(N)$ be the Riemann surface obtained by compactifying the quotient of the upper half-plane by the action of $\Gamma_0(N)$. Then $X_0(N)$ has a canonical structure of algebraic curve over $\mathbf{Q}$; denote its Jacobian by $J_0(N)$. It is equipped with an action of a commutative ring $\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots]$ of Hecke operators. For more details on modular curves, Hecke operators, and modular forms see, e.g., [8].

Now suppose that $f = \sum_{n=1}^{\infty} a_n q^n$ is a modular newform of weight 2 for the congruence subgroup $\Gamma_0(N)$. The Hecke operators also act on f by $T_n(f) = a_n f$. The eigenvalues a_n generate an order $R_f = \mathbf{Z}[\ldots a_n \ldots]$ in a number field K_f. The kernel I_f of the map $\mathbf{T} \to R_f$ sending T_n to a_n is a prime ideal. Following Shimura [15], we associate to f the quotient $A_f = J_0(N)/I_f J_0(N)$ of $J_0(N)$. Then A_f is an abelian variety over $\mathbf{Q}$ of dimension $[K_f : \mathbf{Q}]$, with bad reduction exactly at the primes dividing N.

One-dimensional quotients of $J_0(N)$ have been intensely studied in recent years, both computationally and theoretically. The original conjectures of Birch and Swinnerton-Dyer [1, 2], for elliptic curves over $\mathbf{Q}$, were greatly influenced by computations. The scale of these computations was extended and systematized by Cremona in [6].

In another direction, Wiles [20] and Taylor-Wiles [18] proved a special case of the conjecture of Shimura-Taniyama, which asserts that every elliptic curve over $\mathbf{Q}$ is a quotient of some $J_0(N)$; this allowed them to establish Fermat's Last Theorem. The full Shimura-Taniyama conjecture was later proved by Breuil, Conrad, Diamond, and Taylor in [4]. This illustrates the central role played by quotients of $J_0(N)$.

2 Component Groups of A_f

The Néron model $\mathcal{A}/\mathbf{Z}$ of an abelian variety $A/\mathbf{Q}$ is by definition a smooth commutative group scheme over $\mathbf{Z}$ with generic fiber A such that for any smooth scheme S over $\mathbf{Z}$, the restriction map

$$\mathrm{Hom}_{\mathbf{Z}}(S, \mathcal{A}) \to \mathrm{Hom}_{\mathbf{Q}}(S_{\mathbf{Q}}, A)$$

is a bijection. For more details, including a proof of existence, see, e.g., [5].

Suppose that A_f is a quotient of $J_0(N)$ corresponding to a newform f on $\Gamma_0(N)$, and let $\mathcal{A}_f$ be the Néron model of A_f. For any prime divisor p of N, the closed fiber $\mathcal{A}_{f/\mathbf{F}_p}$ is a group scheme over $\mathbf{F}_p$, which need not be connected. Denote the connected component of the identity by $\mathcal{A}^{\circ}_{f/\mathbf{F}_p}$. There is an exact sequence

$$0 \to \mathcal{A}^{\circ}_{f/\mathbf{F}_p} \to \mathcal{A}_{f/\mathbf{F}_p} \to \Phi_{A_f,p} \to 0$$

with $\Phi_{A_f,p}$ a finite étale group scheme over $\mathbf{F}_p$ called the *component group* of A_f at p.

The category of finite étale group schemes over $\mathbf{F}_p$ is equivalent to the category of finite groups equipped with an action of $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ (see, e.g., [19, §6.4]). The *order* of an étale group scheme $G/\mathbf{F}_p$ is defined to be the order of the group $G(\overline{\mathbf{F}}_p)$. In this paper we describe an algorithm for computing the order of $\Phi_{A_f,p}$, when p exactly divides N.

3 The Algorithm

Let $J = J_0(N)$, fix a newform f of weight-two for $\Gamma_0(N)$, and let A_f be the corresponding quotient of J. Because J is the Jacobian of a curve, it is canonically isomorphic to its dual, so the projection $J \to A_f$ induces a polarization $A_f^{\vee} \to A_f$, where $A_f^{\vee}$ denotes the abelian variety dual of A_f. We define the *modular degree* δ_{A_f} of A_f to be the positive square root of the degree of this polarization. This agrees with the usual notion of modular degree when A_f is an elliptic curve.

A *torus* T over a field k is a group scheme whose base extension to the separable closure k_s of k is a finite product of copies of $\mathbf{G}_m$. Every commutative algebraic group over k admits a unique maximal subtorus, defined over k, whose formation commutes with base extension (see IX §2.1 of [9]). The *character group* of a torus T is the group $\mathcal{X} = \mathrm{Hom}_{k_s}(T, \mathbf{G}_m)$ which is a free abelian group of finite rank together with an action of $\mathrm{Gal}(k_s/k)$ (see, e.g., [19, §7.3]).

We apply this construction to our setting as follows. The closed fiber of the Néron model of J at p is a group scheme over $\mathbf{F}_p$, whose maximal torus we denote by $T_{J,p}$. We define $\mathcal{X}_{J,p}$ to be the character group of $T_{J,p}$. Then $\mathcal{X}_{J,p}$ is a free abelian group equipped with an action of both $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and the Hecke algebra $\mathbf{T}$ (see, e.g., [14]). Moreover, there exists a bilinear pairing

$$\langle\, , \,\rangle : \mathcal{X}_{J,p} \times \mathcal{X}_{J,p} \to \mathbf{Z}$$

called the *monodromy pairing* such that

$$\Phi_{J,p} \cong \mathrm{coker}(\mathcal{X}_{J,p} \to \mathrm{Hom}(\mathcal{X}_{J,p}, \mathbf{Z})).$$

Let $\mathcal{X}_{J,p}[I_f]$ be the intersection of all kernels $\ker(t)$ for t in I_f, and let

$$\alpha_f : \mathcal{X}_{J,p} \to \mathrm{Hom}(\mathcal{X}_{J,p}[I_f], \mathbf{Z})$$

be the map induced by the monodromy pairing. The following theorem of the second author [16], provides the basis for the computation of orders of component groups.

Theorem 1. *With the notation as above, we have the equality*

$$\#\Phi_{A_f,p} = \frac{\#\mathrm{coker}(\alpha_f) \cdot \delta_{A_f}}{\#(\alpha_f(\mathcal{X}_{J,p})/\alpha_f(\mathcal{X}_{J,p}[I_f]))}\,.$$

3.1 Computing the modular degree $\delta_{A,f}$

Using modular symbols (see, e.g., [6]), we first compute the homology group $H_1(X_0(N), \mathbf{Q}; \mathrm{cusps})$. Using lattice reduction, we compute the $\mathbf{Z}$-submodule $H_1(X_0(N), \mathbf{Z}; \mathrm{cusps})$ generated by all Manin symbols (c, d). Then $H_1(X_0(N), \mathbf{Z})$ is the *integer* kernel of the boundary map.

The Hecke ring $\mathbf{T}$ acts on $H_1(X_0(N), \mathbf{Z})$ and also on the linear dual $\mathrm{Hom}(H_1(X_0(N), \mathbf{Z}), \mathbf{Z})$, where $t \in \mathbf{T}$ acts on $\varphi \in \mathrm{Hom}(H_1(X_0(N), \mathbf{Z}), \mathbf{Z})$ by $(t.\varphi)(x) = \varphi(tx)$. We have a natural restriction map

$$r_f : \mathrm{Hom}(H_1(X_0(N), \mathbf{Z}), \mathbf{Z})[I_f] \to \mathrm{Hom}(H_1(X_0(N), \mathbf{Z})[I_f], \mathbf{Z}).$$

Proposition 1. *The cokernel of r_f is isomorphic to the kernel of the polarization $A_f^\vee \to A_f$ induced by the map $J_0(N) \to A_f$.*

Thus the order of the cokernel of r_f is the square of the modular degree δ_f. We pause to note that the degree of any polarization is a square; see, e.g., [13, Thm. 13.3].

Proof. Let $S = S_2(\mathit{\Gamma}_0(N), \mathbf{C})$ be the complex vector space of weight-two modular forms of level N, and set $H = H_1(X_0(N), Z)$. The integration pairing $S \times H \to \mathbf{C}$ induces a natural map

$$\mathit{\Phi}_f : H \to \mathrm{Hom}(S[I_f], \mathbf{C}).$$

Using the classical Abel-Jacobi theorem, we deduce the following commutative diagram, which has exact columns, but whose rows are not exact.

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] \\
H[I_f] \arrow[r] \arrow[d] & H \arrow[r] \arrow[d] & \mathit{\Phi}_f(H) \arrow[d] \\
\mathrm{Hom}(S, \mathbf{C})[I_f] \arrow[r] \arrow[d] & \mathrm{Hom}(S, \mathbf{C}) \arrow[r] \arrow[d] & \mathrm{Hom}(S[I_f], \mathbf{C}) \arrow[d] \\
A_f^{\vee}(\mathbf{C}) \arrow[r] \arrow[d] \arrow[rr, bend right] & J_0(N)(\mathbf{C}) \arrow[r] \arrow[d] & A_f(\mathbf{C}) \arrow[d] \\
0 & 0 & 0
\end{tikzcd}$$

By the snake lemma, the kernel of $A_f^{\vee}(\mathbf{C}) \to A_f(\mathbf{C})$ is isomorphic to the cokernel of the map $H[I_f] \to \mathit{\Phi}_f(H)$. Since

$$\mathrm{Hom}(H/\ker(\mathit{\Phi}_f), \mathbf{Z}) \cong \mathrm{Hom}(H, \mathbf{Z})[I_f],$$

the $\mathrm{Hom}(-, \mathbf{Z})$ dual of the map $H[I_f] \to \mathit{\Phi}_f(H) = H/\ker(\mathit{\Phi}_f)$ is r_f, which proves the proposition.

3.2 Computing the character group $\boldsymbol{\mathcal{X}_{J,p}}$

Let $N = Mp$, where M and p are coprime. If M is small, then the algorithm of Mestre and Oesterlé [12] can be used to compute $\mathcal{X}_{J,p}$. This algorithm constructs the graph of isogenies between $\overline{\mathbf{F}}_p$-isomorphism classes of pairs consisting of a supersingular elliptic curve and a cyclic M-torsion subgroup. In particular, the method is elementary to apply when $X_0(M)$ has genus 0.

In general, the above category of "enhanced" supersingular elliptic curves can be replaced by one of left (or right) ideals of a quaternion order $\mathcal{O}$ of level M in the quaternion algebra over $\mathbf{Q}$ ramified at p. This gives an equivalent category, in which the computation of homomorphisms is efficient. The character group $\mathcal{X}_{J,p}$ is known by Deligne-Rapoport [7] to be canonically isomorphic to the degree zero subgroup $\mathcal{X}(\mathcal{O})$ of the free abelian "divisor group" on the isomorphism classes of enhanced supersingular elliptic curves and of quaternion ideals. Moreover, this isomorphism is compatible with the operation of Hecke operators, which are effectively computable in $\mathcal{X}(\mathcal{O})$ in terms of ideal homomorphisms.

The inner product of two classes in this setting is defined to be the number of isomorphisms between any two representatives. The linear extension to $\mathcal{X}(\mathcal{O})$ gives an inner product which agrees, under the isomorphism, with the monodromy pairing on $\mathcal{X}_{J,p}$. This gives, in particular, an isomorphism $\Phi_{J,p} \cong \mathrm{coker}(\mathcal{X}(\mathcal{O}) \to \mathrm{Hom}(\mathcal{X}(\mathcal{O}), \mathbf{Z}))$, and an effective means of computing $\#\mathrm{coker}(\alpha_f)$ and $\#(\alpha_f(\mathcal{X}_{J,p})/\alpha_f(\mathcal{X}_{J,p}[I_f]))$.

The arithmetic of quaternions has been implemented in Magma [11] by the first author. Additional details and the application to Shimura curves, generalizing $X_0(N)$, can be found in Kohel [10].

3.3 The Galois action on $\Phi_{A_f,p}$

To determine the Galois action on $\Phi_{A_f,p}$, we need only know the action of the Frobenius automorphism Frob_p. However, Frob_p acts on $\Phi_{A_f,p}$ in the same way as $-W_p$, where W_p is the pth Atkin-Lehner involution, which can be computed using modular symbols. Since f is an eigenform, the involution W_p acts as either $+1$ or -1 on $\Phi_{A_f,p}$. Moreover, the operator W_p is determined by an involution on the set of quaternion ideals, so it can be determined explicitly on the character group.

4 Tables

The main computational results of this work are presented below in two tables. The relevant algorithms have been implemented in Magma and will be made part of a future release. They can also be obtained from the second author.

4.1 Component groups at low level

The first table gives the component groups of the quotients A_f of $J_0(N)$ for $N \leq 127$. The column labeled d contains the dimensions of the A_f, and the column labeled $\#\Phi_{A_f,p}$ contains a list of the orders of the component groups of A_f, one for each divisor p of N, ordered by increasing p. An entry of "?" indicates that $p^2 \mid N$, so our algorithm does not apply. A component group order is starred if the $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$-action is nontrivial. More data along these lines can be obtained from the second author.

4.2 Examples of large component groups

Let Ω_{A_f} be the real period of A_f, as defined by J. Tate in [17]. The second author computed the rational numbers $L(A_f, 1)/\Omega_{A_f}$ for every newform f of level $N \leq 1500$. The five largest prime divisors occur in the ratios given in the second table. The Birch and Swinnerton-Dyer conjecture predicts that the large prime divisor of the numerator of each special value must divide the order either of some component group $\Phi_{A_f,p}$ or of the Shafarevich-Tate group of A_f. In each instance $\Phi_{A_f,2}$ is divisible by the large prime divisor, as predicted.

5 Further directions

Further considerations are needed to compute the *group* structure of $\Phi_{A_f,p}$. However, since the action of Frobenius is known, computing the group structure of $\Phi_{A_f,p}$ suffices to determine its structure as a group scheme.

Our methods say nothing about the component group at primes whose *square* divides the level. The free abelian group on classes of nonmaximal orders of index p at a ramified prime gives a well-defined divisor group. Do the resulting Hecke modules determine the component groups for quotients of level p^2M?

Is it possible to define quantities as in Theorem 1 even when the weight of f is *greater than* 2? If so, how are the resulting quantities related to the Bloch-Kato Tamagawa numbers (see [3]) of the higher weight motive attached to f?

Component groups at low level

N	d	$\#\Phi_{A_f,p}$	N	d	$\#\Phi_{A_f,p}$	N	d	$\#\Phi_{A_f,p}$	N	d	$\#\Phi_{A_f,p}$	N	d	$\#\Phi_{A_f,p}$
11	1	5		3	13	76	1	$?,1^*$	96	1	$?,2$		3	7
14	1	$6^*,3$	54	1	$3^*,?$	77	1	$2^*,1^*$		1	$?,2^*$	114	1	$2^*,5^*,1$
15	1	$4^*,4$		1	$3,?$		1	$3^*,2$	97	3	1^*		1	$20,3^*,1^*$
17	1	4	55	1	$2,2^*$		1	$6,3^*$		4	8		1	$6,3,1$
19	1	3		2	$14^*,2$		2	$2,2^*$	98	1	$2^*,?$	115	1	$5^*,1$
20	1	$?,2^*$	56	1	$?,1$	78	1	$16^*,5^*,1$		2	$14,?$		2	$4^*,1^*$
21	1	$4,2^*$		1	$?,1^*$	79	1	1^*	99	1	$?,1^*$		4	$32,4^*$
23	2	11	57	1	$2^*,1^*$		5	13		1	$?,1$	116	1	$?,1^*$
24	1	$?,2^*$		1	$2,2^*$	80	1	$?,2$		1	$?,1^*$		1	$?,2^*$
26	1	$3^*,3$		1	$10,1^*$		1	$?,2^*$		1	$?,1^*$		1	$?,1^*$
	1	$7,1^*$	58	1	$2^*,1^*$	81	2	?	100	1	$?,?$	117	1	$?,1$
27	1	?		1	$10,1^*$	82	1	$2^*,1^*$	101	1	1^*		2	$?,3$
29	2	7	59	5	29		2	$28,1^*$		7	25		2	$?,1^*$
30	1	$4^*,3,1^*$	61	1	1^*	83	1	1^*	102	1	$2^*,2^*,1^*$	118	1	$2^*,1^*$
31	2	5		3	5		6	41		1	$6^*,6,1^*$		1	$19^*,1$
32	1	?	62	1	$4,1^*$	84	1	$?,1^*,2^*$		1	$8,4,1$		1	$10,1^*$
33	1	$6^*,2$		2	$66^*,3$		1	$?,3,2$	103	2	1^*		1	$1,1^*$
34	1	$6,1^*$	63	1	$?,1^*$	85	1	$2^*,1$		6	17	119	4	$9,3^*$
35	1	$3^*,3$		2	$?,3$		2	$2^*,1^*$	104	1	$?,1^*$		5	$48^*,8$
	2	$8,4^*$	64	1	?		2	$6,1^*$		2	$?,2$	120	1	$?,1,1^*$
36	1	$?,?$	65	1	$1^*,1^*$	86	2	$21^*,3$	105	1	$1,1,1$		1	$?,2,1$
37	1	1^*		2	$3^*,3$		2	$55,1^*$		2	$10^*,2^*,2$	121	1	?
	1	3		2	$7,1^*$	87	2	$5,1^*$	106	1	$4^*,1^*$		1	?
38	1	$9^*,3$	66	1	$2^*,3,1^*$		3	$92^*,4$		1	$5^*,1$		1	?
	1	$5,1^*$		1	$4,1^*,1^*$	88	1	$?,1^*$		1	$24,1^*$		1	?
39	1	$2^*,2$		1	$10,5,1$		2	$?,2^*$		1	$3,1^*$	122	1	$4^*,1^*$
	2	$14,2^*$	67	1	1	89	1	1^*	107	2	1^*		2	$39^*,3$
40	1	$?,2$		2	1^*		1	2		7	53		3	$248,1^*$
41	3	10		2	11		5	11	108	1	$?,?$	123	1	$1^*,1^*$
42	1	$8,2^*,1^*$	68	2	$?,2^*$	90	1	$2^*,?,3$	109	1	1		1	$5,1$
43	1	1^*	69	1	$2,1^*$		1	$6,?,1^*$		3	1^*		2	$7,1^*$
	2	7		2	$22^*,2$		1	$4,?,1$		4	9		3	$184^*,4$
44	1	$?,1^*$	70	1	$4,2^*,1^*$	91	1	$1^*,1^*$	110	1	$7^*,1^*,3$	124	1	$?,1^*$
45	1	$?,1^*$	71	3	5		1	$1,1$		1	$3,1^*,1^*$		1	$?,1$
46	1	$10^*,1$		3	7		2	$7,1^*$		1	$5,5,1$	125	2	?
47	4	23	72	1	$?,?$		3	$4^*,8$		2	$16^*,3,1^*$		2	?
48	1	$?,2$	73	1	2	92	1	$?,1^*$	111	3	$10^*,2$		4	?
49	1	?		2	1^*		1	$?,1$		4	$266,2^*$	126	1	$8^*,?,1^*$
50	1	$1^*,?$		2	3	93	2	$4^*,1^*$	112	1	$?,1^*$		1	$2,?,1$
	1	$5,?$	74	2	$9^*,3$		3	$64,2^*$		1	$?,1$	127	3	1^*
51	1	$3,1^*$		2	$95,1^*$	94	1	$2,1^*$		1	$?,1^*$		7	21
	2	$16^*,4$	75	1	$1^*,?$		2	$94^*,1$	113	1	2			
52	1	$?,2^*$		1	$1,?$	95	3	$10,2^*$		2	2			
53	1	1^*		1	$5,?$		4	$54^*,6$		3	1^*			

Large $L(A_f,1)/\Omega_{A_f}$

N	dim	$L(A_f,1)/\Omega_{A_f}$	$\#\Phi_{A_f,p}$
$1154 = 2{\cdot}577$	20	$2^{?}{\cdot}85495047371/17^2$	$2^{?}{\cdot}17^2{\cdot}85495047371, 2^{?}$
$1238 = 2{\cdot}619$	19	$2^{?}{\cdot}7553329019/5{\cdot}31$	$2^{?}{\cdot}5{\cdot}31{\cdot}7553329019, 2^{?}$
$1322 = 2{\cdot}661$	21	$2^{?}{\cdot}57851840099/331$	$2^{?}{\cdot}331{\cdot}57851840099, 2^{?}$
$1382 = 2{\cdot}691$	20	$2^{?}{\cdot}37{\cdot}1864449649/173$	$2^{?}{\cdot}37{\cdot}173{\cdot}1864449649, 2^{?}$
$1478 = 2{\cdot}739$	20	$2^{?}{\cdot}7{\cdot}29{\cdot}1183045463/5{\cdot}37$	$2^{?}{\cdot}5{\cdot}7{\cdot}29{\cdot}37{\cdot}1183045463, 2^{?}$

References

1. B. J. Birch, and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves, I*, J. Reine Angew. Math. **212** (1963), 7–25.
2. B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves, II*, J. Reine Angew. Math. **218** (1965), 79–108.
3. S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, The Grothendieck Festschrift, Vol. I, Birkhäuser Boston, Boston, MA, 1990, 333–400.
4. C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$, in preparation.
5. S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.
6. J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.
7. P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, In P. Deligne and W. Kuyk, eds., *Modular functions of one variable, Vol. II*, Lecture Notes in Math., **349**, Springer, Berlin, 1973, 143–316.
8. F. Diamond and J. Im, *Modular forms and modular curves*, In V. K. Murty, ed., *Seminar on Fermat's Last Theorem*, Amer. Math. Soc., Providence, RI, 1995, 39–133.
9. A. Grothendieck, *Séminaire de géométrie algébrique du Bois-Marie 1967–1969 (SGA 7 I)*, Lecture Notes in Mathematics, **288**, Springer-Verlag, Berlin-New York, 1972
10. D. Kohel, *Hecke module structure of quaternions*, In K. Miyake, ed., *Class Field Theory – Its Centenary and Prospect*, The Advanced Studies in Pure Mathematics Series, Math Soc. Japan, to appear.
11. W. Bosma, J. Cannon, and C. Playoust. *The Magma algebra system I: The user language*, J. Symb. Comp., **24** (1997), no. 3-4, 235–265.
12. J.-F. Mestre, *La méthode des graphes. Exemples et applications*, In *Proceedings of the international conference on class numbers and fundamental units of algebraic number fields*, Nagoya University, Nagoya, 1986, 217–242.
13. J. S. Milne, *Abelian Varieties*, In G. Cornell and J. Silverman, eds., *Arithmetic geometry*, Springer, New York, 1986, 103–150,
14. K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), no. 2, 431–476.
15. G. Shimura, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544.
16. W. A. Stein, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley, 2000.
17. J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1995, Exp. No. 306, 415–440.

18. R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. **141** (1995), no. 3, 553–572.
19. W. C. Waterhouse, *Introduction to affine group schemes*, Graduate Texts in Mathematics, **66**, Springer-Verlag, New York-Berlin, 1979
20. A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. **141** (1995), no. 3, 443–551.

3 Explicit approaches to modular abelian varieties (Ph.D. Thesis)

Explicit approaches to modular abelian varieties

by

William Arthur Stein

B.S. (Northern Arizona University) 1994

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mathematics

in the

Graduate Division
of the
University of California at Berkeley

Committee in charge:

Professor Hendrik Lenstra, Chair
Professor Bjorn Poonen
Professor Bin Yu

Spring 2000

The dissertation of William Arthur Stein is approved:

Chair　　　　　　　　　　　　　　　　　　　　　　　　Date

Date

Date

University of California at Berkeley

Spring 2000

Explicit approaches to modular abelian varieties

Copyright 2000
by
William Arthur Stein

Abstract

Explicit approaches to modular abelian varieties

by

William Arthur Stein

Doctor of Philosophy in Mathematics

University of California at Berkeley

Professor Hendrik Lenstra, Chair

I investigate the Birch and Swinnerton-Dyer conjecture, which ties together the constellation of invariants attached to an abelian variety. I attempt to verify this conjecture for certain specific modular abelian varieties of dimension greater than one. The key idea is to use Barry Mazur's notion of visibility, coupled with explicit computations, to produce lower bounds on the Shafarevich-Tate group. I have not finished the proof of the conjecture in these examples; this would require computing explicit upper bounds on the order of this group.

I next describe how to compute in spaces of modular forms of weight at least two. I give an integrated package for computing, in many cases, the following invariants of a modular abelian variety: the modular degree, the rational part of the special value of the L-function, the order of the component group at primes of multiplicative reduction, the period lattice, upper and lower bounds on the torsion subgroup, and the real volume. Taken together, these algorithms are frequently sufficient to compute the odd part of the conjectural order of the Shafarevich-Tate group of an analytic rank 0 optimal quotient of $J_0(N)$, with N square-free. I have not determined the exact structure of the component group, the order of the component group at primes whose square divides the level, or the exact order of the torsion subgroup in all cases. However, I do provide generalizations of some of the above algorithms to higher weight forms with nontrivial character.

Professor Hendrik Lenstra
Dissertation Committee Chair

To my parents and my grandmother, Annette Maurer.

Contents

List of Figures

List of Tables

List of Symbols

Symbol	**Definition**	**Page**
$A^{\vee}$	dual to A	33
$\boldsymbol{\mathcal{B}}_k(N,\varepsilon)$	module of boundary modular symbols	21
c_A	Manin constant of A	57
m_A	modular degree	50
$\mathbf{e}_i$	ith winding element $X^{i-1}Y^{k-2-(i-1)}\{0,\infty\}$	52
$\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$	module of modular symbols	20
$\overline{\boldsymbol{\mathcal{M}}}_k(N,\varepsilon)$	module of extended modular symbols	59
$M[I]$	$\cap_{a\in I}\ker(a)$	
$P(X,Y)\{\alpha,\beta\}$	higher weight modular symbol	20
$[P(X,Y),(u,v)]$	higher weight Manin symbol	27
$\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$	module of cuspidal modular symbols	21
T_n	nth Hecke operator	29
V_k	module of homogeneous polynomials of degree k	20
W_d	dth Atkin-Lehner involution	23
α_t, β_t	degeneracy maps	24
Θ_f	rational period mapping	47
σ, τ	$\sigma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $\tau = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$	27
Φ_f	analytic period mapping	59
$\Phi_{A,p}$	component group of A at p	71
Ω_A	real volume	66
$\langle\ ,\ \rangle$	integration pairing	22
$*$	star involution	23

Acknowledgements

It is a pleasure to thank my thesis adviser, Hendrik Lenstra, for patiently guiding my intellectual development in innumerable ways. In addition, Robert Coleman, Barry Mazur, and Ken Ribet generously shared their countless ideas and unbounded enthusiasm with me. Robert helped me to understand p-adics, Barry taught me to visualize Shafarevich-Tate groups, and Ken explained congruences, component groups, and tutored me in the elusive art of mathematical exposition.

This thesis would not have been possible without the questions that arose out of exhilarating collaborations with Amod Agashe, Kevin Buzzard, Loïc Merel, David Moulton, Ed Schaefer, Helena Verrill, and Joe Wetherell. I would like to thank Ahmed Abbes, Matt Baker, Brian Conrad, János Csirik, Jordan Ellenberg, Edray Goins, David Jones, Ezra Miller, Bjorn Poonen, and Mak Trifkovic for asking and answering many stimulating questions.

I wish to thank Hendrik Lenstra and Ken Ribet for providing me with financial support during part of the writing of this thesis, Sarah M. Hallam for providing further support, the `Cal@SiliconValley` university fellowship for supporting me throughout my final year, and the MAGMA group for their computational support.

Finally, I would like to thank Bjorn Poonen for thoroughly reading this thesis and pointing out many ways in which it could be improved.

Preface

> The object of numerical computation is theoretical advance.
> *–A. O. L. Atkin, see [5]*

The definition of the spaces of modular forms as functions on the upper half plane satisfying a certain equation is very abstract. The definition of the Hecke operators even more so. Nevertheless, one wishes to carry out explicit investigations into these objects.

We are fortunate that we now have methods available that allow us to transform the vector space of cusp forms of given weight and level into a concrete object, which can be explicitly computed. We have the work of Atkin-Lehner, Birch, Swinnerton-Dyer, Manin, Merel, and many others to thank for this (see [6, 16, 45]). For example, the Eichler-Selberg trace formula, as extended in [30], can be used to compute characteristic polynomials of Hecke operators. One can compute Hecke operators using Brandt matrices and quaternion algebras [32, 52]; another closely related method involves the module of enhanced supersingular elliptic curves [47]. In the course of computing large tables of invariants of elliptic curves in [16], Cremona demonstrated the power of systematic computation using modular symbols.

Various methods often must be used in concert to obtain information about the package of invariants attached to a modular form. For example, computing orders of component groups of optimal quotients of $J_0(N)$ involves computations on the module of supersingular elliptic curves combined with modular symbols techniques (see Chapter 4).

Chapter 1 is an attempt to systematically prove the Birch and Swinnerton-Dyer conjecture for a certain finite list of rank-0 quotients of $J_0(N)$ that have nontrivial Shafarevich-Tate groups. The key idea is to use Barry Mazur's notion of visibility, coupled with explicit computations, to produce lower bounds on the Shafarevich-Tate group. I have not finished the proof of the conjecture in these examples; this would require computing explicit upper bounds on the order of this group. However, I obtain explicit formulas and data that will be helpful in further investigations.

The following three chapters describe the algorithms used in Chapter 1, along with generalizations to eigenforms on $\Gamma_1(N)$ of integral weight greater than two. I have used these algorithms to investigate the Artin Conjecture [12], Serre's conjecture, and many other problems not described in this thesis. I have implemented most of the algorithms that are described in Chapters 2–4 in both MAGMA and C++; this implementation should be available in the standard release of MAGMA in versions 2.7 and greater.

William A. Stein
University of California, Berkeley

Chapter 1

The Birch and Swinnerton-Dyer conjecture

Now that the Shimura-Taniyama conjecture has been proved, many experts consider the Birch and Swinnerton-Dyer conjecture (BSD conjecture) to be one of the main outstanding problems in the field (see [19, pg. 549] and [68, Intro.]). This conjecture ties together many of the arithmetic and analytic invariants of an elliptic curve. At present, there is no general class of elliptic curves for which the full BSD conjecture is known, though a slightly weakened form is known for a fairly broad class of complex multiplication elliptic curves of analytic rank 0 (see [55]), and several deep partial results have been obtained during the last twenty years (see, e.g., [27] and [33]).

Approaches to the BSD conjecture that rely on congruences between modular forms are likely to require a deeper understanding of the analogue of the BSD conjecture for higher-dimensional abelian varieties. As a first step, this chapter presents theorems and explicit computations of some of the arithmetic invariants of modular abelian varieties.

The reader is urged to also read A. Agashe's 2000 Berkeley Ph.D. thesis which cover similar themes. The paper of Cremona and Mazur's [18] paints a detailed experimental picture of the way in which congruences link Mordell-Weil and Shafarevich-Tate groups of elliptic curves.

1.1 The BSD conjecture

By [10] we now know that every elliptic curve over $\mathbf{Q}$ is a quotient of the curve $X_0(N)$, whose complex points are the isomorphism classes of pairs consisting of a (generalized) elliptic curve and a cyclic subgroup of order N. Let $J_0(N)$ denote the Jacobian of $X_0(N)$; this is an abelian variety of dimension equal to the genus of $X_0(N)$ whose points correspond to the degree 0 divisor classes on $X_0(N)$. The survey article [21] is a good guide to the facts and literature about the family of abelian varieties $J_0(N)$.

Following Mazur [41], we make the following definition.

Definition 1.1 (Optimal quotient). An *optimal quotient* of $J_0(N)$ is a quotient A of $J_0(N)$ by an abelian subvariety.

Consider an optimal quotient A such that $L(A,1) \neq 0$. By [35], $A(\mathbf{Q})$ and $\text{Ш}(A)$ are both finite. The BSD conjectureasserts that

$$\frac{L(A,1)}{\Omega_A} = \frac{\#\text{Ш}(A) \cdot \prod_{p|N} c_p}{\#A(\mathbf{Q}) \cdot \#A^\vee(\mathbf{Q})}.$$

Here the Shafarevich-Tate group

$$\text{Ш}(A) := \ker\left(H^1(\mathbf{Q}, A) \to \prod_v H^1(\mathbf{Q}_v, A)\right)$$

is a measure of the failure of the local-to-global principle; the Tamagawa numbers c_p are the orders of the groups of rational points of the component groups of A (see Chapter 4); the real number Ω_A is the measure of $A(\mathbf{R})$ with respect to a basis of differentials having everywhere nonzero good reduction (see Section 3.12.6); and $A^\vee$ is the abelian variety dual to A (see [50, §9]). This chapter makes a small contribution to the long-term goal of verifying the above conjecture for many specific abelian varieties on a case-by-case basis. In a large list of examples, we compute the conjectured order of $\text{Ш}(A)$, up to a power of 2, and then show that $\text{Ш}(A)$ is at least as big as conjectured. We also discuss methods to obtain upper bounds on $\#\text{Ш}(A)$, but do not carry out any computations in this direction. This is the first step in a program to verify the above conjecture for an infinite family of quotients of $J_0(N)$.

1.1.1 The ratio $L(A,1)/\Omega_A$

Extending classical work on elliptic curves, A. Agashe and the author proved the following theorem.

Theorem 1.2. *Let m be the largest square dividing N. The ratio $L(A,1)/\Omega_A$ is a rational number that can be explicitly computed, up to a unit (conjecturally 1) in $\mathbf{Z}[1/(2m)]$.*

Proof. The proof uses modular symbols combined with an extension of the argument used by Mazur in [41] to bound the Manin constant. The modular symbols part of the proof for L-functions attached to newforms of weight $k \geq 2$ is given in Section 3.10; it involves expressing the ratio $L(A,1)/\Omega_A$ as the lattice index of two modules over the Hecke algebra. The bound on the Manin constant is given in Section 3.11. □

The author has computed $L(A,1)/\Omega_A$ for all simple optimal quotients of level $N \leq 1500$; this table can be obtained from the author's web page.

Remark 1.3. The method of proof should also give similar results for special values of twists of $L(A,s)$, just as it does in the case $\dim A = 1$ (see [16, Prop. 2.11.2]).

1.1.2 Torsion subgroup

We can compute upper and lower bounds on $\#A(\mathbf{Q})_{\text{tor}}$, see Section 3.8; these frequently determine $\#A(\mathbf{Q})_{\text{tor}}$.

These methods, combined with the method used to obtain Theorem 1.2, yield the following corollary, which supports the expected cancellation between torsion and c_p coming from the reduction map sending rational points to their image in the component group of A. The corollary also generalizes to higher weight forms, thus suggesting a geometric way to think about reducibility of modular Galois representations.

Corollary 1.4. *Let n be the order of the image of $(0)-(\infty)$ in $A(\mathbf{Q})$, and let m be the largest square dividing N. Then $n \cdot L(A,1)/\Omega_A \in \mathbf{Z}[1/(2m)]$.*

For the proof, see Corollary 3.48 in Chapter 3.

1.1.3 Tamagawa numbers

We prove the following theorem in Chapter 4.

Theorem 1.5. *When $p^2 \nmid N$, the number c_p can be explicitly computed (up to a power of 2).*

We can compute the order c_p of the group of rational points of the component group, but not its structure as a group. When $p^2 \mid N$ it may be possible to compute c_p using the Drinfeld-Katz-Mazur model of $X_0(N)$, but we have not yet done this. There are also good bounds on the primes that can divide c_p when $p^2 \mid N$.

Systematic computations (see Section 4.7.1) using this formula suggest the following conjectural refinement of a result of Mazur [40].

Conjecture 1.6. *Suppose N is prime and A is an optimal quotient of $J_0(N)$ corresponding to a newform f. Then $A(\mathbf{Q})_{\mathrm{tor}}$ is generated by the image of $(0)-(\infty)$ and $c_p = \#A(\mathbf{Q})_{\mathrm{tor}}$. Furthermore, the product of the c_p over all simple optimal quotients corresponding to newforms equals the numerator of $(N-1)/12$.*

I have checked this conjecture for all $N \leq 997$ and, up to a power of 2, for all $N \leq 2113$. The first part is known when A is an elliptic curve (see [48]). Upon hearing of this conjecture, Mazur reportedly proved it when all "q-Eisenstein quotients" are simple. There are three promising approaches to finding a complete proof. One involves the explicit formula of Theorem 1.5; another is based on Ribet's level lowering theorem (see [53]), and a third makes use of a simplicity result of Merel (see [46]).

The formula that lies behind Theorem 1.5 probably has a natural analogue in weight greater than 2. One could then guess that it produces Tamagawa numbers of motifs attached to eigenforms of higher weight; however, we have no idea if this is really the case. These numbers appear in the conjectures of Bloch and Kato, which generalize the BSD conjecture to motifs (see [7]). Anyone wishing to try to compute them should be aware of Neil Dummigan's paper [22], which gives some information about the Tamagawa numbers of motifs attached by Scholl in [57] to modular eigenforms.

1.1.4 Upper bounds on $\#Ш(A)$

V. Kolyvagin (see [34]) and K. Kato (see, e.g., [58]) constructed Euler systems that were used to prove that $Ш(A)$ is *finite* when $L(A,1) \neq 0$. To verify the full BSD conjecture for certain abelian varieties, we must make the Kolyvagin-Kato finiteness bound explicit. Kolyvagin's bounds involve computations with Heegner points, and Kato's involve a study of the Galois representations associated to A.

Kolyvagin's bounds

In [33], Kolyvagin obtains explicit upper bounds for $\#\text{Ш}(A)$ for a certain (finite) list of elliptic curves A by computing the index in $A(K)$ of the subgroup generated by the Heegner point, where K is a suitable imaginary quadratic extension. In [35], Kolyvagin and Logachev generalize Kolyvagin's earlier results; in Section 1.6, "Unsolved problems", they say that: "If one were to compute the height of a Heegner point y [...] considered in the present paper, then one would have succeeded in obtaining an upper bound for $\#\text{Ш}$ for this curve." (By "curve" they mean abelian variety.) This suggests that explicit computations should yield upper bounds on the order of $\text{Ш}(A)$, but that they had not yet figured out how to carry out such computations.

Kato's bounds

Kato has constructed Euler systems coming from K_2-groups of modular curves. These can be used to prove the following theorem (see, e.g., [56, Cor. 3.5.19]).

Theorem 1.7 (Kato). *Suppose E is an elliptic curve over $\mathbf{Q}$ without complex multiplication that E has conductor N, that E has good reduction at p, that p does not divide $2r_E \prod_{q|N} L_q(q^{-1})\#E(\mathbf{Q}_q)_{\text{tor}}$, and the Galois representation $\rho_{E,p} : G_{\mathbf{Q}} \to \text{Aut}(E[p])$ is surjective. Then*

$$\#\text{Ш}(E)_{p^\infty} \textit{ divides } \frac{L(E,1)}{\Omega_E}.$$

Here $L_q(x)$ is the local Euler factor at q and the constant r_E arises in the construction of Kato's Euler system. Rubin suggests that computing r_E is not very difficult (private communication). Appropriate variants of Kato's arguments give similar results for quotients of $J_0(N)$ of arbitrary dimension, though these have not been written down.

1.1.5 Lower bounds on $\#\text{Ш}(A)$

One approach to showing that $\text{Ш}(A)$ is as *at least* as large as predicted by the BSD conjecture is suggested by Mazur's notion of the visible part $\text{Ш}(A)^\circ$ of $\text{Ш}(A)$ (see [18, 43]). Let $A^\vee \subset J_0(N)$ be the dual to A. The *visible part* of $\text{Ш}(A^\vee)$ is the kernel of the natural map $\text{Ш}(A^\vee) \to \text{Ш}(J_0(N))$. Mazur observed that if an element of order p in $\text{Ш}(A^\vee)$ is visible, then it is explained by a "jump in the rank of Mordell-Weil" in the sense that there is another abelian subvariety $B \subset J_0(N)$ such that $p \mid \#(A^\vee \cap B)$ and the rank of B is positive.

Mazur's observation can be turned around: if there is another abelian variety B of positive rank such that $p \mid \#(A^\vee \cap B)$, then, under mild hypotheses (see Theorem 1.8), there is an element of $\text{Ш}(A^\vee)$ of order p. From a computational point of view it is easy to understand the intersections $A^\vee \cap B$; see Section 3.6. From a theoretical point of view, nontrivial intersections "correspond" to congruences between modular forms. Thus the well-developed theory of congruences between modular forms can be used to obtain a lower bound on $\#Ш(A^\vee)$.

Invisible elements of $\#Ш(A^\vee)$

Numerical experiments suggest that as $A^\vee$ varies, $Ш(A^\vee)$ is often *not* visible inside of $J_0(N)$. For example (see Table 1.2), the BSD conjecture predicts the existence of invisible elements of odd order in $Ш(A^\vee)$ for almost half of the 37 optimal quotients of prime level ≤ 2113.

Visibility at higher level

For every integer M (Ribet [54] tells us which M to choose), we can ask whether $Ш(A^\vee)$ maps to 0 under one of the natural maps $A^\vee \to J_0(NM)$; that is, we can ask whether $Ш(A^\vee)$ "becomes visible at level NM." We have been unable to prove in any particular case that $Ш(A^\vee)$ is not visible at level N, but becomes visible at some level NM. See Section 1.4.1 for some computations which strongly indicate that such examples exist.

Visibility in some Jacobian

Johan de Jong proved that if E is an elliptic curve over a number field K and $c \in H^1(K, E)$ then there is a Jacobian J and an imbedding $E \hookrightarrow J$ such that c maps to 0 under the natural map $H^1(K, E) \to H^1(K, J)$ (see Remark 3 in [18]); de Jong's proof appears to generalize when E is replaced by an abelian variety, but time does not permit going into the details here.

1.1.6 Motivation for considering abelian varieties

If A is an elliptic curve, then explaining $Ш(A)$ using only congruences between elliptic curves will probably fail. This is because pairs of non-isogenous elliptic curves with isomorphic p-torsion for large p are, according to E. Kani's conjecture, extremely rare. It is crucial to understand what happens in all dimensions.

Within the range accessible by computer, abelian varieties exhibit more richly textured structure than elliptic curves. For example, there is a visible element of prime order 83341 in the Shafarevich-Tate group of an abelian variety of prime conductor 2333; in contrast, over all optimal elliptic curves of conductor up to 5500, it appears that the largest order of an element of a Shafarevich-Tate group is 7 (see the discussion in [18]).

1.2 Existence of nontrivial visible elements of $Ш(A)$

The reader who wants to see tables of Shafarevich-Tate groups can safely skip to the next section.

Without relying on any unverified conjectures, we will use the following theorem to exhibit abelian varieties A such that the visible part of $Ш(A)$ is nonzero. In the following theorem we do *not* assume that J is the Jacobian of a curve.

Theorem 1.8. *Let A and B be abelian subvarieties of an abelian variety J such that $A \cap B$ is finite and $A(\mathbf{Q})$ is finite. Assume that B has purely toric reduction at each prime at which J has bad reduction. Let p be an odd prime at which J has good reduction, and assume that p does not divide the orders of any of the (geometric) component groups of A and B, or the orders of the torsion subgroups of $(J/B)(\mathbf{Q})$ and $B(\mathbf{Q})$. Suppose further that $B[p] \subset A \cap B$. Then there exists an injection*

$$B(\mathbf{Q})/pB(\mathbf{Q}) \hookrightarrow Ш(A)^{\circ}$$

of $B(\mathbf{Q})/pB(\mathbf{Q})$ into the visible part of Ш(A).

Proof. Let $C = J/A$. The long exact sequence of Galois cohomology associated to the short exact sequence

$$0 \to A \to J \to C \to 0$$

begins

$$0 \to A(\mathbf{Q}) \to J(\mathbf{Q}) \to C(\mathbf{Q}) \xrightarrow{\delta} H^1(\mathbf{Q}, A) \to \cdots .$$

Because $B[p] \subset A$, the map $B \to C$, obtained by composing the inclusion $B \hookrightarrow J$ with $J \to C$, factors through multiplication-by-p, giving the following commutative diagram:

$$\begin{tikzcd}
 & B \arrow[r, "p"] \arrow[d] & B \arrow[d] \\
A \arrow[r] & J \arrow[r] & C.
\end{tikzcd}$$

Because $B(\mathbf{Q})[p] = 0$ and $B(\mathbf{Q}) \cap A(\mathbf{Q}) = 0$, we deduce the following commutative diagram with exact rows:

$$\begin{tikzcd}
 & 0 \arrow[d] & K_1 \arrow[d] & K_2 \arrow[d] & \\
0 \arrow[r] & B(\mathbf{Q}) \arrow[r, "p"] \arrow[d] & B(\mathbf{Q}) \arrow[r] \arrow[d] \arrow[dr, "\pi"] & B(\mathbf{Q})/pB(\mathbf{Q}) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & J(\mathbf{Q})/A(\mathbf{Q}) \arrow[r] \arrow[d] & C(\mathbf{Q}) \arrow[r] & \delta(C(\mathbf{Q})) \arrow[r] & 0 \\
 & K_3, & & &
\end{tikzcd}$$

where K_1 and K_2 are the indicated kernels and K_3 is the cokernel. We have the snake lemma exact sequence

$$0 \to K_1 \to K_2 \to K_3.$$

Because $B(\mathbf{Q})[p] = 0$ and K_2 is a p-torsion group, we have $K_1 = 0$. The quotient $J(\mathbf{Q})/B(\mathbf{Q})$ has no p-torsion because it is a subgroup of $(J/B)(\mathbf{Q})$; also, $A(\mathbf{Q})$ is a finite group of order coprime to p, so $K_3 = J(\mathbf{Q})/(A(\mathbf{Q}) + B(\mathbf{Q}))$ has no p-torsion. Thus $K_2 = 0$.

The above argument shows that $B(\mathbf{Q})/pB(\mathbf{Q})$ is a subgroup of $H^1(\mathbf{Q}, A)$. However, $H^1(\mathbf{Q}, A)$ contains infinitely many elements of order p (see [59]), whereas $Ш(A)[p]$ is a finite group, so we must work harder to deduce that $B(\mathbf{Q})/pB(\mathbf{Q})$ lies in $Ш(A)[p]$. Fix $x \in B(\mathbf{Q})$. We must show that $\pi(x)$ lies in $Ш(A)[p]$; equivalently, that $\mathrm{res}_v(\pi(x)) = 0$ for all places v of $\mathbf{Q}$.

At the archimedean place $v = \infty$, the restriction $\mathrm{res}_v(\pi(x))$ is killed by 2 and the odd prime p, hence $\mathrm{res}_v(\pi(x)) = 0$.

Suppose that v is a place at which J has bad reduction. By hypothesis, B has purely toric reduction, so over the maximal unramified extension $\mathbf{Q}_v^{\mathrm{ur}}$ of $\mathbf{Q}_v$ there is an isomorphism $B \cong \mathbf{G}_m^d/\Gamma$ of $\mathrm{Gal}(\overline{\mathbf{Q}}_v/\mathbf{Q}_v^{\mathrm{ur}})$-modules, for some "lattice" Γ. For example, when $\dim B = 1$, this is the Tate curve representation of B. Let n be the order of the component group of B at v; thus n equals the order of the cokernel of the valuation map $\Gamma \to \mathbf{Z}^d$. Choose a representative $P = (x_1, \dots, x_d) \in \mathbf{G}_m^d$ for the point x. Then nP can be adjusted by elements of Γ so that each of its components $x_i \in \mathbf{G}_m$ has valuation 0. By assumption, p is a prime at which J has good reduction, so $p \neq v$; it follows that there is a point $Q \in \mathbf{G}_m^d(\mathbf{Q}_v^{\mathrm{ur}})$ such that $pQ = nP$. Thus the cohomology class $\mathrm{res}_v(\pi(nx))$ is unramified at v. By [51, Prop. I.3.8],

$$H^1(\mathbf{Q}_v^{\mathrm{ur}}/\mathbf{Q}_v, A(\mathbf{Q}_v^{\mathrm{ur}})) = H^1(\mathbf{Q}_v^{\mathrm{ur}}/\mathbf{Q}_v, \Phi_{A,v}(\overline{\mathbf{F}}_v)),$$

where $\Phi_{A,v}$ is the component group of A at v. Since the component group $\Phi_{A,v}(\overline{\mathbf{F}}_v)$ has order n, it follows that

$$\mathrm{res}_v(\pi(nx)) = n\,\mathrm{res}_v(\pi(x)) = 0.$$

Since the order p of $\mathrm{res}_v(\pi(x))$ is coprime to n, we conclude that $\mathrm{res}_v(\pi(x)) = 0$.

Next suppose that J has good reduction at v and that v is *odd*, in the sense that the residue characteristic of v is odd. To simplify notation in this paragraph, since v is a non-archimedean place of $\mathbf{Q}$, we will also let v denote the odd prime number which is the residue characteristic of v. Let $\mathcal{A}$, $\mathcal{J}$, $\mathcal{C}$, be the Néron models of A, J, and C, respectively (for more on Néron models, see Chapter 4). Let A, J, C, also denote the sheaves on the étale-site over $\mathrm{Spec}(\mathbf{Z}_v)$ determined by the group schemes $\mathcal{A}$, $\mathcal{J}$, and $\mathcal{C}$, respectively. Since v is odd, $1 = e < v - 1$, so we may apply [8, Thm. 7.5.4] to conclude that the sequence of group schemes

$$0 \to \mathcal{A} \to \mathcal{J} \to \mathcal{C} \to 0$$

is exact; in particular, it is exact as a sequence of sheaves on the étale site (see the proof of [8, Thm. 7.5.4]). Thus it is exact on the stalks, so by [49, 2.9(d)] the sequence

$$0 \to \mathcal{A}(\mathbf{Z}_v^{\mathrm{ur}}) \to \mathcal{J}(\mathbf{Z}_v^{\mathrm{ur}}) \to \mathcal{C}(\mathbf{Z}_v^{\mathrm{ur}}) \to 0$$

is exact; by the Néron mapping property the sequence

$$0 \to A(\mathbf{Q}_v^{\mathrm{ur}}) \to J(\mathbf{Q}_v^{\mathrm{ur}}) \to C(\mathbf{Q}_v^{\mathrm{ur}}) \to 0$$

is also exact. Thus $\mathrm{res}_v(\pi(x))$ in unramified, so it arises by inflation from an element of $H^1(\mathbf{Q}_v^{\mathrm{ur}}/\mathbf{Q}_v, A)$. By [51, Prop. I.3.8],

$$H^1(\mathbf{Q}_v^{\mathrm{ur}}/\mathbf{Q}_v, A) \cong H^1(\mathbf{Q}_v^{\mathrm{ur}}/\mathbf{Q}_v, \Phi_{A,v}),$$

where $\Phi_{A,v}$ is the component group of A at v. Since A has good reduction, $\Phi_{A,v} = 0$, hence $\mathrm{res}_v(\pi(x)) = 0$.

If J has bad reduction at $v = 2$, then we already dealt with 2 above. Consider the case when J has good reduction at 2. Because the absolute ramification index e of $\mathbf{Z}_2$ is 1, which is *not* less than $v - 1 = 1$, we can not apply [8, Thm. 7.5.4]. However, we can modify our situation by an isogeny of degree a power of 2, then apply a different theorem as follows. The 2-primary subgroup Ψ of $A \cap B$ is rational as a subgroup over $\mathbf{Q}$, in the sense that the conjugates of any point in Ψ are also contained in Ψ. The abelian varieties $\tilde{J} = J/\Psi$, $\tilde{A} = A/\Psi$, and $\tilde{B} = B/\Psi$ also satisfy the hypothesis of the theorem we are proving. By [8, Prop. 7.5.3(a)], the corresponding sequence of Néron models

$$0 \to \tilde{\mathcal{A}} \to \tilde{\mathcal{J}} \to \tilde{\mathcal{C}} \to 0$$

is exact, so the sequence

$$0 \to \tilde{A}(\mathbf{Q}_v^{\mathrm{ur}}) \to \tilde{J}(\mathbf{Q}_v^{\mathrm{ur}}) \to \tilde{C}(\mathbf{Q}_v^{\mathrm{ur}}) \to 0$$

is exact. Thus the image of $\mathrm{res}_v(\pi(x))$ in $H^1(\mathbf{Q}_v, \tilde{A})$ is unramified. It equals 0, again by [51, Prop. 3.8], since the component group of $\tilde{A}$ at v has order a power of 2 (in fact it is trivial, since $\tilde{A}$ has good reduction at 2), whereas $\pi(x)$ has odd prime order p. Thus $\mathrm{res}_v(\pi(x)) = 0$, since the kernel of $H^1(\mathbf{Q}_v, A) \to H^1(\mathbf{Q}_v, \tilde{A})$ is a finite group of 2-power order. □

1.3 Description of tables

In this section we describe our tables of optimal quotients of $J_0(N)$, which have nontrivial Shafarevich-Tate group. The tables, which can be found on pages 15–18, were computed using a combination of HECKE [64], LiDIA, NTL, PARI, and most successfully MAGMA [9]. The component group computations at non-prime level rely on Kohel's quaternion algebra package, which was also written in MAGMA.

We compute the conjectural order of the Shafarevich-Tate group of an abelian variety A, and then make assertions about the Shafarevich-Tate group of $A^\vee$. This is justified because the order of $Ш(A^\vee)$ equal the order of $Ш(A)$, since both are finite and the Cassells-Tate pairing sets up a nondegenerate duality between them.

1.3.1 Notation

Each optimal quotient A of $J_0(N)$ is denoted by a label, such as **389E**, which consists of a level N and a letter indicating the isogeny class. In the labeling, N is a positive integer and the isogeny class is given by a letter: the first isogeny class is labeled **A**, the second is labeled **B**, the third labeled **C**, and so on. Thus **389E** is the fifth isogeny class of optimal quotient of $J_0(389)$, corresponding to a Galois-conjugacy class of newforms. The isogeny classes that we consider are in bijection with the Galois-conjugacy classes of newforms in $S_2(\Gamma_0(N))$. The classes of newforms are ordered as described in Section 3.5.5.

WARNING: The *odd part* of a rational number x is $x/2^v$, where $v = \mathrm{ord}_2(x)$. In the tables, only the **odd parts** of the arithmetic invariants of A are given.

1.3.2 Table 1.2: Shafarevich-Tate groups at prime level

Table 1.2 was constructed as follows. Using the results of Section 3.10, we computed the odd part of the conjectural order $\#\text{Ш}_{\text{an}}(A)$ of the Shafarevich-Tate group of every optimal quotient of $J_0(p)$ that corresponds to a single Galois conjugacy-class of eigenforms and has analytic rank 0, for p a prime with $p \leq 2161$. We also computed a few sporadic examples of prime level p with $p > 2161$. The sporadic examples appear at the bottom of the table below a horizontal line.

Notation

The columns of the table contain the following information. The abelian varieties A for which $\#\text{Ш}_{\text{an}}(A)$ is greater than 1 are laid out in the first column of Table 1.2. The second column contains the dimensions of the abelian varieties in the first column. The third column contains the *odd part* (i.e., largest odd divisor) of the order of the Shafarevich-Tate group, as predicted by the BSD conjecture. Column four contains the odd parts of the modular degrees of the abelian varieties in the first column.

The fifth column contains an optimal quotient B of $J_0(p)$ of positive analytic rank, such that the ℓ-torsion of $B^\vee$ is contained in $A^\vee$, when one exists, where ℓ is a divisor of $\#\text{Ш}_{\text{an}}(A)$. We computed this intersection using the algorithm described in Section 3.6. Such a B is called an *explanatory factor*. When no such abelian varieties exists, we write "NONE" in the fifth column. The sixth column contains the dimensions of the abelian varieties in the fifth column, and the seventh column contains the odd parts of the modular degrees of the abelian varieties in the fifth column.

Ranks of the explanatory factors

That the explanatory factors have positive analytic rank follows from our modular symbols computation of $L(B,1)/\Omega_B$. This is supported by the table in [11], except in the case **2333A**, where there is a mistake in [11] (see below).

The explanatory factor **389A** is the first elliptic curve of rank 2. The table in [11] suggests that the explanatory factor **1061B** is the first 2-dimensional abelian variety (attached to a newform) whose Mordell-Weil group when tensored with the field of fractions F of the corresponding ring of Fourier coefficients, is of dimension 2 over F. Similarly **1567B** appears to be the first 3-dimensional one of rank 2, and **2333A** is the first 4-dimensional one of rank 2. It thus seems very likely that the ranks of each explanatory factor is exactly 2, though we have not proved this.

Discussion of the data

There are 23 examples in which $\text{Ш}(A)$ is visible and 18 in which $\text{Ш}(A)$ is invisible. The largest visible $\text{Ш}(A)$ found occurs at level 2333 and has order at least 83341^2 (83341 is prime). The largest invisible $\text{Ш}(A)$ occurs in a 112-dimensional abelian variety at level 2111 and has order at least 211^2.

The example **1283C** demonstrates that $\#\text{Ш}_{\text{an}}(A)$ can divide the modular degree, yet be *invisible*. In this case 5 divides $\#\text{Ш}_{\text{an}}(A)$. Since 5 divides the modular degree, it follows that there must be other non-isogenous subvarieties of $J_0(1283)$ that intersect **1283C** in a subgroup of order divisible by 5. In this case, the only such subvariety is **1283A**, which has dimension 2 and whose 5-torsion is contained in **1283C**. However **1283A** has analytic (hence algebraic) rank 0, so $\#\text{Ш}_{\text{an}}(A)$ cannot be visible.

The cases **1483D**, **1567D**, **2029C**, and **2593B** are interesting because *all* of Ш, even though it has two nontrivial p-primary components in each of these cases, is made visible in a single B. In the case **1913E** only the 5-primary component of Ш is visible in **1913A**, but still *both* the 5-primary and 61-primary components of Ш are visible in **1913C**.

Examples **1091C** and **1429B** were first found in [1] and **1913B** in [18].

Errata to Brumer's paper

Contrary to our computations, [11] suggests that **2333A** has rank 0. However, the author pointed the discrepancy out to Brumer who replied:

> I looked in vain for information about θ-relations on 2333 and have concluded that I never ran the interval from 2300 to 2500 or else had lost all info by the time I wrote up the paper. The punchline: 4 relations for 2333 and 2 relations for 2381 (also missing from the table).

1.3.3 Tables 1.3–1.6: New visible Shafarevich-Tate groups

Let n denote the largest odd square dividing the numerator of $L(A,1)/\Omega_A$. Table 1.3 lists those A such that for some $p \mid n$ there exists a quotient B of $J_0(N)$, corresponding to a *newform* and having positive analytic rank, such that $(A^\vee \cap B^\vee)[p] \neq 0$. Our search was systematic up to level 1001, but there might be omitted examples between levels 1001 and 1028. Table 1.4 contains further arithmetic information about each explanatory factor. Table 1.6 gives the quantities involved in the formula of Chapter 4 for Tamagawa numbers, for each of the abelian varieties A in Table 1.3.

Notation

Most of the notation is the same as in Table 1.2. However the additional columns w_q and c_p contain the signs of the Atkin-Lehner involutions and the Tamagawa numbers, respectively. These are given in order, from smallest to largest prime divisor of N.

In each case B has dimension 1. When $4 \mid N$, we write "a" for c_2 to remind us that we did not compute c_2 because the reduction at 2 is additive. Again only *odd parts* of the invariants are given. Section 4.7.2 contains a discussion of the notation used in the headings of Table 1.6.

Remarks on the data

The explanatory factors B of level ≤ 1028 are *exactly* the set of rank 2 elliptic curves of level ≤ 1028.

Table 1.1: Odd invisible $|Ш_E| > 1$, all $N \leq 5500$ (from Table 1 of [18])

E	$\sqrt{|Ш_E|}$	m_E	**F**	m_F	Remarks
2849A	3	$2^5 \cdot 5 \cdot 61$	**NONE**	–	
3364C	7	$2^6 \cdot 3^2 \cdot 5^2 \cdot 7$	**none**	–	
4229A	3	$2^3 \cdot 3 \cdot 7 \cdot 13$	**none**	–	
4343B	3	$2^4 \cdot 1583$	**NONE**	–	
4914N	3	$2^4 \cdot 3^5$	**none**	–	**E** has rational 3-torsion
5054C	3	$2^3 \cdot 3^3 \cdot 11$	**none**	–	
5073D	3	$2^5 \cdot 3 \cdot 5 \cdot 7 \cdot 23$	**none**	–	
5389A	3	$2^2 \cdot 2333$	**NONE**	–	

1.4 Further visibility computations

1.4.1 Does Ш become visible at higher level?

This section is concerned with whether or not the examples of invisible elements of Shafarevich-Tate groups of elliptic curves of level N that are given in [18] become visible in abelian surfaces inside appropriate $J_0(Np)$. We analyze each of the cases in Table 1 of [18]. For the reader's convenience, the part of this table which concerns us is reproduced as Table 1.1. The most interesting examples we give are **2849A** and **5389A**. As in [18], the assertions below assume the truth of the BSD conjecture.

How we found the explanatory curves

We use a naive heuristic observation to find possible explanatory curves of higher level, even though their conductors are out of the range of Cremona's tables. Note that we have not proved that these factors are actually explanatory in any cases, and expect that in some cases they are not.

First we recall some of the notation from Table 1 of [18], which is partially reproduced below. The "NONE" label in the row corresponding to an elliptic curve E indicates that there are elements in $Ш(E)$ whose order does not divide the modular degree of E, and hence they must be invisible. The label "none" indicates only that no suitable explanatory elliptic curves could be found, so $Ш(E)$ is not visible in an *abelian surface* inside $J_0(N)$; it could, however, be visible in the full abelian variety $J_0(N)$.

Studying the Weierstrass equations corresponding to the curves in [18] reveals that the elliptic curves labeled "NONE" have unusually large height, as compared to their conductors. However, the explanatory factors often have unusually small height. Motivated by this purely heuristic observation, we make a table of all elliptic curves of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

with $a_1, a_2, a_3 \in \{-1, 0, 1\}$, $|a_4|, |a_6| < 1000$, and conductor bounded by 50000. The bound on the conductor is required only so that the table will fit within computer storage. This table took a few days to generate.

2849A

Barry Mazur and Adam Logan found the first known example of an *invisible* Shafarevich-Tate group. This was $\text{Ш}(E)$, where E is the elliptic curve **2849A**, which has minimal Weierstrass equation

$$E: \quad y^2 + xy + y = x^3 + x^2 - 53484x - 4843180.$$

Consulting our table of curves of small height, we find an elliptic curve F of conductor $8547 = 2849 \cdot 3$ such that $f_E \equiv f_F \pmod 3$, where f_E and f_F are the newforms attached to E and F, respectively. This is a congruence for *all* eigenvalues a_p attached to E and F. The elliptic curve F is defined by the equation

$$F: \quad y^2 + xy + y = x^3 + x^2 - 154x - 478.$$

Cremona's program `mwrank` reveals that the Mordell-Weil group of F has rank 2. Thus maybe $\text{Ш}(E)$ becomes visible at level 8547. Unfortunately, visibility of $\text{Ш}(E)$ is not implied by Theorem 1.8 because the geometric component group of F at 3 has order divisible by 3.

4343B

Consider the elliptic curve E labeled **4343B**. According to Table 1 of [18], $\text{Ш}(E)$ has order 9, but the modular degree prevents $\text{Ш}(E)$ from being visible in $J_0(4343)$. At level $21715 = 5 \cdot 4343$ there is an elliptic curve F of rank 1 that is congruent to E. Its equation is

$$F: \quad y^2 - xy - y = x^3 - x^2 + 78x - 256.$$

5389A

The last curve labeled "NONE" in the table is curve **5389A**, which has minimal Weierstrass equation

$$y^2 + xy + y = x^3 - 35590x - 2587197.$$

The main theorem of [54] implies that there exists a newform that is congruent modulo 3 to the newform corresponding to **5389A** and of level $3 \cdot 5389$. This is because $(-2)^2 = (3+1)^2 \pmod 3$. However, our table of curves of small height does not contain any curve of conductor $3 \cdot 5389$. Next we observe that $(-2)^2 \equiv (7+1)^2 \pmod 3$, so using Ribet's theorem we can instead augment the level by 7. Our table of small-height curves contains just one (up to isogeny) elliptic curve of conductor 37723, and *luckily* the corresponding newform is congruent modulo 3 to the newform corresponding to **5389A** (away from primes dividing the level)! The Weierstrass equation of this curve is

$$F: \quad y^2 - y = x^3 + x^2 + 34x - 248.$$

According to Cremona's program `mwrank`, the rank of F is 2.

3364C, 4229A, 5073D

Perhaps Ш(E) is already visible in some of the cases in which the curve is labeled "none", because the method fails in most of these cases. Each of the curves **3364C**, **4229A**, and **5073D** is labeled "none". In none of these 3 cases are we able to find an explanatory factor at higher level, within the range of our table of elliptic curves of small height.

4194N, 5054C

The curve **4914N** is labeled "none" and we find the remark "E has rational 3-torsion". There is a congruent curve F of conductor 24570 given by the equation

$$F : \quad y^2 - xy = x^3 - x^2 - 15x - 75,$$

and $F(\mathbf{Q}) = \{0\}$. The curve **5054C** is labeled "none" and its Shafarevich-Tate group contains invisible elements of order 3. We find a congruent curve of level 25270 and rank 1. The equation of the congruent curve is

$$F : y^2 - xy = x^3 + x^2 - 178x + 882.$$

1.4.2 Positive rank example

The abelian varieties with nontrivial Ш(A) that one finds in both ours and Cremona's tables all have rank 0. In this section we outline a computation which sugggests, but does not prove, that there is a positive-rank abelian subvariety A of $J_0(5077)$ such that Ш(A) possesses a nontrivial visible element of order 31.

According to [16], the smallest conductor elliptic curve E of rank 3 is found in $J = J_0(5077)$. The number 5077 is prime, and the isogeny decomposition of J is[1]

$$J \sim A \times B \times E,$$

where each of A, B, and E are abelian subvarieties of J associated to newforms, which have dimensions 205, 216, and 1, respectively. Using Remark 3.38 or [69], we find that the modular degree of E is $1984 = 2^6 \cdot 31$. The sign of the Atkin-Lehner involution on E is the same as its sign on A, so $E[31] \subset A$. We have $E(\mathbf{Q}) \cong \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$, and the component group of E is trivial. The numerator of $(5077-1)/12$ is $3^2 \cdot 47$, so [40] implies that none of the abelian varieties above have 31-torsion. It might be possible to find an analogue of Theorem 1.8 that applies when A has positive rank, and deduce in this case that Ш(A) contains visible elements of order 31.

[1] This decomposition was found in about one minute using the Mestre-Oesterlé method of graphs (see [47]).

Table 1.2: Shafarevich-Tate groups at prime level. (The entries in the columns "mod deg" and "$\#\text{Ш}_{\mathrm{an}}$" are only really the odd parts of "mod deg" and "$\#\text{Ш}_{\mathrm{an}}$".)

A	dim	$\#\text{Ш}_{\mathrm{an}}(A)$	mod deg(A)	**B**	dim	mod deg(B)
389E	20	5^2	5	**389A**	1	5
433D	16	7^2	$3 \cdot 7 \cdot 37$	**433A**	1	7
563E	31	13^2	13	**563A**	1	13
571D	2	3^2	$3^2 \cdot 127$	**571B**	1	3
709C	30	11^2	11	**709A**	1	11
997H	42	3^4	3^2	**997B**	1	3
1061D	46	151^2	$61 \cdot 151 \cdot 179$	**1061B**	2	151
1091C	62	7^2	1	NONE		
1171D	53	11^2	$3^4 \cdot 11$	**1171A**	1	11
1283C	62	5^2	$5 \cdot 41 \cdot 59$	NONE		
1429B	64	5^2	1	NONE		
1481C	71	13^2	$5^2 \cdot 2833$	NONE		
1483D	67	$3^2 \cdot 5^2$	$3 \cdot 5$	**1483A**	1	$3 \cdot 5$
1531D	73	3^2	3	**1531A**	1	3
1559B	90	11^2	1	NONE		
1567D	69	$7^2 \cdot 41^2$	$7 \cdot 41$	**1567B**	3	$7 \cdot 41$
1613D	75	5^2	$5 \cdot 19$	**1613A**	1	5
1621C	70	17^2	17	**1621A**	1	17
1627C	73	3^4	3^2	**1627A**	1	3^2
1693C	72	1301^2	1301	**1693A**	3	1301
1811D	98	31^2	1	NONE		
1847B	98	3^6	1	NONE		
1871C	98	19^2	14699	NONE		
1877B	86	7^2	1	NONE		
1907D	90	7^2	$3 \cdot 5 \cdot 7 \cdot 11$	**1907A**	1	7
1913B	1	3^2	$3 \cdot 103$	**1913A**	1	$3 \cdot 5^2$
1913E	84	$5^4 \cdot 61^2$	$5^2 \cdot 61 \cdot 103$	**1913A,C**	$1, 2$	$3 \cdot 5^2, 5^2 \cdot 61$
1933C	83	$3^2 \cdot 7^2$	$3 \cdot 7$	**1933A**	1	$3 \cdot 7$
1997C	93	17^2	1	NONE		
2027C	94	29^2	29	**2027A**	1	29
2029C	90	$5^2 \cdot 269^2$	$5 \cdot 269$	**2029A**	2	$5 \cdot 269$
2039F	99	$3^4 \cdot 5^2$	$19 \cdot 29 \cdot 7759 \cdot 3214201$	NONE		
2063C	106	13^2	$61 \cdot 139$	NONE		
2089J	91	11^2	$3 \cdot 5 \cdot 11 \cdot 19 \cdot 73 \cdot 139$	**2089B**	1	11
2099B	106	3^2	1	NONE		
2111B	112	211^2	1	NONE		
2113B	91	7^2	1	NONE		
2161C	98	23^2	1	NONE		
2333C	101	83341^2	83341	**2333A**	4	83341
2339C	114	3^8	6791	NONE		
2411B	123	11^2	1	NONE		
2593B	109	$67^2 \cdot 2213^2$	$67 \cdot 2213$	**2593A**	4	$67 \cdot 2213$

Table 1.3: New visible Shafarevich-Tate groups

A	dim	$\#Ш_{\mathrm{an}}$	w_q	c_p	$\#A(\mathbf{Q})$	$\frac{\#A(\mathbf{Q})\cdot L(A,1)}{\Omega_A}$	mod deg(A)	**B**
389E	20	5^2	$-$	97	97	5^2	5	**389A**
433D	16	7^2	$-$	3^2	3^2	7^2	$3\cdot 7\cdot 37$	**433A**
446F	8	11^2	$+-$	$1,3$	3	11^2	$11\cdot 359353$	**446B**
563E	31	13^2	$-$	281	281	13^2	13	**563A**
571D	2	3^2	$-$	1	1	3^2	$3^2\cdot 127$	**571B**
655D	13	3^4	$+-$	$1,1$	1	3^4	$3^2\cdot 19\cdot 515741$	**655A**
664F	8	5^2	$-+$	$a,1$	1	5^2	5	**664A**
681B	1	3^2	$+-$	$1,1$	1	3^2	$3\cdot 5^3$	**681C**
707G	15	13^2	$+-$	$1,1$	1	13^2	$13\cdot 800077$	**707A**
709C	30	11^2	$-$	59	59	11^2	11	**709A**
718F	7	7^2	$+-$	$1,1$	1	7^2	$7\cdot 151\cdot 35573$	**718B**
794G	14	11^2	$+-$	$3,1$	3	11^2	$3\cdot 7\cdot 11\cdot 47\cdot 35447$	**794A**
817E	15	7^2	$+-$	$1,5$	5	7^2	$7\cdot 79$	**817A**
916G	9	11^2	$-+$	$a,1$	1	11^2	$3^9\cdot 11\cdot 17\cdot 239$	**916C**
944O	6	7^2	$+-$	$a,1$	1	7^2	7	**944E**
997H	42	3^4	$-$	83	83	3^4	3^2	**997BC**
1001L	7	7^2	$+-+$	$1,1,1$	1	7^2	$7\cdot 19\cdot 47\cdot 2273$	**1001C**
1028E	14	11^2	$-+$	$a,1$	3	$3^4\cdot 11^2$	$3^{13}\cdot 11$	**1028A**

Table 1.4: Explanatory factors

B	rank	w_q	c_p	$\#A(\mathbf{Q})$	mod deg(A)	Comments
389A	2	$-$	1	1	5	first curve of rank 2
433A	2	$-$	1	1	7	
446B	2	$+-$	$1,1$	1	11	called **446D** in [16]
563A	2	$-$	1	1	13	
571B	2	$-$	1	1	3	
655A	2	$+-$	$1,1$	1	3^2	
664A	2	$-+$	$1,1$	1	5	
681C	2	$+-$	$1,1$	1	3	
707A	2	$+-$	$1,1$	1	13	
709A	2	$-$	1	1	11	
718B	2	$+-$	$1,1$	1	7	
794A	2	$+-$	$1,1$	1	11	
817A	2	$+-$	$1,1$	1	7	
916C	2	$-+$	$3,1$	1	$3 \cdot 11$	
944E	2	$+-$	$1,1$	1	7	
997B	2	$-$	1	1	3	
997C	2	$-$	1	1	3	
1001C	2	$+-+$	$1,3,1$	1	$3^2 \cdot 7$	
1028A	2	$-+$	$3,1$	1	$3 \cdot 11$	intersects **1028E** mod 11

Table 1.5: Factorizations

$\mathbf{446} = 2 \cdot 223$	$\mathbf{655} = 5 \cdot 131$	$\mathbf{664} = 2^3 \cdot 83$	$\mathbf{681} = 3 \cdot 227$
$\mathbf{707} = 7 \cdot 101$	$\mathbf{718} = 2 \cdot 359$	$\mathbf{794} = 2 \cdot 397$	$\mathbf{817} = 19 \cdot 43$
$\mathbf{916} = 2^2 \cdot 229$	$\mathbf{944} = 2^4 \cdot 59$	$\mathbf{1001} = 7 \cdot 11 \cdot 13$	$\mathbf{1028} = 2^2 \cdot 257$

Table 1.6: Component groups

A	dim	p	w_q	$\#\Phi_{X,p}$	$m_{X,p}$	$\#\Phi_{A,p}(\overline{\mathbf{F}}_p)$
389E	20	389	$-$	97	$5 \cdot 97$	97
433D	16	433	$-$	3^2	$3^3 \cdot 7 \cdot 37$	3^2
446F	8	223	$-$	3	$3 \cdot 11 \cdot 359353$	3
		2	$+$	3	$3 \cdot 11$	$3 \cdot 359353$
563E	31	563	$-$	281	$13 \cdot 281$	281
571D	2	571	$-$	1	$3^2 \cdot 127$	1
655D	13	131	$-$	1	$3^2 \cdot 19 \cdot 515741$	1
		5	$+$	1	3^2	$19 \cdot 515741$
664F	8	83	$+$	1	5	1
681B	1	227	$-$	1	$3 \cdot 5^3$	1
		3	$+$	1	$3 \cdot 5^2$	5
707G	15	101	$-$	1	$13 \cdot 800077$	1
		7	$+$	1	13	800077
709C	30	709	$-$	59	$11 \cdot 59$	59
718F	7	359	$-$	1	$7 \cdot 151 \cdot 35573$	1
		2	$+$	1	7	$151 \cdot 35573$
794G	14	397	$-$	3	$3^2 \cdot 7 \cdot 11 \cdot 47 \cdot 35447$	3
		2	$+$	3	$3 \cdot 11$	$3^2 \cdot 7 \cdot 47 \cdot 35447$
817E	15	43	$-$	5	$5 \cdot 7 \cdot 79$	5
		19	$+$	1	7	79
916G	9	229	$+$	1	$3^9 \cdot 11 \cdot 17 \cdot 239$	1
944O	6	59	$-$	1	7	1
997H	42	997	$-$	83	$3^2 \cdot 83$	83
1001L	7	13	$+$	1	$7 \cdot 19 \cdot 47 \cdot 2273$	1
		11	$-$	1	$7 \cdot 19 \cdot 47 \cdot 2273$	1
		7	$+$	1	$7 \cdot 19 \cdot 47$	2273
1028E	14	257	$+$	1	$3^{13} \cdot 11$	1

Chapter 2

Modular symbols

Modular symbols permeate this thesis. In their simplest incarnation, modular symbols provide a finite presentation for the homology group $H_1(X_0(N), \mathbf{Z})$ of the Riemann surface $X_0(N)$. This presentation is equipped with such a rich structure that from it we can deduce the action of the Hecke operators; this is already sufficient information for us to compute a basis for the space $S_2(\Gamma_0(N), \mathbf{C})$ of cusp forms.

We recall the definition of spaces of modular symbols in Sections 2.1–2.2. Then in Section 2.3, we review the duality between modular symbols and modular forms. In Section 2.4, we see that modular symbols are furnished with analogues of each of the standard operators that one finds on spaces of modular forms, and in Section 2.5 we see that the same is true of the degeneracy maps. Section 2.6 describes Manin symbols, which supply a convenient finite presentation for the space of modular symbols. Finally, Section 2.7 introduces the complex torus attached to a newform, which appears in various guises throughout this thesis.

Before continuing, we offer an apology. We will only consider modular symbols that are already equipped with a fixed Dirichlet character. Though fixing a character complicates the formulas, the resulting increase in efficiency is of extreme value in computational applications. Fixing a character allows us to compute in just the part of the space of modular symbols for $\Gamma_1(N)$ that interests us. We apologize for any inconvenience this may cause the less efficiency minded reader.

Acknowledgment. This chapter and the next were greatly influenced by the publications of Cremona [15, 16] and Merel [45], along with the foundational contributions of Manin [38], Mazur [42, 39], and Shokurov [63]. Cremona's book [16] provides a motivated roadmap that guides the reader who wishes to compute with modular symbols in the familiar context of elliptic curves, and Merel's article provides an accessible overview of the action of Hecke operators on higher weight modular symbols, and the connection between modular symbols and related cohomology theories.

2.1 The definition of modular symbols

Fix a positive integer N, an integer $k \geq 2$, and a continuous homomorphism

$$\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$$

such that $\varepsilon(-1) = (-1)^k$. We call N the *level*, k the *weight*, and ε the *Dirichlet character.*

Consider the quotient of the abelian group generated by all formal symbols $\{\alpha, \beta\}$, with $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q}) = \mathbf{Q} \cup \{\infty\}$, by the following relations:

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0,$$

for all $\alpha, \beta, \gamma \in \mathbf{P}^1(\mathbf{Q})$. Let $\boldsymbol{\mathcal{M}}$ be the torsion-free quotient of this group by its torsion subgroup. Because $\boldsymbol{\mathcal{M}}$ is torsion free, $\{\alpha, \alpha\} = 0$ and $\{\alpha, \beta\} = -\{\beta, \alpha\}$.

Remark 2.1. One is motivated to consider these relations by viewing $\{\alpha, \beta\}$ as the homology class of an appropriate path from α to β in the upper half plane.

Let V_{k-2} be the $\mathbf{Z}$-submodule of $\mathbf{Z}[X, Y]$ made up of all homogeneous polynomials of degree $k-2$, and set $\boldsymbol{\mathcal{M}}_k := V_{k-2} \otimes \boldsymbol{\mathcal{M}}$. For $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbf{Q})$ and $P \in V_{k-2}$, let

$$\begin{aligned} gP(X, Y) &= P\left(\det(g) g^{-1} \begin{pmatrix} X \\ Y \end{pmatrix}\right) = P\left(\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}\right) \\ &= P(dX - bY, -cX + aY). \end{aligned}$$

This defines a left action of $\mathrm{GL}_2(\mathbf{Q})$ on V_{k-2}; it is a left action because

$$\begin{aligned} (gh)P(v) &= P(\det(gh)(gh)^{-1} v) = P(\det(h) h^{-1} \det(g) g^{-1} v) \\ &= gP(\det(h) h^{-1} v) = g(hP(v)). \end{aligned}$$

Combining this action with the action of $\mathrm{GL}_2(\mathbf{Q})$ on $\mathbf{P}^1(\mathbf{Q})$ by linear fractional transformations gives a left action of $\mathrm{GL}_2(\mathbf{Q})$ on $\boldsymbol{\mathcal{M}}_k$:

$$g(P \otimes \{\alpha, \beta\}) = g(P) \otimes \{g(\alpha), g(\beta)\}.$$

Finally, for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, let $\varepsilon(g) := \varepsilon(\overline{a})$, where $\overline{a} \in \mathbf{Z}/N\mathbf{Z}$ is the reduction modulo N of a.

Let

$$\mathbf{Z}[\varepsilon] := \mathbf{Z}[\varepsilon(a) : a \in \mathbf{Z}/N\mathbf{Z}]$$

be the subring of $\mathbf{C}$ generated by the values of the character ε.

Definition 2.2 (Modular symbols). The space of *modular symbols* $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$ of level N, weight k and character ε is the largest torsion-free quotient of $\boldsymbol{\mathcal{M}}_k \otimes \mathbf{Z}[\varepsilon]$ by the $\mathbf{Z}[\varepsilon]$-submodule generated by the relations $gx - \varepsilon(g)x$ for all $x \in \boldsymbol{\mathcal{M}}_k$ and all $g \in \Gamma_0(N)$.

Denote by $P\{\alpha, \beta\}$ the image of $P \otimes \{\alpha, \beta\}$ in $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$. For any $\mathbf{Z}[\varepsilon]$-algebra R, let

$$\boldsymbol{\mathcal{M}}_k(N, \varepsilon; R) := \boldsymbol{\mathcal{M}}_k(N, \varepsilon) \otimes_{Z[\varepsilon]} R.$$

See Section 3.1 for an algorithm which can be used to compute $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; \mathbf{Q}(\varepsilon))$.

2.2 Cuspidal modular symbols

Let $\boldsymbol{\mathcal{B}}$ be the free abelian group generated by the symbols $\{\alpha\}$ for all $\alpha \in \mathbf{P}^1(\mathbf{Q})$. There is a left action of $\mathrm{GL}_2(\mathbf{Q})$ on $\boldsymbol{\mathcal{B}}$ given by $g\{\alpha\} = \{g(\alpha)\}$. Let $\boldsymbol{\mathcal{B}}_k := V_{k-2} \otimes \boldsymbol{\mathcal{B}}$, and let $\mathrm{GL}_2(\mathbf{Q})$ act on $\boldsymbol{\mathcal{B}}_k$ by $g(P\{\alpha\}) = (gP)\{g(\alpha)\}$.

Definition 2.3 (Boundary modular symbols). The space $\boldsymbol{\mathcal{B}}_k(N, \varepsilon)$ of *boundary modular symbols* is the largest torsion-free quotient of $\boldsymbol{\mathcal{B}}_k \otimes \mathbf{Z}[\varepsilon]$ by the relations $gx = \varepsilon(g)x$ for all $g \in \Gamma_0(N)$ and $x \in \boldsymbol{\mathcal{B}}_k$.

Denote by $P\{\alpha\}$ the image of $P \otimes \{\alpha\}$ in $\boldsymbol{\mathcal{B}}_k(N, \varepsilon)$. The *boundary map*

$$\delta : \boldsymbol{\mathcal{M}}_k(N, \varepsilon) \to \boldsymbol{\mathcal{B}}_k(N, \varepsilon)$$

is defined by

$$\delta(P\{\alpha, \beta\}) = P\{\beta\} - P\{\alpha\}.$$

Definition 2.4 (Cuspidal modular symbols). The space $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ of *cuspidal modular symbols* is the kernel of δ.

The three spaces defined above fit together in the following exact sequence:

$$0 \to \boldsymbol{\mathcal{S}}_k(N, \varepsilon) \to \boldsymbol{\mathcal{M}}_k(N, \varepsilon) \xrightarrow{\delta} \boldsymbol{\mathcal{B}}_k(N, \varepsilon).$$

2.3 Duality between modular symbols and modular forms

For any positive integer k, any $\mathbf{C}$-valued function f on the complex upper half plane

$$\mathfrak{h} := \{z \in \mathbf{C} : \mathrm{im}(z) > 0\},$$

and any matrix $\gamma \in \mathrm{GL}_2(\mathbf{Q})$, define a function $f|[\gamma]_k$ on $\mathfrak{h}$ by

$$(f|[\gamma]_k)(z) = \det(\gamma)^{k-1} \frac{f(\gamma z)}{(cz + d)^k}.$$

Definition 2.5 (Cusp forms). Let $S_k(N, \varepsilon)$ be the complex vector space of holomorphic functions $f(z)$ on $\mathfrak{h}$ that satisfy the equation

$$f|[\gamma]_k = \varepsilon(\gamma) f$$

for all $\gamma \in \Gamma_0(N)$, and such that f is holomorphic and vanishes at all cusps, in the sense of [21, pg. 42].

Definition 2.6 (Antiholomorphic cusp forms). Let $\overline{S}_k(N, \varepsilon)$ be the space of *antiholomorphic cusp forms*; the definition is as above, except

$$\frac{f(\gamma z)}{(c\overline{z} + d)^k} = \overline{\varepsilon}(\gamma) f(z)$$

for all $\gamma \in \Gamma_0(N)$.

There is a canonical isomorphism of real vector spaces between $S_k(N,\varepsilon)$ and $\overline{S}_k(N,\varepsilon)$ that associates to f the antiholomorphic cusp form defined by the function $z \mapsto \overline{f(z)}$.

Theorem 2.7 (Merel). *There is a pairing*

$$\langle\ ,\ \rangle : (S_k(N,\varepsilon) \oplus \overline{S}_k(N,\varepsilon)) \times \boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathbf{C}) \to \mathbf{C}$$

given by

$$\langle f \oplus g, P\{\alpha,\beta\}\rangle = \int_{\alpha}^{\beta} f(z)P(z,1)dz + \int_{\alpha}^{\beta} g(z)P(\overline{z},1)d\overline{z},$$

where the path from α to β is, except for the endpoints, contained in $\mathfrak{h}$. The pairing is perfect when restricted to $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathbf{C})$.

Proof. Take the ε part of each side of [45, Thm. 3]. □

2.4 Linear operators

2.4.1 Hecke operators

For each positive integer n and each space V of modular symbols or modular forms, there is a *Hecke operator* T_n, which acts as a linear endomorphism of V. For the definition of T_n on modular symbols, see [45, §2]. Alternatively, because we consider only modular symbols with character, the following recipe completely determines the Hecke operators. First, when $n = p$ is prime, we have

$$T_p(x) = \left[\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} + \sum_{r \bmod p} \begin{pmatrix} 1 & r \\ 0 & p \end{pmatrix}\right] x,$$

where the first matrix is omitted if $p \mid N$. If m and n are coprime, then $T_{mn} = T_m T_n$. Finally, if p is a prime, $r \geq 2$ is an integer, ε is the Dirichlet character of associated to V, and k is the weight of V, then

$$T_{p^r} = T_p T_{p^{r-1}} - \varepsilon(p)p^{k-1}T_{p^{r-2}}.$$

Definition 2.8. The *Hecke algebra associated to V* is the subring

$$\mathbf{T} = \mathbf{T}_V = \mathbf{Z}[\ldots T_n \ldots]$$

of $\mathrm{End}(V)$ generated by all Hecke operators T_n, with $n = 1, 2, 3, \ldots$.

Proposition 2.9. *The pairing of Theorem 2.7 respects the action of the Hecke operators, in the sense that $\langle fT, x\rangle = \langle f, Tx\rangle$ for all $T \in \mathbf{T}$, $x \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$, and $f \in S_k(N,\varepsilon) \oplus \overline{S}_k(N,\varepsilon)$.*

Proof. See [45, Prop. 10]. □

2.4.2 The $*$-involution

The matrix $j = \left(\begin{smallmatrix}-1&0\\0&1\end{smallmatrix}\right)$ defines an involution $*$ of $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ given by $x \mapsto x^* = j(x)$. Explicitly,

$$(P(X,Y)\{\alpha,\beta\})^* = P(X,-Y)\{-\alpha,-\beta\}.$$

Because the space of modular symbols is constructed as a quotient, it is not obvious that the $*$-involution is well defined.

Proposition 2.10. *The $*$-involution is well defined.*

Proof. Recall that $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ is the largest torsion-free quotient of the free $\mathbf{Z}[\varepsilon]$-module generated by symbols $x = P\{\alpha,\beta\}$ by the submodule generated by relations $\gamma x - \varepsilon(\gamma)x$ for all $\gamma \in \Gamma_0(N)$. In order to check that the operator $*$ is well defined, it suffices to check, for any $x \in \boldsymbol{\mathcal{M}}_k$, that $*(\gamma x - \varepsilon(\gamma)x)$ is of the form $\gamma' y - \varepsilon(\gamma')y$, for some y in $\boldsymbol{\mathcal{M}}_k$. Note that if $\gamma = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \Gamma_0(N)$, then $j\gamma j^{-1} = \left(\begin{smallmatrix}a&-b\\-c&d\end{smallmatrix}\right)$ is also in $\Gamma_0(N)$ and $\varepsilon(j\gamma j^{-1}) = \varepsilon(\gamma)$. We have

$$\begin{aligned} j(\gamma x - \varepsilon(\gamma)x) &= j\gamma x - j\varepsilon(\gamma)x \\ &= j\gamma j^{-1} j x - \varepsilon(\gamma) j x \\ &= (j\gamma j^{-1})(jx) - \varepsilon(j\gamma j^{-1})(jx). \end{aligned}$$

□

If f is a modular form, let f^* be the holomorphic function $\overline{f(-\overline{z})}$, where the bar denotes complex conjugation. The Fourier coefficients of f^* are the complex conjugates of those of f; though f^* is again a holomorphic modular form, its character is $\overline{\varepsilon}$ instead of ε. The pairing of Theorem 2.7 is the restriction of a pairing on the full spaces without character, and we have the following proposition.

Proposition 2.11. *We have*

$$\langle f^*, x^* \rangle = \overline{\langle f, x \rangle}.$$

Definition 2.12 (Plus-one quotient). The *plus-one quotient* $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)_+$ is the largest torsion-free quotient of $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ by the relations $x^* - x = 0$ for all $x \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$. Similarly, the *minus-one quotient* is the quotient of $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ by all relations $x^* + x = 0$, for $x \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$.

WARNING 2.13. We were forced to make a choice in our definition of the operator $*$. Fortunately, it agrees with that of [16, §2.1.3], but *not* with the choice made in [45, §1.6].

2.4.3 The Atkin-Lehner involutions

In this section we assume that k is even and $\varepsilon^2 = 1$. The assumption on ε is necessary only so that the involution we are about to define preserves $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$. More generally, it is possible to define a map which sends $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ to $\boldsymbol{\mathcal{M}}_k(N,\overline{\varepsilon})$.

To each divisor d of N such that $(d, N/d) = 1$ there is an *Atkin-Lehner involution* W_d of $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$, which is defined as follows. Using the Euclidean algorithm, choose integers x, y, z, w such that

$$dxw - (N/d)yz = 1.$$

Next let $g = \left(\begin{smallmatrix} dx & y \\ Nz & dw \end{smallmatrix}\right)$ and define

$$W_d(x) := \frac{1}{d^{\frac{k-2}{2}}} \cdot g(x).$$

For example, when $d = N$ we have $g = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$. The factor of $d^{\frac{k-2}{2}}$ is necessary to normalize W_d so that it is an involution.

On modular forms there is an Atkin-Lehner involution, also denoted W_d, which acts by $W_d(f) = f|[W_d]_k$. These two like-named involutions are compatible with the integration pairing:

$$\langle W_d(f), x\rangle = \langle f, W_d(x)\rangle.$$

2.5 Degeneracy maps

In this section, we describe natural maps between spaces of modular symbols of different levels.

Fix a positive integer N and a Dirichlet character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$. Let M be a positive divisor of N that is divisible by the conductor of ε, in the sense that ε factors through $(\mathbf{Z}/M\mathbf{Z})^*$ via the natural map $(\mathbf{Z}/N\mathbf{Z})^* \to (\mathbf{Z}/M\mathbf{Z})^*$ composed with some uniquely defined character $\varepsilon' : (\mathbf{Z}/M\mathbf{Z})^* \to \mathbf{C}^*$. For any positive divisor t of N/M, let $T = \left(\begin{smallmatrix} 1 & 0 \\ 0 & t \end{smallmatrix}\right)$ and fix a choice $D_t = \{T\gamma_i : i = 1, \ldots, n\}$ of coset representatives for $\Gamma_0(N)\backslash T\Gamma_0(M)$.

WARNING 2.14. There is a mistake in [45, §2.6]: The quotient "$\Gamma_1(N)\backslash\Gamma_1(M)T$" should be replaced by "$\Gamma_1(N)\backslash T\Gamma_1(M)$".

Proposition 2.15. *For each divisor t of N/M there are well-defined linear maps*

$$\alpha_t : \boldsymbol{\mathcal{M}}_k(N, \varepsilon) \to \boldsymbol{\mathcal{M}}_k(M, \varepsilon'), \qquad \alpha_t(x) = (tT^{-1})x = \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} x$$

$$\beta_t : \boldsymbol{\mathcal{M}}_k(M, \varepsilon') \to \boldsymbol{\mathcal{M}}_k(N, \varepsilon), \qquad \beta_t(x) = \sum_{T\gamma_i \in D_t} \varepsilon'(\gamma_i)^{-1} T\gamma_i x.$$

Furthermore, $\alpha_t \circ \beta_t$ is multiplication by $t^{k-2} \cdot [\Gamma_0(M) : \Gamma_0(N)]$.

Proof. To show that α_t is well defined, we must show that for each $x \in \boldsymbol{\mathcal{M}}_k(N, \varepsilon)$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, that we have

$$\alpha_t(\gamma x - \varepsilon(\gamma)x) = 0 \in \boldsymbol{\mathcal{M}}_k(M, \varepsilon').$$

We have

$$\alpha_t(\gamma x) = \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} \gamma x = \begin{pmatrix} a & tb \\ c/t & d \end{pmatrix} \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} x = \varepsilon'(a) \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} x,$$

so

$$\alpha_t(\gamma x - \varepsilon(\gamma)x) = \varepsilon'(a)\alpha_t(x) - \varepsilon(\gamma)\alpha_t(x) = 0.$$

We next verify that β_t is well defined. Suppose that $x \in \boldsymbol{\mathcal{M}}_k(M, \varepsilon')$ and $\gamma \in \Gamma_0(M)$; then $\varepsilon'(\gamma)^{-1}\gamma x = x$, so

$$\begin{aligned}\beta_t(x) &= \sum_{T\gamma_i \in D_t} \varepsilon'(\gamma_i)^{-1} T\gamma_i \varepsilon'(\gamma)^{-1}\gamma x \\ &= \sum_{T\gamma_i\gamma \in D_t} \varepsilon'(\gamma_i\gamma)^{-1} T\gamma_i\gamma x.\end{aligned}$$

This computation shows that the definition of β_t does not depend on the choice D_t of coset representatives. To finish the proof that β_t is well defined we must show that, for $\gamma \in \Gamma_0(M)$, we have $\beta_t(\gamma x) = \varepsilon'(\gamma)\beta_t(x)$ so that β_t respects the relations that define $\boldsymbol{\mathcal{M}}_k(M, \varepsilon)$. Using that β_t does not depend on the choice of coset representative, we find that for $\gamma \in \Gamma_0(M)$,

$$\begin{aligned}\beta_t(\gamma x) &= \sum_{T\gamma_i \in D_t} \varepsilon'(\gamma_i)^{-1} T\gamma_i\gamma x \\ &= \sum_{T\gamma_i\gamma^{-1} \in D_t} \varepsilon'(\gamma_i\gamma^{-1})^{-1} T\gamma_i\gamma^{-1}\gamma x \\ &= \varepsilon'(\gamma)\beta_t(x).\end{aligned}$$

To compute $\alpha_t \circ \beta_t$, we use that $\#D_t = [\Gamma_0(N) : \Gamma_0(M)]$:

$$\begin{aligned}\alpha_t(\beta_t(x)) &= \alpha_t\left(\sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1} T\gamma_i x\right) \\ &= \sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1}(tT^{-1})T\gamma_i x \\ &= t^{k-2}\sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1}\gamma_i x \\ &= t^{k-2}\sum_{T\gamma_i} x \\ &= t^{k-2} \cdot [\Gamma_0(N) : \Gamma_0(M)] \cdot x.\end{aligned}$$

The scalar factor of t^{k-2} appears instead of t, because t is acting on x as an element of $\mathrm{GL}_2(\mathbf{Q})$ *not* as an an element of $\mathbf{Q}$. □

Definition 2.16 (New and old modular symbols). The subspace $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)^{\mathrm{new}}$ of *new modular symbols* is the intersection of the kernels of the α_t as t runs through all positive divisors of N/M and M runs through positive divisors of M strictly less than N and divisible by the conductor of ε. The subspace $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)^{\mathrm{old}}$ of *old modular symbols* is the subspace generated by the images of the β_t where t runs through all positive divisors of N/M and M runs through positive divisors of M strictly less than N and divisible by the conductor of ε.

WARNING: The new and old subspaces need not be disjoint, as the following example illustrates! This is contrary to the statement on page 80 of [45].

Example 2.17. We justify the above warning. Consider, for example, the case $N = 6$, $k = 2$, and trivial character. The spaces $\boldsymbol{\mathcal{M}}_2(2)$ and $\boldsymbol{\mathcal{M}}_2(3)$ are each of dimension 1, and each is generated by the modular symbol $\{\infty, 0\}$. The space $\boldsymbol{\mathcal{M}}_2(6)$ is of dimension 3, and is generated by the 3 modular symbols $\{\infty, 0\}$, $\{-1/4, 0\}$, and $\{-1/2, -1/3\}$. The space generated by the 2 images of $\boldsymbol{\mathcal{M}}_2(2)$ under the 2 degeneracy maps has dimension 2, and likewise for $\boldsymbol{\mathcal{M}}_2(3)$. Together these images generate $\boldsymbol{\mathcal{M}}_2(6)$, so $\boldsymbol{\mathcal{M}}_2(6)$ is equal to its old subspace. However, the new subspace is nontrivial because the two degeneracy maps $\boldsymbol{\mathcal{M}}_2(6) \to \boldsymbol{\mathcal{M}}_2(2)$ are equal, as are the two degeneracy maps $\boldsymbol{\mathcal{M}}_2(6) \to \boldsymbol{\mathcal{M}}_2(3)$. In particular, the intersection of the kernels of the degeneracy maps has dimension at least 1 (in fact, it equals 1).

Computationally, it appears that something similar to this happens if and only if the weight is 2, the character is trivial, and the level is composite. This behavior is probably related to the nonexistence of a characteristic 0 Eisenstein series of weight 2 and level 1.

The following tempting argument is incorrect; the error lies in the fact that an element of the old subspace is a *linear combination* of $\beta_t(y)$'s for various y's and t's: "If x is in both the new and old subspace, then $x = \beta_t(y)$ for some modular symbol y of lower level. This implies $x = 0$ because

$$0 = \alpha_t(x) = \alpha_t(\beta_t(y)) = t^{k-2} \cdot [\Gamma_0(N) : \Gamma_0(M)] \cdot y."$$

Remark 2.18. The map $\beta_t \circ \alpha_t$ cannot in general be multiplication by a scalar since $\boldsymbol{\mathcal{M}}_k(M, \varepsilon')$ usually has smaller dimension than $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$.

2.5.1 Computing coset representatives

Definition 2.19 (Projective line mod N). Let N be a positive integer. Then the *projective line* $\mathbf{P}^1(N)$ is the set of pairs (a, b), with $a, b \in \mathbf{Z}/N\mathbf{Z}$ and $\gcd(a, b, N) = 1$, modulo the eqivalence relation which identifies (a, b) and (a', b') if and only if $ab' \equiv ba' \pmod{N}$.

Let M be a positive divisor of N and t a divisor of N/M. The following *random* algorithm computes a set D_t of representatives for the orbit space $\Gamma_0(M)\backslash T\Gamma_0(N)$. There are deterministic algorithms for computing D_t, but all of the ones the author has found are *vastly* less efficient than the following random algorithm.

Algorithm 2.20. Let $\Gamma_0(N/t, t)$ denote the subgroup of $\mathrm{SL}_2(\mathbf{Z})$ consisting of matrices that are upper triangular modulo N/t and lower triangular modulo t. Observe that two right cosets of $\Gamma_0(N/t, t)$ in $\mathrm{SL}_2(\mathbf{Z})$, represented by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$, are equivalent if and only if $(a, b) = (a', b')$ as points of $\mathbf{P}^1(t)$ and $(c, d) = (c', d')$ as points of $\mathbf{P}^1(N/t)$. Using the following algorithm, we compute right coset representatives for $\Gamma_0(N/t, t)$ inside $\Gamma_0(M)$.

1. Compute the number $[\Gamma_0(M) : \Gamma_0(N)]$ of cosets.

2. Compute a random element $x \in \Gamma_0(M)$.

3. If x is not equivalent to anything generated so far, add it to the list.

4. Repeat steps (2) and (3) until the list is as long as the bound of step (1).

There is a natural bijection between $\Gamma_0(N)\backslash T\Gamma_0(M)$ and $\Gamma_0(N/t,t)\backslash\Gamma_0(M)$, under which $T\gamma$ corresponds to γ. Thus we obtain coset representatives for $\Gamma_0(N)\backslash T\Gamma_0(M)$ by left multiplying each coset representative of $\Gamma_0(N/t,t)\backslash\Gamma_0(M)$ by T.

2.5.2 Compatibility with modular forms

The degeneracy maps defined above are compatible with the corresponding degeneracy maps $\tilde{\alpha}_t$ and $\tilde{\beta}_t$ on modular forms. This is because the degeneracy maps on modular forms are defined by summing over the same coset representatives D_t. Thus we have the following compatibilities.

$$\begin{aligned}\langle\tilde{\alpha}_t(f),x\rangle &= \langle f,\alpha_t(x)\rangle,\\ \langle\tilde{\beta}_t(f),x\rangle &= \langle f,\beta_t(x)\rangle.\end{aligned}$$

If p is prime to N, then $T_p\alpha_t = \alpha_t T_p$ and $T_p\beta_t = \beta_t T_p$.

2.6 Manin symbols

From the definition given in Section 2.1, it is not obvious that $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ is of finite rank. The Manin symbols provide a finite presentation of $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ that is vastly more useful from a computational point of view.

Definition 2.21 (Manin symbols). The *Manin symbols* are the set of pairs

$$[P(X,Y),(u,v)]$$

where $P(X,Y)\in V_{k-2}$ and $0\le u,v<N$ with $\gcd(u,v,N)=1$.

Define a *right* action of $\mathrm{GL}_2(\mathbf{Q})$ on the free $\mathbf{Z}[\varepsilon]$-module M generated by the Manin symbols as follows. The element $g=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ acts by

$$[P,(u,v)]g = [g^{-1}P(X,Y),(u,v)g] = [P(aX+bY,cX+dY),(au+cv,bu+dv)].$$

Let $\sigma=\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$ and $\tau=\left(\begin{smallmatrix}0&-1\\1&-1\end{smallmatrix}\right)$. Let $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)'$ be the largest torsion-free quotient of M by the relations

$$\begin{aligned}x+x\sigma&=0,\\ x+x\tau+x\tau^2&=0,\\ \varepsilon(\lambda)[P,(u,v)]-[P,(\lambda u,\lambda v)]&=0.\end{aligned}$$

Theorem 2.22. *There is a natural isomorphism $\varphi:\boldsymbol{\mathcal{M}}_k(N,\varepsilon)'\longrightarrow\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ given by*

$$[X^iY^{2-k-i},(u,v)]\mapsto g(X^iY^{k-2-i}\{0,\infty\})$$

where $g=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbf{Z})$ is any matrix such that $(u,v)\equiv(c,d)\pmod N$.

Proof. In [45, §1.2, §1.7] it is proved that $\varphi\otimes_{\mathbf{Z}[\varepsilon]}\mathbf{C}$ is an isomorphism, so φ is injective and well defined. The discussion in Section 2.6.1 below ("Manin's trick") shows that every element in $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ is a $\mathbf{Z}[\varepsilon]$-linear combination of elements in the image, so φ is surjective as well. □

2.6.1 Conversion between modular and Manin symbols

For some purposes it is better to work with modular symbols, and for others it is better to work with Manin symbols. For example, there are descriptions of the Atkin-Lehner involution in terms of both Manin and modular symbols; it appears more efficient to compute this involution using modular symbols. On the other hand, practically Hecke operators can be computed more efficiently using Manin symbols. It is thus essential to be able to convert between these two representations. The conversion from Manin to modular symbols is straightforward, and follows immediately from Theorem 2.22. The conversion back is accomplished using the method used to prove Theorem 2.22; it is known as "Manin's trick", and involves continued fractions.

Given a Manin symbol $[X^iY^{k-2-i},(u,v)]$, we write down a corresponding modular symbol as follows. Choose $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ such that $(c,d) \equiv (u,v) \pmod{N}$. This is possible by Lemma 1.38 of [62, pg. 20]; in practice, finding $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is not completely trivial, but can be accomplished using the extended Euclidean algorithm. Then

$$\begin{aligned}[X^iY^{k-2-i},(u,v)] \quad \longleftrightarrow \quad & \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(X^iY^{k-2-i}\{0,\infty\}) \\ &= (dX-bY)^i(-cX+aY)^{2-k-i}\left\{\frac{b}{d},\frac{a}{c}\right\}.\end{aligned}$$

In the other direction, suppose that we are given a modular symbol $P(X,Y)\{\alpha,\beta\}$ and wish to represent it as a sum of Manin symbols. Because

$$P\{a/b,c/d\} = P\{a/b,0\}+P\{0,c/d\},$$

it suffices to write $P\{0,a/b\}$ in terms of Manin symbols. Let

$$0=\frac{p_{-2}}{q_{-2}}=\frac{0}{1},\ \frac{p_{-1}}{q_{-1}}=\frac{1}{0},\ \frac{p_0}{1}=\frac{p_0}{q_0},\ \frac{p_1}{q_1},\ \frac{p_2}{q_2},\ \dots,\ \frac{p_r}{q_r}=\frac{a}{b}$$

denote the continued fraction convergents of the rational number a/b. Then

$$p_jq_{j-1}-p_{j-1}q_j = (-1)^{j-1} \qquad \text{for } -1\le j\le r.$$

If we let $g_j = \begin{pmatrix} (-1)^{j-1}p_j & p_{j-1} \\ (-1)^{j-1}q_j & q_{j-1}\end{pmatrix}$, then $g_j \in \mathrm{SL}_2(\mathbf{Z})$ and

$$\begin{aligned} P(X,Y)\{0,a/b\} &= P(X,Y)\sum_{j=-1}^{r}\left\{\frac{p_{j-1}}{q_{j-1}},\frac{p_j}{q_j}\right\} \\ &= \sum_{j=-1}^{r} g_j((g_j^{-1}P(X,Y))\{0,\infty\}) \\ &= \sum_{j=-1}^{r}[g_j^{-1}P(X,Y),((-1)^{j-1}q_j,q_{j-1})].\end{aligned}$$

Note that in the jth summand, $g_j^{-1}P(X,Y)$, replaces $P(X,Y)$. Since $g_j \in \mathrm{SL}_2(\mathbf{Z})$ and $P(X,Y)$ has integer coefficients, the polynomial $g_j^{-1}P(X,Y)$ also has integer coefficients, so no denominators are introduced.

The continued fraction expansion $[c_1, c_2, \ldots, c_n]$ of the rational number a/b can be computed using the Euclidean algorithm. The first term, c_1, is the "quotient": $a = bc_1 + r$, with $0 \leq r < b$. Let $a' = b$, $b' = r$ and compute c_2 as $a' = b'c_2 + r'$, etc., terminating when the remainder is 0. For example, the expansion of $5/13$ is $[0, 2, 1, 1, 2]$. The numbers

$$d_i = c_1 + \cfrac{1}{c_2 + \frac{1}{c_3 + \cdots}}$$

will then be the (finite) convergents. For example if $a/b = 5/13$, then the convergents are

$$0/1,\ 1/0,\ d_1 = 0,\ d_2 = \frac{1}{2},\ d_3 = \frac{1}{3},\ d_4 = \frac{2}{5},\ d_5 = \frac{5}{13}.$$

2.6.2 Hecke operators on Manin symbols

Thoerem 2 of [45] gives a description of the Hecke operators T_n directly on the space of Manin symbols. This avoids the expense of first converting a Manin symbol to a modular symbol, computing T_n on the modular symbol, and then converting back. For the reader's convenience, we very briefly recall Merel's theorem here, along with an enhancement due to Cremona.

As in [16, §2.4], define R_p as follows. When $p = 2$,

$$R_2 := \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} \right\}.$$

When p is odd, R_p is the set of 2×2 integer matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with determinant p, and either

1. $a > |b| > 0$, $d > |c| > 0$, and $bc < 0$; or
2. $b = 0$, and $|c| < d/2$; or
3. $c = 0$, and $|b| < a/2$.

Proposition 2.23. *For* $[P(X,Y),(u,v)] \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ *and* p *a prime, we have*

$$\begin{aligned} T_p([P(X,Y),(u,v)]) &= \sum_{g \in R_p} [P(X,Y),(u,v)].g \\ &= \sum_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in R_p} [P(aX+bY, cX+dY), (au+cv, bu+dv)], \end{aligned}$$

where the sum is restricted to matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ *such that* $\gcd(au+cv, bu+dv, N) = 1$.

Proof. For the case $k = 2$ and an algorithm to compute R_p, see [16, §2.4]. The general case follows from [45, Theorem 2] applied to the set $\boldsymbol{\mathcal{S}}$ of [45, §3] by observing that when p is an odd *prime* $\boldsymbol{\mathcal{S}}'_p$ is empty. □

2.6.3 The cuspidal and boundary spaces in terms of Manin symbols

This section is a review of Merel's explicit description of the boundary map in terms of Manin symbols for $\Gamma = \Gamma_1(N)$ (see [45, §1.4]). In the next section, we describe a very efficient way to compute the boundary map.

Let $\mathcal{R}$ be the equivalence relation on $\Gamma\backslash\mathbf{Q}^2$ which identifies the element $[\Gamma\left(\begin{smallmatrix}\lambda u\\ \lambda v\end{smallmatrix}\right)]$ with $\mathrm{sign}(\lambda)^k[\Gamma\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$, for any $\lambda \in \mathbf{Q}^*$. Denote by $B_k(\Gamma)$ the finite dimensional $\mathbf{Q}$-vector space with basis the equivalence classes $(\Gamma\backslash\mathbf{Q}^2)/\mathcal{R}$. The dimension of this space is $\#(\Gamma\backslash\mathbf{P}^1(\mathbf{Q}))$.

Proposition 2.24. *The map*

$$\mu : \boldsymbol{\mathcal{B}}_k(\Gamma) \to B_k(\Gamma), \qquad P\left\{\frac{u}{v}\right\} \mapsto P(u,v)\left[\Gamma\begin{pmatrix}u\\ v\end{pmatrix}\right]$$

is well defined and injective. Here u and v are assumed coprime.

Thus the kernel of $\delta : \boldsymbol{\mathcal{S}}_k(\Gamma) \to \boldsymbol{\mathcal{B}}_k(\Gamma)$ is the same as the kernel of $\mu\circ\delta$.

Proposition 2.25. *Let $P \in V_{k-2}$ and $g = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$. We have*

$$\mu\circ\delta([P,(c,d)]) = P(1,0)[\Gamma\left(\begin{smallmatrix}a\\ c\end{smallmatrix}\right)] - P(0,1)[\Gamma\left(\begin{smallmatrix}b\\ d\end{smallmatrix}\right)].$$

2.6.4 Computing the boundary map

In this section we describe how to compute the map $\delta : \boldsymbol{\mathcal{M}}_k(N,\varepsilon) \to B_k(N,\varepsilon)$ given in the previous section. The algorithm presented here generalizes the one in [16, §2.2]. To compute the image of $[P,(c,d)]$, with $g = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$, we must compute the class of $[\left(\begin{smallmatrix}a\\ c\end{smallmatrix}\right)]$ and of $[\left(\begin{smallmatrix}b\\ d\end{smallmatrix}\right)]$. Instead of finding a canonical form for cusps, we use a quick test for equivalence modulo scalars. In the following algorithm, by the ith standard cusp we mean the ith basis vector for a basis of $B_k(N,\varepsilon)$. The basis is constructed as the algorithm is called successively. We first give the algorithm, then prove the facts used by the algorithm in testing equivalence.

Algorithm 2.26. Given a cusp $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$ this algorithm computes an integer i and a scalar α such that $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$ is equivalent to α times the ith standard cusp. First, using Proposition 2.27 and Algorithm 2.28, check whether or not $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$ is equivalent, modulo scalars, to any cusp found so far. If so, return the index of the representative and the scalar. If not, record $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ in the representative list. Then, using Proposition 2.30, check whether or not $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ is forced to equal zero by the relations. If it does not equal zero, return its position in the list and the scalar 1. If it equals zero, return the scalar 0 and the position 1; keep $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ in the list, and record that it is zero.

In the case considered in Cremona's book [16], the relations between cusps involve only the trivial character, so they do not force any cusp classes to vanish. Cremona gives the following two criteria for equivalence.

Proposition 2.27 (Cremona). *Let $\left(\begin{smallmatrix}u_i\\ v_i\end{smallmatrix}\right)$, $i = 1,2$ be written so that* $\gcd(u_i, v_i) = 1$.

1. *There exists $g \in \Gamma_0(N)$ such that $g\left(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix}\right)$ if and only if*

$$s_1 v_2 \equiv s_2 v_1 \pmod{\gcd(v_1 v_2, N)}, \text{ where } s_j \text{ satisfies } u_j s_j \equiv 1 \pmod{v_j}.$$

2. *There exists $g \in \Gamma_1(N)$ such that $g\left(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix}\right)$ if and only if*

$$v_2 \equiv v_1 \pmod{N} \text{ and } u_2 \equiv u_1 \pmod{\gcd(v_1, N)}.$$

Proof. The first is Proposition 2.2.3 of [16], and the second is Lemma 3.2 of [15]. □

Algorithm 2.28. Suppose $\left(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix}\right)$ are equivalent modulo $\Gamma_0(N)$. This algorithm computes a matrix $g \in \Gamma_0(N)$ such that $g\left(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix}\right)$. Let s_1, s_2, r_1, r_2 be solutions to $s_1 u_1 - r_1 v_1 = 1$ and $s_2 u_2 - r_2 v_2 = 1$. Find integers x_0 and y_0 such that $x_0 v_1 v_2 + y_0 N = 1$. Let $x = -x_0(s_1 v_2 - s_2 v_1)/(v_1 v_2, N)$ and $s_1' = s_1 + x v_1$. Then $g = \begin{pmatrix} u_2 & r_2 \\ v_2 & s_2 \end{pmatrix} \cdot \begin{pmatrix} u_1 & r_1 \\ v_1 & s_1' \end{pmatrix}^{-1}$ sends $\left(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}\right)$ to $\left(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix}\right)$.

Proof. This follows from the proof of Proposition 2.27 in [16]. □

To see how the ε relations, for nontrivial ε, make the situation more complicated, observe that it is possible that $\varepsilon(\alpha) \neq \varepsilon(\beta)$ but

$$\varepsilon(\alpha)\left[\begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\gamma_\alpha \begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\gamma_\beta \begin{pmatrix} u \\ v \end{pmatrix}\right] = \varepsilon(\beta)\left[\begin{pmatrix} u \\ v \end{pmatrix}\right];$$

One way out of this difficulty is to construct the cusp classes for $\Gamma_1(N)$, then quotient out by the additional ε relations using Gaussian elimination. This is far too inefficient to be useful in practice because the number of $\Gamma_1(N)$ cusp classes can be unreasonably large. Instead, we give a quick test to determine whether or not a cusp vanishes modulo the ε-relations.

Lemma 2.29. *Suppose α and α' are integers such that $\gcd(\alpha, \alpha', N) = 1$. Then there exist integers β and β', congruent to α and α' modulo N, respectively, such that $\gcd(\beta, \beta') = 1$.*

Proof. By [62, 1.38] the map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective. By the Euclidean algorithm, there exist integers x, y and z such that $x\alpha + y\alpha' + zN = 1$. Consider the matrix $\left(\begin{smallmatrix} y & -x \\ \alpha & \alpha' \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ and take β, β' to be the bottom row of a lift of this matrix to $\mathrm{SL}_2(\mathbf{Z})$. □

Proposition 2.30. *Let N be a positive integer and ε a Dirichlet character of modulus N. Suppose $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ is a cusp with u and v coprime. Then $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ vanishes modulo the relations*

$$\left[\gamma\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)\right] = \varepsilon(\gamma)\left[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)\right], \qquad \text{all } \gamma \in \Gamma_0(N)$$

if and only if there exists $\alpha \in (\mathbf{Z}/N\mathbf{Z})^*$, with $\varepsilon(\alpha) \neq 1$, such that

$$\begin{aligned} v &\equiv \alpha v \pmod{N}, \\ u &\equiv \alpha u \pmod{\gcd(v, N)}. \end{aligned}$$

Proof. First suppose such an α exists. By Lemma 2.29 there exists $\beta, \beta' \in \mathbf{Z}$ lifting α, α^{-1} such that $\gcd(\beta, \beta') = 1$. The cusp $\left(\begin{smallmatrix}\beta u\\ \beta' v\end{smallmatrix}\right)$ has coprime coordinates so, by Proposition 2.27 and our congruence conditions on α, the cusps $\left(\begin{smallmatrix}\beta u\\ \beta' v\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ are equivalent by an element of $\Gamma_1(N)$. This implies that $\left[\left(\begin{smallmatrix}\beta u\\ \beta' v\end{smallmatrix}\right)\right] = [\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$. Since $\left[\left(\begin{smallmatrix}\beta u\\ \beta' v\end{smallmatrix}\right)\right] = \varepsilon(\alpha)\,[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$, our assumption that $\varepsilon(\alpha) \neq 1$ forces $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)] = 0$.

Conversely, suppose $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)] = 0$. Because all relations are two-term relations, and the $\Gamma_1(N)$-relations identify $\Gamma_1(N)$-orbits, there must exists α and β with

$$\left[\gamma_\alpha \begin{pmatrix}u\\ v\end{pmatrix}\right] = \left[\gamma_\beta \begin{pmatrix}u\\ v\end{pmatrix}\right] \qquad \text{and } \varepsilon(\alpha) \neq \varepsilon(\beta).$$

Indeed, if this did not occur, then we could mod out by the ε relations by writing each $[\gamma_\alpha\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$ in terms of $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$, and there would be no further relations left to kill $[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$. Next observe that

$$\left[\gamma_{\beta^{-1}\alpha}\begin{pmatrix}u\\ v\end{pmatrix}\right] = \left[\gamma_{\beta^{-1}}\gamma_\alpha\begin{pmatrix}u\\ v\end{pmatrix}\right] = \varepsilon(\beta^{-1})\left[\gamma_\alpha\begin{pmatrix}u\\ v\end{pmatrix}\right] = \varepsilon(\beta^{-1})\left[\gamma_\beta\begin{pmatrix}u\\ v\end{pmatrix}\right] = \left[\begin{pmatrix}u\\ v\end{pmatrix}\right].$$

Applying Proposition 2.27 and noting that $\varepsilon(\beta^{-1}\alpha) \neq 1$ shows that $\beta^{-1}\alpha$ satisfies the properties of the "α" in the statement of the proposition we are proving. □

We enumerate the possible α appearing in Proposition 2.30 as follows. Let $g = (v, N)$ and list the $\alpha = v \cdot \frac{N}{g} \cdot a + 1$, for $a = 0, \ldots, g-1$, such that $\varepsilon(\alpha) \neq 0$.

Working in the plus one or minus one quotient. Let s be a sign, either $+1$ or -1. To compute $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)_s$ it is necessary to replace $B_k(N, \varepsilon)$ by its quotient modulo the additional relations $[\left(\begin{smallmatrix}-u\\ v\end{smallmatrix}\right)] = s\,[\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)]$ for all cusps $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$. Algorithm 2.26 can be modified to deal with this situation as follows. Given a cusp $x = \left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$, proceed as in Algorithm 2.26 and check if either $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}-u\\ v\end{smallmatrix}\right)$ is equivalent (modulo scalars) to any cusp seen so far. If not, use the following trick to determine whether the ε and s-relations kill the class of $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$: use the unmodified Algorithm 2.26 to compute the scalars α_1, α_2 and standard indices i_1, i_2 associated to $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}-u\\ v\end{smallmatrix}\right)$, respectively. The s-relation kills the class of $\left(\begin{smallmatrix}u\\ v\end{smallmatrix}\right)$ if and only if $i_1 = i_2$ but $\alpha_1 \neq s\alpha_2$.

2.7 The complex torus attached to a modular form

Fix integers $N \geq 1$, $k \geq 2$, and let ε be a mod N Dirichlet character. For the rest of this section assume that $\varepsilon^2 = 1$.

We construct a lattice in $\mathrm{Hom}(S_k(N, \varepsilon), \mathbf{C})$ that is invariant under complex conjugation and under the action of the Hecke operators. The quotient of $\mathrm{Hom}(S_k(N, \varepsilon), \mathbf{C})$ by this lattice is a complex torus $J_k(N, \varepsilon)$, which is equipped with an action of the Hecke operators and of complex conjugation.

The reader may wish to compare our construction with a closely related construction of Shimura [60]. Shimura observes that the Petersson pairing gives his torus the structure of an abelian variety over $\mathbf{C}$. Note that his torus is, a priori, different than our torus. We do not know if our torus has the structure of abelian variety over $\mathbf{C}$.

When $k = 2$, the torus $J_2(N, \varepsilon)$ is the set of complex points of an abelian variety, which is actually defined over $\mathbf{Q}$; when $k > 2$, the study of these complex tori is of interest in trying to understand the conjectures of Bloch and Kato (see [7]) on motifs attached to modular forms.

Let $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ (respectively, $S = S_k(N, \varepsilon)$) be the space of cuspidal modular symbols (respectively, cusp forms) of weight k, level N, and character ε. The Hecke algebra $\mathbf{T}$ acts in a way compatible with the integration pairing $\langle\, ,\rangle : S \times \boldsymbol{\mathcal{S}} \to \mathbf{C}$. This pairing induces a $\mathbf{T}$-module homomorphism $\Phi : \boldsymbol{\mathcal{S}} \to S^* = \mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C})$, called the *period mapping.* Because $\varepsilon^2 = 1$, the $*$-involution preserves S.

Proposition 2.31. *The period mapping Φ is injective and $\Phi(\boldsymbol{\mathcal{S}})$ is a lattice in S^*.*

Proof. By Theorem 2.7,

$$\boldsymbol{\mathcal{S}} \otimes_{\mathbf{R}} \mathbf{C} \cong \mathrm{Hom}_{\mathbf{C}}(S \oplus \overline{S}, \mathbf{C}).$$

Because $\varepsilon^2 = 1$, we have $S = S_k(N, \varepsilon; \mathbf{R}) \otimes_{\mathbf{R}} \mathbf{C}$. Set $S_{\mathbf{R}} := S_k(N, \varepsilon; \mathbf{R})$ and likewise define $\overline{S}_{\mathbf{R}}$. We have

$$\mathrm{Hom}_{\mathbf{C}}(S \oplus \overline{S}, \mathbf{C}) = \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}} \oplus \overline{S}_{\mathbf{R}}, \mathbf{R}) \otimes_{\mathbf{R}} \mathbf{C}.$$

Let $\boldsymbol{\mathcal{S}}_{\mathbf{R}} = \boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{R})$ and $\boldsymbol{\mathcal{S}}_{\mathbf{R}}^+$ be the subspace fixed under $*$. By Proposition 2.11 we have maps

$$\boldsymbol{\mathcal{S}}_{\mathbf{R}}^+ \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}} \oplus \overline{S}_{\mathbf{R}}, \mathbf{R}) \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, \mathbf{R})$$

and

$$\boldsymbol{\mathcal{S}}_{\mathbf{R}}^- \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}} \oplus \overline{S}_{\mathbf{R}}, i\mathbf{R}) \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, i\mathbf{R}).$$

The map $\boldsymbol{\mathcal{S}}_{\mathbf{R}}^+ \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, \mathbf{R})$ is an isomorphism: the point is that if $\langle \bullet, x\rangle$, for $x \in \boldsymbol{\mathcal{S}}_{\mathbf{R}}^+$, vanishes on $S_{\mathbf{R}}$ then it vanishes on the whole of $S \oplus \overline{S}$. Likewise, the map $\boldsymbol{\mathcal{S}}_{\mathbf{R}}^- \to \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, i\mathbf{R})$ is an isomorphism. Thus

$$\boldsymbol{\mathcal{S}} \otimes \mathbf{R} = \boldsymbol{\mathcal{S}}_{\mathbf{R}} \cong \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, \mathbf{R}) \oplus \mathrm{Hom}_{\mathbf{R}}(S_{\mathbf{R}}, i\mathbf{R}) \cong \mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C}).$$

Finally, we observe that $\boldsymbol{\mathcal{S}}$ is by definition torsion free, which completes the proof. □

The torus $J_k(N, \varepsilon)$ fits into an exact sequence

$$0 \longrightarrow \boldsymbol{\mathcal{S}} \xrightarrow{\ \Phi\ } \mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C}) \longrightarrow J_k(N, \varepsilon) \longrightarrow 0.$$

Let $f \in S$ be a newform and S_f the complex vector space spanned by the Galois conjugates of f. The period map Φ_f associated to f is the map $\boldsymbol{\mathcal{S}} \to \mathrm{Hom}_{\mathbf{C}}(S_f, \mathbf{C})$ obtained by composing Φ with restriction to S_f. Set

$$A_f := \mathrm{Hom}_{\mathbf{C}}(S_f, \mathbf{C})/\Phi_f(\boldsymbol{\mathcal{S}}).$$

We associate to f a subtorus of J as follows. Let $I_f = \mathrm{Ann}_{\mathbf{T}}(f)$ be the annihilator of f in the Hecke algebra, and set

$$A_f^\vee := \mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C})[I_f]/\Phi(\boldsymbol{\mathcal{S}}[I_f])$$

where $\mathrm{Hom}_{\mathbf{C}}(S,\mathbf{C})[I_f] = \cap_{t\in I_f}\ker(t)$.

The following diagram summarizes the tori just defined; its columns are exact but its rows need not be.

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] \\
\boldsymbol{\mathcal{S}}[I_f] \arrow[r] \arrow[d] & \boldsymbol{\mathcal{S}} \arrow[r] \arrow[d] & \Phi_f(\boldsymbol{\mathcal{S}}) \arrow[d] \\
\mathrm{Hom}_{\mathbf{C}}(S,\mathbf{C})[I_f] \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathbf{C}}(S,\mathbf{C}) \arrow[r] \arrow[d] & \mathrm{Hom}_{\mathbf{C}}(S[I_f],\mathbf{C}) \arrow[d] \\
A_f^{\vee} \arrow[r] \arrow[d] & J_k(N,\varepsilon) \arrow[r] \arrow[d] & A_f \arrow[d] \\
0 & 0 & 0
\end{tikzcd} \tag{2.1}$$

2.7.1 The case when the weight is 2

When $k = 2$ and $\varepsilon = 1$ the above is just Shimura's classical association of an abelian variety to a modular form; see [62, Thm. 7.14] and [61]. In this case A_f and $A_f^{\vee}$ are abelian varieties that are defined over $\mathbf{Q}$. Furthermore A_f is an *optimal quotient* of J, in the sense that the kernel of the map $J \to A_f$ is connected. For a summary of the main results in this situation, see Section 4.6.

Chapter 3

Applications of modular symbols

In the previous chapter we introduced several spaces of modular symbols, and observations such as "Manin's trick" suggested that we could compute with them. The duality between modular symbols and modular forms hints that modular symbols might be useful in computing information about modular forms. In the present chapter, we gather together the fruits of our investigation into this connection.

Sections 3.1–3.5 of this chapter give a method to compute the irreducible components of the spaces $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ of modular symbols. In Section 3.6 we observe that computing intersections of certain abelian varieties can be reduced to linear algebra over $\mathbf{Z}$ by viewing the abelian varieties as complex tori and considering the appropriate diagrams. In Sections 3.7, we continue this trend by pointing out that many invariants of the complex torus attached to a modular form can be computed without computing any approximate period lattices. In Section 3.8, we discuss well-known methods for computing both an upper and lower bound on the order of the torsion subgroup of certain abelian varieties. Section 3.9 presents an algorithm for computing the modular degree of the complex torus associated to a newform.

In Section 3.10 we aim squarely at the problem of gathering data related to the Birch and Swinnerton-Dyer conjecture and its generalizations, where we give a formula for the rational numbers $|L(A_f,j)/\Omega_j|$ attached to a newform. In Section 3.11 we compare the ratio computed in the previous section to the one considered in the Birch and Swinnerton-Dyer conjecture; the two numbers differ by a Manin constant, which we bound. Finally, in Section 3.12 we give algorithms for approximating the period lattice and related numerical quantities associated to a newform of arbitrary weight.

3.1 Computing the space of modular symbols

Definition 3.1. Let W be a subspace of a finite-dimensional vector space V. To *compute* the quotient V/W means to give a matrix representing the projection $V \to V/W$, with respect to some basis for V and some basis B for V/W, along with a lift to V of each element of B.

In other words, to compute V/W means to create a reduction function that assigns to each element of V its canonical representative on the "free basis" B.

Let N be a positive integer, fix a mod N Dirichlet character ε, let $K := \mathbf{Q}[\varepsilon]$ be the smallest extension containing the values of ε, and let $\mathcal{O} := \mathbf{Z}[\varepsilon]$.

Algorithm 3.2. Given a positive integer N, a Dirichlet character ε, and an integer $k \geq 2$ this algorithm computes $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)$. We compute the quotient presentation given in Theorem 2.22 in three steps.

1. Let V_1 be the finite-dimensional K-vector space generated by the Manin symbols $[X^iY^{k-2-i}, (u, v)]$ for $i = 0, \dots, k-2$ and $0 \leq u, v < N$ with $\gcd(u, v, N) = 1$. Let W_1 be the subspace of V_1 generated by all differences
$$[X^iY^{k-2-i}, (\lambda u, \lambda v)] - \varepsilon(\lambda)[X^iY^{k-2-i}, (u, v)].$$
Because all relations are two-term, it is easy to compute $V_2 := V_1/W_1$. In computing this quotient, we do not have to explicitly compute the *large* matrix representing the linear map $V_1 \to V_2$, as it can be replaced by a suitable "reduction procedure" involving arithmetic in $\mathbf{Z}/N\mathbf{Z}$.

2. Let σ act on the set of Manin symbols as in Section 2.6; thus
$$[X^iY^{k-2-i}, (u, v)]\sigma = (-1)^i[Y^iX^{k-2-i}, (v, -u)].$$
Let W_2 be the subspace of V_2 generated by the sums $x + x\sigma$ for $x \in V_2$. Because all relations are two-term relations, it is easy to compute $V_3 := V_2/W_2$.

3. Let τ act on Manin symbols as in Section 2.6; thus
$$[X^iY^{k-2-i}, (u, v)]\tau = [(-Y)^i(X-Y)^{k-2-i}, (v, -u-v)].$$
Note that the symbol on the right can be written as a sum of generating Manin symbols. Let W_3 be the subspace of V_3 generated by the sums $x + x\tau + x\tau^2$ where x varies over the images of a basis of V_2 (*not* just a basis for V_3!). Using some form of Gaussian elimination, we compute V_3/W_3. Finally, $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K) \approx V_3/W_3$.

Proof. For $\lambda \in (\mathbf{Z}/N\mathbf{Z})^*$, denote by $\langle\lambda\rangle$ the right action of λ on Manin symbols; thus
$$[X^iY^{k-2-i}, (u, v)]\langle\lambda\rangle = [X^iY^{k-2-i}, (\lambda u, \lambda v)].$$
By Theorem 2.22 the space $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)$ is isomorphic to the quotient of the vector spaces V_1 of Step 1 modulo the relations $x + x\sigma = 0$, $x + x\tau + x\tau^2 = 0$, and $x\langle\lambda\rangle = \lambda x$ as x varies over all Manin symbols and λ varies over $(\mathbf{Z}/N\mathbf{Z})^*$.

As motivation, we note that a naive computation of V_1 modulo the σ, τ, and $\langle\lambda\rangle$ relations using Gaussian elimination is far too inefficient. This is why we compute the quotient in three steps. The complexity of Steps 1 and Steps 2 are negligible. The difficulty occurs in Step 3; at least the relations of this step occur in a space of dimension much smaller than that of V_1.

To see that the algorithm is correct, it is necessary only to observe that σ and τ both commute with all diamond-bracket operators $\langle\lambda\rangle$; this is an immediate consequence of the above formulas. We remark that in Step 3 it is in general *necessary* to compute the quotient by all relations $x + x\tau + x\tau^2$ with x the image of a basis vector for V_2 instead of just x in V_3 because σ and τ do not commute. □

Remark 3.3. In implementing the above algorithm, one should take special care in Steps 1 and 2 because the relations can together force certain of the Manin symbols to equal 0. For example, there might be relations of the form $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$ which together force $x_1 = x_2 = 0$.

Remark 3.4. To compute the plus-one quotient $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)_+$, it is necessary to modify Step 2 of Algorithm 3.2 by including in W_2 the differences $x - xI$ where $I = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and

$$[X^i Y^{k-2-i}, (u, v)]I = (-1)^i [X^i Y^{k-2-i}, (-u, v)].$$

Likewise, to compute the minus-one quotient we include the sums $x + xI$. Note, as in the remarks in the proof of Algorithm 3.2, we can not add in the I relations in Step 1 because I and σ do not commute.

Algorithm 3.5. Given a positive integer N, a Dirichlet character ε, and an integer $k \geq 2$, this algorithm computes the $\mathcal{O}$-modules $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$ and $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$. (We assume as given algorithms for performing standard operations on $\mathcal{O}$-modules.)

1. Using Algorithm 3.2 compute the K-vector space $V := \boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)$.
2. Compute the $\mathcal{O}$-lattice L in V generated by the classes of the finitely many symbols $[X^i Y^{k-2-i}, (u, v)]$ for $i = 0, \dots, k-2$ and $0 \leq u, v < N$ with $\gcd(u, v, N) = 1$. It is only necessary to take one symbol in each ε-equivalence class, so there are $(k - 2 + 1) \cdot \#\mathbf{P}^1(\mathbf{Z}/N\mathbf{Z})$ generating symbols. This computes $\boldsymbol{\mathcal{M}}_k(N, \varepsilon)$.
3. To compute the submodule $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ of L, we use the algorithm of Section 2.6.4 to compute the boundary map $\delta : \boldsymbol{\mathcal{M}}_k(N, \varepsilon; K) \to B_k(N, \varepsilon; K)$. Then $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ is the kernel of δ restricted to the lattice L.

As a check, using the formulas of Section 3.4, we compute the dimension of the space $S_k(N, \varepsilon)$ of cusp forms and compare with the dimension of $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)$ computed in Algorithm 3.5. The latter dimension must equal twice the former one.

3.2 Computing the Hecke algebra

In this section we give an upper bound on the number of Hecke operators needed to generate the Hecke algebra as a $\mathbf{Z}$-module. The bound on Hecke operators is an application of [66], which was described to the author by Ribet and Agashe when $k = 2$ and the level is prime. There are much better bounds on the number of Hecke operators needed to generate the Hecke algebra as a *ring*, but we do not investigate them here.

Let Γ be a subgroup of $\mathrm{SL}_2(\mathbf{Z})$ that contains $\Gamma_1(N)$ for some N. Let $S_k(\Gamma; \mathbf{C})$ be the space of weight-k cuspforms for Γ, and let $\mathbf{T} \subset \mathrm{End}(S_k(\Gamma; \mathbf{C}))$ be the corresponding Hecke algebra. We now give a bound r such that the Hecke operators T_n, with $n \leq r$, generate $\mathbf{T}$ as a $\mathbf{Z}$-module.

For any ring $R \subset \mathbf{C}$, let $S_k(\Gamma; R)$ denotes the space of cuspforms for Γ with Fourier coefficients in R. Since $S_k(\Gamma; \mathbf{C}) = S_k(\Gamma; \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{C}$, it makes sense to define

$$S_k(\Gamma; R) := S_k(\Gamma; \mathbf{Z}) \otimes_{\mathbf{Z}} R$$

for *any* ring R. The following proposition is well known.

Proposition 3.6. *For any ring R, the pairing*

$$\mathbf{T}_R \otimes_R S_k(N;R) \to R$$

that sends (T,f) to $a_1(Tf)$ is a perfect pairing, where $\mathbf{T}_R = \mathbf{T} \otimes_{\mathbf{Z}} R$. Furthermore, we have $(T_n, f) = a_n(f)$, where T_n is the nth Hecke operator.

Let

$$\mu = [\mathrm{SL}_2(\mathbf{Z}) : \Gamma],$$

and denote by $\lceil x \rceil$ the smallest integer $\geq x$.

Theorem 3.7 (Sturm). *Let λ be a prime ideal in the ring $\mathcal{O}$ of integers in some number field. If $f \in S_k(\Gamma;\mathcal{O})$ satisfies $a_n(f) \equiv 0 \pmod{\lambda}$ for $n \leq \lceil \frac{k}{12}\mu \rceil$, then $f \equiv 0 \pmod{\lambda}$.*

Proof. Theorem 1 of [66]. □

Proposition 3.8. *If $f \in S_k(\Gamma)$ satisfies $a_n(f) = 0$ for $n \leq r = \left\lceil \frac{k}{12}\mu \right\rceil$, then $f = 0$.*

Proof. We must show that the composite map $S_k(\Gamma) \hookrightarrow \mathbf{C}[[q]] \to \mathbf{C}[[q]]/(q^{r+1})$ is injective. Because $\mathbf{C}$ is a flat $\mathbf{Z}$-module and $S_k(\Gamma;\mathbf{Z}) \otimes \mathbf{C} = S_k(\Gamma)$, it suffices to show that the map $F : S_k(\Gamma;\mathbf{Z}) \to \mathbf{Z}[[q]]/(q^{r+1})$ is injective. Suppose $F(f) = 0$, and let p be a prime number. Then $a_n(f) = 0$ for $n \leq r$, hence plainly $a_n(f) \equiv 0 \pmod{p}$ for any such n. Theorem 3.7 implies that $f \equiv 0 \pmod{p}$. Duplicating this argument shows that the coefficients of f are divisible by all primes p, so they are 0. □

Theorem 3.9. *As a $\mathbf{Z}$-module, $\mathbf{T}$ is generated by $T_1, \ldots, T_r$, where $r = \lceil \frac{k}{12}\mu \rceil$.*

Proof. Let Z be the submodule of $\mathbf{T}$ generated by $T_1, T_2, \ldots, T_r$. Consider the exact sequence of additive abelian groups $0 \to Z \xrightarrow{i} \mathbf{T} \to \mathbf{T}/Z \to 0$. Let p be a prime and tensor this sequence with $\mathbf{F}_p$ to obtain the exact sequence

$$Z \otimes \mathbf{F}_p \xrightarrow{\bar{i}} \mathbf{T} \otimes \mathbf{F}_p \to (\mathbf{T}/Z) \otimes \mathbf{F}_p \to 0.$$

Put $R = \mathbf{F}_p$ in Proposition 3.6, and suppose that $f \in S_k(N, \mathbf{F}_p)$ pairs to 0 with each of $T_1, \ldots, T_r$. Then by Proposition 3.6, $a_m(f) = a_1(T_m f) = 0$ in $\mathbf{F}_p$ for each m, $1 \leq m \leq r$. Theorem 3.7 then asserts that $f = 0$. Thus the pairing, when restricted to the image of $Z \otimes \mathbf{F}_p$ in $\mathbf{T} \otimes \mathbf{F}_p$, is also perfect. Thus $\dim_{\mathbf{F}_p} \bar{i}(Z \otimes \mathbf{F}_p) = \dim_{\mathbf{F}_p} S_k(N, \mathbf{F}_p) = \dim_{\mathbf{F}_p} \mathbf{T} \otimes \mathbf{F}_p$, so $(\mathbf{T}/Z) \otimes \mathbf{F}_p = 0$; repeating this argument for all p shows that $\mathbf{T}/Z = 0$. □

3.3 Representing and enumerating Dirichlet characters

Recall that a *Dirichlet character* is a homomorphism $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$.

The following lemma is well known.

Lemma 3.10. *If p is an odd prime, then $(\mathbf{Z}/p^n\mathbf{Z})^*$ is a cyclic group. The group $(\mathbf{Z}/2^n\mathbf{Z})^*$ is generated by -1 and 5.*

We use the following representation of Dirichlet characters. Factor N as a product of prime powers: $N = \prod_{i=1}^{r} p_i^{e_i}$ with $p_i < p_{i+1}$ and each $e_i > 0$; then $(\mathbf{Z}/N\mathbf{Z})^* \cong \prod_{i=1}^{r}(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$. If p_i is odd then the lemma implies that $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ is cyclic. If $p_1 = 2$, then $(\mathbf{Z}/p_1^{e_1}\mathbf{Z})^*$ is a product $\langle -1\rangle \times \langle 5\rangle$ of two cyclic groups, both possibly trivial. For each i, we let $a_i \in (\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ be the smallest generator of the ith factor $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$. If $p_1 = 2$, let a_1 and a_2 correspond to the two factors $\langle -1\rangle$ and $\langle 5\rangle$, respectively; then a_3 corresponds to p_2, etc. Here a_i is smallest in the sense that the minimal lift $\tilde{a}_i \in \mathbf{Z}_{>0}$ is smallest. Let n be the exponent of $(\mathbf{Z}/N\mathbf{Z})^*$, and let $\zeta = e^{2\pi i/n} \in \mathbf{C}^*$. To give ε is the same as giving the images of each generator of a_i as a power of ζ. We thus represent ε as a vector of elements of $\mathbf{C}^*$ with respect to a canonically chosen, but unnatural, basis.

Alternatively, the vector representing a character ε can be equivalently viewed as a vector in $(\mathbf{Z}/n\mathbf{Z})^r$, where again n is the exponent of $(\mathbf{Z}/N\mathbf{Z})^*$. Such a vector represents a character if and only if the ith component of the vector has additive order dividing $\varphi(p_i^{e_i})$. If $p_1 = 2$, then there are $r + 1$ entries instead of r entries, and the condition is suitably modified. If a vector $v = [d_1, \ldots , d_r]$ represents a character ε, then each of the Galois conjugate characters is represented by $[md_1, \ldots , md_r]$ where m varies over elements of $(\mathbf{Z}/n\mathbf{Z})^*$.

When performing actual machine computations, we work in the smallest field that contains all of the values of ε. Thus if $d = \gcd(d_1, \ldots , d_r, n)$, then we work in the subfield $\mathbf{Q}(\zeta^d)$, which is cheaper than working in $\mathbf{Q}(\zeta)$.

It is sometimes important to work in characteristic ℓ. Then the notation is as above, except ζ is replaced by a primitive mth root of unity, where m is the prime-to-ℓ part of n. Note that the primitive nth roots of unity in characteristic ℓ need not be conjugate; for example, both 2 and 3 are square roots of -1 in $\mathbf{F}_5$, but they are not conjugate. Thus we must specify ζ as part of the notation when giving a mod ℓ Dirichlet character.

Example 3.11. Suppose $N = p$ is an odd prime. The group of mod p Dirichlet characters (in characteristic 0) is isomorphic to $\mathbf{Z}/(p - 1)\mathbf{Z}$, and two characters a and b are Galois conjugate if and only if there is an element $x \in (\mathbf{Z}/(p-1)\mathbf{Z})^*$ such that $xa = b$. A character is determined up to Galois conjugacy by its order, so the set of classes of mod p Dirichlet characters are in bijection with the set of divisors d of $p - 1 = \#(\mathbf{Z}/p\mathbf{Z})^*$.

Let p be an odd prime. The quadratic mod p character is denoted $[(p - 1)/2]$. The quadratic mod $2p$ character is denoted by $[0, 0, (p - 1)/2]$; the quadratic mod $4p$ character is denoted $[(p - 1)/2, 0, (p - 1)/2]$. If $n \geq 3$, then the exponent of $(\mathbf{Z}/2^n\mathbf{Z})^*$ is 2^{n-2}, so the nontrivial mod 2^n character that factors through $(\mathbf{Z}/4\mathbf{Z})^*$ is denoted $[2^{n-3}, 0]$.

Definition 3.12. The *conductor* of a character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ is the smallest divisor M of N such that ε factors through the natural reduction map $(\mathbf{Z}/N\mathbf{Z})^* \to (\mathbf{Z}/M\mathbf{Z})^*$.

For simplicity, we assume that N is odd. To compute the conductor of ε, let v be the vector in $(\mathbf{Z}/n\mathbf{Z})^r$ that represents ε, as above. Since both $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ and $(\mathbf{Z}/p_i^{d}\mathbf{Z})^*$ are cyclic and the reduction map is surjective, we find that p_i^d, with $d \leq e_i$, divides the conductor of ε if and only if the ith component of v has additive order dividing $\varphi(p_i^d)$. We can thus compute the power of p_i dividing the conductor of ε by computing the smallest d such that $p_i^d \equiv p_i^{d-1}$ modulo the order of the ith component of v.

3.4 The dimension of $S_k(N, \varepsilon)$

An explicit formula for the dimension of $S_k(N, \varepsilon)$ is given in [13], without proof. For the reader's convenience, we reproduce it here.

Theorem 3.13 (Cohen-Oesterlé). *Let $k \geq 2$ be an integer and $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ be a Dirichlet character such that $\varepsilon(-1) = (-1)^k$. Then*

$$\begin{aligned}\dim S_k(N, \varepsilon) = \delta + \frac{k-1}{12} \cdot N \cdot \prod_{p \mid N} \left(1 + \frac{1}{p}\right) \quad &- \quad \frac{1}{2} \cdot \prod_{p \mid N} \lambda(r_p, s_p, p) \\ + \gamma_k \sum_{\{x \in (\mathbf{Z}/N\mathbf{Z})^* \,:\, x^2+1=0\}} \varepsilon(x) \quad &+ \quad \mu_k \sum_{\{x \in (\mathbf{Z}/N\mathbf{Z})^* \,:\, x^2+x+1=0\}} \varepsilon(x).\end{aligned}$$

Let f be the conductor of ε, i.e., the smallest M such that ε factors through $(\mathbf{Z}/M\mathbf{Z})^*$. If $p \mid N$, then r_p (resp. s_p) denotes the exponent of p in the prime factorization of N (resp. f). Furthermore,

$$\lambda(r_p, s_p, p) := \begin{cases} p^{r'} + p^{r'-1} & \text{if } 2s_p \leq r_p = 2r' \\ 2p^{r'} & \text{if } 2s_p \leq r_p = 2r' + 1 \\ 2p^{r_p - s_p} & \text{if } 2s_p > r_p \end{cases}$$

$$\gamma_k := \begin{cases} 0 & \text{if } k \text{ is odd} \\ -\frac{1}{4} & \text{if } k \equiv 2 \pmod 4 \\ \frac{1}{4} & \text{if } k \equiv 0 \pmod 4 \end{cases}$$

$$\mu_k := \begin{cases} 0 & \text{if } k \equiv 1 \pmod 3 \\ -\frac{1}{3} & \text{if } k \equiv 2 \pmod 3 \\ \frac{1}{3} & \text{if } k \equiv 0 \pmod 3 \end{cases}$$

$$\delta := \begin{cases} 1 & \text{if } k = 2 \text{ and } \varepsilon \text{ is trivial} \\ 0 & \text{otherwise} \end{cases}$$

3.5 Decomposing the space of modular symbols

Consider the space $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ of cuspidal modular symbols of level N and character ε over $K = \mathbf{Q}(\varepsilon)$. In this section we describe how to decompose the new part of $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ as a direct sum of $\mathbf{T}$-modules corresponding to the Galois conjugacy classes of newforms with character ε. As an application, we can compute the q-expansions of the normalized cuspidal newforms of level N and character ε. Using the theory of Atkin-Lehner [4] as extended by Li [37], it is then possible to construct a basis for the space $S_k(N, \varepsilon; \mathbf{C})$ of cusp forms.

The algorithm is, for the most part, a straightforward generalization of the method used by Cremona [16] to enumerate the $\mathbf{Q}$-rational weight-two newforms corresponding to modular elliptic curves. Nevertheless, we present several tricks learned in the course of doing computations, which speed up the algorithm. One useful trick that Cremona also made use of is to work in the space dual to modular symbols as described in the next section.

3.5.1 Duality

Let $K = \mathbf{Q}[\varepsilon]$, and let $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp$ denote $\mathrm{Hom}_K(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K),K)$ equipped with its natural right $\mathbf{T}$-action: for $\varphi \in \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp$,

$$(\varphi T)(x) = \varphi(Tx).$$

The natural pairing

$$\langle\, ,\, \rangle : \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp \times \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K) \to K \tag{3.1}$$

given by $\langle \varphi, x\rangle = \varphi(x)$ satisfies $\langle \varphi T, x\rangle = \langle \varphi, Tx\rangle$.

Viewing the elements $T \in \mathbf{T}$ as sitting inside $\mathrm{End}(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K))$, the transpose map $T \mapsto T^t$ allows us to view $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp$ as a left $\mathbf{T}$-module.

Proposition 3.14. *Let $V \subset \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\mathrm{new}}$ be an irreducible new $\mathbf{T}$-submodule and set $I = \mathrm{Ann}_{\mathbf{T}} V$. Then the characteristic polynomial of each T_p on $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp[I]$ is the same as the characteristic polynomial of T_p on V.*

Proof. We may assume for the purposes of proving the proposition that $K = \overline{\mathbf{Q}}$. There is a basis of simultaneous $\mathbf{T}$-eigenvectors for $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\mathrm{new}}$. With respect to this basis, $\mathbf{T}$ acts via diagonal matrices. The systems of eigenvalues coming from the old subspace are distinct from the systems of eigenvalues on the new space. Thus the dimension of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp[I]$ is the same as the dimension of V, instead of being too large. The proposition now follows by noting that the characteristic polynomial of a matrix is the same as the characteristic polynomial of its transpose. □

The degeneracy maps α_t and β_t of Section 2.5 give rise to maps $\alpha_t^\perp$ and $\beta_t^\perp$ between the dual spaces and having the dual properties to those of α_t and β_t. In particular, they commute with the Hecke operators T_p for p prime to N. The new and old subspace of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^\perp$ are defined as in Definition 2.16.

Algorithm 3.15. This algorithm computes a decomposition of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\perp\,\mathrm{new}}$ into irreducible submodules V.

Using Algorithm 3.2 compute $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)$. Then compute the maps β_t using Algorithm 2.20 and intersect the transposes of their kernels in order to obtain $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)^{\perp\,\mathrm{new}}$. Compute the boundary map $\delta : \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K) \to B_k(N,\varepsilon;K)$ using Algorithm 2.26. We cut out the cuspidal submodule $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\perp\,\mathrm{new}}$ using the Hecke operators, Algorithm 3.17, and Proposition 3.14. Set $p = 2$ and perform the following steps.

1. Using Algorithm 3.17, compute a matrix A representing the Hecke operator T_p on $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\perp\,\mathrm{new}}$.

2. Compute and factor the characteristic polynomial F of A.

3. For each irreducible factor f of F compute $V_f = \ker(f(A))$. Then, compute the $+1$ and -1 eigen-subspaces V_f^+ and V_f^- for the star involution. Let W denote one of these two eigen-subspaces, and use the following criteria to determine whether or not W is irreducible:

(a) If p is greater than the Sturm bound (see Theorem 3.9) then W must be irreducible.

(b) If the characteristic polynomial of some element $T \in \mathbf{T}$ acting on W is irreducible, then W is irreducible.

4. If W is irreducible, record W and consider the next factor of the characteristic polynomial in step 3. Otherwise, replace p by the next prime larger than p and replace $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)^{\perp\,\mathrm{new}}$ by W, then repeat the above sequence of steps, beginning with step 1.

3.5.2 Efficient computation of Hecke operators on the dual space

In this section we give a method for computing the action of the Hecke operators $T_p \in \mathbf{T}$ on an invariant subspace $V \subset \boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)^{\perp}$. A naive way to compute the right action of T_p on V is to compute a matrix representing T_p on $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)$, transpose to obtain T_p on $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)^{\perp}$, and then restrict to V using Gaussian elimination. To compute T_p on $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)$, observe that $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)$ has a basis $e_1, \ldots, e_n$, where each e_i is a Manin symbol $[P, (c, d)]$, and that the action of T_p on $[P, (c, d)]$ can be computed using Section 2.6.2.

In practice, $d = \dim V$ will often be much less than n; we now describe how to compute T_p on V in d/n of the time it takes using the above naive method. This is a substantial savings when d is small. Transposing the injection $V \hookrightarrow \boldsymbol{\mathcal{S}}_k(N, \varepsilon; K)^{\perp}$, we obtain a surjection $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; K) \to V^{\perp}$. There exists a subset $e_{i_1}, \ldots, e_{i_d}$ of the e_i whose image forms a basis for $V^{\perp}$. With some care, it is then possible to compute T_p on $V^{\perp}$ by computing T_p on each of $e_{i_1}, \ldots, e_{i_d}$.

In the rest of this section, we describe in terms of matrices a definite way to carry out this computation. Let V be an $n \times m$ matrix whose rows generate an n-dimensional subspace of an m-dimensional space of row vectors. Let T be an $m \times m$-matrix and suppose that V has rank n and that VT is contained in the row space of V. Let E be an $m \times n$ matrix with the property that the $n \times n$ matrix VE is invertible, with inverse D.

Proposition 3.16. *$VT = VTEDV$.*

Proof. Observe that

$$V(EDV) = (VED)V = IV = V.$$

Thus right multiplication by EDV

$$v \mapsto vEDV$$

induces the *identity map* on the row space of V. Since VT is contained in the row space of V, we have

$$(VT)EDV = VT,$$

as claimed. □

We have not computed T, but we can compute T on each basis element $e_1, \ldots, e_d$ of the ambient space–unfortunately, d is extremely large. Our problem: quickly compute the action of T^t on the invariant subspace spanned by the rows of V. Can this be done without having to compute T on all e_i? Yes, the following algorithm shows how using a subset of only $n = \dim V$ of the e_i.

Algorithm 3.17. Let T be any linear transformation which leaves V invariant and for which we can compute $T(e_i)$ for $i = 1, \ldots, d$. This algorithm computes the matrix representing the action of T on V while computing $T(e_i)$ for only $\dim V$ of the i.

Choose any $m \times n$ matrix E whose columns are sparse linear combinations of the e_i and such that VE is invertible. For this we find a set of positions so that elements of the space spanned by the columns of V are determined by the entries in these spots. This is accomplished by row reducing, and setting E equal to the pivot columns. Using Gaussian elimination, compute the inverse D of the $n \times n$ matrix VE. The matrix representing the action of T with respect to V is then

$$V(TE)D = V(TE)(VE)^{-1}.$$

Proof. Let A be any matrix so that VA is the $n \times n$ identity matrix. By the proposition we have

$$VTA = (VTEDV)A = VTED(VA) = VTED = V(TE)D.$$

To see that VTA represents T, observe that by the proposition,

$$\begin{aligned} VTAV &= (VTEDV)AV = (VTEDVA)V \\ &= (VTED)(VA)V = (VTED)V = VT \end{aligned}$$

so that VTA gives the correct linear combination of the rows of V. □

3.5.3 Eigenvectors

Once a $\mathbf{T}$-simple subspace of $\boldsymbol{\mathcal{S}}^*$ has been identified, the following algorithm, which was suggested to the author by H. Lenstra, produces an eigenvector defined over an extension of the base field.

Algorithm 3.18. Let A be an $n \times n$ matrix over an arbitrary field K and suppose that the characteristic polynomial $f(x) = x^n + \cdots + a_1 x + a_0$ of A is irreducible. Let α be a root of $f(x)$ in an algebraic closure $\overline{K}$ of K. Factor $f(x)$ over $K(\alpha)$ as $f(x) = (x - \alpha)g(x)$. Then for any element $v \in K^n$ the vector $g(A)v$ is either 0 or it is an eigenvector of A with eigenvalue α. The vector $g(A)v$ can be computed by finding Av, $A(Av)$, $A(A(Av))$, and then using that

$$g(x) = x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_1 x + c_0,$$

where the coefficients c_i are determined by the recurrence

$$c_0 = -a_0/\alpha, \qquad c_i = (c_{i-1} - a_i)/\alpha.$$

We will prove below that $g(A)v \neq 0$ for all vectors v not in a proper subspace of K^n. Thus with high probability, a "randomly chosen" v will have the property that $g(A)v \neq 0$. Alternatively, if $v_1, \ldots v_n$ form a basis for K^n, then $g(A)v_i$ must be nonzero for some i.

Proof. By the Cayley-Hamilton theorem [36, XIV.3] we have that $f(A) = 0$. Consequently, for any $v \in K^n$, we have $(A - \alpha)g(A)v = 0$ so that $Ag(A)v = \alpha v$. Since f is irreducible it is the polynomial of least degree satisfied by A and so $g(A) \neq 0$. Therefore $g(A)v \neq 0$ for all v not in the proper closed subset $\ker(g(A))$. □

3.5.4 Eigenvalues

In this section we give an algorithm for computing the q-expansion of one of the newforms corresponding to a factor of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\mathrm{new}}$. This is a generalization of the algorithm described in [16, §2.9].

Algorithm 3.19. Given a factor $V \subset \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)^{\perp\,\mathrm{new}}$ as computed by Algorithm 3.15 this algorithm computes the q-expansion of one of the corresponding Galois conjugate newforms.

1. Using Algorithm 3.17 compute the action of the $*$-involution (Section 2.4) on V. Then compute the $+1$ eigenspace $V^+ \subset V$.

2. Find an element $T \in \mathbf{T}$ such that the characteristic polynomial of the matrix A of T acting on V^+ is irreducible. Such a T must exist by the primitive element theorem [36, V.4]. (Note: It is *not* always the case that T can be taken to equal some Hecke operator T_n. The first example with $k=2$ and $\varepsilon=1$ occurs at level $N=512$.)

3. Using Algorithm 3.5.3 compute an eigenvector e for A over an extension of K.

4. Because e is an eigenvector and the pairing given in Equation 3.1 respects the Hecke action, we have that for any Hecke operator T_n and element $w \in \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)$, that

$$a_n\langle e,w\rangle = \langle eT_n,w\rangle = \langle e,T_nw\rangle.$$

Choose w so that $\langle e,w\rangle \neq 0$. Then

$$a_n = \frac{\langle e,T_nw\rangle}{\langle e,w\rangle}.$$

The a_n can now be computed by computing $\langle e,w\rangle$ once and for all, and then computing $\langle e,T_nw\rangle$ for each n. It is best to choose w in such a way that T_nw can be computed quickly.

The beauty of this algorithm is that when w is a Manin symbol $[P(X,Y),(c,d)]$ the computation of $T_pw = \sum_{x\in R_p} wx$ is very quick, requiring us to only sum over the Heilbronn matrices of determinant p once.

In practice we compute only the eigenvalues a_p using the above algorithm, then use the following recurrences to obtain the a_n:

$$\begin{aligned} a_{nm} &= a_n a_m \qquad \text{if } (n,m)=1, \text{ and}\\ a_{p^r} &= a_{p^{r-1}}a_p - \varepsilon(p)p^{k-2}a_{p^{r-2}}. \end{aligned}$$

3.5.5 Sorting and labeling eigenforms

Systematically ordering the factors is essential, so that we can later refer to them. In Section 3.5.4 we saw how to associate to each new factor a sequence a_n of Hecke eigenvalues. These can be used to sort the factors.

Except in the case of weight 2 and trivial character, we use the following ordering. To each eigenvector associate the following sequence of integers

$$\mathrm{tr}(a_1), \mathrm{tr}(a_2),\ \mathrm{tr}(a_3),\ \mathrm{tr}(a_4),\ \mathrm{tr}(a_5),\ \mathrm{tr}(a_6),\ \ldots$$

where the trace is from $K_f = \mathbf{Q}(\ldots a_n \ldots)$ down to $\mathbf{Q}$. Sort the eigenforms by ordering the sequences in dictionary order with minus coming before plus. Since we included $\mathrm{tr}(a_1)$, this ordering gathers together factors of the same dimension. Furthermore, the sequence of traces determines the Galois conjugacy class of f, because the $g = \sum_{n\geq 1} \mathrm{tr}(a_n)q^n$ is the trace of f, hence g lies in the $\mathbf{C}$-vector space spanned by the Galois conjugates of f.

When $k = 2$ and the character is trivial we use a different and somewhat complicated ordering because it extends the notation for elliptic curves that was introduced in the second edition of [16] and has since become standard. Sort the factors of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)^{\mathrm{new}}$ as follows. First by dimension, with smallest dimension first. Within each dimension, sort in binary order, by the signs of the Atkin-Lehner involutions with $-$ corresponding to 0 and $+$ to 1. For example, if there are three Atkin-Lehner involutions then the sign patterns are sorted as follows:

$$+++, -++, +-+, --+, ++-, -+-, +--, ---.$$

Finally, let p be the smallest prime not dividing N. Within each of the Atkin-Lehner classes, sort by the magnitudes of the $K_f/\mathbf{Q}$-trace of a_p breaking ties by letting the positive trace be first. If there are still any ties, repeat the final step with the next smallest prime not dividing N, etc. (Note: It's not clear to the author that ties will always eventually be broken, though in his computation they always have been.)

3.6 Intersections and congruences

Consider a complex torus $J = V/\Lambda$, and let $A = V_A/\Lambda_A$ and $B = V_B/\Lambda_B$ be subtori whose intersection $A \cap B$ is finite.

Proposition 3.20. *There is a natural isomorphism of groups*

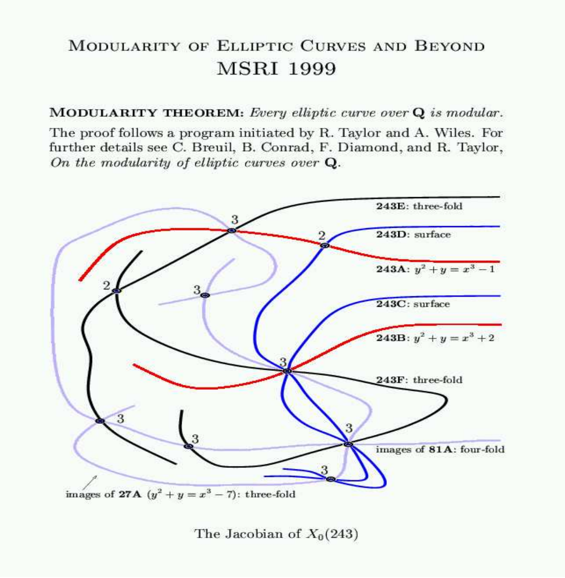
$$A \cap B \cong \left(\frac{\Lambda}{\Lambda_A + \Lambda_B}\right)_{\mathrm{tor}}.$$

Proof. There is an exact sequence

$$0 \to A \cap B \to A \oplus B \to J.$$

Consider the diagram

$$\begin{tikzcd}
 & \Lambda_A \oplus \Lambda_B \arrow[r] \arrow[d] & \Lambda \arrow[r] \arrow[d] & \Lambda/(\Lambda_A + \Lambda_B) \arrow[d] \\
 & V_A \oplus V_B \arrow[r] \arrow[d] & V \arrow[r] \arrow[d] & V/(V_A + V_B) \arrow[d] \\
A \cap B \arrow[r] & A \oplus B \arrow[r] & J \arrow[r] & J/(A+B).
\end{tikzcd}$$



MODULARITY OF ELLIPTIC CURVES AND BEYOND
MSRI 1999

MODULARITY THEOREM: *Every elliptic curve over* $\mathbf{Q}$ *is modular.*

The proof follows a program initiated by R. Taylor and A. Wiles. For further details see C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$.

243E: three-fold
3
2
243D: surface
243A: $y^2 + y = x^3 - 1$
2
3
243C: surface
243B: $y^2 + y = x^3 + 2$
3
243F: three-fold
3
3
3
images of **81A**: four-fold
3
images of **27A** ($y^2 + y = x^3 - 7$): three-fold
The Jacobian of $X_0(243)$

Figure 3.1: T-shirt design

The snake lemma gives an exact sequence

$$0 \to A \cap B \to \Lambda/(\Lambda_A + \Lambda_B) \to V/(V_A + V_B).$$

Since $V/(V_A + V_B)$ is a $\mathbf{C}$-vector space, the torsion part of $\Lambda/(\Lambda_A + \Lambda_B)$ must map to 0. No non-torsion in $\Lambda/(\Lambda_A + \Lambda_B)$ could map to 0, because if it did then $A \cap B$ would not be finite. The lemma follows. □

The following formula for the intersection of n subtori is obtained in a similar way.

Proposition 3.21. *For $i = 1, \ldots, n$ let $A_i = V_i/\Lambda_i$ be a subtorus of $J = V/\Lambda$, and assume that each pairwise intersection $A_i \cap A_j$ is finite. Then*

$$A_1 \cap \cdots \cap A_n \cong \left(\frac{\Lambda \oplus \cdots \oplus \Lambda}{f(\Lambda_1 \oplus \cdots \oplus \Lambda_n)} \right),$$

where $f(x_1, \ldots, x_n) = (x_1 - x_2, x_2 - x_3, x_3 - x_4, \ldots, x_{n-1} - x_n)$.

Remark 3.22. Using this proposition the author constructed the T-shirt design in Figure 3.1.

Example 3.23. L. Kilford of London, England has recently discovered an example at prime level 503 in which "multiplicity one" fails. One verification of his example uses the above proposition. Let E_1, E_2, and E_3 be the three elliptic curves of conductor 503, and for each $i = 1, 2, 3$, let $\mathfrak{m}_i$ be the maximal ideal of $\mathbf{T} \subset \mathrm{End}(J_0(503))$ generated by 2 and all $T_p - a_p(E_i)$, with p prime. Each of the Galois representations $E_i[2]$ is irreducible, and one can check that $\mathfrak{m}_1 = \mathfrak{m}_2 = \mathfrak{m}_3$. If multiplicity one holds, then $E_1[2] = E_2[2] = E_3[2]$ inside of $J_0(503)$. However, this is not the case, as a modular symbols computation in the integral homology $H_1(X_0(N), \mathbf{Z})$ reveals that $E_1 \cap E_2 = \{0\}$.

3.6.1 A strategy for computing congruences

Let N be a positive integer, $k \geq 2$ an integer, and ε a mod N Dirichlet character. Suppose f and g are newforms in $S_k(N, \varepsilon; \overline{\mathbf{Q}})$. The following proposition gives rise to an algorithm for computing most congruences between infinite Fourier expansions.

The advantage of the algorithm is that it only involves finite exact computations and does not rely on the computation of q-expansions. A disadvantage is that congruences between q-expansions need not be reflected by the corresponding modular symbols, so the proposition need not give all congruences. This is illustrated in Example 3.23.

The author first learned about this strategy from the section entitled "First strategy: Computing m-congruences of period lattices" in [18].

Proposition 3.24. *Suppose f and g are newforms in $S_k(N, \varepsilon; \overline{\mathbf{Q}})$. Let I_f and I_g be the corresponding annihilators in the Hecke algebra $\mathbf{T}$. Let $\Lambda = \boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathcal{O})$, and set $\Lambda_f = \Lambda[I_f]$ and $\Lambda_g = \Lambda[I_g]$. If $p \mid \# \left(\frac{\Lambda}{\Lambda_f + \Lambda_g}\right)_{\text{tor}}$ then there is a prime $\wp$ of residue characteristic p such that $f \equiv g \pmod{\wp}$.*

Proof. Consider the exact sequence

$$0 \to \Lambda_f \oplus \Lambda_g \to \Lambda \to \Lambda/(\Lambda_f + \Lambda_g) \to 0$$

where the first map is $(a, b) \mapsto a - b$. Upon tensoring this sequence with $\mathbf{F}_p$ we obtain:

$$Z \to (\Lambda_f \otimes \mathbf{F}_p) \oplus (\Lambda_g \otimes \mathbf{F}_p) \to \Lambda \otimes \mathbf{F}_p \to (\Lambda/(\Lambda_f + \Lambda_g)) \otimes \mathbf{F}_p \to 0,$$

where $Z = \mathrm{Tor}^1(\Lambda/(\Lambda_f + \Lambda_g), \mathbf{F}_p)$. Denote by $\mathrm{im}(\Lambda_f)$ the image of $\Lambda_f \otimes \mathbf{F}_p$ in $\Lambda \otimes \mathbf{F}_p$ and likewise for Λ_g. Our assumption that p divides the torsion part of $\Lambda/(\Lambda_f + \Lambda_g)$ implies that Z is nonzero, so $\mathrm{im}(\Lambda_f)$ and $\mathrm{im}(\Lambda_g)$ have nonzero intersection inside the $\mathbf{F}_p$-vector space $\Lambda \otimes \mathbf{F}_p$. The Hecke algebra $\mathbf{T}$ acts on $\mathrm{im}(\Lambda_f)$ through its action on f, that is, through the quotient $\mathbf{T}/I_f$; similarly, $\mathbf{T}$ acts on $\mathrm{im}(\Lambda_g)$ through $\mathbf{T}/I_g$. Thus $\mathbf{T}$ acts on the nonzero $\mathbf{T} \otimes \mathbf{F}_p$-module $\mathrm{im}(\Lambda_f) \cap \mathrm{im}(\Lambda_g)$ through $\mathbf{T}/(I_f + I_g + p)$. This implies that $I_f + I_g + p$ is not the unit ideal, which is equivalent to the assertion of the proposition. □

3.7 The rational period mapping

Consider a triple (N, k, ε), and let $K = \mathbf{Q}[\varepsilon]$. Let I be an ideal in the Hecke algebra $\mathbf{T}$ associated to (N, k, ε). The rational period mapping associated to I is a map from the space $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)$ of modular symbols to a finite dimensional K-vector space. It is a computable analogue of the classical integration pairing, and is of great value in extracting the rational parts of analytic invariants; e.g., of special values of L-functions. In the next section we use it to compute the image of cuspidal points on $J(N, k, \varepsilon)$.

Definition 3.25. Let $D := \mathrm{Hom}_K(\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K), K)[I]$; the *rational period mapping* is the natural quotient map

$$\Theta_I : \boldsymbol{\mathcal{M}}_k(N, \varepsilon; K) \to \frac{\boldsymbol{\mathcal{M}}_k(N, \varepsilon; K)}{\bigcap \{\ker(\varphi) : \varphi \in D\}}.$$

If $f \in S_k(N,\varepsilon)$ is a newform, set $\Theta_f := \Theta_{I_f}$ where I_f is the annihilator of f in the Hecke algebra.

Algorithm 3.26. This algorithm computes Θ_I. Choose a basis for $W = \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)$ and use it to view W as a space of column vectors equipped with a left action of $\mathbf{T}$. View $W^* = \mathrm{Hom}_K(\boldsymbol{\mathcal{M}}_k(N,\varepsilon;K),K)$ as the space of row vectors of length equal to $\dim \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)$; thus W^* is dual to W via the natural pairing between row and column vectors. The Hecke operators act on W^* on the right. Compute a basis $\varphi_1,\ldots,\varphi_n$ for the K-vector space $W^*[I]$. Then the rational period mapping with respect to this basis is $\varphi_1 \times \cdots \times \varphi_n$; it is given by the matrix whose rows are $\varphi_1,\ldots,\varphi_n$.

Proof. The kernels of $\varphi_1 \times \cdots \times \varphi_n$ and Θ_I are the same. □

Example 3.27. Let I be the annihilator of the newform $f = q - 2q^2 + \cdots \in M_2(37,1;\mathbf{Q})$ corresponding to the elliptic curve **37k2A**. There is a basis for $W = \boldsymbol{\mathcal{M}}_2(37,1;\mathbf{Q})$ such that

$$T_2 = \begin{pmatrix} -1 & 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

The characteristic polynomial of T_2 is $x^2(x+2)^2(x-3)$. Thus $W[I] = \ker(T_2+2)$ is spanned by the column vectors $(1,-1,0,1/2,0)^t$ and $(0,0,1,-1/2,0)^t$, and $W^*[I] = \ker(T_2^t+2)$ is spanned by the row vectors $(1,0,-1,0,0)$ and $(0,1,-1,0,0)$. The rational period mapping is $\Theta_I((a,b,c,d,e)^t) = (a-c, b-c)$.

Lemma 3.28.

$$\dim \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)[I] = \dim \mathrm{Hom}_K(\boldsymbol{\mathcal{M}}_k(N,\varepsilon;K),K)[I].$$

Proof. Let $W = \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)$ and W^* be its dual. Let $a_1,\ldots,a_n$ be a set of generators for I. Choose a basis for W that is compatible with the following filtration:

$$0 \subset (\ker(a_1)\cap\cdots\cap\ker(a_n)) \subset (\ker(a_1)\cap\cdots\cap\ker(a_{n-1})) \subset \cdots \subset \ker(a_1) \subset W.$$

The rank of a matrix equals the rank of its transpose, so the dimension of $\ker(a_1)$ is the same as the dimension of $\ker(a_1^t)$, that is, $\dim W[(a_1)] = \dim W^*[(a_1)]$. Since $\mathbf{T}$ is commutative, a_2 leaves $\ker(a_1)$ invariant; because of how we chose our basis for W, the transpose of $a_2|_{\ker(a_1)}$ is $a_2^t|_{\ker(a_1^t)}$. Thus again, $\dim(\ker(a_2|_{\ker(a_1)}))$ equals $\dim(\ker(a_2^t|_{\ker(a_1^t)}))$. Proceeding inductively, we prove the lemma. □

Corollary 3.29. *Suppose* $\boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)[I] \subset \boldsymbol{\mathcal{S}}_k(N,\varepsilon;K)$, *and let* $P : \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K) \to \mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon;\mathbf{C})[I],\mathbf{C})$ *be the classical period map induced by the integration pairing. Then* $\ker(P) = \ker(\Theta_I)$.

Proof. Since $P(\boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathcal{O}))$ is known to be a finite-covolume $\mathcal{O}$-lattice in the complex vector space $\mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon;\mathbf{C})[I],\mathbf{C})$, the K-dimension of $\mathrm{im}(P)$ equals $2\cdot\dim_{\mathbf{C}} S_k(N,\varepsilon;\mathbf{C})[I]$, which in turn equals $\dim_K \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)[I]$. Thus by Lemma 3.28 the images $\mathrm{im}(P)$ and $\mathrm{im}(\Theta_I)$ have the same dimension, hence $\ker(P)$ and $\ker(\Theta_I)$ also have the same dimension. It thus suffices to prove the inclusion $\ker(\Theta_I) \subset \ker(P)$. Suppose $\Theta_I(x) = 0$; then $\varphi(x) = 0$ for all $x \in W^*[I]$, where $W = \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K)$. Thus $\varphi(x) = 0$ for all $\varphi \in (W \otimes \mathbf{C})^*[I]$. Since the integration pairing that defines P respects the action of $\mathbf{T}$, the composition of P with any linear functional lies in $(W \otimes \mathbf{C})^*[I]$. Thus $P(x) = 0$, as required. □

3.8 The images of cuspidal points

Consider a triple (N,k,ε), and let $K = \mathbf{Q}[\varepsilon]$. Recall that integration defines a period mapping

$$P : \boldsymbol{\mathcal{M}}_k(N,\varepsilon;K) \to \mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon;\mathbf{C}),\mathbf{C}).$$

A *cuspidal point* of

$$J = J(N,k,\varepsilon) := \frac{\mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon;\mathbf{C}),\mathbf{C})}{P(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathcal{O}))}$$

is a point that is in the image under P of $\boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathcal{O})$. It is of great interest to compute the structure of the cuspidal subgroup of J and of the quotients of J. For example, when $k=2$ and $\varepsilon = 1$, the torus J can be identified with $J_0(N)(\mathbf{C})$. In this case, Manin proved (see [38]) that the cuspidal point $\{0,\infty\}$ is a torsion point in $J_0(N)(\mathbf{Q})$, so its order gives a lower bound on $J_0(N)(\mathbf{Q})_{\mathrm{tor}}$.

Algorithm 3.30 (Cuspidal subgroup). Let I be an ideal in the Hecke algebra $\mathbf{T}$. This algorithm computes the cuspidal subgroup of the quotient A_I of J. Using Algorithm 3.5 compute $\boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathcal{O})$ and $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathcal{O})$. Using Algorithm 3.26, compute the rational period mapping Θ_I. Then the cuspidal subgroup is the subgroup of $\Theta_I(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathcal{O}))$ generated by the elements $\Theta_I(x)$ for $x \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathcal{O})$. In particular, the point of $A_I(\mathbf{C})$ corresponding to $X^iY^{k-2-i}\{\alpha,\beta\}$ is the image of $\Theta_I(X^iY^{k-2-i}\{\alpha,\beta\})$ in the quotient of $\Theta_I(\boldsymbol{\mathcal{M}}_k(N,\varepsilon;\mathcal{O}))$ by $\Theta_I(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathcal{O}))$.

Example 3.31. This example continues Example 3.27. The basis chosen is also a basis for $\boldsymbol{\mathcal{M}}_2(37,1;\mathbf{Z})$, so by computing the boundary map, or the integer kernel of $T_2(T_2+2)$, we find that $\boldsymbol{\mathcal{S}}_2(37,1;\mathbf{Z})$ is spanned by $(1,0,0,0,0)$, $(0,1,0,0,0)$, $(0,0,1,0,0)$, and $(0,0,0,1,0)$. Thus $\Theta_I(\boldsymbol{\mathcal{S}}_2(37,1;\mathbf{Z}))$ is generated by $(1,0)$ and $(0,1)$. The modular symbols $\{0,\infty\}$ is represented by $(0,0,0,0,-1)$, so the image of the cusp $(0)-(\infty) \in J_0(37)$ is 0 in **37k2A**.

The rational period mapping associated to **37k2B** (with respect to some basis) is

$$\Theta_I((a,b,c,d,e)^t) = (a - c - 2d + \frac{2}{3}e,\ b + c + 2d - \frac{2}{3}e).$$

Thus $\Theta_I(\boldsymbol{\mathcal{S}}_2(37,1;\mathbf{Z}))$ is generated by $(1,0)$ and $(0,1)$. The image of $\{0,\infty\}$ is is $\frac{2}{3}(1,-1)$, so the image of $(0)-(\infty)$ in **37k2B** has order 3.

3.8.1 Rational torsion

Let f be a newform of weight 2, and suppose $\varepsilon = 1$. Manin proved that $(0)-(\infty)$ defines an element of $J_0(N)(\mathbf{Q})_{\mathrm{tor}}$. Thus the order of the image of $(0)-(\infty)$ provides a lower bounds on $\#A_f(\mathbf{Q})_{\mathrm{tor}}$. In general, many other points in the cuspidal subgroup can be rational. Determining which would give a better lower bound on the rational subgroup; the author has not yet carried out such computations (see, however, [65]).

3.8.2 Upper bound on torsion: Counting points mod p

Let f be a newform of weight 2, and suppose $\varepsilon = 1$. The Hecke algebra $\mathbf{T}$ acts through a quotient $\overline{\mathbf{T}}$ on the subspace of $S_2(\Gamma_0(N))$ spanned by the Galois conjugates of f. Let $\chi_p(X)$ be the characteristic polynomial of the image of T_p in $\overline{\mathbf{T}}$. Suppose $p \nmid N$ and let $N_p = \#A_f(\mathbf{F}_p)$ be the number of points on the mod p reduction of the abelian variety A_f.

Proposition 3.32. *For each prime p not dividing N,*

$$N_p = \chi_p(p+1).$$

Proof. This is probably well-known, but we give a proof (which was suggested to the author by Matt Baker). It follows from the Eichler-Shimura theorem that the following relation holds in the endomorphism ring of $A_f/\mathbf{F}_p$:

$$T_p = \mathrm{Frob} + \mathrm{Ver} = \mathrm{Frob} + p/\mathrm{Frob}\,.$$

Let $\ell \neq p$ be a prime. If the characteristic polynomial of Frob on an ℓ-adic Tate module of $A_f/\mathbf{F}_p$ is $F(t)$, and the characteristic polynomial of T_p on differentials $H^0(A_f/\mathbf{F}_p, \Omega)$ is $f(t)$, then we have $f(t) = x^{-d}F(x)$, where $t = x + (p/x)$ and $d = \dim A_f$. In other words, the relation above gives an easy conversion between f and F. Since it's a general fact that $\#A_f(\mathbf{F}_p) = F(1)$, we have $\#A_f(\mathbf{F}_p) = f(p+1)$. □

The following theorem is proved using formal groups.

Theorem 3.33. *Let A be an abelian variety over $\mathbf{Q}$, with good reduction outside N. Suppose $p \nmid N$. Then the kernel of the reduction map $A(\mathbf{Q})_{\mathrm{tor}} \to A(\mathbf{F}_p)$ is killed by p. If $p > 2$ then the kernel is trivial.*

By taking gcd's we obtain an upper bound on $\#A(\mathbf{Q})_{\mathrm{tor}}$. This upper bound is not in general sharp; in fact, it is unchanged if A is replaced by any isogenous abelian variety. For example, $X_0(11)$ and $X_1(11)$ are isogenous, but have different torsion subgroups.

3.9 The modular degree

Let f be a newform of level N, weight $k \geq 2$ and character ε such that $\varepsilon^2 = 1$. In this section we define and give an algorithm to compute the modular degree of the torus A_f attached to f.

Definition 3.34. The *modular map* is the map $\theta_f : A_f^{\vee} \to A_f$ that is induced by the bottom row of Diagram 2.1 on page 34. The *modular degree* m_f of f (or of A_f) is the degree of this map. If f has weight two, then θ_f is a polarization so by [50, Thm. 13.3] its degree is a perfect square; in this case we *instead* define the modular degree m_f to be the positive square root of the degree of θ_f.

Remark 3.35. When $E/\mathbf{Q}$ be a modular elliptic curve of conductor N that is an optimal quotient of $J_0(N)$, then m_f is the usual modular degree, which is the least degree of a map $X_0(N) \to E$.

Remark 3.36. When $k \neq 2$, the degree of θ_f need not be a perfect square. For example, there is a one-dimensional quotient A_f associated to the unique rational newform

$$f = q + 2q^2 - 8q^3 + 4q^4 + 5q^5 - 16q^6 - 4q^7 + \cdots \in S_4(10)$$

such that the kernel of θ_f is isomorphic to $\mathbf{Z}/10\mathbf{Z}$.

Next, for a newform f let θ'_f be the part of $\#\ker(\theta_f)$ that is coprime to the level. There is a newform in $f \in S_4(\Gamma_0(77))$ such that θ'_f is not a perfect square at 2. For identification purposes, we remark that the field generated by the Fourier coefficients of f has discriminant $2^3 \cdot 3^3 \cdot 2417$.

Algorithm 3.37. Let I_f be the annihilator of f in the Hecke algebra. The modular kernel $\ker(\theta_f)$ is isomorphic to the cokernel of the natural map $\boldsymbol{\mathcal{S}}[I_f] \to \Phi_f(\boldsymbol{\mathcal{S}})$ of Diagram 2.1 on page 34. This cokernel can be computed by replacing Φ_f by the rational period map Θ_{I_f}.

Proof. For concreteness, we give the proof only in the case of weight-two and trivial character. The proof in the general case is similar. Let $S = S_2(\Gamma_0(N), \mathbf{C})$ be the complex vector space of weight-two modular forms of level N, and set $H = H_1(X_0(N), Z)$. The integration pairing $S \times H \to \mathbf{C}$ induces a natural map

$$\Phi_f : H \to \mathrm{Hom}(S[I_f], \mathbf{C}).$$

Using the classical Abel-Jacobi theorem, we deduce the following commutative diagram, which has exact columns, but whose rows are not exact.

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] \\
H[I_f] \arrow[r] \arrow[d] & H \arrow[r] \arrow[d] & \Phi_f(H) \arrow[d] \\
\mathrm{Hom}(S, \mathbf{C})[I_f] \arrow[r] \arrow[d] & \mathrm{Hom}(S, \mathbf{C}) \arrow[r] \arrow[d] & \mathrm{Hom}(S[I_f], \mathbf{C}) \arrow[d] \\
A_f^{\vee}(\mathbf{C}) \arrow[r] \arrow[d] \arrow[rr, bend right=60] & J_0(N)(\mathbf{C}) \arrow[r] \arrow[d] & A_f(\mathbf{C}) \arrow[d] \\
0 & 0 & 0
\end{tikzcd}$$

By the snake lemma, the kernel of $A_f^{\vee}(\mathbf{C}) \to A_f(\mathbf{C})$ is isomorphic to the cokernel of the map $H[I_f] \to \Phi_f(H)$, which proves the proposition. □

Remark 3.38. Suppose E is an optimal quotient of $J_0(p)$, with p prime. The surjectivity result in [48] implies that it is possible to efficiently compute the modular degree using only the method of graphs. For more details, see Chapter 4.

3.10 The rational part of $L(A_f, j)$

Let $k \geq 2$ be an integer, and let $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ be a Dirichlet character such that $\varepsilon^2 = 1$. This assumption on ε is made only for simplicity; there is no fundamental obstruction to considering arbitrary characters. For the remainder of this section we fix a newform $f \in S_k(N, \varepsilon)$. We will compute certain rational numbers associated to f.

The author was motivated to prove the results of this section after seeing Agashe's results in the case $k = 2$ and $\varepsilon = 1$; see [2, Ch. 4].

3.10.1 L-functions

Definition 3.39. The L-series associated to f is the complex-analytic function

$$L(f, s) := \sum_{n=1}^{\infty} a_n n^{-s}.$$

Hecke proved that $L(f, s)$ has an analytic continuation to the whole complex plane. In particular, it makes sense to consider the values $L(f, j)$ where $j \in \{1, 2, \ldots, k-1\}$ is an integer in the "critical strip." The general consesus is that these special values have deep arithmetic significance, in the sense that the quotients $L(f, j)/\omega_{f,j}$ should be algebraic numbers, where $\omega_{f,j}$ is an appropriate period of f, and that these algebraic numbers should encode deep arithmetic properties of the motive attached to f.

For simplicity, especially when doing explicit computations, it is desirable to work exclusively with ratios that are rational numbers instead of algebraic numbers. For this purpose, we consider instead the complex torus A_f attached to f, and introduce

$$L(A_f, s) := \prod_{i=1}^{d} L(f_i, s),$$

where $f_1, \ldots, f_d$ are the distinct Galois-conjugates of f. As we will see, $L(A_f, j)/\Omega_j \in \mathbf{Q}$, where Ω_j will be defined below.

Though the notation $L(A_f, s)$ suggests that there might be a way to attach an L-function to a general complex torus, this is definitely *not* what we have in mind. For our present purposes, the notation $L(A_f, s)$ is nothing more than a convenient shorthand for the product of the L-functions attached to the Galois conjugates of f. However, in the case when $k = 2$ and $\varepsilon = 1$, the L-function $L(A_f, s)$ is known to be the canonical L-series associated to the abelian variety $A_f/\mathbf{Q}$; see, e.g., the discussion in [20, Sec. 7].

3.10.2 Winding elements

Generalizing Mazur and Merel's terminology when $k = 2$, we define winding elements as follows.

Definition 3.40 (Winding element). For $1 \leq i \leq k-1$, the ith *winding element* is

$$\mathbf{e}_i := X^{i-1} Y^{k-2-(i-1)} \{0, \infty\} \in \boldsymbol{\mathcal{M}}_k(N, \varepsilon; \mathbf{Z}).$$

For example, when $k = 2$ there is one winding element $\mathbf{e} = \mathbf{e}_1 = \{0, \infty\}$. See [44, §2.2] for a topologically motivated discussion of the terminology "winding element."

3.10.3 Real and minus volumes

We briefly review the association of a complex torus to a Galois-conjugacy class of newforms. Consider the space $S_k(N, \varepsilon; \mathbf{Z})$ of cusp forms in $S_k(N, \varepsilon)$ whose q-expansion at infinity has Fourier coefficients which lie in $\mathbf{Z}$. Let V be the $\mathbf{C}$-vector space spanned by the Galois conjugates $f_1, \ldots, f_d$ of f, and choose a $\mathbf{Z}$-basis $g_1, \ldots, g_d$ for the intersection $V \cap S_k(N, \varepsilon; \mathbf{Z})$. Then integration via the pairing of Theorem 2.7 against $g_1, \ldots, g_d$ defines a map $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z}) \to \mathbf{C}^d$ whose cokernel is $A_f(\mathbf{C})$. Viewing $A_f(\mathbf{C})$ in this way, the standard measure on $\mathbf{C}^d$ defines a measure on $A_f(\mathbf{C})$.

Because $\varepsilon^2 = 1$, the complex torus $A(\mathbf{C})$ is equipped with an action of complex conjugation. There are two distinguished additive subgroups of $A(\mathbf{C})$: the subgroup $A(\mathbf{R})$ of elements fixed under complex conjugation, and the subgroup $A(\mathbf{C})^-$ of elements sent to their additive inverse by complex conjugation. When j is odd, let Ω_j be the measure of the subgroup fixed under conjugation, and when j is even, let Ω_j be the measure of the subgroup sent to its inverse under conjugation, times i^d, where d is the dimension of A. When j is odd, we call Ω_j the *real volume*; otherwise, we call Ω_j the *minus volume* (see Definition 3.58).

3.10.4 The theorem

We are now prepared to state a theorem that gives a computable expression for the ratio $|L(A_f, j)/\Omega_j|$. This theorem grew out of joint work with Agashe. It generalizes Cremona's method for computation $L(E, 1)/\Omega_E$ when E is an elliptic curve (see [16, §2.8]).

As an immediate corollary of the formula, we see that $|L(A_f, j)/\Omega_j|$ is a rational number. This was already known when $f \in S_2(\Gamma_0(N))$ (see [28, §2]). The author remains ignorant as to whether or not the general corollary was known before, or even if the real numbers Ω_j, exactly as defined here, had been previously considered. However, rationality of certain related period ratios has been known for some time, due to work of Manin, Shimura, and Hatada. For a clear historical summary of these rationality results see Li's MathSciNet review of [29]. See also [42, 44].

We take the absolute value of $L(A_f, j)/\Omega_j$ for simplicity only because at present we do not wish to worry about powers of the 4th root of unity i.

Theorem 3.41. *Let $f \in S_k(N, \varepsilon)$ be a newform, where $k \geq 2$ and $\varepsilon^2 = 1$, and let $j \in \{1, 2, \ldots, k-1\}$ be an integer in the critical strip. Let $\sigma = (-1)^{j-1}$, and let Θ_f be the rational periopd mapping associated to f (see Definition 3.25). Then*

$$\left| \frac{L(A_f, j)}{\Omega_j} \right| = [\Theta_f(\boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z})^\sigma) : \Theta_f(\mathbf{T}\mathbf{e}_j)],$$

where $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z})^\sigma$ denotes the submodule of $\boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z})$ on which the $*$-involution acts as σ, and Ω_j is the real or minus volume of A_f, as in Section 3.10.3. The right hand expression in the formula is a lattice index, whose definition is given below.

Remark 3.42. In the context of the BSD conjecture, $\Omega_{A_f} = \Omega_1 \cdot c_\infty$, where c_∞ is the number of connected components of $A_f(\mathbf{R})$.

The theorem involves lattice indexes, which we define as follows.

Definition 3.43. Let V be a finite-dimensional vector space over $\mathbf{R}$. A *lattice* $L \subset V$ is a free abelian group of rank equal to the dimension of V such that $\mathbf{R}L = V$. If $L, M \subset V$ are lattices, the *lattice index* $[L : M] \in \mathbf{R}$ is the absolute value of the determinant of an automorphism of V taking L isomorphically onto M. For convenience we set $[L : M] = 0$ for any lattice L and additive abelian group M contained in V and of rank strictly smaller than $\dim V$.

The following fact allows us to compute the lattice the index without using complex numbers.

Lemma 3.44. *Suppose $\tau_i : V \to W_i$, $i = 1, 2$, are surjective linear maps such that $\ker(\tau_1) = \ker(\tau_2)$. Let L and M be lattices in V such that $\tau_i(L)$ and $\tau_i(M)$ are both lattices for $i = 1, 2$. Then*

$$[\tau_1(L) : \tau_1(M)] = [\tau_2(L) : \tau_2(M)].$$

Proof. Surjectivity and equality of kernels insures that there is a unique isomorphism $\iota : W_1 \to W_2$ such that $\iota\tau_1 = \tau_2$. Let σ be an automorphism of W_1 such that $\sigma(\tau_1(L)) = \tau_1(M)$. Then

$$\iota\sigma\iota^{-1}(\tau_2(L)) = \iota\sigma\tau_1(L) = \iota\tau_1(M) = \tau_2(M).$$

Since conjugation does not change the determinant,

$$[\tau_2(L) : \tau_2(M)] = |\det(\iota\sigma\iota^{-1})| = |\det(\sigma)| = [\tau_1(L) : \tau_1(M)].$$

□

Proof of Theorem 3.41. Let $\Phi = \Phi_f$ be the period map $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; \mathbf{Z}) \to \mathbf{C}^d$ defined by fixing a basis $f_1, f_2, \ldots, f_d$ of the conjugates of the *newform* f; thus

$$\Phi(x) = (\langle f_1, x\rangle, \langle f_2, x\rangle, \ldots \langle f_d, x\rangle) \in \mathbf{C}^d.$$

We view $\mathbf{C}^d$ as an algebra with unit element $\mathbf{1} = (1, \ldots, 1)$ equipped with an action of the Hecke operators. The operator T_p acts as $(a_p^{(1)}, \ldots, a_p^{(d)})$, where the components $a_p^{(j)}$ are the Galois conjugates of a_p. Let $\mathbf{Z}^d \subset \mathbf{R}^d \subset \mathbf{C}^d$ be the standard submodules.

For brevity, set $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z})$. Let $\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma))$ be the measure of a fundamental domain for the lattice $\Phi(\boldsymbol{\mathcal{S}}^\sigma)$; equivalently, $\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma))$ is the absolute value of the determinant of a basis for $\Phi(\boldsymbol{\mathcal{S}}^\sigma)$. Observe that $\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma)) = [\mathbf{Z}^d : \Phi(\boldsymbol{\mathcal{S}}^\sigma)]$ and $|L(A_f, j)| = [\mathbf{Z}^d : \Phi(\mathbf{e}_j)\mathbf{Z}^d]$.

Let $W \subset \mathbf{C}^d$ be the $\mathbf{Z}$-module spanned by the "columns" of a basis for $S_k(N, \varepsilon; \mathbf{Z})[I_f]$. More precisely, if $g_1, \ldots, g_d$ is a basis, then the nth column is the vector $(a_n(g_1), \ldots, a_n(g_d))$, where $a_n(g_i)$ is the coefficient of q^n in the q-expansion of g_i at infinity. Because Ω_j is computed with respect to a basis for $S_k(N, \varepsilon; \mathbf{Z})[I_f]$,

$$\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma)) = [W : \mathbf{T}\mathbf{1}] \cdot \Omega_j.$$

Observe that $S_k(N,\varepsilon;\mathbf{Z})$ is saturated, in the sense that there are no nontrivial linear relations between the g_i when reduced modulo any prime p. To see this, note that if $\sum a_i g_i \equiv 0 \pmod{p}$, then $\frac{1}{p}\sum a_i g_i \in S_k(N,\varepsilon;\mathbf{Z})$ which, if the a_i are not all 0, is contrary to our assumption that $g_1,\ldots,g_d$ are a $\mathbf{Z}$-basis. Because "row rank = column rank", the same must be true for the "columns" defined in the previous paragraph, so $[\mathbf{Z}^d : W] = 1$. It follows that $[\mathbf{Z}^d : \mathbf{T1}] = [W : \mathbf{T1}]$.

The following calculation combines together the above observations using properties of the lattice index:

$$\begin{aligned}
[\Phi(\boldsymbol{\mathcal{S}}^\sigma) : \Phi(\mathbf{Te}_j)] &= [\Phi(\boldsymbol{\mathcal{S}}^\sigma) : \mathbf{Z}^d]\cdot[\mathbf{Z}^d : \Phi(\mathbf{Te}_j)]\\
&= \frac{1}{[\mathbf{Z}^d : \Phi(\boldsymbol{\mathcal{S}}^\sigma)]}\cdot[\mathbf{Z}^d : \Phi(\mathbf{Te}_j)]\\
&= \frac{1}{\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma))}\cdot[\mathbf{Z}^d : \Phi(\mathbf{e}_j)\mathbf{Z}^d]\cdot[\Phi(\mathbf{e}_j)\mathbf{Z}^d : \Phi(\mathbf{Te}_j)]\\
&= \frac{|L(A_f,j)|}{\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma))}\cdot[\Phi(\mathbf{e}_j)\mathbf{Z}^d : \Phi(\mathbf{Te}_j)]\\
&= \frac{|L(A_f,j)|}{\mu(\Phi(\boldsymbol{\mathcal{S}}^\sigma))}\cdot[\Phi(\mathbf{e}_j)\mathbf{Z}^d : \Phi(\mathbf{e}_j)\mathbf{T1}]\\
&= \frac{|L(A_f,j)|}{|\Omega_j|\cdot[W : \mathbf{T1}]}\cdot[\mathbf{Z}^d : \mathbf{T1}]\\
&= \frac{|L(A_f,j)|}{|\Omega_j|}.
\end{aligned}$$

Theorem 3.41 now follows by using Lemma 3.44, to replace Φ by Θ_f. □

3.10.5 Bounding the denominator of the ratio

In this section we bound the denominators of the ratios appearing in the previous section. We begin with the following lemma, which follows easily from the alternative description of the boundary map given in Proposition 2.25.

Lemma 3.45. *For $j = 2,\ldots,k-2$ the winding element $\mathbf{e}_j$ lies in $\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathbf{Z})$.*

Proof. Recall that $\mathbf{e}_j = P(X,Y)\{0,\infty\}$ where $P(X,Y) = X^{j-1}Y^{k-2-(j-1)}$. Since $2 \le j \le k-2$, it follows that $P(1,0) = P(0,1) = 0$, so Propositison 2.25 implies that $\mathbf{e}_j$ maps to 0 under the boundary map. □

Proposition 3.46. *For $j = 2,\ldots,k-2$,*

$$\frac{L(A_f,j)}{\Omega_j} \in \mathbf{Z}.$$

Proof. This follows from Theorem 3.41 because $\Theta_f(\mathbf{Te}_j) \subset \Theta_f(\boldsymbol{\mathcal{S}}_k(N,\varepsilon;\mathbf{Z})^\sigma)$, so the lattice index is an integer. □

For the rest of this section, we assume for simplicity that $\varepsilon = 1$.

Lemma 3.47. *For $j = 1$ and $j = k - 1$, we have for each $p \nmid N$ that*

$$(T_p - (1 + p^{k-1}))\mathbf{e}_j \in \boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z}).$$

Proof. This is a standard calculation; see, e.g., [16, §2.8] for the case when $k = 2$. □

Proposition 3.48. *Let $j \in \{1, \ldots, k-1\}$, and let n be the order of the image in $A_f(\mathbf{C})$ of the modular symbol $\mathbf{e}_j$, so $n = 1$ if $j \notin \{1, k-1\}$. Then*

$$\frac{L(A_f, j)}{\Omega_j} \in \frac{1}{n}\mathbf{Z}.$$

Proof. Let x denote the image of $\mathbf{e}_j \in A_f(\mathbf{C})$, and set $I = \mathrm{Ann}(x) \subset \mathbf{T}$. Though we write $A_f(\mathbf{C})$ here and below, we will always work within the subgroup of $A_f(\mathbf{C})$ generated by the image of $\boldsymbol{\mathcal{M}}_k(N, \varepsilon; \mathbf{Z})$ under the period map.

First we check that the Hecke operators all act as scalars on x. Since f is a *newform*, the Hecke operators T_p, for $p \mid N$, act as 0 or $\pm p^{k/2-1}$ on f, and hence in the same way on $A_f(\mathbf{C})$ (see, e.g., the end of section 6 of [21]). If $p \nmid N$, Lemma 3.47 shows that $T_p(x) = (1+p^{k-1})x$.

Let $C = \mathbf{Z}x$ denote the cyclic subgroup of $A_f(\mathbf{C})$ generated by x, so n is the order of C. Since the Hecke operators act as scalars on C, we are pleased to find that there is an injection $\mathbf{T}/I \hookrightarrow C$ which sends T_p to $T_p(x)$.

Setting $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}_k(N, \varepsilon; \mathbf{Z})$ and applying Theorem 3.41 we find that

$$\begin{aligned}
\pm\frac{L(A_f, j)}{\Omega_j} &= [\Theta_f(\boldsymbol{\mathcal{S}}^+) : \Theta_f(\mathbf{T}e)] \\
&= [\Theta_f(\boldsymbol{\mathcal{S}}^+) : \Theta_f(I\mathbf{e})] \cdot [\Theta_f(I\mathbf{e}) : \Theta_f(\mathbf{Te})] \\
&= [\Theta_f(\boldsymbol{\mathcal{S}}^+) : I\Theta_f(\mathbf{e})] \cdot [I\Theta_f(\mathbf{e}) : \mathbf{T}\Theta_f(\mathbf{e})] \\
&= \frac{[\Theta_f(\boldsymbol{\mathcal{S}}^+) : I\Theta_f(\mathbf{e})]}{[\mathbf{T}\Theta_f(\mathbf{e}) : I\Theta_f(\mathbf{e})]}.
\end{aligned}$$

To conclude that

$$\frac{[\Theta_f(\boldsymbol{\mathcal{S}}^+) : I\Theta_f(\mathbf{e})]}{[\mathbf{T}\Theta_f(\mathbf{e}) : I\Theta_f(\mathbf{e})]} \in \frac{1}{n}\mathbf{Z}$$

we make two observations. By the construction of $A_f(\mathbf{C})$, the ideal I consists of those elements of $\mathbf{T}$ that send $\Theta_f(\mathbf{e})$ into $\Theta_f(\boldsymbol{\mathcal{S}}^+)$, so $[\Theta_f(\boldsymbol{\mathcal{S}}^+) : I\Theta_f(\mathbf{e})] \in \mathbf{Z}$. Second, there is a surjective map

$$\mathbf{T}/I \to \frac{\mathbf{T}\Theta_f(\mathbf{e})}{I\Theta_f(\mathbf{e})}$$

sending t to $t\Theta_f(\mathbf{e})$, so $[\mathbf{T}\Theta_f(\mathbf{e}) : I\Theta_f(\mathbf{e})]$ divides $n = \#C = \#(\mathbf{T}/I)$. □

Remark 3.49 (Historical notes). In the special case when $k = 2$, the modular symbol $\mathbf{e}_1$ corresponds to $(0) - (\infty) \in J_0(N)$. In this situation, Manin proves at the bottom of page 28 of [38] that $(0) - (\infty) \in J_0(N)(\mathbf{Q})$, and asserts in the footnote to [38, Cor. 3.6] that $(0) - (\infty)$ has finite order. Based on observations such as a special case of the above proposition, he declares: "These explicit formulas have the structure predicted by the Birch-Swinnerton-Dyer conjectures."

The main result of this section was inspired by a weaker result of Agashe, which can be found in Chapter 4 of [2]. Agashe considers only the case $k = 2$ and replaces n by the order of the subgroup of $J_0(N)(\overline{\mathbf{Q}})$ generated by *all* cusps.

3.11 The Manin constant

In this section $k = 2$ and $\varepsilon = 1$; we sometimes omit k and ε from the notation. The assumption that $k = 2$ will be essential, because we do not know how to define a Manin constant in other weights, let alone bound it.

Consider the optimal quotient A of $J_0(N)$ corresponding to a *newform* f on $\Gamma_0(N)$ of weight 2. Let I_A be the kernel of the natural map from the Hecke algebra to $\mathrm{End}(A)$. The *Manin constant* c_A of A is the lattice index

$$c_A := [S_2(\Gamma_0(N); \mathbf{Z})[I_A] : H^0(\mathcal{A}, \Omega_{\mathcal{A}/\mathbf{Z}})]$$

taken inside of $S_2(\Gamma_0(N); \mathbf{Q})$. Though, a priori, c_A is a rational number, the work of [31] implies that $c_A \in \mathbf{Z}$ (see, e.g., [3]).

Generalizing a theorem of Mazur, we prove that c_A is a unit in $\mathbf{Z}[\frac{1}{2m}]$, where m is the largest square dividing N. Essentially no new ideas beyond what Mazur used are involved. We then conjecture that $c_A = 1$, and give supporting numerical evidence.

For related results involving modular "building blocks" for $J_1(N)$, we refer the reader to [26, §4].

3.11.1 The primes that might divide c_A

In the special case $\dim A = 1$, the Manin constant is the classical Manin constant of A, and in [41] Mazur proved that c_A is a unit in $\mathbf{Z}[\frac{1}{2m}]$. We generalize his proof to obtain the analogous result in dimension greater than 1.

Theorem 3.50. *Let A be the new optimal quotient of $J_0(N)$ corresponding to a newform f. Then the Manin constant c_A is a unit in $\mathbf{Z}[\frac{1}{2m}]$, where m is the largest square dividing N.*

Proof. The reader is strongly recommended to keep the proof of Proposition 3.1 in [41] at hand while reading the following argument.

Let π denote the map $J_0(N) \to A$; let $\mathcal{A}$ denote the Néron model of A over $R := \mathbf{Z}[\frac{1}{2m}]$, and $\mathcal{J}$ the Néron model of $J_0(N)$ over R. Let $\mathcal{X}$ be the minimal proper regular model for $X_0(N)$ over R. As in Mazur's proof in [41], consider the diagram

$$H^0(\mathcal{A}, \Omega_{\mathcal{A}}) \xrightarrow{\pi^*} H^0(\mathcal{J}, \Omega_{\mathcal{J}}) \cong H^0(\mathcal{X}, \Omega_{\mathcal{X}}^{\mathrm{reg}}) \xrightarrow{q\text{-exp}} R[[q]]. \tag{3.2}$$

(Note that "$\Omega_{\mathcal{X}}^{\mathrm{reg}}$" is not defined to be the usual sheaf of differentials; see, e.g., the discussion in [40, pg. 67].) The map π^* must be an inclusion, by [41, Cor. 1.1]. To show that the Manin constant is a unit in R, it suffices to check that the image of $H^0(\mathcal{A}, \Omega_{\mathcal{A}})$ in $R[[q]]$ is *saturated*, in the sense that the cokernel is torsion free; indeed, the image of $S_2(\Gamma_0(N); R)[I]$ is saturated and $S_2(\Gamma_0(N); R)[I] \otimes \mathbf{Q} = q\text{-exp}(\pi^*(H^0(\mathcal{A}, \Omega_{\mathcal{A}}))) \otimes \mathbf{Q}$.

For the image of $H^0(\mathcal{A},\Omega_{\mathcal{A}})$ in $R[[q]]$ to be saturated means that the quotient D is torsion free. Let ℓ be a prime not dividing $2m$; tensoring

$$0 \to H^0(\mathcal{A},\Omega_{\mathcal{A}}) \xrightarrow{q\text{-exp}} R[[q]] \to D \to 0$$

with $\mathbf{F}_\ell$ we obtain

$$0 \to D[\ell] \to H^0(\mathcal{A},\Omega_{\mathcal{A}}) \otimes \mathbf{F}_\ell \to \mathbf{F}_\ell[[q]] \to D \otimes \mathbf{F}_\ell \to 0.$$

Here we have used that $\mathrm{Tor}^1(D,\mathbf{F}_\ell)$ is the ℓ-torsion in D, and that $\mathrm{Tor}^1(-,\mathbf{F}_\ell)$ vanishes on the torsion-free group $R[[q]]$. (Alternatively, we could have used the snake lemma.) To show that $D[\ell]=0$, it suffices to prove that the map $\Psi : H^0(\mathcal{A},\Omega_{\mathcal{A}}) \otimes \mathbf{F}_\ell \to \mathbf{F}_\ell[[q]]$ is injective.

Since $\ell \neq 2$ and A is an optimal quotient, [41, Cor 1.1] gives an exact sequence

$$0 \to H^0(\mathcal{A}/\mathbf{Z}_\ell,\Omega_{\mathcal{A}/\mathbf{Z}_\ell}) \to H^0(\mathcal{J}/\mathbf{Z}_\ell,\Omega_{\mathcal{J}/\mathbf{Z}_\ell}) \to H^0(\mathcal{B}/\mathbf{Z}_\ell,\Omega_{\mathcal{B}/\mathbf{Z}_\ell}) \to 0$$

where $\mathcal{B}$ is the Néron model of $\ker(J \to A)$. In particular, $H^0(\mathcal{B}/\mathbf{Z}_\ell,\Omega_{\mathcal{B}/\mathbf{Z}_\ell})$ is torsion free, so

$$\begin{aligned} H^0(\mathcal{A}/\mathbf{Z}_\ell,\Omega_{\mathcal{A}/\mathbf{Z}_\ell}) \otimes \mathbf{F}_\ell \to H^0(\mathcal{J}/\mathbf{Z}_\ell,\Omega_{\mathcal{J}/\mathbf{Z}_\ell}) \otimes \mathbf{F}_\ell &\cong H^0(\mathcal{X}/\mathbf{Z}_\ell,\Omega^{\mathrm{reg}}_{\mathcal{X}/\mathbf{Z}_\ell}) \otimes \mathbf{F}_\ell \\ &\cong H^0(\mathcal{X}/\mathbf{F}_\ell,\Omega^{\mathrm{reg}}_{\mathcal{X}/\mathbf{F}_\ell}) \end{aligned}$$

is injective. (The last isomorphism is by [40, Prop. 3.3, pg. 68].) We also remark that

$$H^0(\mathcal{A},\Omega_{\mathcal{A}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{A}/\mathbf{Z}_\ell,\Omega_{\mathcal{A}/\mathbf{Z}_\ell}) \otimes \mathbf{F}_\ell,$$

because $\mathbf{Z}_\ell$ is torsion free, hence flat over R. Thus the map

$$H^0(\mathcal{A},\Omega_{\mathcal{A}}) \otimes \mathbf{F}_\ell \to H^0(\mathcal{X}/\mathbf{F}_\ell,\Omega^{\mathrm{reg}}_{\mathcal{X}/\mathbf{F}_\ell})$$

is injective.

If $\ell \nmid N$, then injectivity of Ψ now follows from the q-expansion principle, which asserts that the q-expansion map $H^0(\mathcal{X}/\mathbf{F}_\ell,\Omega^{\mathrm{reg}}_{\mathcal{X}/\mathbf{F}_\ell}) \to \mathbf{F}_\ell[[q]]$ is injective.

Suppose ℓ does divide N, and let $\omega \in \ker(\Psi)$. Since $\ell \mid N$ and $\ell \nmid 2m$, we have that $\ell \mid\mid N$; thus $\mathcal{X}/\mathbf{F}_\ell$ breaks up into a union of two irreducible components, and the q-expansion principle implies only that ω vanishes on the irreducible component containing the cusp ∞. However, since A is *new* and corresponds to a *single* eigenform, ω is an eigenvector for the involution W_N (since f and all of its conjugates are). Since W_N permutes the two components, ω must be 0 on all $\mathcal{X}/\mathbf{F}_\ell$. Therefore $\omega = 0$, and hence Ψ is injective. □

3.11.2 Numerical evidence for the $c_A = 1$ conjecture

In the paper [24], the authors show that $c_A = 1$ for 28 two-dimensional optimal quotients of $J_0(N)$ (see Section 3.12.8). The non-square-free levels treated are:

$$N = 3^2\cdot 7,\quad 3^2\cdot 13,\quad 5^3,\quad 3^3\cdot 5,\quad 3\cdot 7^2,\quad 5^2\cdot 7,\quad 2^2\cdot 47,\quad 3^3\cdot 7.$$

In every case, $c_A = 1$.

Conjecture 3.51 (Agashe). *Let A be an optimal quotient of $J_0(N)$, and let c_A be the corresponding Manin constant. Then $c_A = 1$.*

3.12 Analytic invariants

Fix a newform

$$f = \sum_{n\geq 1} a_n q^n \in S_k(N,\varepsilon),$$

and assume that $\varepsilon^2 = 1$.

Remark 3.52. Our assumption that $\varepsilon^2 = 1$ does not imply that f has totally real Fourier coefficients. There is an eigenform in $S_2(24,\varepsilon)$ whose Fourier coefficients are not totally real, where ε is one of the characters of conductor 8.

Let $K_f = \mathbf{Q}(\ldots a_n \ldots)$ and let $f_1, \ldots, f_d$ be the Galois conjugates of f, where $d = [K_f : \mathbf{Q}]$. As in Section 2.7, we consider the complex torus A_f attached to f. In this section we describe how to compute the torus A_f and the special values at the critical integers $1, 2, \ldots, k-1$ of the L function $L(A_f, s)$ associated to A_f. (See 3.39 for the definition of $L(A_f, s)$.)

Let

$$f = \sum_{n\geq 1} a_n q^n \in M_k(N,\varepsilon)$$

be a modular form (we do not assume that f is an eigenform). We recall the integration pairing of Theorem 2.7:

$$\langle\, , \rangle : M_k(N,\varepsilon) \times \boldsymbol{\mathcal{M}}_k(N,\varepsilon) \longrightarrow \mathbf{C}$$

$$\langle f, P\{\alpha,\beta\}\rangle = 2\pi i \int_{\alpha}^{\beta} f(z)P(z,1)dz.$$

Let $I_f \subset \mathbf{T}$ be the kernel of the map $\mathbf{T} \to K_f$ sending T_n to a_n. The integration pairing gives rise to the period mapping

$$\Phi_f : \boldsymbol{\mathcal{M}}_k(N,\varepsilon) \to \mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon)[I_f], \mathbf{C}),$$

and $A_f = \mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon)[I_f], \mathbf{C})/\Phi_f(\boldsymbol{\mathcal{S}}_k(N,\varepsilon))$ is the cokernel.

3.12.1 Extended modular symbols

For the purposes of computing periods, it is advantageous to extend the notion of modular symbols to allows symbols of the form $P\{z,w\}$ where z and w are now arbitrary elements of $\mathfrak{h}^* = \mathfrak{h} \cup \mathbf{P}^1(\mathbf{Q})$. The free abelian group $\overline{\boldsymbol{\mathcal{M}}}_k$ of *extended modular symbols* is spanned by such symbols, and is of uncountable rank over $\mathbf{Z}$. However, it is still equipped with an action of $\Gamma_0(N)$ and we can form the largest torsion-free quotient $\overline{\boldsymbol{\mathcal{M}}}_k(N,\varepsilon)$ of $\overline{\boldsymbol{\mathcal{M}}}_k$ by the relations $\gamma x = \varepsilon(\gamma)x$ for $\gamma \in \Gamma_0(N)$.

The integration pairing extends to $\overline{\boldsymbol{\mathcal{M}}}_k(N,\varepsilon)$. There is a natural embedding $\iota : \boldsymbol{\mathcal{M}}_k(N,\varepsilon) \hookrightarrow \overline{\boldsymbol{\mathcal{M}}}_k(N,\varepsilon)$ which respects the pairing in the sense that $\langle f, \iota(x)\rangle = \langle f, x\rangle$. In many cases it is advantageous to replace $x \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ first by $\iota(x)$, and then by an equivalent sum $\sum y_i$ of symbols $y_i \in \overline{\boldsymbol{\mathcal{M}}}_k(N,\varepsilon)$. The period $\langle f, x\rangle$ is then replaced by the equivalent sum of periods $\sum\langle f, y_i\rangle$. The latter is frequently *much* easier to approximate numerically.

3.12.2 Numerically computing period integrals

Consider a point α in the upper half plane and any one of the (extended) modular symbols $X^mY^{k-2-m}\{\alpha,\infty\}$. Given a cusp form $g=\sum_{n\geq 1} b_nq^n \in S_k(N,\varepsilon)$ and an integer $m\in\{0,1,\ldots,k-2\}$, we find that

$$\langle g,\, X^mY^{k-2-m}\{\alpha,\infty\}\rangle = 2\pi i\int_{\alpha}^{i\infty} g(z)z^m dz = 2\pi i\sum_{n=1}^{\infty} b_n\int_{\alpha}^{i\infty} e^{2\pi i n z}z^m dz. \tag{3.3}$$

The reversal of summation and integration is justified because the imaginary part of α is positive so that the sum converges absolutely. This is made explicit in the following lemma, which can be proved using repeated integration by parts.

Lemma 3.53.

$$\int_{\alpha}^{i\infty} e^{2\pi i n z}z^m dz \;=\; e^{2\pi i n\alpha}\sum_{s=0}^{m}\left(\frac{(-1)^s\alpha^{m-s}}{(2\pi i n)^{s+1}}\prod_{j=(m+1)-s}^{m} j\right). \tag{3.4}$$

The following proposition is the higher weight analogue of [16, Prop. 2.1.1(5)].

Proposition 3.54. *For any $\gamma\in\Gamma_0(N)$, $P\in V_{k-2}$ and $\alpha\in\mathfrak{h}^*$ the following holds:*

$$\begin{aligned} P\{\infty,\gamma(\infty)\} &= P\{\alpha,\gamma(\alpha)\} + (P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\} && (3.5)\\ &= \varepsilon(\gamma)(\gamma^{-1}P)\{\alpha,\infty\} - P\{\gamma(\alpha),\infty\}. && (3.6)\end{aligned}$$

Proof. By definition, if $x\in\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ is a modular symbol and $\gamma\in\Gamma_0(N)$ then $\gamma x=\varepsilon(\gamma)x$; in particular, $\varepsilon(\gamma)\gamma^{-1}x = x$, so

$$\begin{aligned} P\{\infty,\gamma(\infty)\} &= P\{\infty,\alpha\}+P\{\alpha,\gamma(\alpha)\}+P\{\gamma(\alpha),\gamma(\infty)\}\\ &= P\{\infty,\alpha\}+P\{\alpha,\gamma(\alpha)\}+\varepsilon(\gamma)\gamma^{-1}(P\{\gamma(\alpha),\gamma(\infty)\})\\ &= P\{\infty,\alpha\}+P\{\alpha,\gamma(\alpha)\}+\varepsilon(\gamma)(\gamma^{-1}P)\{\alpha,\infty\}\\ &= P\{\alpha,\gamma(\alpha)\}+P\{\infty,\alpha\}-\varepsilon(\gamma)(\gamma^{-1}P)\{\infty,\alpha\}\\ &= P\{\alpha,\gamma(\alpha)\}+(P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\}.\end{aligned}$$

The second equality in the statement of the proposition now follows easily. □

In the classical case of weight two and trivial character, the error term $(P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\}$ vanishes. In general this term does not vanish, instead perturbing the analogues of the formulas found in [16, 2.10].

Algorithm 3.55. Given a triple $\gamma\in\Gamma_0(N)$, $P\in V_{k-2}$ and $g\in S_k(N,\varepsilon)$ (as a q-expansion to some precision) this algorithm computes the period integral $\langle g,\, P\{\infty,\gamma(\infty)\}\rangle$. Express γ as $\left(\begin{smallmatrix} a & b\\ cN & d\end{smallmatrix}\right)\in\Gamma_0(N)$ and take $\alpha=\frac{-d+i}{cN}$ in Proposition 3.54. Replacing γ by $-\gamma$ if necessary, we find that the imaginary parts of α and $\gamma(\alpha)=\frac{a+i}{cN}$ are both equal to $1/(cN)$ which is positive. Equation 3.3 and Lemma 3.53 can now be used to compute the period integrals of Proposition 3.54.

With the goal of computing period lattices in mind, it is reassuring to know that every element of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$ can be written as a linear combination of symbols of the form $P\{\infty,\gamma(\infty)\}$. The author asked Helena Verrill if this is the case and she was eventually able to prove that it is; the proof is given below. In the special case of weight two and trivial character, this is the assertion, which was proved by Manin [38], that the group homomorphism $\Gamma_0(N) \to H_1(X_0(N),\mathbf{Z})$ sending γ to $\{0,\gamma(0)\}$ is surjective. When the weight is greater than two, we have not found any similar group-theoretic statement.

Proposition 3.56. *Any element of $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$ can be written in the form*

$$\sum_{i=1}^{n} P_i\{\infty,\gamma_i(\infty)\}$$

with $P_i \in V_{k-2}$ and $\gamma_i \in \Gamma_0(N)$. Moreover, P_i and γ_i can be chosen so that $\sum \varepsilon(\gamma_i)P_i = \sum \gamma_i^{-1}P_i$.

Proof.[1] First recall the definition of the spaces $\boldsymbol{\mathcal{M}}$, $\boldsymbol{\mathcal{M}}_k = V_{k-2}\otimes\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{M}}_k(N,\varepsilon) = \boldsymbol{\mathcal{M}}_k/I$ (see Section 2.1). Let $I = I_{N,\varepsilon}$ be the ideal in the group ring of $\Gamma_0(N)$ generated by all elements of the form $\varepsilon(\gamma)-\gamma$ for $\gamma\in\Gamma_0(N)$.

Suppose $v \in \boldsymbol{\mathcal{S}}_k(N,\varepsilon)$. Use the relation $\{\alpha,\beta\} = \{\infty,\beta\} - \{\infty,\alpha\} \in \boldsymbol{\mathcal{M}}$ to see that any v is the image of an element $\tilde{v} \in \boldsymbol{\mathcal{M}}_k$ of the form

$$\tilde{v} = \sum_{\beta\in\mathbf{Q}} P_\beta \otimes \{\infty,\beta\} \in \boldsymbol{\mathcal{M}}_k$$

with only finitely many P_β nonzero. The boundary map δ lifts in a natural way to $V_{k-2}\otimes\boldsymbol{\mathcal{M}}$, as illustrated.

$$\begin{tikzcd}
 & I(V_{k-2}\otimes\boldsymbol{\mathcal{M}}) \arrow[r] \arrow[d] & I(V_{k-2}\otimes\boldsymbol{\mathcal{B}}) \arrow[d] \\
 & V_{k-2}\otimes\boldsymbol{\mathcal{M}} \arrow[r, "\tilde{\delta}"] \arrow[d] & V_{k-2}\otimes\boldsymbol{\mathcal{B}} \arrow[d] \\
\boldsymbol{\mathcal{S}}_k(N,\varepsilon) \arrow[r, hook] & \boldsymbol{\mathcal{M}}_k(N,\varepsilon) \arrow[r, "\delta"] & \boldsymbol{\mathcal{B}}_k(N,\varepsilon)
\end{tikzcd}$$

Our assumption that $\delta(v)=0$ implies that $\tilde{\delta}(\tilde{v}) \in I(V_{k-2}\otimes\boldsymbol{\mathcal{B}})$. So there are $Q_{\gamma,\beta}\in V_{k-2}$, for $\gamma\in\Gamma_0(N)$ and $\beta\in\mathbf{P}^1(\mathbf{Q})$, only finitely many nonzero, such that

$$\tilde{\delta}(\tilde{v}) = \sum_{\gamma,\beta}(\varepsilon(\gamma)-\gamma)(Q_{\gamma,\beta}\otimes\{\beta\}).$$

[1] The author thanks Helena Verrill for permission to reproduce her proof here.

We now use a summation trick.

$$\begin{aligned}\sum_{\beta\in\mathbf{Q}}\tilde{\delta}(\tilde{v}) = P_\beta\otimes\{\beta\} - P_\beta\otimes\{\infty\} &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\otimes\{\beta\} - (\gamma Q_{\gamma,\beta})\otimes\{\gamma\beta\}\\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\otimes\{\beta\} - (\gamma Q_{\gamma,\gamma^{-1}\beta})\otimes\{\beta\}\\ &= \sum_{\gamma,\beta}\Big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\Big)\otimes\{\beta\}.\end{aligned}$$

Equating terms we deduce that for $\beta\neq\infty$,

$$P_\beta = \sum_{\gamma}\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}.$$

Using this expression for P_β and that $\varepsilon(\gamma)\gamma^{-1}$ acts trivially on $\boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ we find that

$$\begin{aligned}v = \sum_{\beta}P_\beta\{\infty,\beta\} &= \sum_{\gamma,\beta}\Big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\Big)\{\infty,\beta\}\\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta} - \varepsilon(\gamma)\gamma^{-1}\Big(\gamma Q_{\gamma,\gamma^{-1}\beta}\Big)\{\infty,\beta\}\\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty,\beta\} - \varepsilon(\gamma)Q_{\gamma,\gamma^{-1}\beta}\{\gamma^{-1}\infty,\gamma^{-1}\beta\}\\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty,\beta\} - \varepsilon(\gamma)Q_{\gamma,\beta}\{\gamma^{-1}\infty,\beta\}\\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty,\gamma^{-1}\infty\}.\end{aligned}$$

This is of the desired form. □

Unlike the case of weight two and trivial character, Proposition 3.56 does not give generators for $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$. This is because not every element of the form $P\{\infty,\gamma(\infty)\}$ must lie in $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$. However, if $\gamma P = P$ then $P\{\infty,\gamma(\infty)\}$ does lie in $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$. It would be interesting to know under what circumstances $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$ is generated by symbols of the form $P\{\infty,\gamma(\infty)\}$ with $\gamma P = P$. This sometimes fails for k odd; for example, when $k=3$ the condition $\gamma P = P$ implies that $\gamma\in\Gamma_0(N)$ has an eigenvector with eigenvalue 1, hence is of finite order. When k is even the author can see no obstruction to generating $\boldsymbol{\mathcal{S}}_k(N,\varepsilon)$ using such symbols.

3.12.3 The W_N-trick

In this section we assume that k is even. Consider the involution W_N defined in Section 2.4.3. This is an involution that acts on both modular symbols and modular forms. The follow proposition shows how to compute $\langle g, P\{\infty,\gamma(\infty)\}\rangle$ under certain restrictive assumptions. It generalizes the main result of [17] to higher weight. (Compare also [25].)

Proposition 3.57. *Let $g \in S_k(N, \varepsilon)$ be a cusp form which is an eigenform for the Atkin-Lehner involution W having eigenvalue $w \in \{\pm 1\}$. Then for any $\gamma \in \Gamma_0(N)$ and any $P \in V_{k-2}$, with the property that $\gamma P = \varepsilon(\gamma) P$, we have for any $\alpha \in \mathfrak{h}$ the following formula:*

$$\langle g, P\{\infty, \gamma(\infty)\}\rangle =$$

$$\langle g, w\frac{P(Y,-NX)}{N^{k/2-1}}\{W(\alpha),\infty\} + (P - w\frac{P(Y,-NX)}{N^{k/2-1}})\{i/\sqrt{N},\infty\} - P\{\gamma(\alpha),\infty\}\rangle.$$

Here $W(\alpha) = -1/(N\alpha)$.

Proof. By Proposition 3.54 our condition on P implies that $P\{\infty, \gamma(\infty)\} = P\{\alpha, \gamma(\alpha)\}$. The steps of the following computation are described below.

$$\begin{aligned}
&\langle g, P\{\alpha, \gamma(\alpha)\}\rangle \\
&\quad= \langle g, P\{\alpha, i/\sqrt{N}\} + P\{i/\sqrt{N}, W(\alpha)\} + P\{W(\alpha), \gamma(\alpha)\}\rangle \\
&\quad= \langle g, w\frac{W(P)}{N^{k/2-1}}\{W(\alpha), i/\sqrt{N}\} + P\{i/\sqrt{N}, W(\alpha)\} + P\{W(\alpha), \gamma(\alpha)\}\rangle \\
&\quad= \langle g, (w\frac{W(P)}{N^{k/2-1}} - P)\{W(\alpha), i/\sqrt{N}\} + P\{W(\alpha), \infty\} - P\{\gamma(\alpha), \infty\}\rangle \\
&\quad= \langle g, w\frac{W(P)}{N^{k/2-1}}\{W(\alpha), \infty\} + (P - w\frac{W(P)}{N^{k/2-1}})\{i/\sqrt{N}, \infty\} - P\{\gamma(\alpha), \infty\}\rangle.
\end{aligned}$$

For the first step, we break the path into three paths. In the second step, we apply the W-involution to the first term, and use that the action of W is compatible with the pairing $\langle\, , \rangle$. The third step involves combining the first two terms and breaking up the third. In the final step, we replace $\{W(\alpha), i/\sqrt{N}\}$ by $\{W(\alpha), \infty\} + \{\infty, i/\sqrt{N}\}$ and regroup. □

A good choice for α is $\alpha = \gamma^{-1}\left(\frac{b}{d} + \frac{i}{d\sqrt{N}}\right)$, so that $W(\alpha) = \frac{c}{d} + \frac{i}{d\sqrt{N}}$. This maximizes the minimum of the imaginary parts of α and $W(\alpha)$.

Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$. A polynomial P for which $\gamma(P) = P$ is given by

$$P(X, Y) = (cX^2 + (d-a)XY - bY^2)^{\frac{k-2}{2}}.$$

This formula was obtained by viewing V_{k-2} as the $(k-2)$th symmetric product of the two-dimensional space on which $\Gamma_0(N)$ acts naturally. For example, observe that since $\det(\gamma) = 1$ the symmetric product of two eigenvectors for γ is an eigenvector in V_2 having eigenvalue 1. For the same reason, if $\varepsilon(\gamma) \neq 1$, there is often no polynomial $P(X, Y)$ such that $\gamma(P) = \varepsilon(\gamma)P$. When this is the case, first choose γ so that $\varepsilon(\gamma) = 1$.

Since the imaginary parts of the terms $i/\sqrt{N}$, α and $W(\alpha)$ in the proposition are all relatively large, the sums appearing in Equation 3.3 converge quickly if d is small. Let us emphasize, that *it is* **extremely** *important to choose γ in Proposition 3.57 with d small, otherwise the series will converge* very *slowly.*

3.12.4 Computing the period mapping

Let $I \subset \mathbf{T}$ be the kernel of the map $\mathbf{T} \to K_f$ sending T_n to a_n. As in Section 3.7, let $\Theta_f = \Theta_I$ be the rational period mapping associated to f. We have a commutative diagram

$$\begin{tikzcd}
\boldsymbol{\mathcal{M}}_k(N,\varepsilon) \arrow[rr, "\Phi_f"] \arrow[dr, "\Theta_f"'] & & \mathrm{Hom}_{\mathbf{C}}(S_k(N,\varepsilon)[I],\mathbf{C}) \\
& \dfrac{\boldsymbol{\mathcal{M}}_k(N,\varepsilon)}{\ker(\Phi_f)} \arrow[ur, "i_f"] &
\end{tikzcd}$$

Using Algorithm 3.26, we can compute Θ_f so to compute Φ_f we need to compute i_f. Let $g_1, \ldots, g_d$ be a basis for the $\mathbf{Q}$-vector space $S_k(N,\varepsilon;\mathbf{Q})[I]$. We will compute the period mapping with respect to the basis of $\mathrm{Hom}_{\mathbf{Q}}(S_k(N,\varepsilon;\mathbf{Q})[I],\mathbf{C})$ dual to this basis. Choose elements $x_1, \ldots, x_d \in \boldsymbol{\mathcal{M}}_k(N,\varepsilon)$ with the following properties:

1. Using Proposition 3.54 or Proposition 3.57 it is possible to compute the period integrals $\langle g_i, x_j\rangle$, $i, j \in \{1, \ldots d\}$ efficiently.
2. The $2d$ elements $v + *v$ and $v - *v$ for $v = \Theta_f(x_1), \ldots, \Theta_f(x_d)$ span a space of dimension $2d$.

Given this data, we can compute

$$i_f(v + *v) = 2\mathrm{Re}(\langle g_1, x_i\rangle, \ldots, \langle g_d, x_i\rangle)$$

and

$$i_f(v - *v) = 2i\mathrm{Im}(\langle g_1, x_i\rangle, \ldots, \langle g_d, x_i\rangle).$$

We break the integrals into real and imaginary parts because this increases the precision of our answers. Since the vectors $v_n + *v_n$ and $v_n - *v_n$, $n = 1, \ldots, d$ span $\frac{\boldsymbol{\mathcal{M}}_k(N,\varepsilon)}{\ker(\Phi_f)}$ we have computed i_f.

It is advantageous when possible to find symbols x_i satisfying the conditions of Proposition 3.57. This is usually possible when d is very small, but in practice we have had problems doing this when d is large, for example with **131k2B**, in which case the dimension is 10.

3.12.5 Computing special values

For $s = 1, \ldots, k-1$ we have

$$L(f,s) = \frac{-2\pi^{s-1}i^{s-1}}{(s-1)!} \cdot \langle f, X^{s-1}Y^{k-1-s}\{0,\infty\}\rangle, \tag{3.7}$$

$$L(A_I, s) = \prod_{i=1}^{d} L(f_i, s). \tag{3.8}$$

Let

$$\mathbf{e}_i := X^{i-1}\{0,\infty\}$$

denote the ith *winding element*. In section 3.12.4 we computed the period map Φ_f with respect to a basis $g_1, \ldots, g_d$ for $S_k(N, \varepsilon; \mathbf{Q})[I]$. Upon writing f as a K_f-linear combination $\alpha_1 g_1 + \cdots + \alpha_d g_d$ we find that

$$\begin{aligned}\langle f, \mathbf{e}_i\rangle &= \langle \alpha_1 g_1 + \cdots + \alpha_d g_d, \mathbf{e}_i\rangle\\ &= \alpha_1\langle g_1, \mathbf{e}_i\rangle + \cdots + \alpha_d\langle g_d, \mathbf{e}_i\rangle\\ &= \alpha_1 \Phi_f(\mathbf{e}_i)_1 + \cdots + \alpha_d \Phi_f(\mathbf{e}_i)_d\end{aligned}$$

Here $\Phi_f(\mathbf{e}_i)_j$ denotes the jth coordinate of $\Phi_f(\mathbf{e}_i)$. Finally using Equation 3.7 we compute the special value.

3.12.6 The real and minus volume associated to A_f

Fix a choice of basis $g_1, \ldots, g_d$ for the free $\mathbf{Z}$-module $S_k(N, \varepsilon; \mathbf{Z})[I]$, where I is the annihilator in the Hecke algebra of our fixed newform f.

For any $x \in \boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ we have, by Proposition 2.11,

$$\begin{aligned}\overline{\Phi_f(x)} &= (\overline{\langle g_1, x\rangle}, \ldots, \overline{\langle g_d, x\rangle})\\ &= (\langle g_1^*, x^*\rangle, \ldots, \langle g_d^*, x^*\rangle)\\ &= (\langle g_1, x^*\rangle, \ldots, \langle g_d, x^*\rangle) \in \Phi_f(\boldsymbol{\mathcal{S}}_k(N, \varepsilon)),\end{aligned}$$

so complex conjugation leaves invariant the period lattice

$$\Lambda_f = \Phi_f(\boldsymbol{\mathcal{S}}_k(N, \varepsilon)) \subset \mathrm{Hom}_{\mathbf{C}}(S_k(N, \varepsilon)[I], \mathbf{C}).$$

Fix a $\mathbf{Z}$-basis for $S_k(N, \varepsilon; \mathbf{Z})[I]$, thus making an identification $\mathrm{Hom}_{\mathbf{C}}(S_k(N, \varepsilon)[I], \mathbf{C}) \cong \mathbf{C}^d$. The above observation implies that $A_f(\mathbf{C}) \cong \mathbf{C}^d/\Lambda_f$ is equipped with an action of complex conjugation. Our choice of basis defines a real-valued measure μ on $A_f(\mathbf{C})$, coming from the standard measure on $\mathbf{C}^d$. The measure does not depend on the choice of $\mathbf{Z}$-basis.

Definition 3.58 (Real and minus volume). The *real measure* Ω_f^+ is the measure $\mu(A_f(\mathbf{R}))$. The *minus measure* Ω_f^- is the measure $\mu(A_f(\mathbf{C})^-)$ times i^d, where $A_f(\mathbf{C})^-$ is the set of points in $A_f(\mathbf{C})$ on which complex conjugation acts as -1.

Thus, in connection with Section 3.10.3, $|\Omega_f^+| = |\Omega_1|$ and $|\Omega_f^-| = |\Omega_2|$.

Algorithm 3.59. To compute Ω_f^+ and Ω_f^-, proceed as follows. Using Algorithm 3.26, compute $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)/\mathrm{Ker}(\Phi_f)$. Next, compute a basis for the kernel $(\boldsymbol{\mathcal{S}}_k(N, \varepsilon)/\mathrm{Ker}(\Phi_f))^+$ of the map induced by the $*$-involution. Using Section 3.12.4 compute the image of this basis under i_f; this is a basis for Λ_f^+. The determinant of this latter basis then gives the measure $(\Omega_f^+)^0$ of the identity component $A_f(\mathbf{R})^0$ of $A_f(\mathbf{R})$. Finally $\Omega_f^+ = c_\infty^+ \cdot (\Omega_f^+)^0$, where the number c_∞^+ of real components can be computed using the algorithm in Section 3.12.7

Remark 3.60 (Alternative method). Suppose s is an integer in the set $\{1, \ldots, k-1\}$, and let $\sigma = +$ or $\sigma = -$, depending on whether s is odd or even, respectively. Section 3.10 contains a formula for the ratio $L(A_f, s)/\Omega_f^\sigma$. When this ratio is nonzero, Ω_f^σ can be determined by computing $L(A_f, s)/\Omega_f^\sigma$ and $L(A_f, s)$, using Section 3.12.5.

Remark 3.61. When $k = 2$ and ε is trivial, A_f has the structure of abelian variety over $\mathbf{Q}$. The quantity Ω_f^+ above is related to the quantity Ω_A appearing in the Birch and Swinnerton-Dyer conjecture [67] for A_f. The latter quantity is the measure of $A_f(\mathbf{R})$ with respect to a basis of integral differentials on the Néron model of A_f over $\mathrm{Spec}(\mathbf{Z})$. The two quantities are related by the Manin constant, which the author conjectures is always 1 (see Section 3.11).

3.12.7 The component groups c_∞^+ and c_∞^-

Assume in this section that f has *totally real* Fourier coefficients and continue to assume that $\varepsilon^2 = 1$.

Definition 3.62. Let c_∞^+ be the number of components of the topological space $A_f(\mathbf{R})$. Let c_∞^- be the number of components of $A_f(\mathbf{C})^-$, where $A_f(\mathbf{C})^-$ is the set of points $z \in A_f(\mathbf{C})$ such that $\overline{z} = -z$.

Proposition 3.63. *Let $\overline{C}$ be the map induced by complex conjugation on $\Lambda_f/2\Lambda_f = \Lambda_f \otimes \mathbf{F}_2$. Then*

$$c_\infty^+ = c_\infty^- = 2^{\dim(\ker(\overline{C}-1))-d},$$

where d is the dimension of A_f.

Proof. We must compute the order of the component group

$$\Psi = \frac{A_f(\mathbf{R})}{A_f(\mathbf{R})^0} = \frac{(\mathbf{C}^d/\Lambda_f)^+}{\mathbf{R}^d/\Lambda_f^+},$$

where $\mathbf{R}^d/\Lambda_f^+$ is the identity component because it is the continuous image of the connected set $\mathbf{R}^d$. For $v \in \mathbf{C}^d$ denote by $\overline{v}$ its complex conjugate and by $[v]$ its image in $\mathbf{C}^d/\Lambda_f$. Suppose $[v] \in (\mathbf{C}^d/\Lambda_f)^+$; this means that $[v] = [\overline{v}]$, so since $v + \overline{v} \in \mathbf{R}^d$ we have

$$2[v] = [v] + [\overline{v}] \in \mathbf{R}^d/\Lambda_f^+,$$

so Ψ is annihilated by 2. Thus there is $\lambda \in \Lambda_f$ so that $2v + \lambda \in \mathbf{R}^d$, and so $v + \frac{1}{2}\lambda \in \mathbf{R}^d$, i.e., v can be written as an element of $\frac{1}{2}\Lambda_f$ plus an element of $\mathbf{R}^d$. This means that Ψ is generated by the image of $(\frac{1}{2}\Lambda_f/\Lambda_f)^+$. Thus

$$\Psi \cong \frac{(\frac{1}{2}\Lambda_f/\Lambda_f)^+}{(\frac{1}{2}\Lambda_f \cap \mathbf{R}^d)/\Lambda_f^+} \cong \frac{(\Lambda_f/2\Lambda_f)^+}{\Lambda_f^+/2\Lambda_f^+}$$

Consequently

$$\dim_{\mathbf{F}_2} \Psi = \dim(\Lambda_f/2\Lambda_f)^+ - \dim \Lambda_f^+/2\Lambda_f^+ = \dim(\ker(\overline{C} - 1)) - d.$$

Here $\Lambda_f^+/2\Lambda_f^+$ has dimension d because Λ_f^+ is a lattice in $\mathbf{R}^d$, hence a free $\mathbf{Z}$-module of rank d.

The argument for c_∞^- proceeds in the same way, and results in the same answer because

$$\dim(\ker(\overline{C} - 1)) = \dim(\ker(\overline{C} + 1)).$$

□

To compute C on Λ_f, use Algorithm 3.26 to compute the action of $*$ on

$$\boldsymbol{\mathcal{S}}_k(N, \varepsilon)/\operatorname{Ker}(\Phi_f) \cong \Lambda_f.$$

3.12.8 Examples

Jacobians of genus-two curves

The author is among the the six authors of [24], who gather empirical evidence for the BSD conjecture for Jacobian of genus two curves. Of the 32 Jacobians considered, all but four are optimal quotients of $J_0(N)$ for some N. The methods of this section can be used to compute Ω_f^+ for the Jacobians of these 28 curves. Using explicit models for the genus two curves, the authors of [24] computed the measure of A with respect to a basis for the Néron differentials of A. In all 28 cases our answers agreed to the precision computed. Thus in these cases we have numerically verified that the Manin constant equals 1.

The first example considered in [24] is the Jacobian $A = J_0(23)$ of the modular curve $X_0(23)$. This curve has as a model

$$y^2 + (x^3 + x + 1)y = -2x^5 - 3x^2 + 2x - 2$$

from which one can compute the BSD $\Omega_A = 2.7328....$ The following is an integral basis of cusp forms for $S_2(23)$.

$$\begin{aligned} g_1 &= q - q^3 - q^4 - 2q^6 + 2q^7 + \cdots \\ g_2 &= q^2 - 2q^3 - q^4 + 2q^5 + q^6 + 2q^7 + \cdots \end{aligned}$$

The space $\boldsymbol{\mathcal{M}}_2(23;\mathbf{Q})$ of modular symbols has dimension five and is spanned by $\{-1/19, 0\}$, $\{-1/17, 0\}$, $\{-1/15, 0\}$, $\{-1/11, 0\}$ and $\{\infty, 0\}$. The submodule $\boldsymbol{\mathcal{S}}_2(23;\mathbf{Z})$ has rank four and has as basis the first four of the above five symbols. Choose $\gamma_1 = \left(\begin{smallmatrix} 8 & 1 \\ 23 & 3 \end{smallmatrix}\right)$ and $\gamma_2 = \left(\begin{smallmatrix} 6 & 1 \\ 23 & 4 \end{smallmatrix}\right)$ and let $x_i = \{\infty, \gamma_i(\infty)\}$. Using the W_N-trick (see Section 3.12.3) we compute the period integrals $\langle g_i, x_j\rangle$ using 97 terms of the q-expansions of g_1 and g_2, and obtain

$$\begin{aligned} \langle g_1, x_1\rangle &\sim -1.3543 + 1.0838i, \qquad & \langle g_1, x_2\rangle &\sim -0.5915 + 1.6875i \\ \langle g_2, x_1\rangle &\sim -0.5915 - 0.4801i, \qquad & \langle g_2, x_2\rangle &\sim -0.7628 + 0.6037i \end{aligned}$$

Using 97 terms we already obtain about 14 decimal digits of accuracy, but we do not reproduce them all here. We next find that

$$\langle g_1, x_1 + x_1^*\rangle \sim 2\mathrm{Re}(-1.3543 + 1.0838i) = 2.7086,$$

and so on. Upon writing each generator of $\boldsymbol{\mathcal{S}}_2(23)$ in terms of $x_1 + x_1^*$, $x_1 - x_1^*$, $x_2 + x_2^*$ and $x_2 - x_2^*$ we discover that the period mapping with respect to the basis dual to g_1 and g_2 is (approximately)

$$\begin{aligned} \{-1/19, 0\} &\mapsto (\ \ 0.5915 - 1.6875i, \quad 0.7628 - 0.6037i) \\ \{-1/17, 0\} &\mapsto (-0.5915 - 1.6875i, \quad -0.7628 - 0.6037i) \\ \{-1/15, 0\} &\mapsto (-1.3543 - 1.0838i, \quad -0.5915 + 0.4801i) \\ \{-1/11, 0\} &\mapsto (-1.5256, \qquad\qquad\qquad 0.3425) \end{aligned}$$

Working in $\boldsymbol{\mathcal{S}}_2(23)$ we find $\boldsymbol{\mathcal{S}}_2(23)^+$ is spanned by $\{-1/19, 0\} - \{-1/17, 0\}$ and $\{-1/11, 0\}$. Using the algorithm of Section 3.12.6, we find that there is only one real component so

$$\Omega_I^+ \sim \left|\begin{array}{cc} 1.1831 & 1.5256 \\ -1.5256 & 0.3425 \end{array}\right| = 2.7327...$$

To greater precision we find that $\Omega_f^+ \sim 2.7327505324965$. This agrees with the value in [24]; since the Manin constant is an integer, it must equal 1.

Table 3.1: Volumes associated to level one cusp forms.

k	Ω^+	Ω^-
12	0.002281474899	$0.000971088287i$
16	0.003927981492	$0.000566379403i$
18	0.000286607497	$0.023020042428i$
20	0.008297636952	$0.0005609325015i$
22	0.002589288079	$0.0020245743816i$
24	0.000000002968	$0.0000000054322i$
26	0.003377464512	$0.3910726132671i$
28	0.000000015627	$0.0000000029272i$

Level one cusp forms

In the following two sections we consider several specific examples of tori attached to modular forms of weight greater than two.

Let $k \geq 12$ be an even integer. Associated to each Galois conjugacy class of normalized eigenforms f, there is a torus A_f over $\mathbf{R}$. The real and minus volume of the first few of these tori are displayed in Table 3.1. For weights 24 and 28 we give Ω^-/i so that the columns will line up nicely. In each case, 97 terms of the q-expansion were used.

The volumes appear to be *much* smaller than the volumes of weight two abelian varieties. The dimension of each A_f is 1, except for weights 24 and 28 when the dimension is 2.

CM elliptic curves of weight greater than two

Let f be a rational newform with “complex multiplication”, in the sense that “half” of the Fourier coefficients of f are zero. For our purposes, it is not necessary to define complex multiplication any more precisely. Experimentally, it appears that the associated elliptic A_f has rational j-invariant. As evidence for this we present Table 3.2, which includes the analytic data about every rational CM form of weight four and level $\leq$ 197. The computations of Table 3.2 were done using at least 97 terms of the q-expansion of f. The rationality of j could probably be proved by observing that the CM forces A_f to have extra automorphisms.

In these examples, the invariants c_4 and c_6 are unrecognizable to the author; in contrast, in weight 2 these invariants are (expected to be) integers (see [16, 2.14]).

Some abelian varieties of large dimension

In Table 3.3, we give the volumes of five abelian varieties of dimension greater than 1. In each case, at least 200 terms of the q-expansions were used.

Table 3.2: CM elliptic curves of weight > 2.

E	j	Ω^+	Ω^-	c_4	c_6
9k4A	0	0.2095	$0.1210i$	0.0000	-56626421686.2951
32k4A	1728	0.2283	$0.2283i$	-3339814.8874	0.0000
64k4D	1728	0.1614	$0.1614i$	53437038.1988	0.0000
108k4A	0	0.0440	$0.0762i$	-14699.2655	24463608892439.7456
108k4C	0	0.0554	$0.0960i$	1608.7743	6115643810955.1724
121k4A	-2^{15}	0.0116	$0.0385i$	85659519816.8841	25723073306989527.1216
144k4E	0	0.0454	$0.0262i$	81.1130	-549788016394046.1396
27k6A	0	0.0110	$0.0191i$	0.0000	97856189971744203.7795
32k6A	1728	0.0199	$0.0199i$	-58095643136.7658	8.0094

Table 3.3: Volumes of higher dimensional abelian varieties.

A	dim	Ω^+	Ω^-
79k2B	5	10	$209i$
83k2B	6	22	41
131k2B	10	51	615
11k4A	2	0.0815	0.0212
17k4B	3	0.0047	$0.0007i$

Chapter 4

Component groups of optimal quotients

Let A be an abelian variety over the rational numbers $\mathbf{Q}$. The Birch and Swinnerton-Dyer conjecture supplies a formula for the order of the Shafarevich-Tate group of A. A key step in computing this order is to find each of the Tamagawa numbers c_p of A. The Tamagawa numbers are defined as follows, where the definition of Néron model and component group is given below.

Definition 4.1 (Tamagawa number). Let p be a prime, let $\mathcal{A}$ be a Néron model of A over the p-adic integers $\mathbf{Z}_p$, and let $\Phi_{A,p}$ be the component group of $\mathcal{A}$ at p. Then the *Tamagawa number* c_p of A is the order of the group $\Phi_{A,p}(\mathbf{F}_p)$ of $\mathbf{F}_p$-rational points of $\Phi_{A,p}(\overline{\mathbf{F}}_p)$.

Remark 4.2. We warn the reader that the Tamagawa number is defined in a different way in some other papers. The definitions are equivalent.

In this chapter we present a method for computing the Tamagawa numbers c_p, up to a power of 2, under the hypothesis that A has purely toric reduction at p. Such A are plentiful among the modular abelian varieties; for example, if A is a new optimal quotient of $J_0(N)$ and p exactly divides N, then A is purely toric at p.

In Sections 4.1–4.5 we state and prove an explicit formula involving component groups of fairly general abelian varieties. Then in Section 4.6 we turn to quotients of modular Jacobians $J_0(N)$. We give several tables and issue a conjecture and a question.

The results of this chapter were inspired by a letter that Ribet wrote to Mestre, in which he treats the case when A is an elliptic curve.

4.1 Main results

4.1.1 Néron models and component groups

Let A be an abelian variety over a finite extension K of the p-adic numbers $\mathbf{Q}_p$. Let $\mathcal{O}$ be the ring of integers of K, let $\mathfrak{m}$ be its maximal ideal, and let $k = \mathcal{O}/\mathfrak{m}$ be the residue class field.

Definition 4.3 (Néron model). A *Néron model* of A is a smooth commutative group scheme $\mathcal{A}$ over $\mathcal{O}$ such that A is its generic fiber and $\mathcal{A}$ satisfies the Néron mapping property: the restriction map

$$\mathrm{Hom}_{\mathcal{O}}(S, \mathcal{A}) \longrightarrow \mathrm{Hom}_K(S_K, A)$$

is bijective for all *smooth* schemes S over $\mathcal{O}$.

The Néron mapping property implies that $\mathcal{A}$ is unique up to a unique isomorphism, so we will refer without hesitation to "the" Néron model of A.

The closed fiber $\mathcal{A}_k$ of $\mathcal{A}$ is a group scheme over k, which need not be connected; denote by $\mathcal{A}_k^0$ the connected component containing the identity. There is an exact sequence

$$0 \longrightarrow \mathcal{A}_k^0 \longrightarrow \mathcal{A}_k \longrightarrow \Phi_A \longrightarrow 0,$$

where Φ_A a finite étale group scheme over k. Equivalently, Φ_A may be viewed as a finite abelian group equipped with an action of $\mathrm{Gal}(\overline{k}/k)$.

Definition 4.4 (Component group). The *component group* of an abelian variety $\mathcal{A}$ over a local field K is the group scheme $\Phi_A = A_k/A_k^0$ defined above.

4.1.2 Motivating problem

This chapter is motivated by the problem of computing the groups $\Phi_{A,p}$ attached to quotients A of Jacobians of modular curves $X_0(N)$. When A has semistable reduction, Grothendieck and Mumford described the component group in terms of a monodromy pairing on certain free abelian groups. When $A = J = J_0(N)$ is the Jacobian of $X_0(N)$, this pairing can be explicitly computed, hence the component group Φ_J can also be computed; this has been done in many cases in [40] and [23].

Suppose now that $A = A_f$ is an optimal quotient of $J_0(N)$ that is attached to a newform f, so that the kernel of the map $\pi : J \to A$ is connected. There is a natural map $\pi_* : \Phi_J \to \Phi_A$. We wish to compute the image and the order of the cokernel of π_*.

4.1.3 The main result

We now state our main result more precisely, necessarily supressing some of the definitions of the terms used until later. Suppose $\pi : J \to A$ is an optimal quotient, with J a Jacobian with semistable reduction and A having purely toric reduction. We express the component group of A in terms of the monodromy pairing associated to J.

Let $m_A = \sqrt{\deg(\theta_A)}$, where $\theta_A : A^\vee \to A$ is induced by the canonical principal polarization of J arising from the θ-divisor. Let X_J be the character group of the toric part of the closed fiber of the Néron model of J. Let $\mathcal{L}$ be the saturation of the image of X_A in X_J. The monodromy pairing induces a map $\alpha : X_J \to \mathrm{Hom}(\mathcal{L}, \mathbf{Z})$. Let Φ_X be the cokernel of α and $m_X = [\alpha(X_J) : \alpha(\mathcal{L})]$ be the order of the finite group $\alpha(X_J)/\alpha(\mathcal{L})$. We obtain the equality

$$\frac{\#\Phi_A}{m_A} = \frac{\#\Phi_X}{m_X}.$$

Using the snake lemma, one see that Φ_X is isomorphic to the image of the natural map $\Phi_J \to \Phi_A$, and the above formula implies that the cokernel of the map $\Phi_J \to \Phi_A$ has order m_A/m_X.

If the optimal quotient $J \to A$ arises from a modular form on $\Gamma_0(N)$, then the quantities m_A, m_X and Φ_X can be explicitly computed, hence we can compute $\#\Phi_A$.

4.2 Optimal quotients of Jacobians

Let J be a Jacobian, and let θ_J be the canonical principal polarization arising from the θ-divisor. Recall that an *optimal quotient* of J is an abelian variety A and a surjective map $\pi : J \to A$ whose kernel is an abelian subvariety B of J. Denote by $J^\vee$ and $A^\vee$ the abelian varieties dual to J and A, respectively. Upon composing the dual of π with $\theta_J^\vee = \theta_J$, we obtain a map

$$A^\vee \xrightarrow{\pi^\vee} J^\vee \xrightarrow{\theta_J} J.$$

Proposition 4.5. *The map $A^\vee \to J$ is injective.*

Proof. Since θ_J is an isomorphism it suffices to prove that $\pi^\vee$ is injective. Since the dual of $\pi^\vee$ is $(\pi^\vee)^\vee = \pi$ and π is surjective, the map $\pi^\vee$ must have finite kernel. Thus $A^\vee \to C = \mathrm{im}(\pi^\vee)$ is an isogeny. Let G denote the kernel of this isogeny, and dualize. By [50, §11] we have the following two commutative diagrams:

$$\begin{tikzcd} G \arrow[r] & A^\vee \arrow[r, two heads] \arrow[dr, "\pi^\vee"'] & C \arrow[d] & {} \arrow[r, "\text{dualize}"] & {} & A & C^\vee \arrow[l] & G^\vee \arrow[l] \\ & & J^\vee & & & & J, \arrow[ul, "\pi"] \arrow[u, "\varphi"'] & \end{tikzcd}$$

where $G^\vee$ is the Cartier dual of G. Since $G^\vee$ is finite, $\ker(\varphi)$ is of finite index in $\ker(\pi)$. Since $\ker(\pi)$ is an abelian variety, as a group it is divisible. But a divisible group has no nontrivial finite-index subgroups (divisibility is a property inherited by quotients, and nonzero finite groups are not divisible). Thus $\ker(\varphi) = \ker(\pi)$, so $G^\vee = 0$. It follows that $G = 0$. □

Henceforth we will abuse notation and denote the injection $A^\vee \to J$ by $\pi^\vee$. The kernel of θ_A equals the intersection of $A^\vee$ and $B = \ker(\pi)$, as depicted in the following diagram:

$$\begin{tikzcd} A^\vee \cap B \arrow[r] \arrow[d] & B \arrow[d] \\ A^\vee \arrow[r, hook, "\pi^\vee"] \arrow[dr, "\theta_A"'] & J \arrow[d, "\pi"] \\ & A. \end{tikzcd}$$

Since θ_A is a polarization, the degree $\#\ker(\theta_A)$ of θ_A is a perfect square (see [50, Thm. 13.3]). Recall that the *modular degree* is the integer

$$m_A = \sqrt{\#\ker(\theta_A)}.$$

For an algorithm to compute m_A, see Section 3.9 and Corollary 4.23.

4.3 The closed fiber of the Néron model

Let K be a finite extension of $\mathbf{Q}_p$ with ring of integers $\mathcal{O}$ and residue class field k. Let A be an abelian variety over K and denote its Néron model by $\mathcal{A}$. Let Φ_A be the group of connected components of the closed fiber $\mathcal{A}_k$. This group is a finite étale group scheme over k; equivalently, it is a finite abelian group equipped with an action of $\mathrm{Gal}(\overline{k}/k)$. There is an exact sequence of group schemes

$$0 \to \mathcal{A}_k^0 \to \mathcal{A}_k \to \Phi_A \to 0.$$

The group scheme $\mathcal{A}_k^0$ is an extension of an abelian variety $\mathcal{B}$ of some dimension a by a group scheme $\mathcal{C}$; we have a diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & & & \\
 & \mathcal{T} \arrow[d] & & & \\
0 \arrow[r] & \mathcal{C} \arrow[r] \arrow[d] & \mathcal{A}_k^0 \arrow[r] & \mathcal{B} \arrow[r] & 0 \\
 & \mathcal{U} \arrow[d] & & & \\
 & 0 & & &
\end{tikzcd}$$

with $\mathcal{T}$ a torus of dimension t and $\mathcal{U}$ a unipotent group of dimension u. The abelian variety A is said to have *purely toric reduction* if $t = \dim A$, and have *semistable reduction* if $u = 0$.

Definition 4.6 (Character group of torus). The *character group*

$$X_A = \mathrm{Hom}_{\overline{k}}(\mathcal{T}_{/\overline{k}}, \mathbf{G}_{m/\overline{k}})$$

is a free abelian group of rank t contravariantly associated to A.

As discussed in, e.g., [53], if A is semistable there is a *monodromy pairing* $X_A \times X_{A^\vee} \to \mathbf{Z}$ and an exact sequence

$$0 \to X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

4.4 Rigid uniformization

In this section we review the rigid analytic uniformization of a semistable abelian variety over a finite extension K of the maximal unramified extension $\mathbf{Q}_p^{\mathrm{ur}}$ of $\mathbf{Q}_p$. We use this uniformization to prove that if A has purely toric reduction, and $\phi : A^\vee \to A$ is a symmetric isogeny (as defined below), then

$$\deg(\phi) = (\#\,\mathrm{coker}(X_A \to X_{A^\vee}))^2.$$

We also prove some lemmas about character groups.

It is possible to prove the assertions we will need without recourse to rigid uniformization, as Ahmed Abbes has pointed out to the author.

4.4.1 Raynaud's uniformization

Theorem 4.7 (Raynaud). *If A is a semistable abelian variety, its universal covering (as defined in [14]) is isomorphic to an extension G of an abelian variety B with good reduction by a torus T. The covering map from G to A is a homomorphism, and its kernel is a twisted free abelian group Γ of finite rank.*

This may be summarized by the diagram

$$\begin{tikzcd}
 & \Gamma \arrow[d] & \\
T \arrow[r] & G \arrow[r] \arrow[d] & B \\
 & A, &
\end{tikzcd}$$

which we call the *uniformization cross* of A.

Remark 4.8. The group Γ can be identified with the character group $X_{A^\vee}$ of the previous and latter sections.

The uniformization cross of the dual abelian variety $A^\vee$ is

$$\begin{tikzcd}
 & \Gamma^\vee \arrow[d] & \\
T^\vee \arrow[r] & G^\vee \arrow[r] \arrow[d] & B^\vee \\
 & A^\vee, &
\end{tikzcd}$$

where $\Gamma^\vee = \mathrm{Hom}(T, \mathbf{G}_m)$, where $T^\vee = \mathrm{Hom}(\Gamma, \mathbf{G}_m)$, and the morphisms $\Gamma^\vee \to G^\vee$ and $T^\vee \to G^\vee$ are the one-motif duals of the morphisms $T \to G$ and $\Gamma \to G$, respectively. For more details see, e.g., [14].

To avoid confusion when considering the uniformization of more than one abelian variety, we will often denote the objects T, G, Γ, and B connected with A by T_A, G_A, Γ_A, and B_A, respectively.

Example 4.9 (Tate curve). If $E/\mathbf{Q}_p$ is an elliptic curve with split multiplicative reduction, then the uniformization is $E = \mathbf{G}_m/q^{\mathbf{Z}}$ where $q = q(j)$ is obtained by inverting the expression for j as a function of $q(z) = e^{2\pi i z}$.

4.4.2 Some lemmas

Let $\pi : J \to A$ be an optimal quotient, assume that J has semistable reduction, and that A has purely toric reduction.

Lemma 4.10. *The map $\Gamma_J \to \Gamma_A$ induced by π is surjective.*

Proof. Since G_J is simply connected, π induces a map $G_J \to T_A$ and a map $\Gamma_J \to \Gamma_A$. Because π is surjective and T_A is a torus, the map $G_J \to T_A$ is surjective. Upon applying the snake lemma to the following diagram, we obtain a surjective map from $B = \ker(\pi)$ to $M = \mathrm{coker}(\Gamma_J \to \Gamma_A)$:

$$\begin{tikzcd}
 & \Gamma_J \arrow[r] \arrow[d] & \Gamma_A \arrow[r] \arrow[d] & M \arrow[r] \arrow[d] & 0 \\
 & G_J \arrow[r] \arrow[d] & T_A \arrow[r] \arrow[d] & 0 & \\
B \arrow[r] & J \arrow[r, "\pi"] & A. & &
\end{tikzcd}$$

Since $\pi : J \to A$ is an optimal quotient, the kernel B is connected. Thus M must also be connected. Since M is discrete it follows that $M = 0$. □

Abelian varieties with purely toric reduction

Assume that A has purely toric reduction. Then $B = 0$, and the uniformization cross is simply

$$\begin{tikzcd}
\Gamma \arrow[d] \\
T \arrow[d] \\
A.
\end{tikzcd}$$

Definition 4.11 (Symmetric isogeny). A *symmetric isogeny* $\varphi : A^\vee \to A$ is an isogeny such that the map $\varphi^\vee : A^\vee \to (A^\vee)^\vee = A$ is equal to φ.

Let $\varphi : A^\vee \to A$ be a symmetric isogeny. Denote by $\varphi_t : T^\vee \to T$ and $\varphi_a : \Gamma^\vee \to \Gamma$ the maps induced by φ.

Proposition 4.12. *There is an exact sequence*

$$0 \to \ker(\varphi_t) \to \ker(\varphi) \to \mathrm{coker}(\varphi_a) \to 0,$$

and $\ker(\varphi_t)$ *is the Cartier dual of* $\mathrm{coker}(\varphi_a)$*.*

Proof. Since φ is an isogeny we obtain the following diagram:

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & \Gamma^\vee \arrow[r, "\varphi_a"] \arrow[d] & \Gamma \arrow[r] \arrow[d] & \mathrm{coker}(\varphi_a) \arrow[d] \\
\ker(\varphi_t) \arrow[r] \arrow[d] & T^\vee \arrow[r, "\varphi_t"] \arrow[d] & T \arrow[r] \arrow[d] & 0 \\
\ker(\varphi) \arrow[r] & A^\vee \arrow[r, "\varphi"] & A. &
\end{tikzcd}$$

The snake lemma then gives the claimed exact sequence.

For the second assertion, observe that if we take one-motif duals of every object in the diagram

$$\begin{tikzcd}
 & \Gamma^\vee \arrow[r, "\varphi_a"] \arrow[d] & \Gamma \arrow[r] \arrow[d] & \mathrm{coker}(\varphi_a) \\
\ker(\varphi_t) \arrow[r] & T^\vee \arrow[r, "\varphi_t"] & T &
\end{tikzcd}$$

we obtain the following diagram:

$$\begin{tikzcd}
 & T & T^\vee \arrow[l, "\varphi_a^\vee"'] & \mathrm{coker}(\varphi_a)^\vee \arrow[l] \\
\ker(\varphi_t)^\vee & \Gamma \arrow[l] \arrow[u] & \Gamma^\vee. \arrow[l, "\varphi_t^\vee"] \arrow[u] &
\end{tikzcd}$$

Since φ is symmetric, $\varphi_a^\vee = \varphi_t$, so

$$\ker(\varphi_t) = \mathrm{coker}(\varphi_a)^\vee.$$

□

Lemma 4.13. $\#\ker(\varphi) = \#\,\mathrm{coker}(\varphi_a)^2$

Proof. Use the exact sequence of Proposition 4.12 together with the observation that the order of a finite group scheme equals the order of its Cartier dual. □

4.5 The main theorem

Let $\pi : J \to A$ be an optimal quotient, with J a Jacobian having semistable reduction and A an abelian variety having purely toric reduction. Let X_A, $X_{A^\vee}$, and X_J denote the character groups of the toric parts of the closed fibers of the abelian varieties A, $A^\vee$, and J, respectively.

4.5.1 Description of the component group in terms of the monodromy pairing

Recall that there is a pairing $X_A \times X_{A^\vee} \to \mathbf{Z}$ called the monodromy pairing. We have an exact sequence

$$0 \to X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

If J is a Jacobian then J is canonically self-dual via the θ-polarization, so the monodromy pairing on J can be viewed as a pairing $X_J \times X_J \to \mathbf{Z}$, and there is an exact sequence

$$0 \to X_J \to \mathrm{Hom}(X_J, \mathbf{Z}) \to \Phi_J \to 0.$$

Example 4.14 (Tate curve). Suppose $E = \mathbf{G}_m/q^{\mathbf{Z}}$ is a Tate curve over $\mathbf{Q}_p^{\mathrm{ur}}$. The monodromy pairing on $X_E = q^{\mathbf{Z}}$ is

$$\langle q, q\rangle = \operatorname{ord}_p(q) = -\operatorname{ord}_p(j).$$

Thus Φ_E is cyclic of order $-\operatorname{ord}_p(j)$.

Proof of the main theorem

We now prove the main theorem. Let $\pi : J \to A$ be an optimal quotient, and let $\theta : A^\vee \to A$ denote the induced polarization. Let π_*, π^*, θ_*, and θ^* be the maps induced on character groups by the various functorialities, as indicated in the following two key diagrams:

$$\begin{tikzcd}
A^\vee \arrow[r, hook, "\pi^\vee"] \arrow[dr, "\theta"'] & J \arrow[d, two heads, "\pi"] \\
 & A
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
X_A \arrow[r, hook, "\pi^*"] \arrow[dr, "\theta^*"] & X_J \arrow[d, two heads, "\pi_*"] \\
 & X_{A^\vee}. \arrow[ul, bend left, "\theta_*"]
\end{tikzcd}$$

The surjectivity of π_* was proved in Lemma 4.10. The injectivity of π^* follows because

$$\theta_*\pi_*\pi^* = \theta_*\theta^* = \deg(\theta) \neq 0,$$

and multiplication by a nonzero integer on a free abelian group is injective.

Let

$$\alpha : X_J \to \operatorname{Hom}(\pi^* X_A, \mathbf{Z})$$

be the map defined by the monodromy pairing restricted to $X_J \times \pi^* X_A$.

Lemma 4.15. $\ker(\pi_*) = \ker(\alpha)$

Proof. Suppose $x \in \ker(\pi_*)$, and let $y = \pi^* z$ with $z \in X_A$. Then

$$\langle x, y\rangle = \langle x, \pi^* z\rangle = \langle \pi_* x, z\rangle = 0,$$

so $x \in \ker(\alpha)$. Next let $x \in \ker(\alpha)$. Then for all $z \in X_A$,

$$0 = \langle x, \pi^* z\rangle = \langle \pi_* x, z\rangle,$$

so $\pi_* x$ is in the kernel of the monodromy map

$$X_{A^\vee} \to \operatorname{Hom}(X_A, \mathbf{Z}).$$

Since $X_{A^\vee}$ and $\operatorname{Hom}(X_A, \mathbf{Z})$ are free of the same finite rank and the cokernel is torsion, the monodromy map is injective. Thus $\pi_* x = 0$ and $x \in \ker(\pi_*)$. □

Lemma 4.16. *There is an exact sequence*

$$X_J \to \operatorname{Hom}(\pi^* X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

Proof. Lemma 4.15 gives the following commutative diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & X_J/\ker(\alpha) \arrow[r] \arrow[d, "\cong"] & \mathrm{Hom}(\pi^* X_A, \mathbf{Z}) \arrow[r] \arrow[d, "\cong"] & \mathrm{coker}(\alpha) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & X_{A^\vee} \arrow[r] & \mathrm{Hom}(X_A, \mathbf{Z}) \arrow[r] & \Phi_A \arrow[r] & 0.
\end{tikzcd}$$

By Lemma 4.15, the first vertical map is an isomorphism. The second is an isomorphism because it is induced by the isomorphism $\pi^* : X_A \to \pi^* X_A$. It follows that $\mathrm{coker}(\alpha) \cong \Phi_A$, as claimed. □

Let $\mathcal{L}$ be the *saturation* of $\pi^* X_A$ in X_J; thus $\pi^* X_A$ is a finite-index subgroup of $\mathcal{L}$ and the quotient $X_J/\mathcal{L}$ is torsion free. For L of finite index in $\mathcal{L}$, define the *modular degree* of L to be

$$m_L = [\alpha(X_J) : \alpha(L)],$$

and the *component group* of L to be

$$\Phi_L = \mathrm{coker}(X_J \to \mathrm{Hom}(L, \mathbf{Z})).$$

When $L = \mathcal{L}$ and A is fixed, we often slightly abuse notation and write $m_X = m_{\mathcal{L}}$ and $\Phi_X = \Phi_{\mathcal{L}}$. We think of m_X and Φ_X as the character group "modular degree and component group" of A.

Lemma 4.17. *Choose a subgroup L of finite index in $\mathcal{L}$. The rational number $\dfrac{\#\Phi_L}{m_L}$ is independent of the choice of L.*

Proof. Suppose L' is another finite index subgroup of $\mathcal{L}$, and let $n = [L : L']$. Here n is a rational number, the lattice index of L' in L. Since α is injective when restricted to $\mathcal{L}$, it follows that

$$m_{L'} = [\alpha(X_J) : \alpha(L')] = [\alpha(X_J) : \alpha(L)] \cdot [\alpha(L) : \alpha(L')] = m_L \cdot n.$$

Similarly, $\#\Phi_{L'} = \#\Phi_L \cdot n$. □

Recall that $m_A = \sqrt{\deg(\theta)}$ and

$$\Phi_A \cong \mathrm{coker}(X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z})),$$

where m_A is the modular degree of A and Φ_A is the component group of A.

Theorem 4.18. *For any subgroup L of finite index in $\mathcal{L}$, the following relation holds:*

$$\frac{\#\Phi_A}{m_A} = \frac{\#\Phi_L}{m_L}.$$

Proof. By Lemma 4.17 we may assume that $L = \pi^* X_A$. With this choice of L, Lemma 4.16 asserts that $\Phi_L \cong \Phi_A$. By Lemma 4.15, properties of the index, and Lemma 4.13 we have

$$\begin{aligned} m_L &= [\alpha(X_J) : \alpha(L)] \\ &= [\pi_*(X_J) : \pi_*(L)] \\ &= [X_{A^\vee} : \pi_*(\pi^* X_A)] \\ &= [X_{A^\vee} : \theta^* X_A] \\ &= \#\operatorname{coker}(\theta^*) \\ &= \sqrt{\deg(\theta)} = m_A. \end{aligned}$$

□

Proposition 4.19.

$$\operatorname{image}(\Phi_J \to \Phi_A) \cong \Phi_{\mathcal{L}}.$$

Proof. Since $\pi^* X_A \subset \mathcal{L} \subset X_J$, an application of Lemma 4.16 gives the following commutative diagram with exact rows:

$$\begin{tikzcd} X_J \arrow[r] \arrow[d, equal] & \operatorname{Hom}(X_J, \mathbf{Z}) \arrow[r] \arrow[d] & \Phi_J \arrow[r] \arrow[d] & 0 \\ X_J \arrow[r] \arrow[d, equal] & \operatorname{Hom}(\mathcal{L}, \mathbf{Z}) \arrow[r] \arrow[d] & \Phi_{\mathcal{L}} \arrow[r] \arrow[d] & 0 \\ X_J \arrow[r] & \operatorname{Hom}(\pi^* X_A, \mathbf{Z}) \arrow[r] & \Phi_A \arrow[r] & 0. \end{tikzcd}$$

The map $\operatorname{Hom}(\mathcal{L}, \mathbf{Z}) \to \operatorname{Hom}(\pi^* X_A, \mathbf{Z})$ is an isomorphism, so the map $\Phi_{\mathcal{L}} \to \Phi_A$ is injective. Thus

$$\operatorname{image}(\Phi_J \to \Phi_A) \cong \operatorname{image}(\Phi_J \to \Phi_{\mathcal{L}}).$$

The cokernel of $\operatorname{Hom}(X_J, \mathbf{Z}) \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})$ surjects onto the cokernel of $\Phi_J \to \Phi_{\mathcal{L}}$. Using the exact sequence

$$0 \to \mathcal{L} \to X_J \to X_J/\mathcal{L} \to 0,$$

we find that

$$\operatorname{coker}(\operatorname{Hom}(X_J, \mathbf{Z}) \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})) \subset \operatorname{Ext}^1(X_J/\mathcal{L}, \mathbf{Z}).$$

Because $\mathcal{L}$ is saturated, the quotient $X_J/\mathcal{L}$ is torsion free, so the indicated Ext^1 group vanishes. Thus the map $\Phi_J \to \Phi_{\mathcal{L}}$ is surjective, from which the proposition follows. □

The following corollary follows from Theorem 4.18 and Proposition 4.19.

Corollary 4.20.

$$\#\operatorname{coker}(X_J \to X_A) = \frac{m_A}{m_{\mathcal{L}}}.$$

Remark 4.21. A non-obvious consequence of this corollary is that

$$m_{\mathcal{L}} \mid m_A.$$

4.6 Optimal quotients of $J_0(N)$

We now summarize some facts about $J_0(N)$ that will be used in our numerical computations. Some of these facts were discussed in greater generality in the previous chapters of this thesis.

4.6.1 Modular curves and semistability

Let $X_0(N)$ be the modular curve associated to the subgroup $\Gamma_0(N)$ of $\mathrm{SL}_2(\mathbf{Z})$ that consists of those matrices which are upper triangular modulo N. Initially, $X_0(N)$ is constructed as a Riemann surface as the quotient

$$\Gamma_0(N)\backslash(\{z : z \in \mathbf{C}, \operatorname{Im}(z) > 0\} \cup \mathbf{P}^1(\mathbf{Q})).$$

With some work, we find that $X_0(N)$ has a canonical structure of algebraic curve over $\mathbf{Q}$.

Suppose that p is a prime divisor of N such that N/p is coprime to p. We write $p \,||\, N$. In this situation, it is well-known that the Jacobian $J_0(N)$ of $X_0(N)$ has semistable reduction at p.

4.6.2 Newforms and optimal quotients

The Hecke algebra

$$\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots] \subset \operatorname{End}(J_0(N))$$

is a commutative ring of endomorphisms of $J_0(N)$ of $\mathbf{Z}$-rank equal to the dimension $J_0(N)$. The character group $X_{J_0(N)}$ of $J_0(N)$ at p is equipped with a functorial action of $\mathbf{T}$. The Hecke algebra $\mathbf{T}$ also acts on the complex vector space $S = S_2(\Gamma_0(N), \mathbf{C})$ of cusp forms.

A newform f is an eigenform normalized so that the coefficient of q in the Fourier expansion of f at the cusp ∞ is 1, and such that f is not a modular form of any level $N' \mid N$, with N' a proper divisor of N.

Let f be a newform, and associate to f the ideal I_f of the Hecke algebra $\mathbf{T}$ of elements which annihilate f. Then $\mathcal{O}_f = \mathbf{T}/I_f$ is an order in the ring of integers of the totally real number field K_f obtained by adjoining the Fourier coefficients of f to $\mathbf{Q}$. The quotient

$$A_f = J_0(N)/I_f J_0(N)$$

is an optimal quotient of $J_0(N)$ of dimension equal to $[K_f : \mathbf{Q}]$. It is purely toric at p, since $p \,||\, N$.

4.6.3 Homology and the modular degree

Let $H = H_1(X_0(N), \mathbf{Z})$ be the integral homology of the complex algebraic curve $X_0(N)$. Integration defines a $\mathbf{T}$-equivariant nondegenerate pairing $S \times H \to \mathbf{C}$. This pairing induces a map $\alpha : H \to \operatorname{Hom}_{\mathbf{C}}(S, \mathbf{C})$.

Theorem 4.22. *We have the following commutative diagram of* $\mathbf{T}$*-modules:*

$$\begin{tikzcd}
H[I_f] \arrow[r, hook] \arrow[d, hook] & H \arrow[r, two heads] \arrow[d, hook] & \alpha(H) \arrow[d, hook] \\
\mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C})[I_f] \arrow[r, hook] \arrow[d, two heads] & \mathrm{Hom}_{\mathbf{C}}(S, \mathbf{C}) \arrow[r, two heads] \arrow[d, two heads] & \mathrm{Hom}_{\mathbf{C}}(S[I_f], \mathbf{C}) \arrow[d, two heads] \\
A_f^{\vee}(\mathbf{C}) \arrow[r, hook] \arrow[rr, bend right, "\theta_A"'] & J(\mathbf{C}) \arrow[r, two heads] & A_f(\mathbf{C})
\end{tikzcd}$$

Proof. This can be deduced from [61]. See also Section 2.7. □

Corollary 4.23. $m_A^2 = [\alpha(H) : \alpha(H[I_f])]$.

Proof. Recall that m_A is by definition equal to $\sqrt{\deg(\theta_A)}$. The kernel of an isogeny between complex tori is isomorphic to the cokernel of the induced map on lattices. The corollary now follows from the diagram of Theorem 4.22, which indicates that the index $[\alpha(H) : \alpha(H[I_f])]$ is the cokernel of the map $H[I_f] \to \alpha(H)$.

For more details, see Section 3.9. □

4.6.4 Rational points of the component group (Tamagawa numbers)

Let $\mathrm{Frob}_p : X_J \to X_J$ denote the map induced by the Frobenius automorphism. We have $\mathrm{Frob}_p = -W_p$, where W_p is the map induced by the Atkin-Lehner involution on $J_0(p)$. Let f be a newform, $A = A_f$ the corresponding optimal quotient, and w_p the sign of the eigenvalue of W_p on f.

Proposition 4.24.

$$\Phi_A(\mathbf{F}_p) = \begin{cases} \Phi_A(\overline{\mathbf{F}}_p) & \text{if } w_p = -1, \\ \Phi_A(\overline{\mathbf{F}}_p)[2] & \text{if } w_p = 1. \end{cases}$$

Proof. If $w_p = -1$, then $\mathrm{Frob}_p = 1$ and the $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$-action of $\Phi_A(\overline{\mathbf{F}}_p)$ is trivial. In this case $\Phi(\mathbf{F}_p) = \Phi(\overline{\mathbf{F}}_p)$. Next suppose $w_p = 1$. Recall that we have an exact sequence

$$0 \to X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

Since W_p acts as $+1$ on f, it also acts as $+1$ on each of the four modules A, X_A, $\mathrm{Hom}(X_A, \mathbf{Z})$, and Φ_A. Thus $\mathrm{Frob}_p = -W_p$ acts as -1 on Φ_A. Since the subgroup of 2-torsion elements of a finite abelian group equals the subgroup of elements fixed under -1, it follows that $\Phi_A(\mathbf{F}_p) = \Phi_A(\overline{\mathbf{F}}_p)[2]$. □

WARNING: When we extend this result to the whole of $J_0(N)$, it is necessary to be exceedingly careful! The action of $\mathrm{Frob}_p = T_p$ need not be by ±1, even though it must be by an involution of order 2. For example, the component group of $J_0(65)$ at 5 is cyclic of order 42. The action of Frob_5 is by multiplication by -13. Note that $(-13)^2 = 1 \pmod{42}$. The fixed points of multiplication by -13 is the order 14 subgroup of $\mathbf{Z}/42\mathbf{Z}$.

4.7 Computations

Using the algorithms of Chapter 3, we can enumerate the optimal quotients A_f of $J_0(N)$ and compute the modular degree m_A. The method of graphs (see [47]) and quaternion algebras (see [32]) can be used to compute $X = X_{J_0(N)}$ with its $\mathbf{T}$-action and the monodromy pairing. We can then compute the following three modules: the saturated submodule $\mathcal{L} = \bigcap_{t\in I_f} \ker(t)$ of X, the character group modular degree $m_X = m_{\mathcal{L}}$, and $\Phi_X = \Phi_{\mathcal{L}}$. By Theorem 4.18 we obtain

$$\#\Phi_A = \#\Phi_X \cdot \frac{m_A}{m_X}.$$

Using this method, we have computed $\#\Phi_A$ in a number of cases. We give tables that report on some of these computations in Secton 4.7.2. In the next section we discuss a conjecture and a question, which were both suggested by our numerical computations.

4.7.1 Conjectures and questions

Suppose that $N = pM$ with $(p, M) = 1$. Let

$$H_{\text{new}} = \ker\Big(H_1(X_0(N),\mathbf{Z}) \longrightarrow H_1(X_0(M),\mathbf{Z}) \oplus H_1(X_0(M),\mathbf{Z})\Big),$$

where the map is induced by the two natural degeneracy maps $X_0(N) \to X_0(M)$.

The Hecke algebra $\mathbf{T}$ acts on H_{new}, and also on the submodule $H_{\text{new}}[I_f]$ of those elements that are annihilated by I_f. Integration defines a map $\alpha : H_{\text{new}} \to \operatorname{Hom}(S[I_f], \mathbf{C})$. Define the p-new homology modular degree m_H by

$$m_H^2 = [\alpha(H_{\text{new}}) : \alpha(H_{\text{new}}[I_f])].$$

We expect that there is a very close relationship between m_X and m_H.

Question 4.25. Is m_X equal to m_H?

The following conjecture offers a refinement of some of the results of [40].

Conjecture 4.26 (Refined Eisenstein conjecture). *Let p be a prime and let $f_1, \ldots, f_n$ be a set of representatives for the Galois-conjugacy classes of newforms in $S_2(\Gamma_0(p))$. Let $A_1, \ldots, A_n$ be the optimal quotients associated to $f_1, \ldots, f_n$, respectively. Then for each i, $i = 1, \ldots, n$, we have*

$$\#A_i(\mathbf{Q})_{\text{tor}} = \#\Phi_{A_i}(\overline{\mathbf{F}}_p) = \#\Phi_{A_i}(\mathbf{F}_p).$$

Furthermore,

$$\#\Phi_{J_0(p)}(\overline{\mathbf{F}}_p) = \prod_{i=1}^{d} \#\Phi_{A_i}(\overline{\mathbf{F}}_p).$$

We have verified Conjecture 4.26 for all $p \leq 757$, and, up to a power of 2, for all $p < 2000$.

Remark 4.27. It is tempting to guess that, e.g., the natural map

$$\Phi_{J_0(113)}(\overline{\mathbf{F}}_p) \to \prod_{i=1}^{4} \Phi_{A_i}(\overline{\mathbf{F}}_p)$$

is an isomorphism. Two of the $\Phi_{A_i}(\overline{\mathbf{F}}_p)$ have order 2, so the product $\prod \Phi_{A_i}(\overline{\mathbf{F}}_p)$ can not be a cyclic group. However, the groups $\Phi_{J_0(p)}(\overline{\mathbf{F}}_p)$ are known to be cyclic for all primes p.

4.7.2 Tables

We have computed component groups of many optimal quotients A_f of $J_0(N)$. In this section we provide tables, which hint at the data we have gathered. Our notation for optimal quotients is described in Section 1.3.1. See also Table 1.6.

Table 4.1: Component groups at low level

Table 4.1 gives the component groups of the quotients A_f of $J_0(N)$ for $N \leq 106$. The column labeled d contains the dimensions of the A_f, and the column labeled $\#\Phi_{A,p}$ contains a list of the orders of the component groups of A_f, one for each divisor p of N, ordered by increasing p. An entry of "?" indicates that $p^2 \mid N$, so our algorithm does not apply. A component group order is starred if the $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$-action is nontrivial.

Table 4.2–4.3: Big component groups

Using the algorithms described in Section 3.10, we computed the rational numbers $L(A,1)/\Omega_A$ for every optimal quotient A that is attached to a newform of level ≤ 1500. There are exactly 5 optimal quotients A such that the numerator of $L(A,1)/\Omega_A$ is nonzero and divisible by a prime $> 10^9$. The Birch and Swinnerton-Dyer conjecture predicts that these large prime divisors must divide either $\#\Phi_A$ or the Shafarevich-Tate group of A. This is the case, as Table 4.3 shows.

Table 4.4: Quotients of $J_0(N)$

Table 4.4 contains all of the invariants involved in the computation of component groups for each of the newform optimal quotients of levels 65, 66, 68, and 69.

Table 4.5: Quotients of $J_0(p)^-$

We computed the quantities m_A, m_X and Φ_X for each abelian variety $A = A_f$ associated to a newform of prime level p with $p \leq 757$. The results are as follows:

1. In all cases $m_A = m_X$, so the map $\Phi_J \to \Phi_A$ is surjective.
2. $\Phi_A = 1$ whenever the sign of the Atkin-Lehner involution w_p on A is 1.
3. $\prod \#\Phi_A(\overline{\mathbf{F}}_p) = \#\Phi_J(\overline{\mathbf{F}}_p)$

Table 4.5 lists those A of level ≤ 631 for which $w_p = -1$, along with the order of the corresponding component group.

Table 4.1: Component groups at low level

A	d	$\#\Phi_{A,p}$	A	d	$\#\Phi_{A,p}$	A	d	$\#\Phi_{A,p}$	A	d	$\#\Phi_{A,p}$
11A	1	5	**51A**	1	$3, 1^*$	**72A**	1	$?, ?$	**90C**	1	$4, ?, 1$
14A	1	$6^*, 3$	**51B**	2	$16^*, 4$	**73A**	1	2	**91A**	1	$1^*, 1^*$
15A	1	$4^*, 4$	**52A**	1	$?, 2^*$	**73B**	2	1^*	**91B**	1	$1, 1$
17A	1	4	**53A**	1	1^*	**73C**	2	3	**91C**	2	$7, 1^*$
19A	1	3	**53A**	1	1^*	**74A**	2	$9^*, 3$	**91D**	3	$4^*, 8$
20A	1	$?, 2^*$	**53B**	3	13	**74B**	2	$95, 1^*$	**92A**	1	$?, 1^*$
21A	1	$4, 2^*$	**54A**	1	$3^*, ?$	**75A**	1	$1^*, ?$	**92B**	1	$?, 1$
23A	2	11	**54B**	1	$3, ?$	**75B**	1	$1, ?$	**93A**	2	$4^*, 1^*$
24A	1	$?, 2^*$	**55A**	1	$2, 2^*$	**75C**	1	$5, ?$	**93B**	3	$64, 2^*$
26A	1	$3^*, 3$	**55B**	2	$14^*, 2$	**76A**	1	$?, 1^*$	**94A**	1	$2, 1^*$
26B	1	$7, 1^*$	**56A**	1	$?, 1$	**77A**	1	$2^*, 1^*$	**94B**	2	$94^*, 1$
27A	1	$?$	**56B**	1	$?, 1^*$	**77B**	1	$3^*, 2$	**95A**	3	$10, 2^*$
29A	2	7	**57A**	1	$2^*, 1^*$	**77C**	1	$6, 3^*$	**95B**	4	$54^*, 6$
30A	1	$4^*, 3, 1^*$	**57B**	1	$2, 2^*$	**77D**	2	$2, 2^*$	**96A**	1	$?, 2$
31A	2	5	**57C**	1	$10, 1^*$	**78A**	1	$16^*, 5^*, 1$	**96B**	1	$?, 2^*$
32A	1	$?$	**58A**	1	$2^*, 1^*$	**79A**	1	1^*	**97A**	3	1^*
33A	1	$6^*, 2$	**58B**	1	$10, 1^*$	**79B**	5	13	**97B**	4	8
34A	1	$6, 1^*$	**59A**	5	29	**80A**	1	$?, 2$	**98A**	1	$2^*, ?$
35A	1	$3^*, 3$	**61A**	1	1^*	**80B**	1	$?, 2^*$	**98B**	2	$14, ?$
35B	2	$8, 4^*$	**61B**	3	5	**81A**	2	$?$	**99A**	1	$?, 1^*$
36A	1	$?, ?$	**62A**	1	$4, 1^*$	**82A**	1	$2^*, 1^*$	**99B**	1	$?, 1$
37A	1	1^*	**62B**	2	$66^*, 3$	**82B**	2	$28, 1^*$	**99C**	1	$?, 1^*$
37B	1	3	**63A**	1	$?, 1^*$	**83A**	1	1^*	**99D**	1	$?, 1^*$
38A	1	$9^*, 3$	**63B**	2	$?, 3$	**83B**	6	41	**100A**	1	$?, ?$
38B	1	$5, 1^*$	**64A**	1	$?$	**84A**	1	$?, 1^*, 2^*$	**101A**	1	1^*
39A	1	$2^*, 2$	**65A**	1	$1^*, 1^*$	**84B**	1	$?, 3, 2$	**101B**	7	25
39B	2	$14, 2^*$	**65B**	2	$3^*, 3$	**85A**	1	$2^*, 1$	**102A**	1	$2^*, 2^*, 1^*$
40A	1	$?, 2$	**65C**	2	$7, 1^*$	**85B**	2	$2^*, 1^*$	**102B**	1	$6^*, 6, 1^*$
41A	3	10	**66A**	1	$2^*, 3, 1^*$	**85C**	2	$6, 1^*$	**102C**	1	$8, 4, 1$
42A	1	$8, 2^*, 1^*$	**66B**	1	$4, 1^*, 1^*$	**86A**	2	$21^*, 3$	**103A**	2	1^*
43A	1	1^*	**66C**	1	$10, 5, 1$	**86B**	2	$55, 1^*$	**103B**	6	17
43B	2	7	**67A**	1	1	**87A**	2	$5, 1^*$	**104A**	1	$?, 1^*$
44A	1	$?, 1^*$	**67B**	2	1^*	**87B**	3	$92^*, 4$	**104B**	2	$?, 2$
45A	1	$?, 1^*$	**67C**	2	11	**88A**	1	$?, 1^*$	**105A**	1	$1, 1, 1$
46A	1	$10^*, 1$	**68A**	2	$?, 2^*$	**88B**	2	$?, 2^*$	**105B**	2	$10^*, 2^*, 2$
47A	4	23	**69A**	1	$2, 1^*$	**89A**	1	1^*	**106A**	1	$4^*, 1^*$
48A	1	$?, 2$	**69B**	2	$22^*, 2$	**89B**	1	2	**106B**	1	$5^*, 1$
49A	1	$?$	**70A**	1	$4, 2^*, 1^*$	**89C**	5	11	**106C**	1	$24, 1^*$
50A	1	$1^*, ?$	**71A**	3	5	**90A**	1	$2^*, ?, 3$	**106D**	1	$3, 1^*$
50B	1	$5, ?$	**71B**	3	7	**90B**	1	$6, ?, 1^*$			

Table 4.2: Big $L(A,1)/\Omega_A$

A	dim	N	$L(A,1)/\Omega_A \cdot$ Manin constant
1154E	20	$2 \cdot 577$	$2^? \cdot 85495047371/17^2$
1238G	19	$2 \cdot 619$	$2^? \cdot 7553329019/5 \cdot 31$
1322E	21	$2 \cdot 661$	$2^? \cdot 57851840099/331$
1382D	20	$2 \cdot 691$	$2^? \cdot 37 \cdot 1864449649/173$
1478J	20	$2 \cdot 739$	$2^? \cdot 7 \cdot 29 \cdot 1183045463/5 \cdot 37$

Table 4.3: Big component groups

A	p	w	$\#\Phi_X$	m_X	$\#\Phi_A(\overline{\mathbf{F}}_p)$
1154E	2	$-$	17^2	2^{24}	$2^? \cdot 17^2 \cdot 85495047371$
	577	$+$	1	$2^{26} \cdot 85495047371$	$2^?$
1238G	2	$-$	$5 \cdot 31$	2^{26}	$2^? \cdot 5 \cdot 31 \cdot 7553329019$
	619	$+$	1	$2^{28} \cdot 7553329019$	$2^?$
1322E	2	$-$	331	2^{28}	$2^? \cdot 331 \cdot 57851840099$
	661	$+$	1	$2^{32} \cdot 57851840099$	$2^?$
1382D	2	$-$	173	2^{29}	$2^? \cdot 37 \cdot 173 \cdot 1864449649$
	691	$+$	1	$2^{31} \cdot 37 \cdot 1864449649$	$2^?$
1478J	2	$-$	$5 \cdot 37$	2^{31}	$2^? \cdot 5 \cdot 7 \cdot 29 \cdot 37 \cdot 1183045463$
	739	$+$	1	$2^{33} \cdot 7 \cdot 29 \cdot 1183045463$	$2^?$

Table 4.4: Component groups of quotients of $J_0(N)$

A	dim	p	w_p	$\#\Phi_X$	m_X	m_A	$\#\Phi_A$
65A	1	5	$+$	1	2	2	1
		13	$+$	1	2		1
65B	2	5	$+$	3	2^2	2^2	3
		13	$-$	3	2^2		3
65C	2	5	$-$	7	2^2	2^2	7
		13	$+$	1	2^2		1
66A	1	2	$+$	1	2	2^2	2
		3	$-$	3	2^2		3
		11	$+$	1	2^2		1
66B	1	2	$-$	2	2	2^2	2^2
		3	$+$	1	2^2		1
		11	$+$	1	2^2		1
66C	1	2	$-$	1	2	$2^2\cdot 5$	$2\cdot 5$
		3	$-$	1	2^2		5
		11	$-$	1	$2^2\cdot 5$		1
68A	2	17	$+$	2	$2\cdot 3$	$2\cdot 3$	2
69A	1	3	$-$	2	2	2	2
		23	$+$	1	2		1
69B	2	3	$+$	2	2	$2\cdot 11$	$2\cdot 11$
		23	$-$	2	$2\cdot 11$		2

Table 4.5: Component groups of quotients of $J_0(p)^-$

A	d	$\#\Phi_A$	A	d	$\#\Phi_A$	A	d	$\#\Phi_A$	A	d	$\#\Phi_A$
11A	1	5	**157B**	7	13	**313A**	2	1	**487B**	2	3
17A	1	2^2	**163C**	7	3^3	**313C**	12	$2 \cdot 13$	**487C**	3	1
19A	1	3	**167B**	12	83	**317B**	15	79	**487D**	16	3^3
23A	2	11	**173B**	10	43	**331D**	16	$5 \cdot 11$	**491C**	29	$5 \cdot 7^2$
29A	2	7	**179A**	1	1	**337B**	15	$2^2 \cdot 7$	**499C**	23	83
31A	2	5	**179C**	11	89	**347D**	19	173	**503B**	1	1
37B	1	3	**181B**	9	$3 \cdot 5$	**349B**	17	29	**503C**	1	1
41A	3	$2 \cdot 5$	**191B**	14	$5 \cdot 19$	**353A**	1	2	**503D**	3	1
43B	2	7	**193C**	8	2^4	**353B**	3	2	**503F**	26	251
47A	4	23	**197C**	10	7^2	**353D**	14	$2 \cdot 11$	**509B**	28	127
53B	3	13	**199A**	2	1	**359D**	24	179	**521B**	29	$2 \cdot 5 \cdot 13$
59A	5	29	**199C**	10	$3 \cdot 11$	**367B**	19	61	**523C**	26	$3 \cdot 29$
61B	3	5	**211A**	2	5	**373C**	17	31	**541B**	24	$3^2 \cdot 5$
67A	1	1	**211D**	9	7	**379B**	18	$3^2 \cdot 7$	**547C**	25	$7 \cdot 13$
67C	2	11	**223C**	12	37	**383C**	24	191	**557B**	1	1
71A	3	5	**227B**	2	1	**389A**	1	1	**557D**	26	139
71B	3	7	**227C**	2	1	**389E**	20	97	**563A**	1	1
73A	1	2	**227E**	10	113	**397B**	2	1	**563E**	31	281
73C	2	3	**229C**	11	19	**397C**	5	11	**569B**	31	$2 \cdot 71$
79B	5	13	**233A**	1	2	**397D**	10	3	**571A**	1	1
83B	6	41	**233C**	11	29	**401B**	21	$2^2 \cdot 5^2$	**571B**	1	1
89B	1	2	**239B**	17	$7 \cdot 17$	**409B**	20	$2 \cdot 17$	**571C**	2	1
89C	5	11	**241B**	12	$2^2 \cdot 5$	**419B**	26	$11 \cdot 19$	**571D**	2	1
97B	4	2^3	**251B**	17	5^3	**421B**	19	$5 \cdot 7$	**571F**	4	1
101B	7	5^2	**257B**	14	2^6	**431B**	1	1	**571I**	18	$5 \cdot 19$
103B	6	17	**263B**	17	131	**431D**	3	1	**577A**	2	3
107B	7	53	**269C**	16	67	**431F**	24	$5 \cdot 43$	**577B**	2	1
109A	1	1	**271B**	16	$3^2 \cdot 5$	**433A**	1	1	**577C**	3	1
109C	4	3^2	**277B**	3	1	**433B**	3	1	**577D**	18	2^4
113A	1	2	**277D**	9	23	**433D**	16	$2^2 \cdot 3^2$	**587C**	31	293
113B	2	2	**281B**	16	$2 \cdot 5 \cdot 7$	**439C**	25	73	**593B**	1	2
113D	3	7	**283B**	14	47	**443C**	1	1	**593C**	2	1
127B	7	$3 \cdot 7$	**293B**	16	73	**443E**	22	$13 \cdot 17$	**593E**	27	$2 \cdot 37$
131B	10	$5 \cdot 13$	**307A**	1	1	**449B**	23	$2^4 \cdot 7$	**599C**	37	$13 \cdot 23$
137B	7	$2 \cdot 17$	**307B**	1	1	**457C**	20	$2 \cdot 19$	**601B**	29	$2 \cdot 5^2$
139A	1	1	**307C**	1	1	**461D**	26	$5 \cdot 23$	**607D**	31	101
139C	7	23	**307D**	1	1	**463B**	22	$7 \cdot 11$	**613C**	27	$3 \cdot 17$
149B	9	37	**307E**	2	3	**467C**	26	233	**617B**	28	$2 \cdot 7 \cdot 11$
151B	3	1	**307F**	9	17	**479B**	32	239	**619B**	30	103
151C	6	5^2	**311B**	22	$5 \cdot 31$	**487A**	2	1	**631B**	32	$3 \cdot 5 \cdot 7$

Bibliography

[1] A. Agashé, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. **328** (1999), no. 5, 369–374.

[2] A. Agashe, *The Birch and Swinnerton-Dyer formula for modular abelian varieties of analytic rank* 0, U. C. Berkeley Ph.D. thesis (2000).

[3] A. Agashe and W. A. Stein, *On the generalized manin constant for quotients of* $J_0(N)$, in preparation.

[4] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[5] B. Birch, *Atkin and the Atlas Lab*, Proceedings of the conference in honor of A. O. L. Atkin held at the University of Illinois, Chicago, IL, September 1995, Amer. Math. Soc., Providence, RI, 1998, pp. 13–20.

[6] B. J. Birch and W. Kuyk (eds.), *Modular functions of one variable. IV*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 476.

[7] S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, The Grothendieck Festschrift, Vol. I, Birkhäuser Boston, Boston, MA, 1990, pp. 333–400.

[8] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.

[9] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system I: The user language*, J. Symb. Comp. **24** (1997), no. 3-4, 235–265,
http://www.maths.usyd.edu.au:8000/u/magma/.

[10] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*, or Wild* 3*-adic exercises*, (2000),
http://www.math.harvard.edu/HTML/Individuals/Richard_Taylor.html.

[11] A. Brumer, *The rank of* $J_0(N)$, Astérisque (1995), no. 228, 3, 41–68, Columbia University Number Theory Seminar (New York, 1992).

[12] K. Buzzard and W. A. Stein, *Modularity of some icosahedral Galois representations*, in preparation.

[13] H. Cohen and J. Oesterlé, *Dimensions des espaces de formes modulaires*, (1977), 69–78. Lecture Notes in Math., Vol. 627.

[14] R. Coleman, *The monodromy pairing*, Asian Math. Journal (1999).

[15] J. E. Cremona, *Modular symbols for $\Gamma_1(N)$ and elliptic curves with everywhere good reduction*, Math. Proc. Cambridge Philos. Soc. **111** (1992), no. 2, 199–218.

[16] ______, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[17] ______, *Computing periods of cusp forms and modular elliptic curves*, Experiment. Math. **6** (1997), no. 2, 97–107.

[18] J. E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, to appear in Experiment. Math.

[19] H. Darmon, *Wiles' theorem and the arithmetic of elliptic curves*, Modular forms and Fermat's last theorem (Boston, MA, 1995), Springer, New York, 1997, pp. 549–569.

[20] H. Darmon and L. Merel, *Winding quotients and some variants of Fermat's last theorem*, J. Reine Angew. Math. **490** (1997), 81–100.

[21] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[22] N. Dummigan, *Period ratios of modular forms*, to appear in Math. Annalen.

[23] B. Edixhoven, *L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein"*, Astérisque (1991), no. 196–197, 7–8, 159–170 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[24] E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. of Comp. (2000).

[25] D. Goldfeld, *On the computational complexity of modular symbols*, Math. Comp. **58** (1992), no. 198, 807–814.

[26] J. González and J-C. Lario, **Q**-*curves and their Manin ideals*, preprint (2000).

[27] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

[28] B. H. Gross, *L-functions at the central critical point*, Motives (Seattle, WA, 1991), Amer. Math. Soc., Providence, RI, 1994, pp. 527–535.

[29] K. Hatada, *Multiplicity one theorem and modular symbols*, J. Math. Soc. Japan **33** (1981), no. 3, 445–470.

[30] H. Hijikata, *Explicit formula of the traces of Hecke operators for* $\Gamma_0(N)$, J. Math. Soc. Japan **26** (1974), no. 1, 56–82.

[31] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

[32] D. R. Kohel, *Hecke module structure of quaternions*, preprint (1998).

[33] V. A. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 429–436.

[34] ———, *On the structure of Shafarevich-Tate groups*, Algebraic geometry (Chicago, IL, 1989), Springer, Berlin, 1991, pp. 94–121.

[35] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of* Ш *over totally real fields*, Math. USSR Izvestiya **39** (1992), no. 1, 829–853.

[36] S. Lang, *Algebra*, third ed., Addison-Wesley Publishing Co., Reading, Mass., 1993.

[37] W-C. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.

[38] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66.

[39] B. Mazur, *Courbes elliptiques et symboles modulaires*, Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 414, Springer, Berlin, 1973, pp. 277–294. Lecture Notes in Math., Vol. 317.

[40] ———, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[41] ———, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[42] ———, *On the arithmetic of special values of L functions*, Invent. Math. **55** (1979), no. 3, 207–240.

[43] ———, *Visualizing elements of order three in the Shafarevich-Tate group*, preprint (1999).

[44] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61.

[45] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations (Berlin), Springer, 1994, pp. 59–94.

[46] ———, *L'accouplement de Weil entre le sous-groupe de Shimura et le sous-groupe cuspidal de* $J_0(p)$, J. Reine Angew. Math. **477** (1996), 71–115.

[47] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242.

[48] J.-F. Mestre and J. Oesterlé, *Courbes de Weil semi-stables de discriminant une puissance m-ième*, J. Reine Angew. Math. **400** (1989), 173–184.

[49] J. S. Milne, *Étale cohomology*, Princeton University Press, Princeton, N.J., 1980.

[50] ———, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[51] ———, *Arithmetic duality theorems*, Academic Press Inc., Boston, Mass., 1986.

[52] A. Pizer, *An algorithm for computing modular forms on* $\Gamma_0(N)$, J. Algebra **64** (1980), no. 2, 340–390.

[53] K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), no. 2, 431–476.

[54] ———, *Raising the levels of modular representations*, Séminaire de Théorie des Nombres, Paris 1987–88, Birkhäuser Boston, Boston, MA, 1990, pp. 259–271.

[55] K. Rubin, *The "main conjectures" of Iwasawa theory for imaginary quadratic fields*, Invent. Math. **103** (1991), no. 1, 25–68.

[56] ———, *Euler Systems*, Princeton University Press, Spring 2000, Annals of Mathematics Studies **147**, http://math.Stanford.EDU/~rubin/weyl.html.

[57] A. J. Scholl, *Motives for modular forms*, Invent. Math. **100** (1990), no. 2, 419–430.

[58] ———, *An introduction to Kato's Euler systems*, Galois Representations in Arithmetic Algebraic Geometry, Cambridge University Press, 1998, pp. 379–460.

[59] I. R. Shafarevich, *Exponents of elliptic curves*, Dokl. Akad. Nauk SSSR (N.S.) **114** (1957), 714–716.

[60] G. Shimura, *Sur les intégrales attachées aux formes automorphes*, J. Math. Soc. Japan **11** (1959), 291–311.

[61] ———, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544.

[62] ———, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

[63] V. V. Šokurov, *Modular symbols of arbitrary weight*, Funkcional. Anal. i Priložen. **10** (1976), no. 1, 95–96.

[64] W. A. Stein, `HECKE`*: The modular symbols calculator*, Software (available online) (1999).

[65] G. Stevens, *Arithmetic on modular curves*, Birkhäuser Boston Inc., Boston, Mass., 1982.

[66] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin, 1987, pp. 275–280.

[67] J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1995, pp. Exp. No. 306, 415–440.

[68] C. Viola, *Arithmetic theory of elliptic curves. Lectures given at the 3rd session of the Centro Internazionale Matematico Estivo (CIME).*, Springer-Verlag, Berlin, 1997 (English).

[69] D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. **28** (1985), no. 3, 372–384.

Index

4 A Mod Five Approach To Modularity Of Icosahedral Galois Representations, with K. Buzzard

PACIFIC JOURNAL OF MATHEMATICS
Vol. 203, No. 2, 2002

A MOD FIVE APPROACH TO MODULARITY OF ICOSAHEDRAL GALOIS REPRESENTATIONS

Kevin Buzzard and William A. Stein

We give eight new examples of icosahedral Galois representations that satisfy Artin's conjecture on holomorphicity of their L-function. We give in detail one example of an icosahedral representation of conductor $1376 = 2^5 \cdot 43$ that satisfies Artin's conjecture. We briefly explain the computations behind seven additional examples of conductors $2416 = 2^4 \cdot 151$, $3184 = 2^4 \cdot 199$, $3556 = 2^2 \cdot 7 \cdot 127$, $3756 = 2^2 \cdot 3 \cdot 313$, $4108 = 2^2 \cdot 13 \cdot 79$, $4288 = 2^6 \cdot 67$, and $5373 = 3^3 \cdot 199$. We also generalize a result of Sturm on computing congruences between eigenforms.

Introduction.

Consider a continuous irreducible Galois representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C})$$

with $n > 1$. Inspired by his reciprocity law, Artin conjectured in [1] that $L(\rho, s)$ has an analytic continuation to the whole complex plane. Many of the known cases of this conjecture were obtained by proving the apparently stronger assertion that ρ is *automorphic*, in the sense that the L-function of ρ is equal to the L-function of a certain automorphic representation (whose L-function is known to have analytic continuation). In the special case where $n = 2$ and ρ is in addition assumed to be odd, the automorphic representation in question should be the one associated to a classical weight 1 modular eigenform, and in fact there is conjectured to be a bijection between such ρ and the set of all weight 1 cuspidal newforms, which should preserve L-functions. It is this bijection that we are concerned with in this paper, so assume for the rest of the paper that $n = 2$ and ρ is odd.

In this special case, the construction of [7] shows how to construct a continuous irreducible odd 2-dimensional representation from a weight 1 newform, and the problem is to go the other way. Say that a representation is *modular* if it arises in this way.

If the image of ρ is solvable, then ρ is known to be modular [11, 18]; if the image is not solvable, then $\mathrm{Im}(\rho)$ in $\mathrm{PGL}_2(\mathbf{C})$ is isomorphic to the alternating group A_5, and the modularity of ρ is, in general, unknown. We call such a 2-dimensional representation an “icosahedral representation”. The published literature contains only eight examples (up to twist) of odd icosahedral Galois representations that are known to satisfy Artin’s conjecture: One of conductor $800 = 2^5 \cdot 5^2$ (see [3]), and seven of conductors: 2083, $2^2 \cdot 487$, $2^2 \cdot 751$, $2^2 \cdot 887$, $2^2 \cdot 919$, $2^5 \cdot 73$, and $2^5 \cdot 193$ (see [8]).

After the first draft of this paper was written, the preprint [4] appeared, which contains a general theorem that yields infinitely many (up to twist) modular icosahedral representations. However, we feel that our work, although much less powerful, is still of some worth, because it gives an effective computational approach to proving that certain mod 5 representations are modular, without computing any spaces of weight 1 forms or using effective versions of the Chebotarëv density theorem. We also note that the main theorem of [4] does not apply to any of the examples considered in the present paper. Very recently, the preprint [17] appeared, which gives new local conditions under which an icosahedral representation can be proved to be modular. In particular, [17] also proves that the first three examples in the present paper, of conductors 1376, 2416, 3184, are modular; these correspond to the first, third, and fourth equations at the end of [17]. However, [17] does not apply to our remaining five examples. Finally, we note that this paper also contains a result (Corollary 1.7) generalizing the main results of [16], which makes explicit computations with mod p modular forms much more practical.

Let ρ be a continuous odd icosahedral representation. We briefly summarise our approach for verifying modularity of ρ. As all the representations we consider are unramified at 5, one can use the main theorem of [5] to reduce the problem to showing that the mod 5 reduction of ρ is modular. We do this by using a computer to find a candidate mod 5 modular form at weight 5 and then, using the table of icosahedral extensions of $\mathbf{Q}$ in [8] and what we know about the 5-adic representation attached to our candidate form, we deduce that the mod 5 representation attached to our candidate form must be the reduction of ρ. In particular, this paper gives a computational method for checking the modularity of certain mod 5 representations whose conductors are not too large.

We now explain something about a problematic point in this approach, which is to verify that a given modular form which has been obtained by a computation actually gives rise to an explicit mod 5 representation which has been given by another computation. In each of our examples it is easy to compute a few Hecke operators and be morally convinced that this is the case; it is far more difficult to prove this. Effective variants of the Chebotarëv density theorem require that we check vastly more traces of Frobenius than is practical. Our approach was as follows. Let f be one of the forms that we computed. We firstly used the compatibility of the Local and Global Langlands correspondences for GL_2 and some twisting tricks to deduce that the kernel of the projective mod 5 representation associated to f must correspond to an A_5-extension of $\mathbf{Q}$. We then used the theory of companion forms and a careful local analysis of the representations associated to the forms to deduce strong local results about these A_5-extensions. Finally we used Table 2 of [8] to prove that in each case the A_5-extension was precisely the one we wanted it to be.

We carried out this program for icosahedral representations of the following conductors: $\mathbf{1376} = 2^5 \cdot 43$, $\mathbf{2416} = 2^4 \cdot 151$, $\mathbf{3184} = 2^4 \cdot 199$, $\mathbf{3556} = 2^2 \cdot 7 \cdot 127$, $\mathbf{3756} = 2^2 \cdot 3 \cdot 313$, $\mathbf{4108} = 2^2 \cdot 13 \cdot 79$, $\mathbf{4288} = 2^6 \cdot 67$, and $\mathbf{5373} = 3^3 \cdot 199$.

This paper is divided into three sections. In Section 1, we give in detail our proof that the icosahedral representation of minimal conductor 1376 satisfies Artin's conjecture. The subsections of Section 1 follow the plan outlined above. Section 2 summarizes the data necessary to deduce Artin's conjecture for all eight of our examples. Finally, Section 3 contains a brief review of modular symbols, and contains some tables of running times.

1. Modularity of an icosahedral representation of conductor $1376 = 2^5 \cdot 43$.

In this section we prove the following theorem.

Theorem 1.1. *The icosahedral representations whose corresponding icosahedral extension is the splitting field of* $x^5 + 2x^4 + 6x^3 + 8x^2 + 10x + 8$ *are modular.*

Let K be the splitting field of $h = x^5+2x^4+6x^3+8x^2+10x+8$. The Galois group of K is A_5, so we obtain a homomorphism $G_{\mathbf{Q}} \to A_5 \subset \mathrm{PGL}_2(\mathbf{C})$; let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{C})$ be a minimal lift, minimal in the sense that the Artin conductor of ρ is minimal. By Table A_5 of [3], the conductor of ρ is $N = 1376 = 2^5 \cdot 43$. Since $h \equiv (x-1)(x^2-x+1)(x^2-x+2) \pmod 5$, and $\mathrm{disc}(h)$ is coprime to 5, any Frobenius element at 5 in $\mathrm{Gal}(K/\mathbf{Q})$ has order 2.

We use the notation of Tables 3.1 and 3.2 of [3, p. 46], which gives a complete classification of the way that ramified primes can behave in such representations. In our case the ramified primes are 2 and 43. From Table 3.2 of [3] we see that the type of ρ at 2 is 17 and the type at 43 is 2. The level N Dirichlet character $\widetilde{\varepsilon} = \det(\rho)$ factors as $\widetilde{\varepsilon} = \widetilde{\varepsilon}_2 \cdot \widetilde{\varepsilon}_{43}$ where $\widetilde{\varepsilon}_2$ is a character of conductor dividing 2^5 and $\widetilde{\varepsilon}_{43}$ is a character of conductor 43. We can work out these characters explicitly as we know the type of ρ at 2 and 43—indeed, there is a character associated to each type in Buhler's table, which unfortunately is not tabulated. An easy local computation shows that $\widetilde{\varepsilon}_{43}$ has order 3, and fortunately Buhler's level 800 example also was of type 17 at 2 (see the first line of [3, Table 3.2]), hence by [3, p. 80] $\widetilde{\varepsilon}_2$ is the unique character of conductor 4 and order 2. We think of these characters now has having values in $\mathbf{Q}(\zeta_3) \subseteq \overline{\mathbf{Q}}$, where ζ_3 is a primitive cube root of unity.

If ρ is modular, then there is a weight 1 newform $f_? \in S_1(N, \widetilde{\varepsilon}; \overline{\mathbf{Q}})$ that gives rise to ρ. Suppose for the moment that ρ is modular, so that $f_?$ exists. The Eisenstein series E_4 of level 1 and weight 4 is congruent to 1 modulo 5, so $E_4 \cdot \overline{f}_? \in S_5(N, \widetilde{\varepsilon}; \overline{\mathbf{Q}})$ reduces modulo a prime above 5 to a form which is an eigenform for all Hecke operators T_q for $q \neq 5$ prime, with the same eigenvalues mod 5 as $f_?$, and hence is a mod 5 weight 5 eigenform giving rise to the mod 5 reduction of ρ. Using a computer, we can search for such a mod 5 eigenform. In practice one computes a $\mathbf{Z}[\zeta_3]$-lattice in $S_5(N, \widetilde{\varepsilon}; \mathbf{Q}(\zeta_3))$ and then reduces the lattice modulo 5; we refer to the resulting quotient space as $S_5(N, \varepsilon; \mathbf{F}_{25})$, abusing notation slightly, where ε denotes the reduction of $\widetilde{\varepsilon}$. (Similarly we write ε_2 and ε_{43} to be the reductions of $\widetilde{\varepsilon}_2$ and $\widetilde{\varepsilon}_{43}$.) We search for an eigenform f in this mod 5 space of modular forms, whose existence is assured if we believe Artin's conjecture.

Instead of multiplying $\overline{f}_?$ by E_4, we could have multiplied it by an appropriate Eisenstein series of weight 1 and level 5. We used E_4 because the dimension of $S_5(N, \varepsilon; \overline{\mathbf{F}}_5)$ is 696 whereas the dimension of the relevant space $S_2(5 \cdot N, \varepsilon_{43})$ of weight 2 cusp forms is 1040.

1.1. Searching for the newform f. Using modular symbols we compute the space $S_5(1376, \varepsilon; \mathbf{F}_{25})$. By computing the kernels of various Hecke operators on this space, we find f. In the following computations, we represent nonzero elements of $\mathbf{F}_{25}$ as powers of a generator α of $\mathbf{F}_{25}^*$, which satisfies

$$\alpha^2 + 4\alpha + 2 = 0.$$

If 2 is the least common multiple of the degrees of the factors of the polynomial h modulo an unramified prime p, then $\mathrm{Frob}_p \in \mathrm{Gal}(K/\mathbf{Q})$ has order 2, hence trace 0. The first three such p are $19, 31, 97$. We computed the mod 5 reduction $S_5(1376, \varepsilon; \mathbf{F}_{25}) = \boldsymbol{\mathcal{S}}_5(1376, \varepsilon; \mathbf{F}_{25})^+$ of the $\mathbf{Z}_5[\zeta_3]$-lattice of modular symbols of level 1376 and character ε, where complex conjugation acts as $+1$. The intersection V of the kernels of T_{19}, T_{31}, and T_{97} inside $\boldsymbol{\mathcal{S}}_5(1376, \varepsilon; \mathbf{F}_{25})^+$ has dimension 8, and no doubt all the eigenforms in this space give rise to ρ or one of its twists. One of the eigenvalues of T_3 on this space is α^{16}, and the kernel V_1 of $T_3 - \alpha^{16}$ is 2-dimensional over $\mathbf{F}_{25}$. The Hecke operator T_5 acted as a diagonalizable matrix on V_1, with eigenvalues α^{10} and α^{22}, so the corresponding two systems of eigenvalues must correspond to mod 5 modular eigenforms, and furthermore we must have found all mod 5 modular eigenforms $\sum a_n q^n$ of this level, weight and character, such that $a_{19} = a_{31} = a_{97} = 0$ and $a_3 = \alpha^{16}$.

Remark 1.2. The careful reader might wonder how we know that the systems of mod 5 eigenvalues really do correspond to mod 5 modular forms, and not to perhaps some strange mod 5 torsion in the space of modular symbols.

Table 1. Eigenvalues of f.

2	0	59	4	137	0	227	α^{10}	313	0	419	3	509	α^{8}
3	α^{16}	61	α^{14}	139	α^{22}	229	0	317	0	421	α^{20}	521	α^{10}
5	α^{22}	67	α^{4}	149	α^{4}	233	α^{14}	331	α^{14}	431	4	523	α^{14}
7	α^{14}	71	α^{20}	151	1	239	0	337	0	433	α^{4}	541	α^{20}
11	4	73	α^{2}	157	α^{14}	241	α^{2}	347	α^{16}	439	α^{20}	547	α^{22}
13	α^{14}	79	α^{20}	163	0	251	α^{2}	349	α^{4}	443	0	557	3
17	α^{14}	83	α^{4}	167	α^{22}	257	3	353	0	449	0	563	1
19	0	89	α^{10}	173	4	263	α^{16}	359	0	457	0	569	α^{16}
23	α^{16}	97	0	179	α^{2}	269	2	367	α^{22}	461	0	571	α^{22}
29	α^{8}	101	α^{8}	181	α^{14}	271	α^{8}	373	0	463	α^{10}	577	α^{14}
31	0	103	α^{14}	191	α^{10}	277	0	379	3	467	0	587	α^{20}
37	α^{10}	107	0	193	4	281	α^{16}	383	3	479	0	593	0
41	1	109	α^{10}	197	0	283	0	389	1	487	α^{8}	599	α^{22}
43	α^{10}	113	2	199	3	293	3	397	α^{16}	491	α^{2}	601	0
47	1	127	0	211	0	307	α^{4}	401	0	499	α^{20}	607	α^{16}
53	α^{22}	131	2	223	0	311	α^{22}	409	2	503	α^{2}	613	2

However, we eliminated this possibility by computing the dimension of the full space of mod 5 modular symbols where complex conjugation acts as $+1$, and checking that it equals 696, the dimension of $S_5(1376, \widetilde{\varepsilon}, \mathbf{C})$, which we computed using the formula in [6].

Let f be the eigenform in V_1 that satisfies $a_5 = \alpha^{22}$; the q-expansion of f begins

$$f = q + \alpha^{16}q^3 + \alpha^{22}q^5 + \alpha^{14}q^7 + \alpha^{14}q^9 + 4q^{11} + \cdots .$$

Further eigenvalues are given in Table 1. The primes p in the table such that $a_p = 0$ are exactly those predicted by considering the splitting behavior of h. This is strong evidence that ρ is modular, and also that our modular symbols algorithms have been correctly implemented.

1.2. Twisting into $\mathrm{GL}(2, \mathbf{F}_5)$**.** Although there is a representation $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}(2, \mathbf{F}_{25})$ attached to the weight 5 mod 5 eigenform f, it is difficult to say anything about its image without further work. We use a trick to show that the image of ρ_f is small. Firstly, for a character $\chi : G_{\mathbf{Q}} \to \overline{\mathbf{F}}_5$, let $\widetilde{\chi}$ denote its Teichmüller lift to $\overline{\mathbf{Q}}_5$. Consider the $\mathbf{Z}$-algebra of Hecke operators acting on $S_5(N, \widetilde{\varepsilon}; \overline{\mathbf{Q}}_5)$. By choosing a minimal prime under the maximal ideal of this algebra corresponding to f, we see that there is a characteristic 0 eigenform $\widetilde{f} \in S_5(N, \widetilde{\varepsilon}; \overline{\mathbf{Q}}_5)$ lifting f.

The component ε_{43} of ε at 43 is represented by the map sending $(1, 3) \in (\mathbf{Z}/2^5\mathbf{Z})^* \times (\mathbf{Z}/43\mathbf{Z})^*$ to $2\alpha + 1$ and sending the subgroup $(\mathbf{Z}/2^5\mathbf{Z})^* \times \{1\}$

Table 2. Eigenvalues of $g = f \otimes \varepsilon_{43}$.

2	$*$	59	4	137	0	227	3	313	0	419	3	509	1	617	0
3	1	61	2	139	2	229	0	317	0	421	4	521	3	619	4
5	$*$	67	4	149	4	233	2	331	2	431	4	523	2	631	4
7	2	71	4	151	1	239	0	337	0	433	4	541	4	641	4
11	4	73	3	157	2	241	3	347	1	439	4	547	2	643	1
13	2	79	4	163	0	251	3	349	4	443	0	557	3	647	4
17	2	83	4	167	2	257	3	353	0	449	0	563	1	653	1
19	0	89	3	173	4	263	1	359	0	457	0	569	1	659	2
23	1	97	0	179	3	269	2	367	2	461	0	571	2	661	2
29	1	101	1	181	2	271	1	373	0	463	3	577	2	673	1
31	0	103	2	191	3	277	0	379	3	467	0	587	4	677	4
37	3	107	0	193	4	281	1	383	3	479	0	593	0	683	0
41	1	109	3	197	0	283	0	389	1	487	1	599	2	691	1
43	$*$	113	2	199	3	293	3	397	1	491	3	601	0	701	2
47	1	127	0	211	0	307	4	401	0	499	4	607	1	709	4
53	2	131	2	223	0	311	2	409	2	503	3	613	2	719	4

to 1. Note that 3 is a primitive root mod 43, and that $2\alpha+1$ has order 3. The complementary character ε_2 is defined by $\varepsilon = \varepsilon_2 \cdot \varepsilon_{43}$. The twist $\widetilde{g} = \widetilde{f} \otimes \widetilde{\varepsilon}_{43}$ is, by [14, Prop. 3.64], an eigenform in $S_5(43N, \widetilde{\varepsilon}_2; \overline{\mathbf{Q}}_5)$, and its reduction is a form $g \in S_5(43N, \varepsilon_2; \mathbf{F}_{25})$. The eigenvalues $a_p(g) = a_p(f)\varepsilon_{43}(p)$, for the first few $p \nmid 5N$, are given in Table 2.

Proposition 1.3. *Let $g = f \otimes \varepsilon_{43}$. Then $a_p(g) \in \mathbf{F}_5$ for all $p \nmid 5N$.*

Proof. Consider an eigenform $\widetilde{f} \in S_5(N, \widetilde{\varepsilon}; \overline{\mathbf{Q}}_5)$ lifting f as above. Associated to $\widetilde{f}$ there is an automorphic representation $\pi = \otimes'_v \pi_v$ of $\mathrm{GL}(2, \mathbf{A})$, where $\mathbf{A}$ is the adèle ring of $\mathbf{Q}$. Because $43 \mid\mid N$, and 43 divides the conductor of ε, we see that the local component π_{43} of π at 43 must be ramified principal series. By the compatibility of the local and global Langlands correspondence, proved by Deligne, Langlands and Carayol, we see that $\rho_{\widetilde{f}}|_{D_{43}} \sim \left(\begin{smallmatrix}\Psi_1 & 0\\ 0 & \Psi_2\end{smallmatrix}\right)$ with, without loss of generality, Ψ_2 unramified. We have $(\Psi_1 \cdot \Psi_2)|_{I_{43}} = \widetilde{\varepsilon}|_{I_{43}} = \widetilde{\varepsilon}_{43}$, therefore, $\rho_{\widetilde{f}}|_{I_{43}} \sim \left(\begin{smallmatrix}\widetilde{\varepsilon}_{43} & 0\\ 0 & 1\end{smallmatrix}\right)$.

Now twist $\widetilde{f}$ by $\widetilde{\varepsilon}_{43}^{-1}$; we find that $\rho_{\widetilde{f} \otimes \widetilde{\varepsilon}_{43}^{-1}}|_{I_{43}} \sim \left(\begin{smallmatrix}1 & 0\\ 0 & \widetilde{\varepsilon}_{43}^{-1}\end{smallmatrix}\right)$. In particular, there is an eigenform $\widetilde{f}' \in S_5(N, \widetilde{\varepsilon}_2\widetilde{\varepsilon}_{43}^{-1}; \overline{\mathbf{Q}}_5)$ whose associated Galois representation is the twist by $\widetilde{\varepsilon}_{43}^{-1}$ of that of $\widetilde{f}$ (recall that $N = 1376$ so 43 divides N exactly once). Let f' denote the mod 5 reduction of $\widetilde{f}'$. Then one checks easily that $f' \in S_5(N, \varepsilon_2\varepsilon_{43}^{-1}; \mathbf{F}_{25}) = S_5(N, \varepsilon^5; \mathbf{F}_{25})$.

For all primes $p \nmid 5N$ we have $a_p(f') = \varepsilon_{43}(p)^{-1}a_p(f)$. In particular, we have $a_p(f') = 0$ for $p = 19, 31$. Also, $\varepsilon_{43}(3) = \alpha^8$ and $\varepsilon_{43}(5) = \alpha^8$, so

$$a_3(f') = \alpha^{16}/\alpha^8 = \alpha^8 = (\alpha^{16})^5$$

$$a_5(f') = \alpha^{22}/\alpha^8 = \alpha^{14} = (\alpha^{22})^5.$$

Now if σ is the nontrivial automorphism of $\mathbf{F}_{25}$, then $\sigma(f')$ and f both lie in $S_5(1376, \varepsilon; \mathbf{F}_{25})$ and have the same a_p for $p = 3, 5, 19, 31$, so they are equal because we found f by computing the unique eigenform with given a_p for $p = 3, 5, 19, 31$. So $g = f \otimes \varepsilon_{43} = \sigma(f) \otimes \varepsilon_{43}^2$. Thus for all $p \nmid 5N$, we see that $a_p(g) = a_p(f)^5\varepsilon_{43}^2$ has fifth power $a_p(g)^5 = a_p(f)^{25}\varepsilon_{43}^{10} = a_p(f)\varepsilon_{43} = a_p(g)$. □

1.3. Proof that ρ_g is unramified at 5. We begin with a generalization of [16]. Let $M > 4$ be an integer, and let $h = \sum_{n \geq 1} c_n q^n$ be a normalized cuspidal eigenform of some weight $k \geq 1$, level M and character χ, defined over some field of characteristic not dividing M. Even though the base field might not have characteristic zero, we may still define the conductor of χ to be the smallest divisor f of M such that χ factors through $(\mathbf{Z}/f\mathbf{Z})^\times$. Let I be a set of primes, with the property that for all p in I, one of the following conditions hold:

(i) p divides M but p does not divide $M/\operatorname{cond}(\chi)$, or

(ii) p divides M exactly once, and h is p-new, in the sense that there is no eigenform h' of level M/p such that the T_n-eigenvalues of h and h' agree for all n prime to p.

Let C denote the orbit of the cusp ∞ in $X_1(M)$ under the action of the group generated by w_p for $p \in I$, and the Diamond operators $\langle d \rangle_M$. The orbit of ∞ under the Diamond operators has size $\phi(M)/2$, and each w_p increases the size of the orbit by a factor of 2. In this situation, we have:

Lemma 1.4. *The first t terms of the q-expansion of h at any cusp in C are determined by M, k, χ, c_p for p in I, and c_n for $1 \leq n \leq t$.*

Remark 1.5. Our proof is just a translation of Corollary 4.6.18 of [13] into the language of moduli problems (Miyake's argument technically is only valid over the complex numbers).

Proof. If $J \subseteq I$ is any subset, and w_J denotes the product of w_p for $p \in J$, then $h|w_J$ is an eigenform for all the Diamond operators, and this observation reduces the proof of the lemma to showing that for $p \in I$, if $h|w_p = \sum_m d_m q^m$, then d_j for $1 \leq j \leq n$ and d_q for all $q \in I$ are determined by M, k, χ, p, c_j for $1 \leq j \leq n$ and c_q for all $q \in I$.

We first deal with primes p of the form (i). Say $M = p^m R$, where R is prime to p. Thinking of h as a rule for attaching k-fold differentials to elliptic curves equipped with points of order p^m and R, we have by definition that

$$h(\mathbf{G}_m/q^{\mathbf{Z}}, \zeta, \zeta_R) = \left(\sum c_n q^n\right)(dt/t)^k,$$

where $\zeta = \zeta_{p^m}$ and ζ_R are fixed p^mth and Rth roots of unity in $\mathbf{G}_m$ which correspond to the cusp ∞, and dt/t is the canonical differential on the Tate curve $\mathbf{G}_m/q^{\mathbf{Z}}$. We normalize things such that

$$h(\mathbf{G}_m/q^{p^m\mathbf{Z}}, q, \zeta_R) = \left(\sum d_n q^n\right)(dt/t)^k,$$

and remark that because h is an eigenvector for the Diamond operators, we do not have to worry too much about whether this corresponds to the standard normalization of the w_p-operator.

We recall that the operator pU_p in this setting can be thought of as being defined by the rule:

$$(pU_p h)(E, P, Q) = \sum_C \pi^* h(E/C, \overline{P}, \overline{Q}),$$

where C runs through the subgroups of E of order p which have trivial intersection with $\langle P\rangle$, and π denotes the canonical projection $E \to E/C$. We see that

$$\begin{aligned}(pc_p)^m\left(\sum d_n q^n\right)(dt/t)^k &= (p^m U_{p^m} h)(\mathbf{G}_m/q^{p^m\mathbf{Z}}, q, \zeta_R)\\ &= \sum_{c=0}^{p^m-1} \pi^* h(\mathbf{G}_m/\langle q^{p^m}, \zeta q^c\rangle, q, \zeta_R),\end{aligned}$$

where π denotes the canonical projection from $\mathbf{G}_m/\langle q^{p^m}\rangle$ to the appropriate quotient. This last sum can be written as a double sum

$$\begin{aligned}&\sum_{c\in(\mathbf{Z}/p^m\mathbf{Z})^\times} \pi^* h(\mathbf{G}_m/\langle q^{p^m}, \zeta q^c\rangle, q, \zeta_R) + \sum_{a=0}^{p^{m-1}-1} \pi^* h(\mathbf{G}_m/\langle q^{p^m}, \zeta q^{pa}\rangle, q, \zeta_R)\\
&= \sum_{b\in(\mathbf{Z}/p^m\mathbf{Z})^\times} \pi^* h(\mathbf{G}_m/\langle q^{p^m}, \zeta^{-b} q\rangle, q, \zeta_R)\\
&\quad + p^{m-1}\pi^* U_{p^{m-1}} h(\mathbf{G}_m/\langle q^{p^m}, \zeta^{p^{m-1}}\rangle, q, \zeta_R)\\
&= \sum_{b\in(\mathbf{Z}/p^m\mathbf{Z})^\times} \pi^* h(\mathbf{G}_m/\langle \zeta^{-b} q\rangle, \zeta^b, \zeta_R)\\
&\quad + (pc_p)^{m-1}\pi^* h(\mathbf{G}_m/\langle q^{p^m}, \zeta^{p^{m-1}}\rangle, q, \zeta_R)\\
&= \sum_b \chi_p(b) \sum_{n\geq 1} c_n(\zeta^{-b}q)^n (dt/t)^k + p^k (pc_p)^{m-1}\pi^* h(\mathbf{G}_m/\langle q^{p^{m+1}}\rangle, q^p, \zeta_R^p),\end{aligned}$$

where we have written $\chi = \chi_R\chi_p$, for χ_R a character of level R and χ_p a character of level p^m. We deduce that

$$\begin{aligned}&(pc_p)^m \left(\sum d_n q^n\right)(dt/t)^k - p^k(pc_p)^{m-1}\chi_R(p)\pi^* h(\mathbf{G}_m/\langle q^{p^{m+1}}\rangle, q^p, \zeta_R)\\&= \left(\sum_n \left(\sum_b \chi_p(b)\zeta^{-bn}\right) c_n q^n\right)(dt/t)^k\\&= W(\chi_p)\left(\sum_{p\nmid n}\chi_p(-n)^{-1}c_n q^n\right)(dt/t)^k\end{aligned}$$

where $W(\chi_p) = \sum_{b\in(\mathbf{Z}/p^m\mathbf{Z})^\times}\chi_p(b)\zeta^b$ can be checked to be nonzero because the conductor of χ_p is p^m. Hence

$$\begin{aligned}&(pc_p)^m \sum_n d_n q^n - p^k(pc_p)^{m-1}\chi_R(p)\sum_n d_n q^{np}\\&= W(\chi_p)\chi_p(-1)\sum_{p\nmid n}\chi_p(n)^{-1}c_n q^n.\end{aligned}$$

Equating coefficients of q we deduce that $W(\chi_p)\chi_p(-1) = (pc_p)^m d_1$, and because $h|w_p$ is an eigenform for T_n for all n prime to p, with eigenvalues determined by χ and c_n, we deduce that we can determine d_n for n prime to p from c_n. It remains to establish what d_p is, and equating coefficients of q^p in the above equation gives us that $(pc_p)^m d_p = p^k(pc_p)^{m-1}\chi_R(p)d_1$ and hence that d_p is determined by χ and c_p. Note that as a consequence we see that $d_p/d_1 = p^{k-1}\chi_R(p)/c_p$, a classical formula if the base field is the complexes.

Now we deal with primes of the form (ii) (note that we never use this case in the rest of the paper). We think of h as a rule associating k-fold differentials to triples (E, C, Q) where C a cyclic subgroup of order p and Q a point of order $R = M/p$. Because h is p-new, the trace of h down to $X_1(M/p)$ must be zero, and hence we see that for any elliptic curve E equipped with a point Q of order R,

$$\sum_C \pi^* h(E/C, E[p]/C, \overline{Q}) = 0.$$

As before, normalize things so that

$$h(\mathbf{G}_m/q^{\mathbf{Z}}, \mu_p, \zeta_R) = \left(\sum_n c_n q^n\right)(dt/t)^k$$

and

$$h(\mathbf{G}_m/q^{p\mathbf{Z}}, \langle q\rangle, \zeta_R) = \left(\sum_n d_n q^n\right)(dt/t)^k.$$

The fact that the trace of h is zero implies that

$$(pU_p)h(\mathbf{G}_m/q^{p\mathbf{Z}},\langle q\rangle,\zeta_R)+\pi^*h(\mathbf{G}_m/q^{\mathbf{Z}},\mu_p,\zeta_R)=0,$$

and hence that

$$c_p\sum d_nq^n+p^{k-1}\sum c_nq^n=0$$

from which we deduce that the d_n can be read off from c_p and the c_n. □

Remark 1.6. The size of C is $\phi(M)\cdot 2^{|I|-1}$, and the usefulness of this lemma is that if h_1 and h_2 are two normalized eigenforms of the same level, weight and character as above, both new at all primes in I, and the coefficients of q^n in the q-expansions of h_1 and h_2 agree for $n\in I$ and $n\leq t$, then h_1-h_2 has a zero of order at least $t+1$ at all cusps in C, and in particular if $\phi(M)\cdot 2^{|I|-1}(t+1)>\frac{k}{24}[\mathrm{SL}_2(\mathbf{Z}):\Gamma_1(M)]=\deg(\omega^k)$ on $X_1(M)$ then $h_1=h_2$. Using the fact that $[\Gamma_0(M):\Gamma_1(M)]=\phi(M)$, we deduce:

Corollary 1.7. *Let h_1 and h_2 be two normalized eigenforms as above. If the coefficients of q^n in the q-expansions of h_1 and h_2 agree for all primes in I and for all $n\leq\frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}):\Gamma_0(M)]/2^{|I|}$ then $h_1=h_2$.*

Remark 1.8. One can certainly do better than this corollary in many cases. For example, when $n>1$ and p^n exactly divides both the level of an eigenform and the conductor of its character, then one can compute the q-expansion of the eigenform at many "middle cusps" too, and hence increase the size of C in the result above. The general result however is rather messy to state and prove, and so for simplicity we have chosen to prove only what we needed in the cases we were interested in.

We now go back to the explicit situation we are concerned with. Although g is an eigenform of level $59168=2^5\cdot 43^2$, we can still consider the corresponding representation $\rho_g:G_{\mathbf{Q}}\to\mathrm{GL}(2,\mathbf{F}_5)$, and then directly analyze its ramification.

Proposition 1.9. *The representation ρ_g is unramified at 5.*

Proof. Continuing the modular symbols computations as above, we find that V_1 is spanned by the two eigenforms

$$\begin{aligned}f\ &=q+\alpha^{16}q^3+\alpha^{22}q^5+\alpha^{14}q^7+\alpha^{14}q^9+4q^{11}+\cdots\\ f_1&=q+\alpha^{16}q^3+\alpha^{10}q^5+\alpha^{14}q^7+\alpha^{14}q^9+4q^{11}+\cdots.\end{aligned}$$

For $p\neq 5$ and $p\leq 997$, we have $a_p(f_1)=a_p(f)$. To check that $a_p(f)=a_p(f_1)$ for all $p\neq 5$, it suffices to show that the difference $f-f_1$ has q-expansion involving only powers of q^5; for this we use the θ-operator $q\frac{d}{dq}:S_5(1376,\varepsilon;\mathbf{F}_{25})\to S_{11}(1376,\varepsilon;\mathbf{F}_{25})$. Since θ sends normalized eigenforms to normalized eigenforms, it suffices to check that the subspace of $S_{11}(1376,\varepsilon;\mathbf{F}_{25})$ generated by $\theta(f)$ and $\theta(f_1)$ has dimension 1. Corollary 1.7

implies that it suffices to verify that the coefficients $a_p(\theta(f))$ and $a_p(\theta(f_1))$ are equal for all

$$p \leq \frac{11}{12} \cdot [\mathrm{SL}_2(\mathbf{Z}) : \Gamma_0(1376)] \cdot \frac{1}{2} = 968.$$

The eigenform f must be new because we computed it by finding the intersections of the kernels of Hecke operators T_p with $p \nmid 1376$; if f were an oldform then the intersection of the kernels of these Hecke operators would necessarily have dimension greater than 1. Because it takes less than a second to compute each $a_p(\theta(f))$, we were easily able to verify that the space generated by $\theta(f)$ and $\theta(f_1)$ has dimension 1.

Remark 1.10. In this example (but not some of the other seven examples!) it is possible to avoid appealing to Corollary 1.7 by using one of the following two alternative methods:

1) Define θ directly on modular symbols and compute it. On modular symbols, the analogue of the θ operator seems to be multiplication by $X^pY - XY^p$; thus, if $p = k = 5$ then $\theta(X^3\{0,\infty\}) = (X^8Y - X^4Y^5)\{0,\infty\}$. The main point in the proof is that one can check easily from the definitions that $T_q\theta = q\theta T_q$ for q a good prime, and hence this map theta must correspond with the "classical" theta up to a constant; one should perhaps worry that this constant could be zero, but in practice given an f one can check explicitly that $\theta(f) \neq 0$ by direct computation.
2) Compute the intersection

$$\bigcap_{p \geq 2} \ker(T_p - pa_p(f)) \subset S_{11}(1376, \varepsilon; \mathbf{F}_{25}).$$

Since $\theta(f)$ and $\theta(f_1)$ both lie in the intersection, the moment the dimension of a partial intersection is 1, it follows that $\theta(f - f_1) = 0$.

We successfully carried out both alternatives. For the second, we find that after intersecting kernels for $p \leq 11$, the dimension is already 1. The first of these two methods took much less time than the second.

Next we use that $\theta(f - f_1) = 0$ to show that ρ_g is unramified, thus finishing the proof of the proposition. Since f is ordinary, Deligne's theorem (see [9, §12]) implies that

$$\rho_f|_{D_5} \sim \begin{pmatrix} \gamma & * \\ 0 & \delta \end{pmatrix} \qquad \text{over } \overline{\mathbf{F}}_5$$

with γ and δ unramified characters, $\gamma(\mathrm{Frob}_5) = \varepsilon(5)/a_5 = \alpha^8/\alpha^{22} = \alpha^{10}$, and $\delta(\mathrm{Frob}_5) = \alpha^{22}$. Since $a_p(f_1) = a_p(f)$, for $p \neq 5$, we have

$$\rho_f|_{D_5} \sim \rho_{f_1}|_{D_5} \sim \begin{pmatrix} \gamma' & * \\ 0 & \delta' \end{pmatrix}$$

with $\gamma'(\mathrm{Frob}_5) = \alpha^8/\alpha^{10} = \alpha^{22}$ and $\delta'(\mathrm{Frob}_5) = \alpha^{10}$; in particular, $\gamma' = \delta$. Thus $\rho_f|_{D_5}$ contains $\gamma \oplus \delta$, so $\rho_f|_{D_5} \sim \gamma \oplus \delta$ and hence there is a choice of basis so that $* = 0$. □

1.4. The image of $\operatorname{proj} \rho_g$.

Proposition 1.11. *The image of* $\operatorname{proj} \rho_g$ *is* A_5.

Proof. The image H of $\operatorname{proj} \rho_g$ in $\mathrm{PGL}_2(\mathbf{F}_5)$ is easily checked to lie in $\mathrm{PSL}_2(\mathbf{F}_5) \cong A_5$ because of what we know about the determinant of ρ_g. Hence H is a subgroup of A_5 that contains an element of order 2 (complex conjugation) and an element of order 3 (for example, $\rho_g(\mathrm{Frob}_7)$ has characteristic polynomial $x^2 - 2x - 1$). This proves that H is isomorphic to either S_3, A_4, or A_5. Let L be the number field cut out by H. If L were an S_3-extension, then there would be a quadratic extension contained in it which is unramified outside $2 \cdot 5 \cdot 43$; it is furthermore unramified at 5 by the previous section and unramified at 43 because I_{43} has order 3. Thus it is one of the three quadratic fields unramified outside 2. In particular, the trace of Frob_p would be zero for all primes in a certain congruence class modulo 8. However, there are primes p congruent to 3, 5, and 7 mod 8 such that $a_p(g) \neq 0$, e.g., 3, 7, and 13.

If H were isomorphic to A_4, then let M denote the cyclic extension of degree 3 over $\mathbf{Q}$ contained in L. Now M is unramified at 2 and 5, and hence is the subfield of $\mathbf{Q}(\zeta_{43})$ of degree 3. Choose $p \nmid 1376 \cdot 5$ that is inert in M, i.e., so that p is not a cube mod 43. The order of $\rho_g(\mathrm{Frob}_p)$ in $\mathrm{GL}_2(\mathbf{F}_5)$ must be divisible by 3. However, a quick check using Table 2 shows that this is not the case for $p = 3$. □

1.5. Bounding the ramification at 2 and 43.

Let L be the fixed field of $\ker(\operatorname{proj}(\rho_g))$. We have just shown that $\mathrm{Gal}(L/\mathbf{Q})$ is isomorphic to A_5. By a root field for L, we mean a non-Galois extension of $\mathbf{Q}$ of degree 5 whose Galois closure is L.

Proposition 1.12. *The discriminant of a root field for L divides* $(43\cdot 8)^2 = 344^2$*, and in particular, L must be mentioned in Table* 1 *of* [8, pg. 122].

Proof. The analysis of the local behavior of ρ_f at 43 given in Proposition 1.3 shows that the inertia group at 43 in $\mathrm{Gal}(L/\mathbf{Q})$ has order 3. Using Table 3.1 of [3], we see that if $\mathrm{Gal}(L/\mathbf{Q}) \cong A_5$ then it must be of type 2 at 43, and hence the discriminant of a root field of L, that is, of a non-Galois extension of $\mathbf{Q}$ of degree 5 whose Galois closure is L, must be 43^2 at 43.

At 2 the behavior of ρ is more subtle and we shall not analyze it fully. But we can say that, because ρ has arisen from a form of level $1376 = 2^5 \cdot 43$, we must be either of type 5 or one of types 14–17, in the notation of Table 3.2 of [3]. In particular, the discriminant at 2 of a root field for L will be at most 2^6.

Finally, L is unramified at all other primes, because ρ is. Hence the discriminant of a root field for L, assuming that $\mathrm{Gal}(L/\mathbf{Q}) \cong A_5$, divides $(43.8)^2 = 344^2$. □

We know that L is an icosahedral extension of $\mathbf{Q}$ with discriminant dividing $43^2 \cdot 2^6$. Table 1 of [8, p. 122] contains all icosahedral extensions, such that the discriminant of a root field is bounded by 2083^2. The table must contain L; there is only one icosahedral extension with discriminant dividing $43^2 \cdot 2^6$, so $L = K$.

1.6. Obtaining a classical weight one form. We have shown that a twist of the icosahedral representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \mathbf{C})$, obtained by lifting $G_{\mathbf{Q}} \to \mathrm{Gal}(K/\mathbf{Q}) \approx A_5$, has a mod 5 reduction $\rho_g : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_5)$ that is modular. Since ρ ramifies at only finitely many primes, and ρ is unramified at 5 with distinct eigenvalues, [5] implies that ρ arises from a classical weight 1 newform.

2. More examples.

The data necessary to deduce modularity of each of our eight icosahedral examples is summarized in Tables 3–6.

Table 3. Data on icosahedral representations mod 5.

N	h	$\mathrm{ord}(\mathrm{Frob}_5)$	p with $a_p = 0$	ε	$\dim S_5(N, \varepsilon)$
1376	$[2, 6, 8, 10, 8]$	2	$19, 31, 97$	$[2, 1, 3]$	696
2416	$[0, -2, 2, 5, 6]$	2	$53, 97, 127$	$[2, 1, 3]$	1210
3184	$[5, 8, -20, -21, -5]$	2	$31, 89, 97$	$[2, 1, 3]$	1594
3556	$[3, 9, -6, -4, -40]$	3	$19, 29, 89$	$[1, 2, 3]$	2042
3756	$[0, -3, 10, 30, -18]$	3	$17, 61, 67$	$[1, 2, 3]$	2506
4108	$[4, 3, 9, 4, 5]$	3	$17, 23, 31, 89$	$[1, 3, 2]$	2234
4288	$[4, 5, 8, 3, 2]$	3	$19, 23, 47$	$[1, 2, 3]$	2164
5373	$[2, 1, 7, 23, -11]$	2	$7, 23, 37, 79, 89$	$[2, 3]$	2394

The notation in Table 3 is as follows. The first column contains the conductor. The second column contains a 5-tuple $[a_4, a_3, a_2, a_1, a_0]$ such that the A_5-extension is the splitting field of the polynomial $h = x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$. The column labeled $\mathrm{ord}(\mathrm{Frob}_5)$ contains the order of the image of Frob_5 in A_5. The next column, which is labeled "p with $a_p = 0$", contains the first few p such that a_p is easily seen to equal 0 by considering the splitting of h mod p. The ε column contains the character of the representation, where the notation is as follows. Write $(\mathbf{Z}/N\mathbf{Z})^*$ as a product of cyclic groups corresponding to the prime divisors of N in ascending order, and then the tuples give the orders of the images of these

Table 4. The newform f and the companion form bound.

N	f	bound
1376	$q+\alpha^{16}q^3+\alpha^{22}q^5+\alpha^{14}q^7+\alpha^{14}q^9+4q^{11}+\alpha^{14}q^{13}+\cdots$	968
2416	$q+3q^3+\alpha^{22}q^5+\alpha^{16}q^7+\alpha^{4}q^{11}+\alpha^{2}q^{13}+\alpha^{16}q^{15}+\cdots$	1672
3184	$q+\alpha^{16}q^3+3q^5+\alpha^{22}q^7+\alpha^{14}q^9+3q^{11}+\alpha^{22}q^{13}+\cdots$	2200
3556	$q+\alpha^{16}q^3+\alpha^{14}q^5+\alpha^{10}q^7+\alpha^{14}q^9+\alpha^{2}q^{11}+\alpha^{22}q^{13}+\cdots$	1408
3756	$q+\alpha^{14}q^3+\alpha^{14}q^5+3q^7+\alpha^{4}q^9+\alpha^{16}q^{11}+\alpha^{10}q^{13}+\cdots$	1727
4108	$q+\alpha^{16}q^3+\alpha^{11}q^5+\alpha^{20}q^7+\alpha^{14}q^9+\alpha^{10}q^{11}+4q^{13}+\cdots$	1540
4288	$q+3q^3+\alpha^{14}q^5+\alpha^{20}q^7+3q^9+\alpha^{20}q^{11}+\alpha^{16}q^{13}+\cdots$	2992
5373	$q+\alpha^{16}q^2+\alpha^{14}q^4+4q^5+3q^8+\alpha^{4}q^{10}+2q^{11}+\cdots$	3300

cyclic factors; when $8 \mid N$, there are two cyclic factors corresponding to the prime 2. Finally, the last column records the dimension of $S_5(\Gamma_1(N),\varepsilon)$.

The notation in Table 4 is as follows. The first column contains the conductor. The second column contains an eigenform that was found by first intersecting the kernels of the Hecke operators T_p with p as in Table 3, and then locating an eigenform. In each case, a companion form was found, by computing $a_p(f)$ for $p \leq$ bound, where bound is the bound from Corollary 1.7.

Table 5 shows that the fixed field of the image of each $\operatorname{proj}(\rho_g)$ is icosahedral. The first column contains the conductor N. The second column contains a twist g of f such that $a_p(g) \in \mathbf{F}_5$ for all $p \nmid 5N$. The third column contains a Frob_p such that $\operatorname{proj}(\rho_g(\mathrm{Frob}_p))$ has order 3, along with the characteristic polynomial of $\rho_g(\mathrm{Frob}_p)$. As in the proof of Proposition 1.11, the other two boxes give data that allows us to deduce that the fixed field of the image of $\operatorname{proj}(\rho_g)$ is icosahedral. The case 5373 must be treated separately, because there are three possibilities M_1, M_2, and M_3 for the cubic field M of the analogue of Proposition 1.11. For M_1 we find a prime p such that

$$(p^2 \mod 9,\;\; p^{66} \mod 199) \notin \{(1,1),(4,1),(7,1)\}$$

with $\rho_g(\mathrm{Frob}_p)$ of order not divisible by 3; for this, $p=2$ suffices, since the characteristic polynomial of $\rho_g(\mathrm{Frob}_2)$ is $(x+2)^2$ and $(p^2 \bmod 9,\; p^{66} \bmod 199) = (4,106)$. For M_2 we find a prime p such that

$$(p^2 \mod 9,\;\; p^{66} \mod 199) \notin \{(1,1),(4,92),(7,106)\}$$

with $\rho_g(\mathrm{Frob}_p)$ of order not divisible by 3; again, $p=2$ suffices. For M_3 we find a prime p such that

$$(p^2 \mod 9,\;\; p^{66} \mod 199) \notin \{(1,1),(4,106),(7,92)\}$$

with $\rho_g(\mathrm{Frob}_p)$ of order not divisible by 3; here, $p=13$ suffices, as the characteristic polynomial of $\rho_g(\mathrm{Frob}_p)$ is $(x+4)^2$ and $(p^2 \bmod 9,\;\; p^{66} \bmod 199) = (7,106)$.

Table 5. Verification that the image of $\mathrm{proj}(\rho_g)$ is A_5.

Find a Frobenius element with projective order 3.

N	g	proj. order 3	charpoly
1376	$f \otimes \varepsilon_{43}$	Frob_7	$x^2 - 2x - 1$
2416	$f \otimes \varepsilon_{151}$	Frob_{19}	$x^2 + 2x - 1$
3184	$f \otimes \varepsilon_{199}$	Frob_7	$x^2 + 3x + 4$
3556	$f \otimes \varepsilon_{127}$	Frob_{13}	$x^2 + 3x + 4$
3756	$f \otimes \varepsilon_{313}$	Frob_{23}	$x^2 + 2x + 4$
4108	$f \otimes \varepsilon_{13}$	Frob_{29}	$x^2 + 3x + 4$
4288	$f \otimes \varepsilon_{67}$	Frob_{11}	$x^2 + x + 1$
5373	$f \otimes \varepsilon_{199}$	Frob_{11}	$x^2 + 3x + 4$

Not S_3: For all $t \in T$, find unramified p s.t. $t \not\equiv \square \mod p$ and $a_p(g) \neq 0$.

N	T	p
1376	$\{-1, -2\}$	3, 7
2416	$\{-1, -2\}$	3, 7
3184	$\{-1, -2\}$	3, 7
3556	$\{-1, -2, -7, -14\}$	3, 13, 3, 11
3756	$\{-1, -2, -3, -6\}$	7, 7, 11, 13
4108	$\{-1, -2, -79, -158\}$	3, 7, 3, 7
4288	$\{-1, -2\}$	3, 7
5373	$\{-3\}$	11

Not A_4: Unramified p, not cube mod ℓ, order of $\rho_g(\mathrm{Frob}_p)$ not divisible by 3.

N	ℓ	p	$\mathrm{charpoly}(\rho_g(\mathrm{Frob}_p))$
1376	43	3	$(x+2)^2$
2416	151	7	$(x+2)^2$
3184	199	3	$(x+2)^2$
3556	127	3	$(x+2)^2$
3756	313	11	$(x+2)^2$
4108	13	3	$(x+2)^2$
4288	67	7	$(x+3)^2$
5373	—		(see text)

Table 6 gives upper bounds on the ramification of the fixed field of the image of $\mathrm{proj}(\rho_g)$. These bounds were deduced using Table 3.1 of [3] by restricting the possible "types" using information about the character ε. Note that though the bounds are not sharp, e.g., the discriminant of the

Table 6. Bounding the discriminant of the fixed field of $\mathrm{proj}(\rho_g)$.

N	Bound on discriminant
1376	$2^6 \cdot 43^2$
2416	$2^6 \cdot 151^2$
3184	$2^6 \cdot 199^2$
3556	$2^2 \cdot 7^2 \cdot 127^2$
3756	$2^2 \cdot 3^2 \cdot 313^2$
4108	$2^2 \cdot 13^2 \cdot 79^2$
4288	$2^6 \cdot 67^2$
5373	$3^4 \cdot 199^2$

representation of conductor 2416 is $2^4 \cdot 151^2$, they are all less than 2083^2, so the corresponding field must appear in Table 2 of [8].

3. Computing mod p modular forms.

3.1. Higher weight modular symbols. The second author developed software that computes the space of weight k modular symbols $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$, for $k \geq 2$ and arbitrary ε. See [12] for the standard facts about higher weight modular symbols, and [15] for a description of how to compute with them.

Let $K = \mathbf{Q}(\varepsilon)$ be the field generated by the values of ε. The cuspidal modular symbols $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$ are a finite dimensional vector space over K, which is generated by all linear combinations of higher weight modular symbols

$$X^i Y^{k-2-i}\{\alpha, \beta\}$$

that lie in the kernel of an appropriate boundary map. There is an involution $*$ that acts on $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)$, and $\boldsymbol{\mathcal{S}}_k(N, \varepsilon)^+ \otimes_K \mathbf{C}$ is isomorphic, as a module over the Hecke algebra, to the space $S_k(N, \varepsilon; \mathbf{C})$ of cusp forms.

Fix $k = 5$. In each case considered in this paper, there is a prime ideal λ of the ring of integers $\mathcal{O}$ of K such that $\mathcal{O}/\lambda \cong \mathbf{F}_{25}$. Let $\mathcal{L}$ be the $\mathcal{O}$-module generated by all modular symbols of the form $X^i Y^{3-i}\{\alpha, \beta\}$, and let

$$\boldsymbol{\mathcal{S}}_5(N, \varepsilon; \mathbf{F}_{25}) = (\mathcal{L} \cap \boldsymbol{\mathcal{S}}_5(N, \varepsilon)) \otimes_{\mathcal{O}} \mathbf{F}_{25}.$$

This is the space that we computed. The Hecke algebra acts on $\boldsymbol{\mathcal{S}}_5(N,\varepsilon; \mathbf{F}_{25})$, so when we find an eigenform we find a maximal ideal of the Hecke algebra.

As an extra check on our computation of $\boldsymbol{\mathcal{S}}_5(N, \varepsilon; \mathbf{F}_{25})$, we computed the dimension of $S_5(N, \varepsilon; \mathbf{C})$ using both the formula of [6] and the Hijikata trace formula (see [10]) applied to the identity Hecke operator.

3.2. Complexity. We implemented the modular symbols algorithms mentioned above in Magma (see [2]) because of its robust support for linear algebra over small finite fields.

The following table gives a flavor of the complexity of the machine computations appearing in this paper. The table indicates how much CPU time on a Sun Ultra E450 was required to compute all data for the given level, including the matrices T_p on the 2-dimensional spaces, for $p < 2000$. For example, the total time for level $N = 1376$ was 6 minutes and 58 seconds.

N	time (minutes)
1376	6:58
2416	10:42
3184	14:16
3556	19:55
3756	27:47
4108	23:11
4288	15:18
5376	24:49

Acknowledgment. Some of the computing equipment was purchased by the second author using a UC Berkeley Vice Chancellor Research Grant. Additional computer runs were made on the Sun Ultra E450 of the Computational Algebra Group at the University of Sydney. Allan Steel was very helpful in optimizing our code.

References

[1] E. Artin, *Über eine neue Art von L-reihen*, Abh. Math. Sem. in Univ. Hamburg, **3**(**1**) (1923/1924), 89-108.

[2] W. Bosma, J. Cannon and C. Playoust, *The Magma algebra system* I: *The user language*, J. Symb. Comp., **24**(**3-4**) (1997), 235-265, CMP 1 484 478, Zbl 0898.68039, http://www.maths.usyd.edu. au:8000/u/magma/.

[3] J.P. Buhler, *Icosahedral Galois representations*, Springer-Verlag, Berlin, 1978, Lecture Notes in Mathematics, Vol. 654, MR 58 #22019, Zbl 0374.12002.

[4] K. Buzzard, M. Dickinson, N. Shepherd-Barron and R. Taylor, *On icosahedral Artin representations*, Duke Math. J., **109**(**2**) (2001), 283-318, CMP 1 845 181.

[5] K. Buzzard and R. Taylor, *Companion forms and weight one forms*, Ann. of Math. (2), **149**(**3**) (1999), 905-919, MR 2000j:11062, Zbl 0965.11019.

[6] H. Cohen and J. Oesterlé, *Dimensions des espaces de formes modulaires*, Lecture Notes in Math., **627** (1977), 69-78, MR 57 #12396, Zbl 0371.10020.

[7] P. Deligne and J-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. (4), **7** (1974), 507-530, MR 52 #284, Zbl 0321.10026.

[8] G. Frey (ed.), *On Artin's conjecture for odd* 2*-dimensional representations*, Springer-Verlag, Berlin, 1994, MR 95i:11001, Zbl 0801.00004.

[9] B.H. Gross, *A tameness criterion for Galois representations associated to modular forms* (mod p), Duke Math. J., **61**(**2**) (1990), 445-517, MR 91i:11060, Zbl 0743.11030.

[10] H. Hijikata, *Explicit formula of the traces of Hecke operators for* $\Gamma_0(N)$, J. Math. Soc. Japan, **26**(**1**) (1974), 56-82, MR 49 #2552, Zbl 0266.12009.

[11] R.P. Langlands, *Base Change for* GL(2), Princeton University Press, Princeton, N.J., 1980, MR 82a:10032, Zbl 0444.22007.

[12] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations (Berlin), Springer, Lecture Notes in Math., **1585** (1994), 59-94, MR 96h:11032, Zbl 0844.11033.

[13] T. Miyake, *Modular Forms*, Springer-Verlag, Berlin, 1989, Translated from the Japanese by Yoshitaka Maeda, MR 90m:11062, Zbl 0701.11014.

[14] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, **1**, MR 95e:11048, Zbl 0872.11023.

[15] W.A. Stein, *Explicit approaches to modular abelian varieties*, U.C. Berkeley, Ph.D. thesis (2000).

[16] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984-1985), Springer, Berlin, Lecture Notes in Math., **1240** (1987), 275-280, MR 88h:11031, Zbl 0615.10035.

[17] R. Taylor, *On icosahedral Artin representations* II, in preparation.

[18] J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. (N.S.), **5**(**2**) (1981), 173-175, MR 82j:12015, Zbl 0475.12016.

Received June 28, 2000 and revised October 18, 2000.

Department of Mathematics
Imperial College
180 Queen's Gate
London, SW7 2BZ, England
E-mail address: buzzard@ic.ac.uk

Department of Mathematics
Harvard University
Cambridge, MA 02138
E-mail address: was@math.harvard.edu

This paper is available via http://www.pacjmath.org/2002/203-2-2.html.

5 There are genus one curves over Q of every odd index

J. reine angew. Math. **547** (2002), 139—147

Journal für die reine und
angewandte Mathematik
© Walter de Gruyter
Berlin · New York 2002

There are genus one curves over $\mathbb{Q}$ of every odd index

By *William A. Stein*[*] at Harvard University

Abstract. The index of a genus one curve X over a field K is the smallest degree of an extension L of K such that $X(L)$ is nonempty. Let K be a number field. We prove that for every integer r not divisible by 8, there is a genus one curve X over K of index r. Our proof involves an analysis of Kolyvagin's Euler system of Heegner points combined with explicit computations on the modular curve $X_0(17)$.

1. Introduction

How complicated are curves of genus one? One possible measure of the complexity of a curve is the smallest degree of an extension of the base field in which the curve has a point. Consider a curve X of positive genus g over a number field K. The canonical divisor class on X contains a K-rational effective divisor of degree $2g-2$, so the greatest common divisor of the degrees of the extension fields in which X has a rational point divides $2g-2$. When $g = 1$ this is no condition at all!

In the 1950s, S. Lang and J. Tate asked in [11] whether, given a positive integer r, there exists a genus one curve X such that r is the smallest of all degrees of extensions of K over which X has a point. Using Kolyvagin's Euler system of Heegner points, we answer their question in the affirmative, under the hypothesis that r is odd. The curves we produce are torsors for the elliptic curve $X_0(17)$, though our methods apply to a more general class of genus one curves. The following theorem is proved in Section 5.4.

Theorem 1.1. *Let K be a number field and let r be an integer not divisible by* 8. *Then there are infinitely many genus one curves over K of index r.*

In Section 2 we recall standard facts about indexes of genus one curves. Section 3 contains a brief discussion of Heegner points, and summarizes the relevant results about Kolyvagin's Euler system from [18]. In Section 4, which forms the heart of our paper, we prove a nonvanishing result for Kolyvagin's cohomology classes. Finally, in Section 5, we

[*] Supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship.

prove Theorem 1.1 by combining a general result about Galois representations with explicit computations on $X_0(17)$.

Acknowledgement. The author would like to thank H. Lenstra for introducing him to this problem, K. Buzzard for teaching him about Kolyvagin's Euler system, K. Rubin and M. Flach for extensive comments, D. Y. Logachev, C. O'Neil, and K. Ribet for inspiring conversations, and N. Elkies and G. Grigorov for useful comments.

2. Indexes of genus one curves

Let E be an elliptic curve over an arbitrary field k. The Galois cohomology group $H^1(k,E) = H^1\big(\mathrm{Gal}(k^{\mathrm{sep}}/k), E(k^{\mathrm{sep}})\big)$ classifies the isomorphism classes of torsors (principal homogeneous spaces) for E over k.

Definition 2.1 (Index of cohomology class). The *index* of $c \in H^1(k,E)$, denoted $\mathrm{ind}(c)$, is the greatest common divisor of the degrees of the separable extensions K of k for which $\mathrm{res}_K(c) = 0$.

The torsor X corresponding to c is a genus one curve over k equipped with an action of E. Furthermore, $X(K) \neq \emptyset$ exactly when $\mathrm{res}_K(c) = 0$, so

$$\mathrm{ind}(c) = \gcd\{[K:k] \colon X(K) \neq \emptyset\}.$$

Thus $\mathrm{ind}(c)$ generates the image of the degree map $\deg\colon \mathrm{Div}_k(X) \to \mathbb{Z}$. We now define $\mathrm{ind}(X)$ so that $\mathrm{ind}(X) = \mathrm{ind}(c)$.

Definition 2.2 (Index of curve). The *index* of an algebraic curve over k is the cardinality of the cokernel of the degree map.

Any canonical divisor is an element of $\mathrm{Div}_k(X)$ of degree $2g-2$, where g is the genus of X, so $\mathrm{ind}(X)$ divides $2g-2$. As mentioned in the introduction, when $g = 1$ this is no condition; in fact, E. Artin conjectured, and Lang and Tate proved in [11], pg. 670, that for every integer r there is some genus one curve X over some field L such that $\mathrm{ind}(X) = r$. The construction of [11] requires the existence of an L-rational point of order r on the elliptic curve $E = \mathrm{Jac}(X)$. The torsion subgroups of elliptic curves are "uniformly bounded", so for K a fixed number field and for almost all r, the results of [11] do not imply the existence of genus one curves over K of index r.

Let E be an elliptic curve over a number field K, and let r be a positive integer. Is there an element of $H^1(K,E)$ of index r? In [21], Shafarevich proved that $H^1(K,E)$ contains infinitely many elements of every *order* (see also [5], §27 where Cassels sketches an alternative approach to proving Shafarevich's theorem). However, this does not answer the question of Artin, because the order need not equal the index as Cassels remarked in [4], where he found an elliptic curve E and a class $c \in H^1(\mathbb{Q},E)$ such that c has order 2 and index 4.

2.1. Elementary facts about the index. We pause to state some basic facts about the order and index, which we will use later. Fix an elliptic curve E over a number field K, and let c and c' be elements of $H^1(K,E)$.

Proposition 2.3. $\mathrm{ord}(c) \mid \mathrm{ind}(c)$, *and they have the same prime factors.*

Proof. See [11], §2, Prop. 5. □

Lemma 2.4. *There is an extension L of K such that* $[L:K] = \mathrm{ind}(c)$ *and* $\mathrm{res}_L(c) = 0$.

Proof. See the paragraph before the corollary in [11], §2. □

Proposition 2.5. *Suppose* c' *has order coprime to* c. *Then* $\mathrm{ind}(c + c') = \mathrm{ind}(c) \cdot \mathrm{ind}(c')$.

Proof. If M is a field that splits $c + c'$, then M also splits $\mathrm{ord}(c')(c + c') = \mathrm{ord}(c')c$, so M splits c. Likewise, M splits c', so $\mathrm{ind}(c) \cdot \mathrm{ind}(c') \mid \mathrm{ind}(c + c')$. For the other divisibility, note that by Lemma 2.4, there are extensions L and L' such that $[L:K] = \mathrm{ind}(c)$, $[L':K] = \mathrm{ind}(c')$, and $\mathrm{res}_L(c) = \mathrm{res}_{L'}(c') = 0$. Then the compositum $L.L'$ splits $c + c'$ and $[L.L':K] = \mathrm{ind}(c) \cdot \mathrm{ind}(c')$. Thus $\mathrm{ind}(c + c')$ divides $\mathrm{ind}(c) \cdot \mathrm{ind}(c')$. □

Remark 2.6. In [12], Lichtenbaum proved that $\mathrm{ind}(c) \mid \mathrm{ord}(c)^2$ for any $c \in H^1(K, E)$, and Cassels proved in [3] that if $c \in Ш(E/K)$, then $\mathrm{ord}(c) = \mathrm{ind}(c)$.

If E is an elliptic curve over $\mathbb{Q}$ such that $\# Ш(E/\mathbb{Q}) = \# E(\mathbb{Q})_{\mathrm{tor}} = 1$, then the results mentioned above do not rule out the possibility that every element of $H^1(\mathbb{Q}, E)$ has index a perfect square. We prove, under the assumption that $L(E, 1) \neq 0$, that there is an integer B such that $H^1(\mathbb{Q}, E)$ contains infinitely many elements of index n, for every integer n that is coprime to B (see Theorem 3.1). For example, in Section 5 we prove that one can take $B = 2$ for the elliptic curve $X_0(17)$.

3. Kolyvagin's Euler system

In this section, we recall the definition of Heegner points and several basic results about the system of cohomology classes Kolyvagin attaches to these points. We also state the main theorem of this paper.

3.1. Kolyvagin classes. Let E be an elliptic curve over $\mathbb{Q}$ of conductor N, and denote by $X_0(N)$ the modular curve that classifies cyclic isogenies of degree N. By [1], there is a surjective map $\pi: X_0(N) \to E$. (Note that for the proof of Theorem 1.1 we do not need any modularity theorems, because we take $E = X_0(17)$.) Let K be a quadratic imaginary extension of $\mathbb{Q}$ in which all primes dividing N split, and let D_K be the discriminant and $\mathcal{O}$ the ring of integers of K. Since all primes dividing N split, there is an ideal $\mathfrak{a} \subset \mathcal{O}$ such that $\mathcal{O}/\mathfrak{a}$ is cyclic of order N. Let H be the Hilbert class field of K, and $x_H \in X_0(N)(H)$ be the Heegner points corresponding to $(\mathbb{C}/\mathcal{O}, \mathfrak{a}^{-1}/\mathcal{O})$. Set $y_H = \pi(x_H) \in E(H)$, $y_K = \mathrm{tr}_{H/K}(y_H) \in E(K)$, and $y = y_K - y_K^{\tau} \in E(K)^{-}$, where τ denotes complex conjugation. Assume that $L(E, 1) \neq 0$, so by [2] and [15] there are infinitely many ways in which to choose K as above so that y has infinite order. Under this nonvanishing hypothesis on $L(E, 1)$, Kolyvagin proves in [10] that the groups $E(\mathbb{Q})$ and $Ш(E/\mathbb{Q})$ are both finite.

In the course of his proof, Kolyvagin considers more general Heegner points $y_\ell \in E(\bar{\mathbb{Q}})$, for appropriate primes ℓ, and from these constructs cohomology classes $c_{\ell, p^n} \in H^1(\mathbb{Q}, E)[p^n]$ that are used to bound the orders of certain Selmer groups associated to E. We will study Kolyvagin's classes further and prove that for each prime p not in an explicit finite set and each positive integer n, there are infinitely many primes ℓ such that

$$\operatorname{ord}(c_{\ell,p^n}) = \operatorname{ind}(c_{\ell,p^n}) = p^n.$$

We thus obtain the following theorem, which will be proved in Section 4.2.

Theorem 3.1. *Let E be an elliptic curve over $\mathbb{Q}$ such that $L(E,1) \neq 0$. Then there is an integer B such that, for all integers r coprime to B, there are infinitely many $c \in H^1(\mathbb{Q},E)$ such that* $\operatorname{ord}(c) = \operatorname{ind}(c) = r$.

Remark 3.2. Cathy O'Neil [16] has investigated the obstruction to $\operatorname{ord}(c) = \operatorname{ind}(c)$. We show that when E has analytic rank 0, this obstruction vanishes for infinitely many c.

3.2. Basic properties of Kolyvagin's Euler system. In [18], Rubin gives a concise account of Kolyvagin's proof of finiteness of $\text{Ш}(E/\mathbb{Q})[p^\infty]$, under the simplifying assumption that p is odd. Though Kolyvagin's argument works even when $p = 2$, for simplicity, we rely exclusively on Rubin's paper.

Let K be a quadratic imaginary field as above, chosen in such a way that the associated Heegner point y_K has infinite order. Fix embeddings of $\bar{\mathbb{Q}}$ into $\mathbb{C}$ and into each p-adic field $\bar{\mathbb{Q}}_p$. Let τ denote complex conjugation, and for any $\mathbb{Z}[\tau]$-module A, let A^+ and A^- denote the kernel of $\tau - 1$ and $\tau + 1$, respectively. For the remainder of this section, we assume that p is an odd prime, and if $K = \mathbb{Q}(\sqrt{-3})$ that $p \geqq 5$. If ℓ is a prime that is inert in K, let K_ℓ denote the completion of K at the unique prime lying over ℓ. If L is a finite Galois extension of $\mathbb{Q}$, let $\operatorname{Frob}_\ell(L/\mathbb{Q})$ denote the conjugacy class of some Frobenius element of a prime lying over ℓ. For each prime $\ell \nmid N$, let $a_\ell = \ell + 1 - \#E(\mathbb{F}_\ell)$ be the ℓth Fourier coefficient of the newform attached to E.

Definition 3.3. For each place v of $\mathbb{Q}$, let

$$m_v = \#H^1\big(\mathbb{Q}_v^{\text{unr}}/\mathbb{Q}_v, E(\mathbb{Q}_v^{\text{unr}})\big).$$

By [14], I.3.8, each m_v is finite and $m_v = 1$ for all but finitely many v, so

$$m(p) = \sup\{\operatorname{ord}_p(m_v)\colon \text{all places } v \text{ of } \mathbb{Q}\}$$

is well defined, and $m(p) = 0$ for almost all p.

Let n be a positive integer.

Proposition 3.4. *Let p be a prime that does not divide the class number of K and for which $m(p) = 0$. Suppose $\ell \nmid pD_KN$ and* $\operatorname{Frob}_\ell\big(K(E[p^n])/\mathbb{Q}\big) = [\tau]$. *Then there is an element $c_{\ell,p^n} \in H^1(\mathbb{Q},E)[p^n]$ such that the order of* $\operatorname{res}_\ell(c_{\ell,p^n})$ *in $H^1(\mathbb{Q}_\ell,E)[p^n]$ is equal to the order of the image of y in $E(K_\ell)/p^nE(K_\ell)$, and the index of c_{ℓ,p^n} divides p^n.*

Proof. The existence of c_{ℓ,p^n} and statement about its order is proved in [18], Prop. 5, where c_{ℓ,p^n} is constructed from Heegner points on $X_0(N)$. For the index bound, note that in the proof of [18], Prop. 5, when $p \nmid [H : K]$, Rubin constructs a class

$$c' \in H^1\big(K'/K, E(K')\big)[p^r]^+,$$

where $r = n + m(p)$ and K' is the unique extension of K of degree p^r in a certain class field of K. Since p is odd, the restriction map res: $H^1(\mathbb{Q}, E)[p^r] \to H^1(K, E)[p^r]^+$ is an isomorphism. Rubin takes $c_{\ell,p^n} = \mathrm{res}^{-1}(c')$. Since c_{ℓ,p^n} splits over the degree $2p^r$ extension K' of $\mathbb{Q}$, the index of c_{ℓ,p^n} divides $2p^r$. But c_{ℓ,p^n} has odd order and, by Proposition 2.3, $\mathrm{ind}(c_{\ell,p^n})$ has the same prime factors as $\mathrm{ord}(c_{\ell,p^n})$, so $\mathrm{ind}(c_{\ell,p^n})$ divides p^r. □

Remark 3.5. The author does not know whether or not the proposition is true if p is allowed to divide the class number of K.

4. Nonvanishing of cohomology classes

In this section, we prove a nonvanishing result about the cohomology classes c_{ℓ,p^n} of Proposition 3.4, then use it to deduce Theorem 3.1.

4.1. Local nonvanishing. Let E be as above. For any point $x \in E(K)$, let $K([p^n]^{-1}x)$ denote the field obtained by adjoining the coordinates of all p^nth roots of x to K. Without imposing further hypothesis, this field need not be Galois over $\mathbb{Q}$.

Lemma 4.1. *If $x \in E(K)^+ \cup E(K)^-$, then $K([p^n]^{-1}x)$ is Galois over $\mathbb{Q}$.*

Proof. Since $G_{\mathbb{Q}}$ acts on x by ± 1, the subgroup $\mathbb{Z}x$ is $G_{\mathbb{Q}}$-invariant. Since $[p^n]: E \to E$ is a $\mathbb{Q}$-rational isogeny the inverse image $[p^n]^{-1}\mathbb{Z}x$ is also $G_{\mathbb{Q}}$-invariant, so $K([p^n]^{-1}x) = K([p^n]^{-1}\mathbb{Z}x)$ is Galois over $\mathbb{Q}$. □

Definition 4.2. An odd prime p is *firm* for E if $m(p) = 0$, there are no nontrivial $\mathbb{Q}$-rational cyclic subgroups of $E[p^\infty]$, and $H^1\big(K(E[p^n])/K, E[p^n]\big) = 0$ for all $n \geqq 1$.

Remark 4.3. The set of primes that are not firm is finite, by Serre's theorem [19] and the theory of complex multiplication.

Let p be an odd prime that is firm for E. The following proposition produces infinitely many primes ℓ such that we have control over the orders of the image in $E(K_\ell)/p^nE(K_\ell)$ of a global point. It will be used as input to Proposition 3.4 to produce cohomology classes of known index. The proof, which involves an application of the Chebotarëv density theorem, follows a strategy similar to that used in the proof of Kolyvagin's theorem on page 135 of [18].

Proposition 4.4. *Let p be a prime that is firm for E, and let $x \in E(K)^{\pm}$. Then there is a set of primes ℓ of positive Dirichlet density such that* $\mathrm{Frob}_\ell\big(K(E[p^n])/\mathbb{Q}\big) = [\tau]$ *and the orders of the images of x in $E(K)/p^nE(K)$ and in $E(K_\ell)/p^nE(K_\ell)$ are the same.*

Proof. Let p^a be the order of the image of x in $E(K)/p^nE(K)$. If $a = 0$, then there is nothing to prove, so assume that $a > 0$. If ℓ is a prime such that the orders of the images of $p^{a-1}x$ in $E(K)/p^nE(K)$ and $E(K_\ell)/p^nE(K_\ell)$ both equal p, then the images of x in $E(K)/p^nE(K)$ and $E(K_\ell)/p^nE(K_\ell)$ both have order p^a. It thus suffices to prove the proposition in the case when the order of the image of x in $E(K)/p^nE(K)$ is p.

Let $L = K(E[p^n])$, suppose ℓ is a prime such that $\mathrm{Frob}_\ell(L/\mathbb{Q}) = [\tau]$, and let λ be one of the prime ideals of L that lies over ℓ. We have a diagram

$$\begin{tikzcd}
E(K)/p^nE(K) \arrow[r, hook] \arrow[d] & H^1(K, E[p^n]) \arrow[r, hook] \arrow[d] & \mathrm{Hom}(G_L, E[p^n]) \arrow[d] \\
E(K_\ell)/p^nE(K_\ell) \arrow[r] & H^1(K_\ell, E[p^n]) \arrow[r] & \mathrm{Hom}(G_{L_\lambda}, E[p^n]).
\end{tikzcd}$$

Let $\varphi\colon G_L \to E[p^n]$ be the element of $\mathrm{Hom}(G_L, E[p^n])$ that x maps to. The top row is injective, because p is firm, so it suffices to show that the image φ_ℓ of φ in $\mathrm{Hom}(G_{L_\lambda}, E[p^n])$ is nonzero.

Let M be the fixed field of the kernel of φ. Since M is the compositum of the two Galois extensions $K([p^n]^{-1}x)$ and $\mathbb{Q}(E[p^n])$ of $\mathbb{Q}$, it is also Galois (see Lemma 4.1). Because $\mathrm{Frob}_\ell(M/\mathbb{Q})|_L = [\tau]$, there is an element $\sigma \in \mathrm{Gal}(M/L)$ such that

$$\mathrm{Frob}_\ell(M/\mathbb{Q}) = [\sigma\tau].$$

The order of $\sigma\tau$ equals the degree of $M_{\lambda'}$ over $\mathbb{Q}_\ell$, where λ' is a prime of M lying over ℓ. If $\varphi_\ell = 0$, then $M_{\lambda'} = L_\lambda = K_\ell$, so $\sigma\tau$ would have order 2.

The image of φ is a nonzero subgroup H of $E[p^n]$, which is defined over $\mathbb{Q}$ since $x \in E(K)^{\pm}$. If every $\sigma \in \mathrm{Gal}(M/L)$ has the property that $\sigma\tau$ has order 2, then $H \subset E[p^n]^-$. This contradicts our assumption that p is firm, since H is a nontrivial cyclic subgroup of $E[p^\infty]$. Thus there exists $\sigma \in \mathrm{Gal}(M/L)$ such that $\sigma\tau$ has order different than 2. For this σ and for any prime ℓ such that $\mathrm{Frob}_\ell(M/\mathbb{Q}) = [\sigma\tau]$, we see that $\varphi_\ell \neq 0$. The Chebotarëv density theorem provides a positive density of such ℓ. □

4.2. Proof of Theorem 3.1. Let E be an elliptic curve over $\mathbb{Q}$ such that $L(E,1) \neq 0$. Let K be one of the infinitely many imaginary quadratic fields such that the associated Heegner point y has infinite order. Let B_K be an integer that is divisible by 2 and

- the primes p such that $y \in pE(K)$,
- the primes p that are not firm,
- the order $\#E(K)_{\mathrm{tor}}$, and
- the class number of K.

If $K = \mathbb{Q}(\sqrt{-3})$, assume in addition that 3 divides B_K.

Fix a prime $p \nmid B_K$. Since $E(K)$ has rank 1 (see, e.g., [9], Thm. 1.3) and $p \nmid \#E(K)_{\mathrm{tor}}$, the image of y in $E(K)/p^nE(K)$ has order p^n. By Proposition 4.4 there are infinitely many primes ℓ such that $\mathrm{Frob}_\ell\big(K(E[p^n])/\mathbb{Q}\big) = [\tau]$ and the image of y in $E(K_\ell)/p^nE(K_\ell)$ has order p^n. For these ℓ, Proposition 4.4 produces infinitely many cohomology classes c_{ℓ,p^n} having order and index both equal to p^n. (Note that if $\ell \neq \ell'$ then $c_{\ell,p^n} \neq c_{\ell',p^n}$.)

Let B be the greatest common divisor of the set of integers B_K, as K varies over all quadratic imaginary extensions such that the associated Heegner point has infinite order. For each prime power p^n that does not divide B, we have produced infinitely many $c \in H^1(\mathbb{Q}, E)$ having order and index both equal to p^n. If the orders of c and c' are coprime, then $\operatorname{ord}(c + c') = \operatorname{ord}(c) \cdot \operatorname{ord}(c')$ and, by Proposition 2.5,

$$\operatorname{ind}(c + c') = \operatorname{ind}(c) \cdot \operatorname{ind}(c').$$

This proves the theorem. □

5. Computing the bound B_K

In this section we compute, in some cases, the the bound B_K that appears in Section 4.2. First we prove a general theorem about semistable elliptic curves. Next we compute the index of a Heegner point, and finally in Section 5.4 we prove Theorem 1.1.

5.1. Galois representations attached to isolated curves. The following proposition sometimes permits us to compute the integer B_K, which appears in Section 4.2.

Proposition 5.1. *Let E be a semistable elliptic curve over $\mathbb{Q}$ of conductor N, let p be an odd prime, and let K be a quadratic imaginary field such that* $\gcd(D_K, pN) = 1$. *Assume that $p \nmid \operatorname{ord}_\ell(j(E))$, for each prime $\ell | N$, and that E admits no isogenies of degree p. Then $p \nmid \# E(K)_{\text{tor}}$ and p is firm for E.*

Before giving the proof, we summarize its main ingredients. First, we observe that the assertion that $m(p) = 0$ (see Definition 3.3) uses a standard result that relates unramified Galois cohomology to component groups. Next, we use the semistability and isogeny hypotheses to deduce that $\rho_{E,p}$ is surjective. Then we use standard group cohomology to deduce that p is firm.

Proof. Let ℓ be a prime. By [14], I.3.8,

$$H^1(\mathbb{Q}_\ell^{\text{unr}}/\mathbb{Q}_\ell, E(\mathbb{Q}_\ell^{\text{unr}})) \cong H^1(\overline{\mathbb{F}}_\ell/\mathbb{F}_\ell, \Phi_{E,\ell}(\overline{\mathbb{F}}_\ell)),$$

where $\Phi_{E,\ell}$ is the component group of E at ℓ. If $\ell \nmid N$, there is nothing further to prove, so assume $\ell | N$. Since E is semistable, $\# \Phi_{E,\ell}(\overline{\mathbb{F}}_\ell) = -\operatorname{ord}_\ell(j)$. By hypothesis, $p \nmid \operatorname{ord}_\ell(j)$. Thus $m(p) = 0$.

Since E admits no isogenies of degree p, the Galois representation

$$\rho_{E,p}: G_{\mathbb{Q}} \to \mathrm{GL}(2, E[p])$$

is irreducible, and there are no nontrivial $\mathbb{Q}$-rational cyclic subgroups of $E[p^\infty]$. Since E is semistable, work of Serre [19], Prop. 21 and [20], §3.1 implies that $\rho_{E,p}$ is surjective. Thus $p \nmid \# E(K)_{\text{tor}}$ because a point in $E(\overline{\mathbb{Q}})$ of order p must generate an extension of $\mathbb{Q}$ of degree at least $p^2 - 1 \geqq 3$.

The field K and $\mathbb{Q}(E[p])$ are linearly disjoint, since $\gcd(D_K, pN) = 1$, so

$$H^1\big(K(E[p])/K, E[p]\big) \cong H^1\big(\mathbb{Q}(E[p])/\mathbb{Q}, E[p]\big) \approx H^1\big(\mathrm{GL}(2, \mathbb{F}_p), \mathbb{F}_p^2\big).$$

The group $H = H^1\big(K(E[p^n])/K, E[p^n]\big)$ has exponent a power of p. If an element α in $\mathrm{Gal}\big(K(E[p^n])/K\big) \subset \mathrm{GL}_2(\mathbb{Z}/p^n\mathbb{Z})$ is scalar, then every element of H has order dividing $\alpha - 1$. This is because the scalar is central, so the morphism of pairs it induces is both the identity and multiplication by α. It is necessary only to choose α such that $\gcd(\alpha - 1, p) = 1$. Since p is odd, -1 is a nonidentity element of

$$\mathrm{Aut}(E[p]) = \mathrm{Gal}\big(K(E[p])/K\big).$$

Every automorphism lifts, so -1 lifts to some g in $\mathrm{Gal}\big(K(E[p^n])/K\big) \subset \mathrm{Aut}(E[p^n])$. Then $g^{p^{n-1}} = -1$ in $\mathrm{Aut}(E[p^n])$, so $-1 \in \mathrm{Gal}\big(K(E[p^n])/K\big)$ and every element of H has order dividing 2. (To show that $g^{p^{n-1}} = -1$, we use that $\mathrm{ord}_p\begin{pmatrix} p^n \\ k \end{pmatrix} = n + \mathrm{ord}_p\left(\frac{1}{k}\right)$.) □

5.2. The number B_K for $X_0(17)$. In this section, we show that for $E = X_0(17)$ and $K = \mathbb{Q}(\sqrt{-2})$, we have $B_K = 2$. This is accomplished by showing that the index $[E(K) : \mathbb{Z}y]$ is a power of 2. The elliptic curve $E = X_0(17)$ given by the Weierstrass equation

$$y^2 + xy + y = x^3 - x^2 - x - 14$$

satisfies the hypothesis of Proposition 5.1 for each odd prime p. Since the j-invariant of E is $3^3 \cdot 11^3/17^4$, every odd prime p is firm for E and $\# E(K)_{\mathrm{tor}}$ is a power of 2.

The conductor 17 of E splits in K, and the quadratic twist E' of E by K is the curve $y^2 = x^3 - 44x + 7120$, which is labeled **1088K4** in [7]. Using MAGMA (or `mwrank`), one finds that $E'(\mathbb{Q}) \cong \mathbb{Z}P \times \mathbb{Z}/2$, where $P = (-3, 85) \in E'(\mathbb{Q})$ has infinite order. Since the rank of E' is 1, we set $K = \mathbb{Q}(\sqrt{-2})$ in Section 4.2. Then B_K is divisible only by 2 and the index $[E(K) : \mathbb{Z}y]$. This index can only change by a power of 2 if y is replaced by y_K, so we instead consider the index $[E(K) : \mathbb{Z}y_K]$. The cokernel of the natural map $E(\mathbb{Q}) \oplus E'(\mathbb{Q}) \to E(K)$ is a 2-group and $E(\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z}$, so $[E(K) : \mathbb{Z}y_K]$ is a power of 2 times $h(y_K)/h(P)$, where h is the Néron-Tate canonical height on E_K. By the Gross-Zagier formula (see [8], Thm. 6.3),

$$h(y_K) = \frac{u^2|D|^{\frac{1}{2}}}{\|\omega_f\|} L'_{E'}(1) L_E(1),$$

where $D = -8$ is the discriminant of K, $u = 1$ is half the number of units, and $\|\omega_f\|$ is the Peterson norm of the newform f corresponding to E. Generators for the period lattice of E are $\omega_1 \sim 1.547079$ and $\omega_2 \sim 0.773539 + 1.372869i$; taking the determinant gives $\|\omega_f\| \sim 2.123938$. Furthermore, again from [7], we find that $L_E(1) \sim 0.386769$ and $L'_{E'}(1) \sim 2.525026$, so $h(y_K) \sim 1.300533$. Using a computer, we find that $h(P) \sim 1.300533$ as well, so $[E(K) : \mathbb{Z}y_K]$ is a power of two.

5.3. Elements of index 2 and 4. The torsion subgroup of $E = X_0(17)$ is isomorphic to $\mathbb{Z}/4\mathbb{Z}$, so [11], pg. 670 implies that there are infinitely many elements of $H^1(\mathbb{Q}, E)$ having order and index equal to 2, and also infinitely many having order and index equal to 4.

5.4. Proof of Theorem 1.1. To prove Theorem 1.1, we combine the above computations with Theorem 3.1, and an observation about the local properties of Kolyvagin's classes c_{ℓ, p^n}.

Proof of Theorem 1.1. Let $E = X_0(17)$ as above, and let K be an arbitrary number field. Let p^n be either an odd prime power, or 2, or 4. The computations of the previous section combined with Theorem 3.1 prove that there are infinitely many elements c_{ℓ,p^n} of $H^1(\mathbb{Q}, E)$ whose index and order both equal p^n. Let A be the subgroup of $H^1(\mathbb{Q}, E)$ generated by these classes. The kernel B of $\mathrm{res}_K\colon A \to H^1(K, E)$ is finite, so the set $\mathscr{S}$ of primes ℓ such that $\mathrm{res}_\ell(c) \neq 0$ for some $c \in B$ is finite. By Proposition 3.4, we have $\mathrm{res}_v(c_{\ell,p^n}) = 0$ for all places $v \neq \ell$, so the subgroup A' of A generated by all c_{ℓ,p^n} with $\ell \notin \mathscr{S}$ has trivial intersection with B. Thus $\mathrm{res}_K(A')$ consists of infinitely many classes in $H^1(K, E)$ having order and index both equal to p^n, and the theorem now follows from Proposition 2.3. □

References

[1] *C. Breuil*, *B. Conrad*, *F. Diamond*, and *R. Taylor*, On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises, J. Amer. Math. Soc. **14** (2001) no. 4, 843–939.

[2] *D. Bump*, *S. Friedberg*, and *J. Hoffstein*, Eisenstein series on the metaplectic group and nonvanishing theorems for automorphic L-functions and their derivatives, Ann. Math. (2) **131** (1990), no. 1, 53–127.

[3] *J. W. S. Cassels*, Arithmetic on curves of genus 1. IV. Proof of the Hauptvermutung, J. reine angew. Math. **211** (1962), 95–112.

[4] *J. W. S. Cassels*, Arithmetic on curves of genus 1. V. Two counterexamples, J. London Math. Soc. **38** (1963), 244–248.

[5] *J. W. S. Cassels*, Diophantine equations with special reference to elliptic curves, J. London Math. Soc. **41** (1966), 193–291.

[6] *J. E. Cremona*, Algorithms for modular elliptic curves, second ed., Cambridge University Press, Cambridge 1997.

[7] *J. E. Cremona*, Elliptic curves of conductor $\leqq$ 12000, `http://www.maths.nott.ac.uk/personal/jec/ftp/data/`.

[8] *B. Gross* and *D. Zagier*, Heegner points and derivatives of L-series, Invent. Math. **84** (1986), no. 2, 225–320.

[9] *B. H. Gross*, Kolyvagin's work on modular elliptic curves, L-functions and arithmetic (Durham 1989), Cambridge Univ. Press, Cambridge (1991), 235–256.

[10] *V. A. Kolyvagin*, On the structure of Shafarevich-Tate groups, Algebraic geometry (Chicago, IL, 1989), Springer, Berlin (1991), 94–121.

[11] *S. Lang* and *J. Tate*, Principal homogeneous spaces over abelian varieties, Amer. J. Math. **80** (1958), 659–684.

[12] *S. Lichtenbaum*, Duality theorems for curves over p-adic fields, Invent. Math. **7** (1969), 120–136.

[13] *W. G. McCallum*, Kolyvagin's work on Shafarevich-Tate groups, L-functions and arithmetic (Durham 1989), Cambridge Univ. Press, Cambridge (1991), 295–316.

[14] *J. S. Milne*, Arithmetic duality theorems, Academic Press Inc., Boston, Mass., 1986.

[15] *M. R. Murty* and *V. K. Murty*, Non-vanishing of L-functions and applications, Birkhäuser Verlag, Basel 1997.

[16] *C. O'Neil*, The Period-Index Obstruction for Elliptic Curves, J. Number Th., to appear.

[17] *K. A. Ribet* and *W. A. Stein*, Lectures on Serre's conjectures, IAS/Park City Math. Ser. **9** (2001).

[18] *K. Rubin*, The work of Kolyvagin on the arithmetic of elliptic curves, Arithmetic of complex manifolds (Erlangen 1988), Springer, Berlin (1989), 128–136.

[19] *J.-P. Serre*, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, Invent. Math. **15** (1972), no. 4, 259–331.

[20] *J.-P. Serre*, Travaux de Wiles (et Taylor, ...). I, Astérisque **237**, Exp. No. 803, 5 (1996), 319–332, Séminaire Bourbaki, Vol. 1994/95.

[21] *I. R. Shafarevich*, Exponents of elliptic curves, Dokl. Akad. Nauk SSSR (N.S.) **114** (1957), 714–716.

e-mail: was@math.harvard.edu

Eingegangen 6. November 2000, in revidierter Fassung 29. Juni 2001

6 Cuspidal Modular Symbols Are Transportable, with H. Verrill

London Mathematical Society ISSN 1461–1570

CUSPIDAL MODULAR SYMBOLS ARE TRANSPORTABLE

WILLIAM A. STEIN AND HELENA A. VERRILL

Abstract

Modular symbols of weight 2 for a congruence subgroup Γ satisfy the identity $\{\alpha, \gamma(\alpha)\} = \{\beta, \gamma(\beta)\}$ for all α, β in the extended upper half plane and $\gamma \in \Gamma$. The analogue of this identity is false for modular symbols of weight greater than 2. This paper provides a definition of transportable modular symbols, which are symbols for which an analogue of the above identity holds, and proves that every cuspidal symbol can be written as a transportable symbol. As a corollary, an algorithm is obtained for computing periods of cuspforms.

Introduction

It is well known that modular symbols of weight 2 for a congruence subgroup Γ satisfy the identity $\{\alpha, \gamma(\alpha)\} = \{\beta, \gamma(\beta)\}$ for all α, β in the extended upper half plane and $\gamma \in \Gamma$. The analogue of this identity is, in general, false for modular symbols of weight greater than 2. To investigate further, we define transportable modular symbols, which are symbols that can be expressed in such a way that an analogue of the above identity holds. We then prove that every cuspidal symbol is transportable. As a corollary, we obtain an algorithm for computing periods of cuspforms.

In Section 1 we review the definition of modular symbols. In Section 2 we define transportable modular symbols, and prove our main theorem. Section 3 contains an application of our transportability result to the computation of periods of modular forms. Finally, Section 4 contains two examples in which we verify the assertion of Theorem 2.4 and apply the period computation algorithm.

1. *Modular symbols*

In Section 1.1 we recall the definition of modular symbols given in [**5**]; then in Section 1.2 we introduce a slight generalization of the definition. Let N and k be positive integers with $k \geqslant 2$, and let $\varepsilon : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be a Dirichlet character modulo N.

1.1. *Definition*

Let $\mathcal{M}$ be the abelian group generated by all symbols $\{\alpha, \beta\}$ with $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$, modulo the relations $\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$, and modulo any torsion. Let V_{k-2} denote the group of homogeneous polynomials in $\mathbb{Z}[X, Y]$ of degree $k - 2$.

Received 22 March 2001, revised 29 August 2001*; published 24 September 2001*.
2000 Mathematics Subject Classification 20C33, 20C44, 20G05, 20G40
© 2001, William A. Stein and Helena A. Verrill

Each element $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z})$ acts on the left on V_{k-2} by

$$\gamma(P(X, Y)) = P(dX - bY, -cX + aY),$$

and on $\mathcal{M}_k = V_{k-2} \otimes \mathcal{M}$ by

$$\gamma(P \otimes \{\alpha, \beta\}) = \gamma(P) \otimes \{\gamma(\alpha), \gamma(\beta)\}.$$

Fix a Dirichlet character $\varepsilon : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, and denote by $\mathbb{Z}[\varepsilon]$ the ring generated by the image of ε. We also view ε as a homomorphism $\Gamma_0(N) \to \mathbb{C}^*$ by setting $\varepsilon \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \varepsilon(d)$.

The space $\mathcal{M}_k(N, \varepsilon)$ of *modular symbols* of level N and character ε is the quotient of the $\mathbb{Z}[\varepsilon]$-module $\mathcal{M}_k \otimes \mathbb{Z}[\varepsilon]$ by the $\mathbb{Z}[\varepsilon]$-submodule generated by $\gamma(x) - \varepsilon(\gamma)x$ for all $x \in \mathcal{M}_k$, for all $\gamma \in \Gamma_0(N)$, and by any torsion. Denote by $P\{\alpha, \beta\}$ the image of $P \otimes \{\alpha, \beta\}$ in $\mathcal{M}_k(N, \varepsilon)$. The $\mathbb{Q}[\varepsilon]$-vector space

$$\mathcal{M}_k(N, \varepsilon; \mathbb{Q}) = \mathcal{M}_k(N, \varepsilon) \otimes_{\mathbb{Z}} \mathbb{Q}$$

contains $\mathcal{M}_k(N, \varepsilon)$.

Let $\mathcal{B}$ be the free abelian group generated by all symbols $\{\alpha\}$, for $\alpha \in \mathbb{P}^1(\mathbb{Q})$. Define a left action of $\mathrm{SL}(2, \mathbb{Z})$ on $\mathcal{B}_k = V_{k-2} \otimes \mathcal{B}$ by

$$\gamma(P \otimes \{\alpha\}) = \gamma(P) \otimes \{\gamma\alpha\}.$$

The space $\mathcal{B}_k(N, \varepsilon)$ of *boundary symbols* is the quotient of $\mathcal{B}_k \otimes \mathbb{Z}[\varepsilon]$ by the $\mathbb{Z}[\varepsilon]$-submodule generated by $\gamma(x) - \varepsilon(\gamma)x$ for all $x \in \mathcal{B}_k$, for all $\gamma \in \Gamma_0(N)$, and by any torsion. The subspace $\mathcal{S}_k(N, \varepsilon)$ of *cuspidal symbols* is the kernel of the map $\delta : \mathcal{M}_k(N, \varepsilon) \to \mathcal{B}_k(N, \varepsilon)$ given by $\delta(P\{\alpha, \beta\}) = P\{\beta\} - P\{\alpha\}$.

When $\varepsilon = 1$ is the trivial character, we shall also write $\mathcal{M}_k(\Gamma_0(N))$ for $\mathcal{M}_k(N, 1)$, and similarly for $\mathcal{S}_k$ and $\mathcal{B}_k$.

1.2. *Extended modular symbols*

It is useful to extend the notion of modular symbols to allow symbols of the form $P\{z, w\}$ where z and w are arbitrary elements of $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q})$.

Definition 1 (Extended modular symbols). The group $\overline{\mathcal{M}}_k$ of *extended modular symbols* is the free abelian group with basis the set of all symbols $P\{z, w\}$ with $z, w \in \mathfrak{h}^*$, subject to the relations $P\{u, v\} + P\{v, w\} + P\{w, u\} = 0$.

Note that $\overline{\mathcal{M}}_k$ is of uncountable rank over $\mathbb{Z}$. It is equipped with an action of $\Gamma_0(N)$; we let $\overline{\mathcal{M}}_k(N, \varepsilon)$ be the largest torsion-free quotient of $\overline{\mathcal{M}}_k$ by the relations $\gamma x = \varepsilon(\gamma)x$ for $\gamma \in \Gamma_0(N)$.

2. *Transportable modular symbols*

In Section 2.1 we define transportable modular symbols, and we prove an elementary proposition that motivates the definition. Section 2.2, which is the heart of this paper, contains a proof that every cuspidal modular symbol is transportable.

2.1. *Definition*

Definition 2 (Transportable). A modular symbol is *transportable* if it can be written in the form

$$\sum_{i=1}^{m} P_i\{\infty, \gamma_i(\infty)\},$$

for $\gamma_i \in \Gamma_0(N)$ and $P_i \in V_{k-2}$ with

$$\sum_{i=1}^{m} P_i\{\infty, \gamma_i(\infty)\} = \sum_{i=1}^{m} P_i\{\alpha, \gamma_i(\alpha)\}$$

for all $\alpha \in \mathfrak{h}^*$, where the equality takes place in $\overline{\mathcal{M}}_k(N, \varepsilon)$.

When $k = 2$, the identity $\{\infty, \gamma(\infty)\} = \{\alpha, \gamma(\alpha)\}$ holds for any $\alpha \in \mathfrak{h}^*$, so in weight 2 there is a plentiful supply of transportable modular symbols.

Proposition 2.1. *For any $\gamma \in \Gamma_0(N)$, $P \in V_{k-2}$ and $\alpha \in \mathfrak{h}^*$,*

$$\begin{aligned} P\{\infty, \gamma(\infty)\} &= P\{\alpha, \gamma(\alpha)\} + \big(P - \varepsilon(\gamma)\gamma^{-1}P\big)\{\infty, \alpha\} \\ &= \varepsilon(\gamma)\big(\gamma^{-1}P\big)\{\alpha, \infty\} - P\{\gamma(\alpha), \infty\}. \end{aligned} \tag{1}$$

In particular,

$$P\{\infty, \gamma(\infty)\} = P\{\alpha, \gamma(\alpha)\} \Leftrightarrow P = \varepsilon(\gamma)\gamma^{-1}P. \tag{2}$$

Proof. If $x \in \mathcal{M}_k(N, \varepsilon)$ is a modular symbol and $\gamma \in \Gamma_0(N)$, then $\gamma x = \varepsilon(\gamma)x$, where, as usual, ε is viewed as a homomorphism $\Gamma_0(N) \to \mathbb{C}^*$ via $\varepsilon\big(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\big) = \varepsilon(d)$. In particular, $\varepsilon(\gamma)\gamma^{-1}x = x$, so

$$\begin{aligned} P\{\infty, \gamma(\infty)\} &= P\{\infty, \alpha\} + P\{\alpha, \gamma(\alpha)\} + P\{\gamma(\alpha), \gamma(\infty)\} \\ &= P\{\infty, \alpha\} + P\{\alpha, \gamma(\alpha)\} + \varepsilon(\gamma)\gamma^{-1}(P\{\gamma(\alpha), \gamma(\infty)\}) \\ &= P\{\infty, \alpha\} + P\{\alpha, \gamma(\alpha)\} + \varepsilon(\gamma)\big(\gamma^{-1}P\big)\{\alpha, \infty\} \\ &= P\{\alpha, \gamma(\alpha)\} + P\{\infty, \alpha\} - \varepsilon(\gamma)\big(\gamma^{-1}P\big)\{\infty, \alpha\} \\ &= P\{\alpha, \gamma(\alpha)\} + \big(P - \varepsilon(\gamma)\gamma^{-1}P\big)\{\infty, \alpha\}. \end{aligned}$$

The remaining statements of the proposition now follow easily. □

Example 2.2. In some cases it is easy to give a formula for symbols that are obviously transportable. Suppose that $k \geqslant 2$ is an even integer. If P is a polynomial such that $\gamma(P) = P$ for some $\gamma \in \Gamma_0(N)$, then $P\{\infty, \gamma(\infty)\}$ is transportable. Given $\gamma \in \Gamma_0(N)$, an example of such a P is

$$P(X, Y) = \big(cX^2 + (d-a)XY - bY^2\big)^{(k-2)/2}.$$

We found this polynomial by viewing V_{k-2} as the $(k-2)$th symmetric product of the 2-dimensional space on which $\Gamma_0(N)$ acts naturally. If γ, which has determinant 1, has eigenvalues α and α^{-1}, then the eigenvalues of the $(k-2)$-fold symmetric product of γ are given by α^{k-2-2j} for $0 \leqslant j \leqslant k-2$. Although we have not been able to find a counterexample, the authors see no reason to believe that transportable symbols of the form given in this example always span $\mathcal{S}_k(N; \mathbb{Q})$.

More generally, given any sequence of matrices $\gamma_1, \ldots, \gamma_n$ in $\Gamma_0(N)$, it is a simple matter of linear algebra to give transportable symbols of the form $\sum_{i=1}^{n} P_i\{\infty, \gamma_i\infty\}$. This follows from Lemma 2.3, which shows that this symbol is transportable exactly when $(P_1, \ldots, P_n)$ is in the kernel of the map $\bigoplus_{i=1}^{n}(1 - \gamma_i^{-1})$ from $\bigoplus_{i=1}^{n} V_{k-2}$ to V_{k-2}.

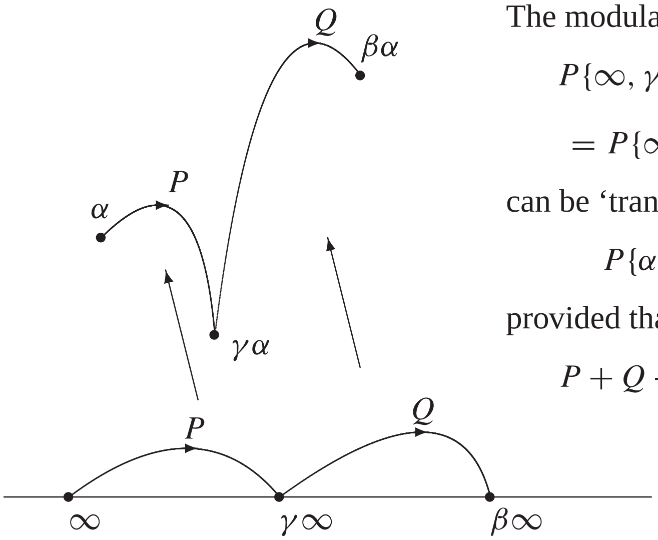
2.2. *Characterization of transportable modular symbols*

Lemma 2.3. *A modular symbol in $\mathcal{M}_k(N, \varepsilon; \mathbb{Q})$ is transportable if and only if it can be written in the form $\sum_{i=1}^m P_i\{\infty, \gamma_i(\infty)\}$ with*

$$\sum P_i = \sum \varepsilon(\gamma_i)\gamma_i^{-1}P_i.$$

Proof. This follows from Proposition 2.1. □

Figure 1 illustrates Lemma 2.3 with a trivial-character example.



The modular symbol

$$P\{\infty, \gamma\infty\} + Q\{\gamma\infty, \beta\infty\}$$

$$= P\{\infty, \gamma\infty\}+Q\{\infty, \beta\infty\}-Q\{\infty, \gamma\infty\}$$

can be 'transported' to

$$P\{\alpha, \gamma\alpha\} + Q\{\gamma\alpha, \beta\alpha\},$$

provided that

$$P + Q - Q = \gamma^{-1}P + \beta^{-1}Q - \gamma^{-1}Q.$$

Figure 1: 'Transporting' a transportable modular symbol.

Theorem 2.4. *A modular symbol is transportable if and only if it is cuspidal.*

Proof. By Lemma 2.3, every transportable modular symbols is cuspidal, so we must prove that every cuspidal symbol is transportable.

Let $I = I_{N,\varepsilon}$ be the ideal in the group ring of $\Gamma_0(N)$ generated by all elements of the form $\varepsilon(\gamma) - \gamma$ for $\gamma \in \Gamma_0(N)$. Suppose that $v \in \mathcal{S}_k(N, \varepsilon)$. Use the relation $\{\alpha, \beta\} = \{\infty, \beta\} - \{\infty, \alpha\} \in \mathcal{M}$ to see that any v is the image of an element $\tilde{v} \in \mathcal{M}_k$ of the form

$$\tilde{v} = \sum_{\beta\in\mathbb{Q}} P_\beta \otimes \{\infty, \beta\} \in \mathcal{M}_k$$

with only finitely many P_β nonzero. For later convenience, we set $P_\infty = 0$, and take sums over all $\beta \in P^1(\mathbb{Q})$. The boundary map δ lifts in a natural way to $\mathcal{M}_k = V_{k-2} \otimes \mathcal{M}$, as illustrated.

$$\begin{tikzcd}
 & I(V_{k-2} \otimes \mathcal{M}) \arrow[r] \arrow[d] & I(V_{k-2} \otimes \mathcal{B}) \arrow[d] \\
 & V_{k-2} \otimes \mathcal{M} \arrow[r, "\tilde{\delta}"] \arrow[d] & V_{k-2} \otimes \mathcal{B} \arrow[d] \\
\mathcal{S}_k(N, \varepsilon) \arrow[r, hook] & \mathcal{M}_k(N, \varepsilon) \arrow[r, "\delta"] & \mathcal{B}_k(N, \varepsilon)
\end{tikzcd}$$

Bearing in mind torsion, our assumption that $\delta(v) = 0$ implies that for some nonzero $M \in \mathbb{Z}$, we have $M\tilde{\delta}(\tilde{v}) \in I(V_{k-2} \otimes \mathcal{B})$. So there are $Q_{\gamma,\beta} \in V_{k-2}$, for $\gamma \in \Gamma_0(N)$ and $\beta \in \mathbb{P}^1(\mathbb{Q})$, only finitely β many nonzero, such that

$$M\tilde{\delta}(\tilde{v}) = \sum_{\gamma,\beta} \big(\varepsilon(\gamma) - \gamma\big)\big(Q_{\gamma,\beta} \otimes \{\beta\}\big).$$

We now use a summation trick.

$$\begin{aligned} M\tilde{\delta}(\tilde{v}) &= M\sum_{\beta}\big(P_\beta \otimes \{\beta\} - P_\beta \otimes \{\infty\}\big) \\ &= \sum_{\gamma,\beta}\big(\varepsilon(\gamma)Q_{\gamma,\beta} \otimes \{\beta\} - (\gamma Q_{\gamma,\beta}) \otimes \{\gamma\beta\}\big) \\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta} \otimes \{\beta\} - (\gamma Q_{\gamma,\gamma^{-1}\beta}) \otimes \{\beta\} \\ &= \sum_{\gamma,\beta}\big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\big) \otimes \{\beta\}. \end{aligned}$$

This shows that

$$M\sum_{\beta}\big(P_\beta \otimes \{\beta\} - P_\beta \otimes \{\infty\}\big) = \sum_{\gamma,\beta}\big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\big) \otimes \{\beta\}. \tag{3}$$

Equating terms, we deduce that for $\beta \neq \infty$,

$$MP_\beta = \sum_{\gamma}\big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\big). \tag{4}$$

Using this expression for P_β, as well as the fact that $\varepsilon(\gamma)\gamma^{-1}$ acts trivially on $\mathcal{M}_k(N,\varepsilon)$, we find that

$$\begin{aligned} Mv = M\sum_{\beta}P_\beta\{\infty,\beta\} &= \sum_{\gamma,\beta}\big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\big)\{\infty,\beta\} \\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta} - \varepsilon(\gamma)\gamma^{-1}\big((\gamma Q_{\gamma,\gamma^{-1}\beta})\{\infty,\beta\}\big) \\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty,\beta\} - \varepsilon(\gamma)Q_{\gamma,\gamma^{-1}\beta}\big\{\gamma^{-1}\infty,\gamma^{-1}\beta\big\} \\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty,\beta\} - \varepsilon(\gamma)Q_{\gamma,\beta}\big\{\gamma^{-1}\infty,\beta\big\} \\ &= \sum_{\gamma,\beta}\varepsilon(\gamma)Q_{\gamma,\beta}\big\{\infty,\gamma^{-1}\infty\big\}. \end{aligned} \tag{5}$$

Equating coefficients of $\{\infty\}$ in Equation 3, we have

$$-M\sum_{\beta}P_\beta = \sum_{\gamma}\big(\varepsilon(\gamma)Q_{\gamma,\infty} - \gamma Q_{\gamma,\gamma^{-1}\infty}\big),$$

which, combining with Equation 4, and recalling that $P_\infty = 0$, means that

$$-\sum_{\gamma,\beta\neq\infty}\big(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\gamma^{-1}\beta}\big) = \sum_{\gamma}\big(\varepsilon(\gamma)Q_{\gamma,\infty} - \gamma Q_{\gamma,\gamma^{-1}\infty}\big),$$

and hence

$$\sum_{\gamma,\beta}\bigl(\varepsilon(\gamma)Q_{\gamma,\beta} - \gamma Q_{\gamma,\beta}\bigr) = 0.$$

Using the expression

$$v = -\tfrac{1}{M}\sum_{\beta,\gamma}\varepsilon(\gamma)Q_{\gamma,\beta}\{\infty, \gamma^{-1}\infty\}$$

obtained from Equation 5, we see that this is the condition for v to be transportable. □

Corollary 2.5. *Fix $\alpha \in \mathfrak{h}^*$. Every element of $\mathcal{S}_k(N,\varepsilon)$ is a sum of modular symbols of the form $P\{\alpha, \gamma(\alpha)\}$.*

Proof. Let $x \in \mathcal{S}_k(N,\varepsilon)$. Proposition 2.1 implies that x is transportable, so there exist P_i and γ_i such that

$$x = \sum P_i\{\infty, \gamma_i(\infty)\} = \sum P_i\{\beta, \gamma_i(\beta)\}$$

for any $\beta \in \mathfrak{h}^*$. Taking $\beta = \alpha$ proves the corollary. □

Remark 2.6.

1. When $k = 2$, the corollary follows from [4, Section 1], which asserts that map $\Gamma_0(N) \to \mathcal{S}_2(\Gamma_0(N)) = H_1(X_0(N),\mathbb{Z})$ sending γ to $\{\alpha, \gamma(\alpha)\}$ is a surjective group homomorphism.
2. In Proposition 2.7 below, we shall prove more generally that every element of $\mathcal{M}_k(N,\varepsilon)$ is a sum of modular symbols of the form $P\{\alpha, \gamma(\alpha)\}$, as long as we allow α to vary over $\mathbb{P}^1(\mathbb{Q})$.

2.3. *What space do the symbols $P\{\infty, \gamma(\infty)\}$ span?*

Suppose that N and k are positive integers, with k even.

Definition 3. For any $\alpha \in \mathbb{P}^1(\mathbb{Q})$, let $\mathcal{W}_\alpha$ denote the subspace of $\mathcal{M}_k(\Gamma_0(N);\mathbb{Q})$ spanned by symbols of the form $P\{\alpha, \gamma(\alpha)\}$, for $P \in V_{k-2}$ and $\gamma \in \Gamma_0(N)$.

Corollary 2.5 draws our attention to $\mathcal{W}_\infty$. Since $\mathcal{W}_\infty$ contains $\mathcal{S}_k(\Gamma_0(N))$, it is natural to ask how much bigger it is. As mentioned in Remark 2.6, when $k = 2$, Manin proved that for any $\alpha \in \mathbb{P}^1(\mathbb{Q})$, we have $\mathcal{W}_\alpha = \mathcal{W}_\infty = \mathcal{S}_2(\Gamma_0(N);\mathbb{Q})$. We now compute $\mathcal{W}_\alpha$ for any weight $k > 2$.

Proposition 2.7. *Suppose that $k > 2$. Then the space $\mathcal{W}_\alpha$ is equal to the inverse image under the boundary map δ of the one-dimensional subspace $V_{k-2}\{\alpha\} \subset \mathcal{B}_k(\Gamma_0(N);\mathbb{Q})$. Hence* $\dim \mathcal{W}_\alpha = \dim \mathcal{S}_k(\Gamma_0(N);\mathbb{Q}) + 1$ *and* $\mathcal{M}_k(\Gamma_0(N);\mathbb{Q}) = \sum_{\alpha\in\mathbb{P}^1(\mathbb{Q})}\mathcal{W}_\alpha$.

Proof. In [5, Section 1.4], Merel shows that $V_{k-2}\{\alpha\}$ has dimension 1 (see the proof of [5, Section 1.4, Proposition 4]), and that $P(X,Y)\{u/v\}$ is nonzero if $P(u,v) \neq 0$.

Corollary 2.5 implies that $\mathcal{W}_\alpha$ contains the kernel $\mathcal{S}_k(\Gamma_0(N))$ of the boundary map δ. It thus suffices to show that $\delta(\mathcal{W}_\alpha) = V_{k-2}\{\alpha\}$. For $P \in V_{k-2}$ and $\gamma \in \Gamma_0(N)$, we have

$$\delta(P\{\alpha, \gamma(\alpha)\}) = P\{\gamma(\alpha)\} - P\{\alpha\} = (\gamma^{-1}P - P)\{\alpha\} \in V_{k-2}\{\alpha\},$$

so $\delta(\mathcal{W}_\alpha) \subset V_{k-2}\{\alpha\}$.

For $\gamma = \left(\begin{smallmatrix} 1 & 0 \\ N & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$\begin{aligned}\delta\big(X^{k-3}Y\{\alpha, \gamma(\alpha)\}\big) &= \big(\gamma^{-1}(X^{k-3}Y) - X^{k-3}Y\big)\{\alpha\} \\ &= \big(X^{k-3}(NX+Y) - X^{k-3}Y\big)\{\alpha\} \\ &= NX^{k-2}\{\alpha\}.\end{aligned}$$

If $\alpha \neq 0$, then, as mentioned above, $X^{k-2}\{\alpha\} \neq 0$. (If $\alpha = 0$, use XY^{k-3} and $\gamma = \left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$ instead.) Because there is a nonzero element in $\delta(\mathcal{W}_\alpha)$ and $V_{k-2}\{\alpha\}$ has dimension 1, it follows that $\delta(\mathcal{W}_\alpha) = V_{k-2}\{\alpha\}$. The final claim of the proposition is true because $\mathcal{B}_k(\Gamma_0(N);\mathbb{Q}) = \sum_{\alpha\in\mathbb{P}^1(\mathbb{Q})} V_{k-2}\{\alpha\}$. □

Corollary 2.8. *Fix $\alpha \in \mathbb{P}^1(\mathbb{Q})$. Then $\mathcal{W}_\alpha = \mathcal{M}_k(\Gamma_0(N);\mathbb{Q})$ if and only if $N = 1$.*

Proof. When $N = 1$, γ can be any element of $\mathrm{SL}_2(\mathbb{Z})$, so the assertion is clear. Next, suppose that $\mathcal{W}_\alpha = \mathcal{M}_k(\Gamma_0(N);\mathbb{Q})$. If $k = 2$, then by [**4**, Section 1], $\mathcal{W}_\alpha = \mathcal{S}_k(\Gamma_0(N);\mathbb{Q})$, so $N = 1$ since there is always a weight 2 Eisenstein series when $N > 1$. Next, suppose that $k > 2$. By [**5**, Section 1.4, Proposition 5], δ is surjective and by [**5**, Section1.4, Proposition 5] the dimension of the image of δ equals $\#\Gamma_0(N)\backslash\mathbb{P}^1(\mathbb{Q})$. Combining Proposition 2.7 with our assumption that $\mathcal{W}_\alpha = \mathcal{M}_k(\Gamma_0(N);\mathbb{Q})$ implies that $\#\Gamma_0(N)\backslash\mathbb{P}^1(\mathbb{Q}) = 1$, so $N = 1$, as claimed. □

3. *Application to computing periods of newforms*

The authors were led to introduce transportable modular symbols in order to study the error term $(P - \varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\}$ of equation 1 of Proposition 2.1 in the context of computing periods of newforms. There are many ways to compute periods of newforms, but we hope that the method given below will be of value in some contexts.

Section 3.1 contains an algorithm for computing periods that relies on Theorem 2.4. We present a potentially more efficient method in Section 3.2.

3.1. *An algorithm for computing periods*

Let $f = \sum a_n q^n \in S_k(N,\varepsilon)$ be a cuspform, and let $x \in \mathcal{M}_k(N,\varepsilon)$ be a modular symbol. Then $\langle f, x\rangle$ is a linear combination of integrals of the form

$$\big\langle f, X^mY^{k-2-m}\{\alpha,\infty\}\big\rangle = 2\pi i\int_\alpha^{i\infty} f(z)z^m dz, \tag{6}$$

(see [**5**, Section 1.5]), where $\alpha \in \mathfrak{h}^*$ and the integer m satisfies $0 \leqslant m \leqslant k-2$. If $\alpha \in \mathfrak{h}$, then the imaginary part of α is positive, so

$$2\pi i\int_\alpha^{i\infty} f(z)z^m dz = \sum_{n\geqslant 1} a_n c_n,$$

where

$$c_n = 2\pi i\int_\alpha^{i\infty} z^m e^{2\pi i n z} dz.$$

The reversal of summation and integration is justified because the sum converges absolutely. We compute the c_n using the following formula, which we obtain using repeated integration by parts.

Lemma 3.1.

$$\int_{\alpha}^{i\infty} e^{2\pi inz} z^m dz = e^{2\pi in\alpha} \sum_{s=0}^{m} \left\{ \frac{(-1)^s \alpha^{m-s}}{(2\pi in)^{s+1}} \cdot \prod_{j=(m+1)-s}^{m} j \right\}.$$

If α has large imaginary part, the c_n will rapidly converge to 0 as $n \to \infty$. However, the reversal of summation and integration above need not be valid when α is a real number, so for computational purposes we are led to express periods in terms of integrals with end points that are in $\mathfrak{h}$. When $k = 2$, this is easy because of the identity $\{\infty, \gamma(\infty)\} = \{\alpha, \gamma(\alpha)\}$, which is valid for any $\alpha \in \mathfrak{h}^*$. However, this identity can fail when $k > 2$; the failure is made precise in Proposition 2.1.

Since we can take the real part of α to be greater than 0, each of the terms on the right-hand side of Equation 1 can be computed using the sum given by Lemma 3.1.

We showed in Section 2 that every cuspidal modular symbol can be expressed as a sum of symbols of the form $P\{\infty, \gamma(\infty)\}$. Periods of modular symbols of this form can then be computed using the following algorithm.

Algorithm 3.2. Given a triple $\gamma \in \Gamma_0(N)$, $P \in V_{k-2}$ and $g \in S_k(N, \varepsilon)$, this algorithm computes the period integral $\langle g,\ P\{\infty, \gamma(\infty)\}\rangle$.

Express γ as $\left(\begin{smallmatrix} a & b \\ cN & d \end{smallmatrix}\right) \in \Gamma_0(N)$, and set $\alpha = (-d + i)/cN$ in Proposition 2.1.

Replacing γ by $-\gamma$ if necessary, we find that the imaginary parts of α and $\gamma(\alpha) = (a + i)/cN$ are both equal to the positive number $1/cN$.

Equation 6 and Lemma 3.1 can now be used to compute the period integrals provided by Proposition 2.1.

3.2. *The W_N-trick*

In this section, in order to obtain a potentially more efficient way of computing periods than Algorithm 3.2, we generalize the method of Cremona [3] to even integer weight $k \geqslant 2$. In Algorithm 3.2, with $\gamma = \left(\begin{smallmatrix} a & b \\ cN & d \end{smallmatrix}\right)$, the endpoints of the corresponding integrals have imaginary part $1/cN$. However, using the following trick, one can increase the imaginary part of all the endpoints involved to at least $1/d\sqrt{N}$, which is sometimes a significant improvement.

Recall that the Atkin–Lehner involution $W = W_N$ is induced by the matrix $\left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$; it acts on modular forms by sending a cuspform $f \in S_k(N, \varepsilon)$ to the form

$$f|_W(z) = N^{-k/2} z^{-k} f(-1/(Nz)) \in S_k\big(N, \varepsilon^{-1}\big).$$

If f is an eigenvector for W, then necessarily $\varepsilon = \varepsilon^{-1}$. For the rest of this section, we assume that $\varepsilon^2 = 1$. Then W acts on $\mathcal{M}_k(N, \varepsilon)$ by

$$W\Big(P(X, Y)\{\alpha, \beta\}\Big) = \frac{P(Y, -NX)}{N^{k/2-1}} \left\{-\frac{1}{N\alpha}, -\frac{1}{N\beta}\right\},$$

and this action is compatible with the integration pairing.

Proposition 3.3. *Let $g \in S_k(N, \varepsilon)$ be a cuspform that is an eigenform for the Atkin–Lehner involution W having eigenvalue w. Then for any transportable modular symbol $\sum_{j=1}^{m} P_j\{\infty, \gamma_j(\infty)\}$ with $\gamma_j \in \Gamma_0(N)$ and $P_j \in V_{k-2}$, we have for any $\alpha \in \mathfrak{h}$ the following formula:*

$$\left\langle g, \sum_{j=1}^{m} P_j\{\infty, \gamma_j(\infty)\}\right\rangle = \left\langle g, \sum_{j=1}^{m} w\frac{P_j(Y, -NX)}{N^{k/2-1}}\{W(\alpha), \infty\} + \sum_{j=1}^{m}\left(P_j - w\frac{P_j(Y, -NX)}{N^{k/2-1}}\right)\left\{\frac{i}{\sqrt{N}}, \infty\right\} - \sum_{j=1}^{m} P_j\left\{\gamma_j(\alpha), \infty\right\}\right\rangle.$$

Here $W(\alpha) = -1/(N\alpha)$.

If $\gamma_j = \left(\begin{smallmatrix} a_j & b_j \\ c & d \end{smallmatrix}\right)$, where c and d are fixed integers that do not depend on j, then

$$\left\langle g, \sum_{j=1}^{m} P_j\{\infty, \gamma_j(\infty)\}\right\rangle = \left\langle g, \sum_{j=1}^{m} w\frac{P_j(Y, -NX)}{N^{k/2-1}}\left\{\frac{c}{d} + \frac{i}{d\sqrt{N}}, \infty\right\} + \sum_{j=1}^{m}\left(P_j - w\frac{P_j(Y, -NX)}{N^{k/2-1}}\right)\left\{\frac{i}{\sqrt{N}}, \infty\right\} - \sum_{j=1}^{m} P_j\left\{\frac{b_j}{d} + \frac{i}{d\sqrt{N}}, \infty\right\}\right\rangle.$$

Proof. By Proposition 2.1, our condition of transportability implies that we have

$$\sum_{j=1}^{m} P_j\{\infty, \gamma_j(\infty)\} = \sum_{j=1}^{m} P_j\{\alpha, \gamma_j(\alpha)\}.$$

The steps of the following computation are described below.

$$\begin{aligned}
&\langle g, P_j\{\alpha, \gamma_j(\alpha)\}\rangle \\
&\quad = \left\langle g, P_j\left\{\alpha, \frac{i}{\sqrt{N}}\right\} + P_j\left\{\frac{i}{\sqrt{N}}, W(\alpha)\right\} + P_j\{W(\alpha), \gamma_j(\alpha)\}\right\rangle \\
&\quad = \left\langle g, w\frac{W(P_j)}{N^{k/2-1}}\left\{W(\alpha), \frac{i}{\sqrt{N}}\right\} + P_j\left\{\frac{i}{\sqrt{N}}, W(\alpha)\right\} + P_j\{W(\alpha), \gamma_j(\alpha)\}\right\rangle \\
&\quad = \left\langle g, \left(w\frac{W(P_j)}{N^{k/2-1}} - P_j\right)\left\{W(\alpha), \frac{i}{\sqrt{N}}\right\} + P_j\{W(\alpha), \infty\} - P_j\{\gamma_j(\alpha), \infty\}\right\rangle \\
&\quad = \left\langle g, w\frac{W(P_j)}{N^{k/2-1}}\{W(\alpha), \infty\} + \left(P_j - w\frac{W(P_j)}{N^{k/2-1}}\right)\left\{\frac{i}{\sqrt{N}}, \infty\right\} - P_j\{\gamma_j(\alpha), \infty\}\right\rangle.
\end{aligned}$$

In the first step, we break the path from α to $\gamma_j(\alpha)$ into three paths. In the second step, we apply the W-involution to the first term, and use the fact that the action of W is compatible with the pairing $\langle\,,\,\rangle$. The third step involves combining the first two terms, and breaking up the third. In the final step, we replace $\{W(\alpha), i/\sqrt{N}\}$ by $\{W(\alpha), \infty\} + \{\infty, i/\sqrt{N}\}$, and regroup. Taking the sum of both sides of the expression over j from 1 to m gives the first result of the proposition.

Now, following Cremona [2, Section 2.10.8], in order to simultaneously maximize the imaginary parts of $\gamma_j(\alpha)$ and $W(\alpha)$, we take

$$\alpha = \gamma_1^{-1}\left(\frac{b_1}{d} + \frac{i}{d\sqrt{N}}\right).$$

In this case we have

$$W(\alpha) = \frac{c}{d} + \frac{i}{d\sqrt{N}}$$

and

$$\gamma_j(\alpha) = \frac{b_j}{d} + \frac{i}{d\sqrt{N}}.$$

The second formula then follows. □

Remark 3.4. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$. Since the imaginary parts of the terms $i/\sqrt{N}$, $\gamma_j(\alpha)$ and $W(\alpha)$ in the second part of the proposition are all relatively large, the sums appearing in Equation 6 converge relatively quickly if d is small. However, we emphasize that *it is extremely important to choose γ_j in Proposition 3.3 with d small; otherwise, the series will converge very slowly.*

4. *Examples*

The example of Section 4.1 illustrates some of the results of this paper for the weight-12 modular form Δ, and Section 4.2 concerns a nonrational form of level 11 and weight 4. The computations below were done using the first author's implementation of the algorithms of [6] in Magma [1].

4.1. *The weight*-12 *form* Δ

Let $f = \Delta = q \cdot \prod(1-q^n)^{24}$ be the unique normalized eignform in $S_{12}(1)$. The space $\mathcal{M}_{12}(1;\mathbb{Q})$ of modular symbols has dimension 3, and is spanned by $a_1 = X^{10}\{0,\infty\}$, $a_2 = X^8Y^2\{0,\infty\}$, and $a_3 = X^9Y\{0,\infty\}$, and the cuspidal subspace $\mathcal{S}_{12}(1;\mathbb{Q})$ has dimension 2, and is spanned by a_2 and a_3.

As explained in Example 2.2, there is a transportable modular symbol associated to each nonidentity element $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. The transportable symbol

$$\left(2X^2 + 2XY - Y^2\right)^5\left\{\infty, \tfrac{1}{2}\right\} = -300X^9Y\{0,\infty\}$$

is attached to $\left(\begin{smallmatrix} 1 & 1 \\ 2 & 3 \end{smallmatrix}\right)$, and

$$-4665600X^8Y^2\{0,\infty\} - 87300X^9Y\{0,\infty\}$$

is attached to $\left(\begin{smallmatrix} -8 & 5 \\ 19 & -12 \end{smallmatrix}\right)$. Together, these two transportable symbols span $\mathcal{S}_{12}(1;\mathbb{Q})$.

The period map Φ_f sends $X^iY^{10-i}\{0,\infty\}$ to $2\pi i\int_0^\infty z^i f(z)\,\mathrm{dz}$. These integrals are, up to scalars, special values of $L(f,s)$ at critical integers, so they could be computed using any of the standard methods. In any case, we obtain an approximation for the period map:

$$\begin{aligned} \Phi_f(a_1) &\sim 0.0374412812, \\ \Phi_f(a_2) &\sim -0.0159703242, \\ \Phi_f(a_3) &\sim -0.0232962319i. \end{aligned}$$

The period lattice Λ of f is spanned by $\Phi_f((1/14)a_2)$ and $\Phi_f((1/48)a_3)$. (The fractions appear because $\mathcal{S}_{12}(1;\mathbb{Z})$ has basis $(1/14)a_2$ and $(1/48)a_3$.) Since $\mathbb{C}/\Lambda$ is a one-dimensional torus, it makes sense to consider the corresponding elliptic curve over $\mathbb{C}$. This is the elliptic curve $y^2 = x^3 - 27c_4x - 54c_6$, where $c_4 \sim 28091951348793344.58$ and $c_6 \sim -4.70682548 \times 10^{24}$. The j-invariant of this curve is approximately 2592849.394270. Is j a transcendental number?

4.2. *Level* 11, *weight* 4

The unique normalized eigenform in $S_4(\Gamma_0(11))$ is

$$f = q + \alpha q^2 + (-4\alpha + 3)q^3 + (2\alpha - 6)q^4 + (8\alpha - 7)q^5 + \cdots,$$

where $\alpha^2 - 2\alpha - 2 = 0$. The space $\mathcal{M}_4(\Gamma_0(11);\mathbb{Q})$ has basis

$$\begin{aligned}
a_1 &= X^2\{0,\infty\},\\
a_2 &= \left(64X^2 + 16XY + Y^2\right)\left\{-\tfrac{1}{8}, 0\right\},\\
a_3 &= \left(49X^2 + 14XY + Y^2\right)\left\{-\tfrac{1}{7}, 0\right\},\\
a_4 &= \left(25X^2 + 10XY + Y^2\right)\left\{-\tfrac{1}{5}, 0\right\},\\
a_5 &= \left(100X^2 + 20XY + Y^2\right)\left\{-\tfrac{1}{10}, 0\right\},\\
a_6 &= Y^2\{\infty, 0\}.
\end{aligned}$$

The subspace $\mathcal{S}_4(\Gamma_0(11);\mathbb{Q})$ has basis $b_1 = a_2 - a_6$, $b_2 = a_3 - a_6$, $b_3 = a_4 - a_6$, $b_4 = a_5 - a_6$.

As explained in Example 2.2, there is a transportable modular symbol associated to each nonidentity element $\gamma \in \Gamma_0(11)$. For example the transportable symbol

$$\left(11X^2 - 11XY + Y^2\right)\left\{\infty, \tfrac{10}{11}\right\} = 11(a_5 - a_6)$$

is associated to $\gamma = \left(\begin{smallmatrix} 10 & -1 \\ 11 & -1 \end{smallmatrix}\right)$. The symbol

$$-\tfrac{5}{4}b_1 + \tfrac{5}{4}b_2 - \tfrac{1}{4}b_3 + \tfrac{1}{4}b_4$$

is the transportable symbol associated to $\left(\begin{smallmatrix} 5 & -1 \\ 11 & -2 \end{smallmatrix}\right)$. The symbol

$$-\tfrac{9}{8}b_1 - \tfrac{19}{8}b_2 + \tfrac{19}{8}b_3 + \tfrac{99}{8}b_4$$

is associated to $\left(\begin{smallmatrix} 4 & 1 \\ 11 & 3 \end{smallmatrix}\right)$, and

$$-\tfrac{27}{8}b_1 + \tfrac{11}{8}b_2 + \tfrac{9}{8}b_3 + \tfrac{49}{8}b_4$$

is associated to $\left(\begin{smallmatrix} 3 & -2 \\ 11 & -7 \end{smallmatrix}\right)$. Together, these four transportable symbols span $\mathcal{S}_4(\Gamma_0(11);\mathbb{Q})$.

In order to illustrate Section 2.3, we remark that symbols of the form $P\{\infty, \gamma(\infty)\}$ do not span all of $\mathcal{M}_4(\Gamma_0(11);\mathbb{Q})$, but they do span a space bigger than $\mathcal{S}_4(\Gamma_0(11);\mathbb{Q})$. Corollary 2.5 implies that their span contains $\mathcal{S}_4(\Gamma_0(11);\mathbb{Q})$; however, the symbol $Y^2\{\infty, 1/11\}$ does not lie in $\mathcal{S}_4(\Gamma_0(11);\mathbb{Q})$.

Acknowledgements. The authors would like to thank L. Merel and J. Cremona for helpful advice, I. Kiming for improving the proof of Theorem 2.4, and the referee for several helpful remarks.

References

1. W. Bosma, J. Cannon and C. Playoust, 'The Magma algebra system. I. The user language', *J. Symbolic Comput.* 24 (1997) 235–265. 179

2. J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd edn (Cambridge University Press, Cambridge, 1997). 179

3. J. E. Cremona, 'Computing periods of cusp forms and modular elliptic curves', *Experiment. Math.* 6 (1997) 97–107. 177

4. J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. 36 (1972) 19–66. 175, 176

5. L. Merel, *Universal Fourier expansions of modular forms. On Artin's conjecture for odd* 2*-dimensional representations* (Springer, 1994) 59–94. 170, 175, 175, 176, 176, 176

6. W. A. Stein, 'Explicit approaches to modular abelian varieties', Ph.D. thesis, University of California, Berkeley (2000). 179

William A. Stein   was@math.harvard.edu

Department of Mathematics
Harvard University
One Oxford Street
Cambridge, MA 02138
USA

http://modular.fas.harvard.edu

Helena A. Verrill   verrill@math.uni-hannover.de

Institute for Mathematics
University of Hannover
Welfengarten 1
30167 Hannover
Germany

http://hverrill.net

7 Lectures on Serres conjectures, with K. Ribet

Lectures on Serre's conjectures

Kenneth A. Ribet
William A. Stein

Contents

IAS/Park City Mathematics Series
Volume 00, May 15, 2004

Lectures on Serre's conjectures[1]

Kenneth A. Ribet
William A. Stein

[1] Math Department, MC 3840; Berkeley, CA 94720-3840.
E-mail address: `ribet@math.berkeley.edu, was@math.berkeley.edu`.

1991 *Mathematics Subject Classification.* 11

Key words and phrases. Serre's conjectures, modular forms, Galois representations

The second author was supported by a Cal@SiliconValley fellowship.

©May 15, 2004 American Mathematical Society

Preface

We shall begin by discussing some examples of mod ℓ representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. We'll try to motivate Serre's conjectures by referring first to the case of representations that are unramified outside ℓ; these should come from cusp forms on the full modular group $\mathrm{SL}(2,\mathbf{Z})$. In another direction, one might think about representations coming from ℓ-division points on elliptic curves, or more generally from ℓ-division points on abelian varieties of "GL_2-type." Amazingly, Serre's conjectures imply that all odd irreducible two-dimensional mod ℓ representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ may be realized in spaces of ℓ-division points on such abelian varieties. The weak Serre conjecture states that all such representations come from modular forms, and then it takes only a bit of technique to show that one can take the modular forms to have weight two (if one allows powers of ℓ in the level).

Since little work has been done toward proving the weak Serre conjecture, these notes will focus on the bridge between the weak and the strong conjectures. The latter states that each ρ as above comes from the space of cusp forms of a specific weight and level, with these invariants between determined by the local behavior of ρ at ℓ and at primes other than ℓ (respectively). To motivate the strong conjecture, and to begin constructing the bridge, we discuss the local behavior of those ρ that do come from modular forms. For the most part, we look only at forms of weight $k \geq 2$ whose levels N are prime to ℓ. For these forms, the behavior of ρ at ℓ is described in detail in [**32**], where theorems of P. Deligne and Fontaine are recalled. (In [**32**, §6], B. Edixhoven presents a proof of Fontaine's theorem.) Further, the behavior of ρ at primes $p \neq \ell$ may be deduced from H. Carayol's theorems [**11, 12**], which relate the behavior at p of the ℓ-adic representations attached to f with the p-adic component of the automorphic representation of $\mathrm{GL}(2)$ that one associates with f. (The behavior of ρ at ℓ in the case where ℓ divides N is analyzed in [**89**].)

In [**102**], Serre associates to each ρ a level $N(\rho)$ and a weight $k(\rho)$. These invariants are defined so that $N(\rho)$ is prime to ℓ and so that $k(\rho)$ is an integer greater than 1. As Serre anticipated, if ρ arises from a modular form of weight k and level N, and if k is at least 2 and N is prime to ℓ, then one has $k(\rho) \leq k$ and $N(\rho) \mid N$. To find an f for which $N = N(\rho)$ and $k = k(\rho)$ is to "optimize" the level and weight of a form giving ρ. As Edixhoven explains in his article [**32**], weight optimization follows in a somewhat straightforward manner from the theorems of Deligne and Fontaine alluded to above, Tate's theory of θ-cycles, and Gross's theorem on companion forms [**46**] (see also [**17**]). Moreover, it is largely the case that weight and level optimization can be performed independently.

In [**12**], Carayol analyzes the level optimization problem. He shows, in particular, that the problem breaks down into a series of sub-problems, all but one of which he treats by appealing to a single lemma, the lemma of [**12**, §3]. The remaining sub-problem is the one that intervenes in establishing the implication "Shimura-Taniyama $\Longrightarrow$ Fermat." This problem has been discussed repeatedly [**83, 84, 86, 87**]. In Section 3.10, we will explain the principle of [**86**].

The case $\ell = 2$ is the only remaining case for which the level optimization problem has not been resolved. The proof in [**26, 87**] of level optimization for $\ell \geq 3$ does not fully exploit multiplicity one results, but appears to completely break down when $\ell = 2$. In the recent paper [**9**], Kevin Buzzard observed that many new cases of multiplicity one are known and that this can be used to obtain new level optimization results when $\ell = 2$.

In view of these remarks it might be appropriate for us to summarize in a few sentence what is known about the implication "weak Serre conjecture $\Longrightarrow$ strong Serre conjecture." As explained in [**26**], for $\ell \geq 5$ the weak conjecture of Serre implies the strong conjecture about the optimal weight, level, and character. For $\ell = 3$, the weak conjecture implies the strong conjecture, except in a few well-understood situations, where the order of the character must be divisible by ℓ when the level is optimal. The difficulty disappears if one works instead with Katz's definition of a mod ℓ modular form, where the character is naturally defined only mod ℓ. The situation is less complete when $\ell = 2$, but quite favorable. When $\ell = 2$ the situation concerning the weight is explained by Edixhoven in [**32**]: the results of [**17**] do not apply and those of [**46**] rely on unchecked compatibilities.

A certain amount of work has been done on the Hilbert modular case, i.e., the case where $\mathbf{Q}$ is replaced by a totally real number field F. For this work, the reader may consult articles of Frazer Jarvis [**52, 53, 54**], Kazuhiro Fujiwara [**45**], and Ali Rajaei [**79**]. The authors are especially grateful to Fujiwara for sending them a preliminary version of his manuscript, "Level optimization in the totally real case." However, these notes will treat only the classical case $F = \mathbf{Q}$.

This paper emerged out of a series of lectures that were delivered by the first author at the 1999 IAS/Park City Mathematics Institute. The second author created the text based on the lectures and added examples, diagrams, an exercise section, and the index. Brian Conrad contributed the appendix, which describes a construction of Shimura.

For other expository accounts of Serre's conjectures, the reader may consult the articles of Edixhoven [**33, 34, 35**], H. Darmon [**22**], and R. Coleman [**15**].

The authors would like to thank K. Buzzard and Serre for many useful comments on various drafts of this paper, B. Conrad for providing the appendix, M. Emerton for his enlightening lecture on Katz's definition of modular forms, N. Jochnowitz for suggestions that improved the exposition in Section 2.2, and L. Kilford for help in finding examples of mod 2 representations in Section 3.7.

Kenneth A. Ribet
William A. Stein
University of California, Berkeley

CHAPTER 1

Introduction to Serre's conjecture

1.1. Introduction

Let's start with an elliptic curve $E/\mathbf{Q}$. Nowadays, it's a familiar activity to consider the Galois representations defined by groups of division points of E. Namely, let n be a positive integer, and let $E[n]$ be the kernel of multiplication by n on $E(\overline{\mathbf{Q}})$. The group $E[n]$ is free of rank two over $\mathbf{Z}/n\mathbf{Z}$. After a choice of basis, the action of $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $E[n]$ is given by a homomorphism

$$\rho_n : \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \operatorname{Aut}(E[n]) \approx \operatorname{GL}(2, \mathbf{Z}/n\mathbf{Z}).$$

This homomorphism is unramified at each prime p that is prime to the product of n with the conductor of E (see Exercise 15). For each such p, the element $\rho_n(\operatorname{Frob}_p)$ is a 2×2 matrix that is well defined up to conjugation. Its determinant is $p \bmod n$; its trace is $a_p \bmod n$, where a_p is the usual "trace of Frobenius" attached to E and p, i.e., the quantity $1 + p - \#E(\mathbf{F}_p)$. In his 1966 article [**107**], G. Shimura studied these representations and the number fields that they cut out for the curve $E = J_0(11)$. (This curve was also studied by Serre [**91**, pg. 254].) He noticed that for prime values $n = \ell$, the representations ρ_n tended to have large images. In [**93**], J-P. Serre proved that for any fixed elliptic curve E, not having complex multiplication, the indices

$$[\operatorname{GL}(2, \mathbf{Z}/n\mathbf{Z}) : \rho_n(\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}))]$$

are bounded independently of n. In Shimura's example, Serre proved that

$$[\operatorname{GL}(2, \mathbf{Z}/\ell\mathbf{Z}) : \rho_\ell(\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}))] = 1$$

for all $\ell \neq 5$ (see [**93**, §5.5.1]).

In this article, we will be concerned mainly with two-dimensional representations over finite fields. To that end, we restrict attention to the case where $n = \ell$ is prime. The representation ρ_ℓ is "modular" in the familiar sense that it's a representation of a group over a field in positive characteristic. The theme of this course is that it's modular in a different and deeper sense: it comes from a modular form! Indeed, according to a recent preprint of Breuil, Conrad, Diamond and Taylor (see [**7, 19, 114, 117**]), the Shimura-Taniyama conjecture is now a theorem—all elliptic curves over $\mathbf{Q}$ are modular!! Thus if N is the conductor of E, there is a weight-two newform $f = \sum_{n=1}^{\infty} c_n q^n$ ($q = e^{2\pi i z}$) on $\Gamma_0(N)$ with the property that $a_p = c_p$ for all p prime to N. Accordingly, ρ_ℓ is connected up with modular forms via the relation $\operatorname{tr}(\rho_\ell(\operatorname{Frob}_p)) \equiv c_p \pmod{\ell}$, valid for all but finitely many primes p.

The Shimura-Taniyama conjecture asserts that for each positive integer N there is a bijection between isogeny classes of elliptic curves A over $\mathbf{Q}$ of conductor N and rational newforms f on $\Gamma_0(N)$ of weight two. Given A, the Shimura-Taniyama conjecture produces a modular form $f = \sum_{n=1}^{\infty} c_n q^n$ whose Dirichlet series is equal to the L-series of A:

$$\sum_{n=1}^{\infty} \frac{c_n}{n^s} = L(f,s) =\!=\!=\!=\!=\!=\!= L(A,s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}\ .$$

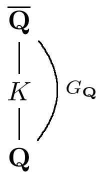
The integers a_n encode information about the number of points on A over various finite fields $\mathbf{F}_p$. If p is a prime not dividing N, then $a_p = p + 1 - \#A(\mathbf{F}_p)$; if $p \mid N$,

$$a_p = \begin{cases} -1 & \text{if } A \text{ has non-split multiplicative reduction at } p \\ 1 & \text{if } A \text{ has split multiplicative reduction at } p \\ 0 & \text{if } A \text{ has additive reduction at } p. \end{cases}$$

The integers a_n are obtained recursively from the a_p as follows:

- $a_{p^r} = \begin{cases} a_{p^{r-1}} a_p - p a_{p^{r-2}} & \text{if } p \nmid N \\ a_p^r & \text{if } p \mid N \end{cases}$
- $a_{nm} = a_n a_m, \qquad$ if $(n,m) = 1$.

The conjectures made by Serre in [**102**], which are the subject of this paper, concern representations $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$. We always require (usually tacitly) that our representations are continuous. The continuity condition just means that the kernel of ρ is open, so that it corresponds to a finite Galois extension K of $\mathbf{Q}$. The representation ρ then embeds $\mathrm{Gal}(K/\mathbf{Q})$ into $\mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$. Since K is a finite extension of $\mathbf{Q}$, the image of ρ is contained in $\mathrm{GL}(2, \mathbf{F})$ for some finite subfield $\mathbf{F}$ of $\overline{\mathbf{F}}_\ell$.



$$\begin{tikzcd} G_{\mathbf{Q}} \arrow[d, two heads] \arrow[dr, "\rho"] & \\ \mathrm{Gal}(K/\mathbf{Q}) \arrow[r, hook] & \mathrm{GL}(2, \overline{\mathbf{F}}_\ell) \end{tikzcd}$$

For various technical reasons, the original conjectures of Serre insist that ρ be irreducible. It is nevertheless fruitful to consider the reducible case as well (see [**111**]).

The conjectures state (in particular) that each continuous irreducible ρ that satisfies a necessary parity condition "arises from" (or is associated with) a suitable modular form mod ℓ. To explain what's going on, let's start with

$$\Delta := \sum_{n=1}^{\infty} \tau(n) q^n = q \prod_{i=1}^{\infty} (1 - q^i)^{24},$$

the unique (normalized) cusp form of weight 12 on $\mathrm{SL}(2, \mathbf{Z})$. In [**92**], Serre conjectured the existence of a "strictly compatible" family of ℓ-adic representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ whose L-function is the L-function of Δ, namely

$$L(\Delta, s) = \sum_{n=1}^{\infty} \tau(n) n^{-s} = \prod_p (1 - \tau(p) p^{-s} + p^{11-2s})^{-1},$$

where the product is taken over all prime numbers p. The conjectured ℓ-adic representations were constructed soon after by Deligne [**24**]. Specifically, Deligne constructed, for each prime ℓ, a representation

$$\rho_{\ell^\infty} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{Z}_\ell),$$

unramified outside ℓ, such that for all primes $p \neq \ell$,

$$\operatorname{tr}(\rho_{\ell^\infty}(\mathrm{Frob}_p)) = \tau(p), \qquad \det(\rho_{\ell^\infty}(\mathrm{Frob}_p)) = p^{11}.$$

On reducing ρ_{ℓ^∞} mod ℓ, we obtain a representation

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell)$$

with analogous properties. (Equalities are replaced by congruences mod ℓ.) In other words, the ρ_ℓ for Δ are just like the ρ_ℓ for an elliptic curve E, except that the integers a_p are replaced by the corresponding values of the τ-function. The determinant of ρ_ℓ is the 11th power of the mod ℓ cyclotomic character $\chi : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$, i.e., the character giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the group of ℓth roots of unity in $\overline{\mathbf{Q}}$ (see Section 1.5).

More generally, take a weight $k \geq 12$ and suppose that $f = \sum_n c_n q^n$ is a nonzero weight-k cusp form for $\mathrm{SL}(2, \mathbf{Z})$ that satisfies $f|T_n = c_n f$ for all $n \geq 1$, T_n being the nth Hecke operator on the space of cusp forms of weight k for $\mathrm{SL}(2, \mathbf{Z})$ (see Section 1.5). Then the complex numbers c_n ($n \geq 1$) are algebraic integers. Moreover, the field $E := \mathbf{Q}(\ldots c_n \ldots)$ generated by the c_n is a totally real number field (of finite degree over $\mathbf{Q}$). Thus the c_n all lie in the integer ring $\mathcal{O}_E$ of E. For each ring homomorphism $\varphi : \mathcal{O}_E \to \overline{\mathbf{F}}_\ell$, one finds a representation

$$\rho = \rho_\varphi : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell),$$

unramified outside ℓ, such that

$$\operatorname{tr}(\rho(\mathrm{Frob}_p)) = \varphi(c_p), \qquad \det(\rho(\mathrm{Frob}_p)) = p^{k-1}$$

for all $p \neq \ell$. We have $\det \rho = \chi^{k-1}$. Of course, there is no guarantee that ρ is irreducible. We can (and do) suppose that ρ is semisimple by replacing it by its semisimplification. Then ρ is determined up to isomorphism by the displayed trace and determinant conditions; this follows from the Cebotarev density theorem and the Brauer-Nesbitt theorem [**21**], which states that semisimple representations are determined by their characteristic polynomials.

It is important to note that k is necessarily an even integer; otherwise the space $S_k(\mathrm{SL}(2, \mathbf{Z}))$ of weight-k cusp forms on $\mathrm{SL}(2, \mathbf{Z})$ is easily seen to be 0. Thus the determinant χ^{k-1} of ρ is an odd power of χ. In particular, $\det \rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$ is unramified outside ℓ and takes the value -1 on complex conjugations $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. It's a nice exercise to check that, conversely, all continuous homomorphisms with these properties are odd powers of χ (see Exercise 1).

In the early 1970s, Serre conjectured that all homomorphisms that are "like ρ" come from cusp forms of some weight on $\mathrm{SL}(2, \mathbf{Z})$. Namely, let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$$

be a continuous, irreducible representation that is (1) unramified outside ℓ and (2) of odd determinant, in the sense that $\det \rho(c) = -1 \in \overline{\mathbf{F}}_\ell$ for complex conjugations $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. In a May, 1973 letter to Tate, Serre conjectured that ρ is of the form ρ_φ. This means that there is a weight $k \geq 12$, an eigenform $f \in S_k(\mathrm{SL}(2, \mathbf{Z}))$, and a homomorphism $\varphi : \mathcal{O}_E \to \overline{\mathbf{F}}_\ell$ (where $\mathcal{O}_E$ is the ring of integers of the field generated by the coefficients of f) so that ρ_φ and ρ are isomorphic.

To investigate Serre's conjecture, it is fruitful to consider the operation $\rho \mapsto \rho \otimes \chi$ on representations. This "twisting" operation preserves the set of representations that come from modular forms. Indeed, let $\theta = q\frac{d}{dq}$ be the classical differential operator $\sum a_n q^n \mapsto \sum n a_n q^n$. According to Serre and Swinnerton-Dyer [**61, 94, 112**], if f is a mod ℓ form of weight k, then θf is a mod ℓ form of weight $k+\ell+1$. Then if ρ is associated to f, $\rho \otimes \chi$ is associated with θf. According to a result of Atkin, Serre and Tate (see [**97**, Th. 3] and Section 2.1), if ρ comes from an eigenform in some space $S_k(\mathrm{SL}(2,\mathbf{Z}))$, then a suitable twist $\rho \otimes \chi^i$ of f comes from a form of weight $\leq \ell+1$.

Serre's conjecture thus has the following consequence: each two-dimensional irreducible odd representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\overline{\mathbf{F}}_\ell$ that is unramified outside ℓ has a twist (by a power of χ) coming from an eigenform on $\mathrm{SL}(2,\mathbf{Z})$ of weight at most $\ell+1$. In particular, suppose that $\ell < 11$. Then the spaces $S_k(\mathrm{SL}(2,\mathbf{Z}))$ with $k \leq \ell+1$ are all 0; as a result, they contain no nonzero eigenforms! The conjecture that all ρ are modular (of level 1) thus predicts that there are *no* representations of the type contemplated if ℓ is 2, 3, 5 or 7. In support of the conjecture, the non-existence statement was proved for $\ell = 2$ by J. Tate in a July, 1973 letter to Serre [**113**]. Soon after, Serre treated the case $\ell = 3$ by methods similar to those of Tate. (See [**113**, p. 155] for a discussion and a reference to a note in Serre's Œuvres.) Quite recently, Sharon Brueggeman considered the case $\ell = 5$; she proved that the conjectured result follows from the Generalized Riemann Hypothesis (see [**8**]). In another direction, Hyunsuk Moon generalized Tate's result and proved that there are only finitely many isomorphism classes of continuous semisimple Galois representations $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_4(\overline{\mathbf{F}}_2)$ unramified outside 2 such that field $K/\mathbf{Q}$ corresponding to the kernel of ρ is totally real (see [**76**]). Similar work in this direction has been carried out by Joshi [**58**], under additional local hypothesis.

Serre discussed his conjecture with Deligne, who pointed out a number of surprising consequences. In particular, suppose that one takes a ρ coming from an eigenform f' of some weight and of level $N > 1$. On general grounds, ρ has the right to be ramified at primes p dividing N as well as at the prime ℓ. Suppose that, by accident as it were, ρ turned out to be unramified at all primes $p \mid N$. Then the conjecture would predict the existence of a level-1 form f' (presumably of the same weight as f) whose mod ℓ representation was isomorphic to ρ. For example, if $N = \ell^\alpha$ is a power of ℓ, then the conjecture predicts that ρ arises from a form f' of level 1. How could one manufacture the f'?

The passage $f \rightsquigarrow f'$ comes under the rubric of "level optimization". When you take a representation ρ that comes from high level N, and it seems as though that representation comes from a lower level N', then to "optimize the level" is to cough up a form of level N' that gives ρ.

Deligne pointed out also that Serre's conjecture implies that representations ρ over $\overline{\mathbf{F}}_\ell$ are required to "lift" to λ-adic representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. In the recent articles [**80**] and [**81**], R. Ramakrishna used purely Galois cohomological techniques to prove results in this direction.

1.2. The weak conjecture of Serre

In the mid 1980s, Gerhard Frey began lecturing on a link between Fermat's Last Theorem and elliptic curves (see [**42, 43**]). (Earlier, Hellegouarch had also considered links between Fermat's Last Theorem and elliptic curves; see the MathSciNet review and Appendix of [**48**].) As is now well known, Frey proposed that if $a^\ell + b^\ell$ was a perfect ℓth power, then the elliptic curve $y^2 = x(x-a^\ell)(x+b^\ell)$ could be proved to be non-modular. Soon after, Serre pointed out that the non-modularity contemplated by Frey would follow from suitable level-optimization results concerning modular forms [**101**]. To formulate such optimization results, Serre went back to the tentative conjecture that he had made 15 years earlier and decided to study representations $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ that are not necessarily unramified at ℓ. The results, of course, were the conjectures of [**102**].

An important consequence of these conjectures is the so-called "weak conjecture of Serre." As background, we recall that Hecke eigenforms on congruence subgroups of $\mathrm{SL}(2, \mathbf{Z})$ give rise to two-dimensional representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. If we set things up correctly, we get representations over $\overline{\mathbf{F}}_\ell$. More specifically, take integers $k \geq 2$ and $N \geq 1$; these are the weight and level, respectively. Let $f = \sum a_n q^n$ be a normalized Hecke eigenform in the space $S_k(\Gamma_1(N))$ of complex weight-k cusp forms on the subgroup $\Gamma_1(N)$ of $\mathrm{SL}(2, \mathbf{Z})$. Thus f is nonzero and it satisfies $f|T_n = a_n f$ for all $n \geq 1$. Further, there is a character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ so that $f|\langle d\rangle = \varepsilon(d)f$ for all $d \in (\mathbf{Z}/N\mathbf{Z})^*$, where $\langle d\rangle$ is the diamond-bracket operator. Again, let $\mathcal{O}$ be the ring of integers of the field $\mathbf{Q}(\ldots a_n \ldots)$ generated by the a_n; this field is a number field that is either totally real or a CM field. Consider a ring homomorphism $\varphi : \mathcal{O} \to \overline{\mathbf{F}}_\ell$ as before. Associated to this set-up is a representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ with properties that connect it up with f (and φ). First, the representation is unramified at all p not dividing ℓN. Next, for all such p, we have

$$\mathrm{tr}(\rho(\mathrm{Frob}_p)) = a_p, \qquad \det(\rho(\mathrm{Frob}_p)) = p^{k-1}\varepsilon(p);$$

the numbers a_p and $p^{k-1}\varepsilon(p)$, literally in $\mathcal{O}$, are mapped tacitly into $\overline{\mathbf{F}}_\ell$ by φ. The representation ρ is determined up to isomorphism by the trace and determinant identities that are displayed, plus the supplemental requirement that it be semisimple. We are interested mainly in the (generic) case in which ρ is irreducible; in that case, it is of course semisimple.

The construction of ρ from f, k and φ was described in [**24**]. In this article, Deligne sketches a method that manufactures for each non-archimedean prime λ of E a representation $\tilde{\rho}_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, E_\lambda)$, where E_λ denotes the completion of E at λ. Given φ, we let $\lambda = \ker\varphi$ and find a model of $\tilde{\rho}_\lambda$ over the ring of integers $\mathcal{O}_\lambda$ of E_λ. Reducing $\tilde{\rho}_\lambda$ modulo λ, we obtain a representation over the finite field $\mathcal{O}_\lambda/\lambda\mathcal{O}_\lambda$, and φ embeds this field into $\overline{\mathbf{F}}_\ell$.

In fact, as Shimura has pointed out, the machinery of [**24**] can be avoided if one seeks only the mod λ representation attached to f (as opposed to the full λ-adic representation $\tilde{\rho}_\lambda$). As the first author pointed out in [**87**], one can use congruences among modular forms to find a form of weight two and level $N\ell^2$ that gives rise to ρ. Accordingly, one can find ρ concretely by looking within the group of ℓ-division points of a suitable abelian variety over $\mathbf{Q}$: the variety $J_1(\ell^2 N)$, which is defined in Section 2.3 and in Conrad's Appendix.

Which representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arise in this way (as k, N, f and φ all vary)? As in the case $N = 1$ (i.e., that where $\Gamma_1(N) = \mathrm{SL}(2, \mathbf{Z})$), any ρ that comes from modular forms is an odd representation: we have $\det(\rho(c)) = -1$ when $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a complex conjugation. To see this, we begin with the fact that $\varepsilon(-1) = (-1)^k$, which generalizes (1.4); this follows from the functional equation that relates $f(\frac{az+b}{cz+d})$ to $f(z)$ when $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is an element of $\Gamma_0(N)$ (see Exercise 7). On the other hand, using the Cebotarev density theorem, we find that $\det\rho = \chi^{k-1}\varepsilon$, where χ is again the mod ℓ cyclotomic character and where ε is regarded now as a map $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \overline{\mathbf{F}}_\ell^*$ in the obvious way, namely by composing $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \overline{\mathbf{F}}_\ell^*$

with the mod N cyclotomic character. The value on c of the latter incarnation of ε is the number $\varepsilon(-1) = (-1)^k$. Since $\chi(c) = -1$, we deduce that $(\det \rho)(c) = -1$, as was claimed.

Serre's weak conjecture states that, conversely, every irreducible odd representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ is modular in the sense that it arises from some f and φ.

A concrete consequence of the conjecture is that all odd irreducible 2-dimensional Galois representations ρ come from abelian varieties over $\mathbf{Q}$. Given ρ, one should be able to find a totally real or CM number field E, an abelian variety A over $\mathbf{Q}$ of dimension $[E : \mathbf{Q}]$ that comes equipped with an action of the ring of integers $\mathcal{O}$ of E, and a ring homomorphism $\varphi : \mathcal{O} \to \overline{\mathbf{F}}_\ell$ with the following property: Let $\lambda = \ker \varphi$. *Then the representation* $A[\lambda] \otimes_{\mathcal{O}/\lambda} \overline{\mathbf{F}}_\ell$ *is isomorphic to* ρ. (In comparing $A[\lambda]$ and ρ, we use $\varphi : \mathcal{O}/\lambda \hookrightarrow \overline{\mathbf{F}}_\ell$ to promote the 2-dimensional $A[\lambda]$ into a representation over $\overline{\mathbf{F}}_\ell$.)

Much of the evidence for the weak conjecture concerns representations taking values in $\mathrm{GL}(2, \mathbf{F}_q)$ where the finite field $\mathbf{F}_q$ has small cardinality. In his original article [**102**, §5], Serre's discusses a large number of examples of such representations. Serre uses theorems of Langlands [**68**] and Tunnell [**115**] to establish his weak conjecture for odd irreducible representations with values in $\mathrm{GL}(2, \mathbf{F}_2)$ and $\mathrm{GL}(2, \mathbf{F}_3)$. Further, he reports on numerical computations of J.-F. Mestre that pertain to representations over $\mathbf{F}_4$ (and trivial determinant). Additionally, Serre remarks [**102**, p. 219] that the weak conjecture is true for those representations with values in $\mathrm{GL}(2, \overline{\mathbf{F}}_p)$ that are dihedral in the sense that their projective images (in $\mathrm{PGL}(2, \overline{\mathbf{F}}_p)$) are dihedral groups. (See also [**29**, §5] for a related argument.) This section of Serre's paper concludes with examples over $\mathbf{F}_9$ and $\mathbf{F}_7$.

More recently, representations over the fields $\mathbf{F}_4$ and $\mathbf{F}_5$ were treated, under somewhat mild hypotheses, by Shepherd-Barron and Taylor [**105**]. For example, Shepherd-Barron and Taylor show that $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_5)$ is isomorphic to the 5-torsion representaton of an elliptic curve over $\mathbf{Q}$ provided that $\det \rho$ is the mod 5 cyclotomic character. Because elliptic curves over $\mathbf{Q}$ are modular, it follows that ρ is modular.

1.3. The strong conjecture

Fix an odd irreducible Galois representation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell).$$

As discussed above, the weak conjecture asserts that ρ is modular, in the sense that there exists integers N and k such that ρ comes from some $f \in S_k(\Gamma_1(N))$. The *strong conjecture* goes further and gives a recipe for integers $N(\rho)$ and $k(\rho)$, then asserts that ρ comes from $S_{k(\rho)}(\Gamma_1(N(\rho)))$. In any particular instance, the strong conjecture is, a priori, easier to verify or disprove than the weak conjecture because $S_{k(\rho)}(\Gamma_1(N(\rho)))$ is a finite-dimensional vector space that can be computed (using, e.g., the algorithm in [**73**]). The relation between the weak and strong conjectures is analogous to the relation between the assertion that an elliptic curve is modular of some level and the assertion that an elliptic curve A is modular of a specific level, the conductor of A.

For each prime p, let $I_p \subset G_{\mathbf{Q}}$ denote an inertia group at p. The optimal level is a product

$$N(\rho) = \prod_{p \neq \ell} p^{n(p)},$$

where $n(p)$ depends only on $\rho|_{I_p}$. The optimal weight $k(\rho)$ depends only on $\rho|_{I_\ell}$. The integer $n(p)$ is a conductor in additive notation. In particular, $n(p) = 0$ if and only if ρ is unramified at p.

View ρ as a homomorphism $G_{\mathbf{Q}} \to \operatorname{Aut}(V)$, where V is a two-dimensional vector space over $\overline{\mathbf{F}}_\ell$. It is natural to consider the subspace of inertia invariants:

$$V^{I_p} := \{v \in V : \rho(\sigma)v = v, \text{ all } \sigma \in I_p\}.$$

For example, $V^{I_p} = V$ if and only if ρ is unramified at p. Define

$$n(p) := \dim(V/V^{I_p}) + \operatorname{Swan}(V),$$

where the wild term $\operatorname{Swan}(V)$ is the Swan conductor

$$\operatorname{Swan}(V) := \sum_{i=1}^{\infty} \frac{1}{[G_0 : G_i]} \dim(V/V^{G_i}) \geq 0.$$

Here $G_0 = I_p$ and the $G_i \subset G_0$ are the higher ramification groups.

Suppose that ρ arises from a newform $f \in S_k(\Gamma_1(N))$. A theorem of Carayol [**12**], which was proved independently by Livné [**70**, Prop. 0.1], implies that $N(\rho) \mid N$. It is productive to study the quotient $N/N(\rho)$. Let $\mathcal{O}$ be the ring of integers of the field generated by the Fourier coefficients of f and let $\varphi : \mathcal{O} \to \overline{\mathbf{F}}_\ell$ be the map such that $\varphi(a_p) = \operatorname{tr}(\rho(\operatorname{Frob}_p))$. Let λ be a prime of $\mathcal{O}$ lying over ℓ and E_λ be the completion of $\operatorname{Frac}(\mathcal{O})$ at λ. Deligne [**24**] attached to the pair f, λ a representation

$$\rho_\lambda : G_{\mathbf{Q}} \to \operatorname{GL}(2, E_\lambda) = \operatorname{Aut}(\tilde{V})$$

where $\tilde{V}$ is a two-dimensional vector space over E_λ. The representation ρ_λ can be conjugated so that its images lies inside $\operatorname{GL}(2, \mathcal{O}_\lambda)$; the reduction of ρ_λ modulo λ is then ρ. The following diagram summarizes the set up:

$$\begin{tikzcd}
 & & \operatorname{GL}(2, E_\lambda) = \operatorname{Aut}(\tilde{V}) \\
G_{\mathbf{Q}} \arrow[r] \arrow[urr, bend left, "\rho_\lambda"] \arrow[drr, bend right, "\rho"'] & \operatorname{GL}(2, \mathcal{O}_\lambda) \arrow[ur, hook] \arrow[dr, "\varphi"] & \\
 & & \operatorname{GL}(2, \overline{\mathbf{F}}_\ell) = \operatorname{Aut}(V)
\end{tikzcd}$$

Let $m(p)$ be the power of p dividing the conductor of ρ_λ. In [**12**], Carayol proves that $m(p) = \operatorname{ord}_p N$. As above, $m(p) = \dim(\tilde{V}/\tilde{V}^{I_p}) + (\text{wild term})$, and the wild term is the same as for ρ. The power of p dividing $N/N(\rho)$ is $\dim(\tilde{V}/\tilde{V}^{I_p}) - \dim(V/V^{I_p}) = \dim V^{I_p} - \dim \tilde{V}^{I_p}$. Though $\tilde{V}$ and V are vector spaces over different fields, we can compare the dimensions of their inertia invariant subspaces. The formula

$$\operatorname{ord}_p(N) = n(p) + (\dim V^{I_p} - \dim \tilde{V}^{I_p}) \tag{1.1}$$

indicates how this difference is the deviation of N from the optimal level locally at p. This is the description of $n(p)$ that is used in proving that if ρ is modular at all, then it is possible to refine N and k to eventually discover that ρ arises from a newform in $S_{k(\rho)}(\Gamma_1(N(\rho)))$. After much work (see [**26, 87**]) it has been shown that *for $\ell > 2$ the weak and strong conjectures are equivalent.* See [**9**] for equivalence in many cases when $\ell = 2$.

Rearranging (1.1) into

$$n(p) = \mathrm{ord}_p(N) - (\dim V^{I_p} - \dim \tilde{V}^{I_p})$$

provides us with a way to read off $N(\rho)$ from $\mathrm{ord}_p(N)$, $\dim V^{I_p}$, and $\dim \tilde{V}^{I_p}$. If $f \in S_k(\Gamma_1(N))$ gives rise to ρ and $\ell \nmid N$, then $k(\rho) \leq k$. In contrast, if we allow powers of ℓ in the level then the weight k can always be made equal to 2.

1.4. Representations arising from an elliptic curve

Equations for elliptic curves can be found in the Antwerp tables [**4**] and the tables of Cremona [**20**]. For example, consider the elliptic curve B given by the equation $y^2+y = x^3+x^2-12x+2$. This is the curve labeled **141A1** in [**20**]; it has conductor $N = 3 \cdot 47$ and discriminant $\Delta = 3^7 \cdot 47$. There is a newform $f \in S_2(\Gamma_0(141))$ attached to B. Because N is square free, the elliptic curve B is *semistable*, in the sense that B has multiplicative reduction at each prime.

The curve B is isolated in its isogeny class; equivalently, for every ℓ the representation

$$\rho_\ell : G_{\mathbf{Q}} \to \mathrm{Aut}(B[\ell]) \approx \mathrm{GL}(2, \mathbf{F}_\ell)$$

is irreducible (see Exercise 4 and Exercise 5). We will frequently consider the representations ρ_ℓ attached to B. The following proposition shows that because B is semistable, each ρ_ℓ is surjective [**93**].

Proposition 1.1. *If A is a semistable elliptic curve over $\mathbf{Q}$ and ℓ is a prime such that ρ_ℓ is irreducible, then ρ_ℓ is surjective.*

Proof. Serre proved this when ℓ is odd; see [**93**, Prop. 21], [**103**, §3.1]. If ρ_2 isn't surjective, then by [**93**, Prop. 21(b)] and Theorem 2.10 it's unramified outside 2. This contradicts [**113**]. □

To give a flavor of Serre's invariants, we describe $N(\rho_\ell)$ and $k(\rho_\ell)$ for the representations ρ_ℓ attached to B. (Note that we still have not defined $k(\rho)$.) At primes p of bad reduction, after a possible unramified quadratic extension of $\mathbf{Q}_p$, the elliptic curve B is a Tate curve. This implies that for $p \neq \ell$, the representation ρ_ℓ is unramified at p if and only if $\mathrm{ord}_p(\Delta) \equiv 0 \pmod{\ell}$; for more details, see Section 2.4.

The optimal level $N(\rho_\ell)$ is a divisor of $3 \cdot 47$; it is divisible only by primes for which ρ_ℓ is ramified, and is not divisible by ℓ. The representation ρ_ℓ is unramified at 3 if and only if $\ell \mid \mathrm{ord}_3(\Delta) = 7$, i.e., when $\ell = 7$. Furthermore, ρ_ℓ is always ramified at 47. First suppose $\ell \notin \{3, 47\}$. If in addition $\ell \neq 7$ then $N(\rho_\ell) = 3 \cdot 47$, and $k(\rho_\ell) = 2$ since $\ell \nmid 3 \cdot 47$. If $\ell = 7$ then $N(\rho_\ell) = 47$, and again $k(\rho_\ell) = 2$. The remaining cases are $\ell = 3$ and $\ell = 47$. If $\ell = 47$ then $N(\rho_\ell) = 3$, and because $\ell - 1$ is the order of the cyclotomic character, $k(\rho_\ell) \equiv 2 \pmod{47-1}$; Serre's recipe then gives $k(\rho_\ell) = 2 + (47 - 1) = 48$. Similarly, when $\ell = 3$, we have $N(\rho_\ell) = 47$ and $k(\rho_\ell) = 2 + (3 - 1) = 4$. The following table summarizes the Serre invariants:

Table 1.4. The Serre invariants of ρ_ℓ

ℓ	$N(\rho_\ell)$	$k(\rho_\ell)$
3	47	4
7	47	2
47	3	48
all other ℓ	141	2

To verify the strong conjecture of Serre for $\ell = 3, 47$, we use a standard tradeoff of level and weight, which relates eigenforms in $S_2(\Gamma_0(141); \mathbf{F}_\ell)$ to eigenforms in $S_{2+\ell-1}(\Gamma_0(141/\ell); \mathbf{F}_\ell)$ (see Section 3.1). The only exceptional prime is $\ell = 7$, for which the minimal weight $k(\rho)$ is 2. The strong conjecture of Serre predicts the existence of an eigenform $g \in S_2(\Gamma_0(47))$ that gives rise to ρ_ℓ. Our initial instinct is to look for an elliptic curve A of conductor 47 such that $A[\ell] \cong B[\ell]$, as $G_{\mathbf{Q}}$-modules. In fact, there are no elliptic curves of conductor 47. This is because $S_2(\Gamma_0(47))$ is four dimensional, having basis the Galois conjugates of a single eigenform $g = \sum c_n q^n$. The Fourier coefficients c_n of g generate the full ring of integers in the field K obtained from $\mathbf{Q}$ by adjoining a root of $h = x^4 - x^3 - 5x^2 + 5x - 1$. The discriminant $1957 = 19 \cdot 103$ of K equals the discriminant of h, so a root of h generates the full ring of integers. The eigenvalue c_2 satisfies $h(c_2) = 0$. Since $h \equiv (x+2)(x^3 + 4x^2 + x + 3) \pmod 7$, there is a prime λ lying over 7 such that $\mathcal{O}/\lambda \cong \mathbf{F}_7$; the isomorphism is given by $c_2 \mapsto -2 \bmod 7$. As a check, note that $\#B(\mathbf{F}_2) = 5$ so $a_2 = 3 - 5 = -2 = \varphi(c_2)$. More generally, for $p \nmid 7 \cdot 141$, we have $\varphi(c_p) \equiv a_p \bmod 7$. This equality of traces implies that the representation $\rho_{g,\lambda}$ is isomorphic to $\rho = \rho_{A,7}$, so A is modular of level 47.

1.5. Background material

In this section, we review the cyclotomic character, Frobenius elements, modular forms, and Tate curves. We frequently write $G_{\mathbf{Q}}$ for $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Many of these basics facts are also summarized in [**23**].

1.5.1. The cyclotomic character

The mod ℓ *cyclotomic character* is defined by considering the group $\boldsymbol{\mu}_\ell$ of ℓth roots of unity in $\overline{\mathbf{Q}}$; the action of the Galois group $G_{\mathbf{Q}}$ on the cyclic group $\boldsymbol{\mu}_\ell$ gives rise to a continuous homomorphism

$$\chi_\ell : G_{\mathbf{Q}} \to \mathrm{Aut}(\boldsymbol{\mu}_\ell). \tag{1.2}$$

Since $\boldsymbol{\mu}_\ell$ is a cyclic group of order ℓ, its group of automorphisms is canonically the group $(\mathbf{Z}/\ell\mathbf{Z})^* = \mathbf{F}_\ell^*$. We emerge with a map $G_{\mathbf{Q}} \to \mathbf{F}_\ell^*$, which is the character in question.

Let A be an elliptic curve and ℓ be a prime number. The Weil pairing e_ℓ (see [**109**, III.8]) sets up an isomorphism of $G_{\mathbf{Q}}$-modules

$$e_\ell : \bigwedge^2 A[\ell] \xrightarrow{\cong} \boldsymbol{\mu}_\ell. \tag{1.3}$$

The determinant of the representation $\rho_{A,\ell}$ is the mod ℓ cyclotomic character χ_ℓ.

Suppose now that $c \in G_{\mathbf{Q}}$ is the automorphism "complex conjugation." Then the determinant of $\rho_{A,\ell}(c)$ is $\chi_\ell(c)$. Now c operates on roots of unity by the map

$\zeta \mapsto \zeta^{-1}$, since roots of unity have absolute value 1. Accordingly,

$$\det \rho_{A,\ell}(c) = -1; \tag{1.4}$$

one says that $\rho_{A,\ell}$ is *odd*. If $\ell \neq 2$, then $\rho_{A,\ell}(c)$ is conjugate over $\overline{\mathbf{F}}_\ell$ to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ (Exercise 7). If $\ell = 2$ then the characteristic polynomial of $\rho_{A,\ell}(c)$ is $(x+1)^2$ so $\rho_{A,\ell}(c)$ is conjugate over $\overline{\mathbf{F}}_\ell$ to either the identity matrix or $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

1.5.2. Frobenius elements

Let K be a number field. The Galois group $\mathrm{Gal}(K/\mathbf{Q})$ leaves the ring $\mathcal{O}_K$ of integers of K invariant, so that one obtains an induced action on the ideals of $\mathcal{O}_K$. The set of prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$ lying over p (i.e., that contain p) is permuted under this action. For each $\mathfrak{p}$, the subgroup $D_\mathfrak{p}$ of $\mathrm{Gal}(K/\mathbf{Q})$ fixing $\mathfrak{p}$ is called the *decomposition group* of $\mathfrak{p}$. Meanwhile, $\mathbf{F}_\mathfrak{p} := \mathcal{O}_K/\mathfrak{p}$ is a finite extension of $\mathbf{F}_p$. The extension $\mathbf{F}_\mathfrak{p}/\mathbf{F}_p$ is necessarily Galois; its Galois group is cyclic, generated by the Frobenius automorphism $\varphi_p : x \mapsto x^p$ of $\mathbf{F}_\mathfrak{p}$. There is a natural surjective map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$; its injectivity is equivalent to the assertion that p is unramified in $K/\mathbf{Q}$. Therefore, whenever this assertion is true, there is a unique $\sigma_\mathfrak{p} \in D_\mathfrak{p}$ whose image in $\mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ is φ_p. The automorphism $\sigma_\mathfrak{p}$ is then a well-defined element of $\mathrm{Gal}(K/\mathbf{Q})$, the Frobenius automorphism for $\mathfrak{p}$. The various $\mathfrak{p}$ are all conjugate under $\mathrm{Gal}(K/\mathbf{Q})$ and that the Frobenius automorphism for the conjugate of $\mathfrak{p}$ by g is $g\sigma_\mathfrak{p}g^{-1}$. In particular, the various $\sigma_\mathfrak{p}$ are all conjugate; this justifies the practice of writing σ_p for any one of them and stating that σ_p is well defined up to conjugation.

We next introduce the concept of Frobenius elements in $G_\mathbf{Q} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let p again be a prime and let $\mathfrak{p}$ now be a prime of the ring of integers of $\overline{\mathbf{Q}}$ lying over p. To $\mathfrak{p}$ we associate: (1) its residue field $\mathbf{F}_\mathfrak{p}$, which is an algebraic closure of $\mathbf{F}_p$, and (2) a decomposition subgroup $D_\mathfrak{p}$ of $G_\mathbf{Q}$. There is again a surjective map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$. The Frobenius automorphism φ_p topologically generates the target group. We shall use the symbol Frob_p to denote any preimage of φ_p in any $D_\mathfrak{p}$ corresponding to a prime lying over p, and refer to Frob_p as a Frobenius element for p in $G_\mathbf{Q}$. This element is doubly ill defined. The ambiguity in Frob_p results from the circumstance that $\mathfrak{p}$ needs to be chosen and from the fact that $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ has a large kernel, the inertia subgroup $I_\mathfrak{p}$ of $D_\mathfrak{p}$. The usefulness of $\mathrm{Frob}_\mathfrak{p}$ stems from the fact that the various $\mathfrak{p}$ are all conjugate, so that likewise the subgroups $D_\mathfrak{p}$ and $I_\mathfrak{p}$ are conjugate. Thus if ρ is a homomorphism mapping $G_\mathbf{Q}$ to some other group, the kernel of ρ contains one $I_\mathfrak{p}$ if and only if it contains all $I_\mathfrak{p}$. In this case, one says that ρ is *unramified* at p; the image of Frob_p is then an element of the target that is well defined up to conjugation.

Consider an elliptic curve A over $\mathbf{Q}$ and let ℓ be a prime number. The fixed field of $\rho_{A,\ell}$ is a finite Galois extension $K_\ell/\mathbf{Q}$ whose Galois group G_ℓ is a subgroup of $\mathrm{GL}(2, \mathbf{F}_\ell)$. A key piece of information about the extension $K_\ell/\mathbf{Q}$ is that its discriminant is divisible at most by ℓ and primes dividing the conductor of A. In other words, if $p \neq \ell$ is a prime number at which A has good reduction, then $K_\ell/\mathbf{Q}$ is unramified at ℓ (see Exercise 15); one says that the representation $\rho_{A,\ell}$ is unramified at p. Whenever this occurs, the construction described above produces a Frobenius element σ_p in G_ℓ that is well defined up to conjugation.

Fix again an elliptic curve A and a prime number ℓ, and let $\rho_{A,\ell} : G_\mathbf{Q} \to \mathrm{GL}(2, \mathbf{F}_\ell)$ be the associated representation. For each prime p not dividing ℓ at which A has good reduction the Frobenius $\sigma_p = \rho_{A,\ell}(\mathrm{Frob}_p)$ is well defined only up to conjugation. Nevertheless, the trace and determinant of σ_p are well defined. The determinant of $\rho_{A,\ell}$ is the mod ℓ cyclotomic character χ, so $\sigma_p = \chi(\mathrm{Frob}_p) = p \in \mathbf{F}_\ell$. On the other hand, one has the striking congruence

$$\mathrm{tr}(\rho_{A,\ell}(\mathrm{Frob}_p)) = p + 1 - \#\tilde{A}(\mathbf{F}_p) \pmod{\ell}.$$

1.5.3. Modular forms

We now summarize some background material concerning modular forms. Serre's book [**96**] is an excellent introduction (it treats only $N = 1$). One might also read the survey article [**27**] or consult any of the standard references [**65, 66, 75, 108**].

The *modular group* $\mathrm{SL}(2, \mathbf{Z})$ is the group of 2×2 invertible integer matrices. For each positive integer N, consider the subgroup

$$\Gamma_1(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbf{Z}) \ : \ N \mid c \quad \text{and} \quad a \equiv d \equiv 1 \pmod{N} \right\}.$$

Let $\mathfrak{h}$ be the complex upper half plane. A *cusp form* of integer weight $k \geq 1$ and level N is a holomorphic function $f(z)$ on $\mathfrak{h}$ such that

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N); \tag{1.5}$$

we also require that $f(z)$ vanishes at the cusps (see [**108**, §2.1]). We denote by $S_k(\Gamma_1(N))$ the space of weight-k cusp forms of level N. It is a finite dimensional complex vector space. When $k \geq 2$ a formula for the dimension can be found in [**108**, §2.6].

Modular forms are usually presented as convergent Fourier series

$$f(z) = \sum_{n=1}^{\infty} a_n q^n$$

where $q := e^{2\pi i z}$. This is possible because the matrices $\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$ lie in $\Gamma_1(N)$ so that $f(z+b) = f(z)$ for all integers b. For the forms that most interest us, the complex numbers a_n are algebraic integers.

The space $S_k(\Gamma_1(N))$ is equipped with an action of $(\mathbf{Z}/N\mathbf{Z})^*$; this action is given by

$$f(z) \mapsto f|\langle \overline{d} \rangle(z) := (cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right)$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbf{Z})$ is any matrix such that $d \equiv \overline{d} \pmod{N}$. The operator $\langle d \rangle = \langle \overline{d} \rangle$ is referred to as a "diamond-bracket" operator.

For each integer $n \geq 1$, the nth *Hecke operator* on $S_k(\Gamma_1(N))$ is an endomorphism T_n of $S_k(\Gamma_1(N))$. The action is generally written on the right: $f \mapsto f|T_n$. The various T_n commute with each other and are interrelated by identities that express a given T_n in terms of the Hecke operators indexed by the prime factors of n. If $p \nmid N$ is a prime define the operator T_p on $S_k(\Gamma_1(N))$ by

$$f|T_p(z) = \sum_{n=1}^{\infty} a_{np} q^n + p^{k-1} \sum_{n=1}^{\infty} a_n (f|\langle p \rangle) q^{np}.$$

For $p \mid N$ prime, define T_p by

$$f|T_p(z) = \sum_{n=1}^{\infty} a_{np} q^n.$$

The *Hecke algebra* associated to cusp forms of weight k on $\Gamma_1(N)$ is the subring

$$\mathbf{T} := \mathbf{Z}[\ldots T_n \ldots \langle d \rangle \ldots] \subset \mathrm{End}(S_k(\Gamma_1(N)))$$

generated by all of the T_n and $\langle d \rangle$. It is finite as a module over $\mathbf{Z}$ (see Exercise 20). The diamond-bracket operators are really Hecke operators, in the sense that they lie in the ring generated by the T_n; thus $\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots]$.

An *eigenform* is a nonzero element $f \in S_k(\Gamma_1(N))$ that is a simultaneous eigenvector for every element of the Hecke algebra $\mathbf{T}$. Writing $f = \sum a_n q^n$ we find that $a_n = a_1 c_n$ where c_n is the eigenvalue of T_n on f. Since f is nonzero, a_1 is also nonzero, so it is possible to multiply f by a_1^{-1}. The resulting *normalized eigenform* wears its eigenvalues on its sleeve: $f = \sum c_n q^n$. Because f is an eigenform, the action of the diamond bracket operators defines a character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$; we call ε the *character* of f.

Associated to an eigenform $f \in S_k(\Gamma_1(N))$ we have a system $(\ldots a_p \ldots)$, $p \nmid N$, of eigenvalues. We say that f is a *newform* if this system of eigenvalues is not the system of eigenvalues associated to an eigenform $g \in S_k(\Gamma_1(M))$ for some level $M \mid N$ with $M \neq N$. Newforms have been extensively studied (see [**2, 13, 69, 75**]); the idea is to understand where systems of eigenvalues first arise, and then reconstruct the full space $S_k(\Gamma_1(N))$ from newforms of various levels.

1.5.4. Tate curves

The Tate curve is a p-adic analogue of the exponentiation of the representation $\mathbf{C}/\Lambda$ of the group of an elliptic curve over $\mathbf{C}$. In this section we recall a few facts about Tate curves; for further details, see [**110**, V.3].

Let K be a finite extension of $\mathbf{Q}_p$; consider an elliptic curve E/K with *split multiplicative* reduction, and let j denote the j-invariant of E. By formally inverting the well-known relation

$$j(q(z)) = \frac{1}{q(z)} + 744 + 196884 q(z) + \cdots$$

between the complex functions $q(z) = e^{2\pi i z}$ and $j(z)$, we find an element $q \in K^*$ with $j = j(q)$ and $|q| < 1$. There is a $\mathrm{Gal}(\overline{\mathbf{Q}}_p/K)$-equivariant isomorphism $E(\overline{\mathbf{Q}}_p) \cong \overline{\mathbf{Q}}_p^*/q^{\mathbf{Z}}$. The Tate curve, which we suggestively denote by $\mathbf{G}_m/q^{\mathbf{Z}}$, is a scheme whose $\overline{\mathbf{Q}}_p$ points equal $\overline{\mathbf{Q}}_p^*/q^{\mathbf{Z}}$.

As a consequence, the group of n-torsion points on the Tate curve is identified with the $\mathrm{Gal}(\overline{\mathbf{Q}}_p/K)$-module $\{\zeta_n^a (q^{1/n})^b : 0 \leq a, b \leq n-1\}$; here ζ_n is a primitive nth root of unity and $q^{1/n}$ is a fixed nth root of q in $\overline{\mathbf{Q}}_p$. In particular, the subgroup generated by ζ_n is invariant under $\mathrm{Gal}(\overline{\mathbf{Q}}_p/K)$, so the local Galois representation on $E[n]$ is reducible. It is also known that the group of connected component of the reduction of the Néron model of E over $\overline{\mathbf{F}}_p$ is a cyclic group whose order is $\mathrm{ord}_p(q)$. The situation is summarized by the following table (taken from [**88**]):

Complex case	p-adic case
$\mathbf{C}/\Lambda$	no p-adic analogue
$\Big\downarrow$ exponential map $e^{2\pi i z}$	no exponential available
$\mathbf{C}^*/q^{\mathbf{Z}}$	$K^*/q^{\mathbf{Z}}$.

Remark 1.2. When E has non-split multiplicative reduction over K, there is an unramified extension L over which E aquires split multiplicative reduction.

1.5.5. Mod ℓ modular forms

There are several excellent papers to consult when learning about mod ℓ modular forms. The papers of Serre [**95**] and Swinnerton-Dyer [**112**] approach the subject from the point of view of Galois representations. Katz's paper [**59**] is very geometric. Edixhoven's paper [**32**] contains a clear description of the basic facts. See also Jochnowitz's paper [**56**].

CHAPTER 2

Optimizing the weight

In [**102**, §2] Serre associated to an odd irreducible Galois representation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$$

two integers $N(\rho)$ and $k(\rho)$, which are meant to be the minimal level and weight of a form giving rise to ρ.

Conjecture 2.1 (Strong conjecture of Serre)**.** Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be an odd irreducible Galois representation arising from a modular form. Then ρ arises from a modular form of level $N(\rho)$ and weight $k(\rho)$.

In this chapter, we are concerned with $k(\rho)$. We consider a mod ℓ representation ρ that arises from an eigenform of level N not divisible by ℓ. Using results of Fontaine and Deligne, we motivate Serre's recipe for $k(\rho)$. In [**32**], Edixhoven also defines an "optimal" weight, which sometimes differs from Serre's $k(\rho)$. Our definition is an "average" of the two; for example, we introduce a tiny modification of $k(\rho)$ when $\ell = 2$. We appologize for any confusion this may cause the reader.

Using various arguments involving the Eichler-Shimura correspondence and Tate's θ-cycles, Edixhoven showed in [**31**] that there must exist another form of weight at most $k(\rho)$, also of level N, which gives rise to ρ. Some of Edixhoven's result rely on unchecked compatibilities that are assumed in [**46**]; however, when $\ell \neq 2$ these results were obtained unconditionally by Coleman and Voloch in [**17**]. We sketch some of Edixhoven's arguments to convey the flavor of the subject.

Remark 2.2 (Notation)**.** We pause to describe a notational shorthand which we will employ extensively in this chapter. If $\rho : G \to \mathrm{Aut}(V)$ is a two-dimensional representation over a field $\mathbf{F}$, we will frequently write

$$\rho \sim \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

to mean that there is a basis for V with respect to which

$$\rho(x) = \begin{pmatrix} \alpha(x) & \beta(x) \\ \gamma(x) & \delta(x) \end{pmatrix} \in \mathrm{GL}_2(\mathbf{F})$$

for all $x \in G$. If we do not wish to specify one of the entries we will simply write $*$. Thus "$\rho \sim \left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$" means that ρ possesses a one-dimensional invariant subspace, and the action on the quotient is trivial.

2.1. Representations arising from forms of low weight

We first consider irreducible Galois representations associated to newforms of low weight. Fix a prime ℓ and suppose $f = \sum a_n q^n$ is a newform of weight k and level N, such that $\ell \nmid N$ and $2 \leq k \leq \ell+1$. Let $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ denote the character of f. Fix a homomorphism φ from the ring of integers $\mathcal{O}$ of $\mathbf{Q}(\ldots a_n \ldots)$ to $\overline{\mathbf{F}}_\ell$. To abbreviate, we often write a_n for $\varphi(a_n)$; thereby thinking of a_n as an element $\overline{\mathbf{F}}_\ell$. Let $\rho = \rho_{f,\varphi} : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be the two-dimensional semisimple odd Galois representation attached to f and φ, and assume that ρ is irreducible.

The recipe for $N(\rho)$ depends on the local behavior of ρ at primes p other than ℓ; the recipe for $k(\rho)$ depends on the restriction $\rho|_{I_\ell}$ of ρ to the inertia group at ℓ. Motivated by questions of Serre, Fontaine and Deligne described $\rho|_{I_\ell}$ in many situations. We distinguish two cases: the ordinary case and the non-ordinary case, which we call the *"supersingular case."*

2.1.1. The ordinary case

Deligne (see [**46**, Prop. 12.1]) considered the *ordinary case*, in which ρ arises from a weight-k newform f with $a_\ell(f) \neq 0 \in \overline{\mathbf{F}}_\ell$. He showed that ρ has a one-dimensional unramified quotient β, so $\rho|_{D_\ell} \sim \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$ with $\beta(I_\ell) = 1$ and $\alpha\beta = \chi^{k-1}\varepsilon$. The mod N character ε is also unramified at ℓ because $\ell \nmid N$. Since the mod ℓ cyclotomic character χ has order $\ell - 1$ and $\rho|_{I_\ell} \sim \begin{pmatrix} \chi^{k-1} & * \\ 0 & 1 \end{pmatrix}$, the value of k modulo $\ell - 1$ is determined by $\rho|_{I_\ell}$. In the case when k is not congruent to 2 modulo $\ell - 1$, the restriction $\rho|_{I_\ell}$ determines the minimal weight $k(\rho)$. We will discuss the remaining case in Section 2.2.

2.1.2. The supersingular case and fundamental characters

Fontaine (see [**32**, §6]) investigated the supersingular case, in which ρ arises from a newform f with $a_\ell(f) = 0 \in \overline{\mathbf{F}}_\ell$. We call such a newform f *supersingular.* To describe the restriction $\rho|_{I_\ell}$ of ρ to the inertia group at ℓ, we introduce the fundamental characters of the tame inertia group. Fix an algebraic closure $\overline{\mathbf{Q}}_\ell$ of the field $\mathbf{Q}_\ell$ of ℓ-adic numbers; let $\mathbf{Q}_\ell^{\mathrm{nr}} \subset \overline{\mathbf{Q}}_\ell$ denote the maximal unramified extension of $\mathbf{Q}_\ell$, and $\mathbf{Q}_\ell^{\mathrm{tm}} \subset \overline{\mathbf{Q}}_\ell$ the maximal tamely ramified extension of $\mathbf{Q}_\ell^{\mathrm{nr}}$. The extension $\mathbf{Q}_\ell^{\mathrm{tm}}$ is the compositum of all finite extensions of $\mathbf{Q}_\ell^{\mathrm{nr}}$ in $\overline{\mathbf{Q}}_\ell$ of degree prime to ℓ. Letting D_ℓ denote the decomposition group, I_ℓ the inertia group, I_t the tame inertia group, and I_w the wild inertia group, we have the following diagram:

$$\begin{tikzcd}
\overline{\mathbf{Q}}_\ell \arrow[d, dash, "I_w"] \arrow[ddd, dash, bend right=40, "D_\ell"'] \arrow[dd, dash, bend left=60, "I_\ell"] \\
\mathbf{Q}_\ell^{\mathrm{tm}} \arrow[d, dash, "I_t"] \\
\mathbf{Q}_\ell^{\mathrm{nr}} \arrow[d, dash, "\widehat{\mathbf{Z}}"] \\
\mathbf{Q}_\ell
\end{tikzcd}$$

It is a standard fact (see, e.g., [**44**, §8]) that the extensions $\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell})$, for all n not divisible by ℓ, generate $\mathbf{Q}_\ell^{\mathrm{tm}}$. For n not divisible by ℓ, the nth roots of unity $\boldsymbol{\mu}_n$ are contained in $\mathbf{Q}_\ell^{\mathrm{nr}}$. Kummer theory (see [**3**]) gives, for each n, a canonical isomorphism

$$\mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell})/\mathbf{Q}_\ell^{\mathrm{nr}}) \xrightarrow{\sim} \boldsymbol{\mu}_n, \qquad \sigma \mapsto \frac{\sigma(\sqrt[n]{\ell})}{\sqrt[n]{\ell}}.$$

Each isomorphism lifts to a map $I_\ell \to \boldsymbol{\mu}_n$ that factors through the tame quotient I_t of I_ℓ. The groups $\boldsymbol{\mu}_n = \boldsymbol{\mu}_n(\overline{\mathbf{Q}}_\ell)$ lie in the ring of integers $\overline{\mathbf{Z}}_\ell$ of $\overline{\mathbf{Q}}_\ell$. Composing any of the maps $I_t \to \boldsymbol{\mu}_n$ with reduction modulo the maximal ideal of $\overline{\mathbf{Z}}_\ell$ gives a mod ℓ character $I_t \to \overline{\mathbf{F}}_\ell^*$, as illustrated:

$$\begin{tikzcd}
I_t \arrow[rrr] \arrow[dr] & & & \overline{\mathbf{F}}_\ell^* \\
& \boldsymbol{\mu}_n(\overline{\mathbf{Q}}_\ell) \arrow[r, "\cong"] & \boldsymbol{\mu}_n(\overline{\mathbf{F}}_\ell). \arrow[ur, hook] &
\end{tikzcd}$$

Let $n = \ell^\nu - 1$ with $\nu > 0$. The map $I_t \to \boldsymbol{\mu}_n$ defines a character $\varepsilon : I_\ell \to \mathbf{F}_{\ell^\nu}^*$. Composing with each of the ν field embeddings $\mathbf{F}_{\ell^\nu} \to \overline{\mathbf{F}}_\ell$ gives the ν *fundamental characters* of level ν:

$$\begin{tikzcd}
I_t \arrow[dr] & & \overline{\mathbf{F}}_\ell. \\
& \mathbf{F}_{\ell^\nu}^* \subset \mathbf{F}_{\ell^\nu} \arrow[ur, shift left=2] \arrow[ur] \arrow[ur, shift right=2, "\nu \text{ maps}"'] &
\end{tikzcd}$$

For example, the unique fundamental character of level 1 is the mod ℓ cyclotomic character (see Exercise 16). When $\nu = 2$, there are two fundamental characters, denoted Ψ and Ψ'; these satisfy $\Psi^\ell = \Psi'$ and $(\Psi')^\ell = \Psi$.

Let A be an elliptic curve over $\mathbf{Q}_\ell$ with good supersingular reduction. In [**93**], Serre proved that the representation

$$I_t \to \mathrm{Aut}(A[\ell]) \subset \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$$

is the direct sum of the two fundamental characters Ψ and Ψ'. One of the characters is

$$I_t \to \mathbf{F}_{\ell^2}^* \subset \mathrm{GL}(2, \mathbf{F}_\ell)$$

where $\mathbf{F}_{\ell^2}^*$ is contained in $\mathrm{GL}(2, \mathbf{F}_\ell)$ as a non-split Cartan subgroup of $\mathrm{GL}(2, \mathbf{F}_\ell)$. More precisely, $\mathbf{F}_{\ell^2}^*$ is embedded in $\mathrm{GL}(2, \mathbf{F}_\ell)$ via the action of the multiplicative group of a field on itself after a choice of basis. More generally, in unpublished joint work, Fontaine and Serre proved in 1979 that if f is a supersingular eigenform of weight $k \leq \ell$, then $\rho|_{I_\ell} : I_\ell \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ factors through I_t and is a direct sum of the two character Ψ^{k-1} and $(\Psi')^{k-1}$. Note that k is determined by this representation, because it is determined modulo $\ell^2 - 1$.

2.2. Representations of high weight

Let D_ℓ be a decomposition group at ℓ and consider a representation $\rho : D_\ell \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ that arises from a newform f of possibly large weight k. Let ρ^{ss} denote the *semisimplification* of ρ; so $\rho^{\mathrm{ss}} = \rho$ if ρ is irreducible, otherwise ρ^{ss} is a direct sum of two characters α and β. The following lemma of Serre (see [**93**, Prop. 4]) asserts that ρ^{ss} is tamely ramified.

Lemma 2.3. *Any semisimple representation ρ is tame, in the sense that $\rho(I_w) = 0$.*

Proof. Since the direct sum of tame representations is tame, we may assume that ρ is simple.

The wild inertia group I_w is the profinite Sylow ℓ-subgroup of I_ℓ: it is a Sylow ℓ-subgroup because each finite Galois extension of $\mathbf{Q}_\ell^{\mathrm{tm}}$ has degree a power of ℓ, and the order of I_t is prime to ℓ; it is unique, because it is the kernel of $\mathrm{Gal}(\overline{\mathbf{Q}}_\ell/\mathbf{Q}_\ell) \to \mathrm{Gal}(\overline{\mathbf{Q}}_\ell/\mathbf{Q}_\ell^{\mathrm{tm}})$, hence normal.

Because ρ is continuous, the image of D_ℓ is finite and we view ρ as a representation on a vector space W over a finite extension of $\mathbf{F}_\ell$. The subspace

$$W^{I_w} = \{w \in W : \sigma(\tau)w = w \text{ for all } \tau \in I_w\}$$

is invariant under D_ℓ. It is nonzero, as can be seen by writing the finite set W as a disjoint union of its orbits under I_w: since I_w is a pro-ℓ-group, each orbit has size either 1 or a positive power of ℓ. The orbit $\{0\}$ has size 1, and $\#W$ is a power of ℓ, so there must be at least $\ell - 1$ other singleton orbits $\{w\}$; for each of these, $w \in W^{I_w}$.

Since ρ is simple and W^{I_w} is a nonzero D_ℓ-submodule, it follows that $W^{I_w} = W$, as claimed. □

The restriction $\rho^{\mathrm{ss}}|_{I_\ell}$ is abelian and semisimple, so it is given by a pair of characters $\alpha, \beta : I_\ell \to \overline{\mathbf{F}}_\ell^*$. Let n be an integer not divisible by ℓ, and consider the tower of fields

$$\begin{tikzcd}
\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell}) \arrow[d, dash, "\boldsymbol{\mu}_n"] \arrow[dd, dash, bend right, "G"'] \\
\mathbf{Q}_\ell^{\mathrm{nr}} \arrow[d, dash, "\widehat{\mathbf{Z}}=\langle \mathrm{Frob}_\ell \rangle"] \\
\mathbf{Q}_\ell
\end{tikzcd}$$

in which $G = \mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell})/\mathbf{Q}_\ell)$, $\boldsymbol{\mu}_n \cong \mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell})/\mathbf{Q}_\ell^{\mathrm{nr}})$, and $\mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}/\mathbf{Q}_\ell)$ is topologically generated by a Frobenius element at ℓ. Choose a lift $g \in G$ of Frob_ℓ, and consider an element $h \in \boldsymbol{\mu}_n$ corresponding to an element $\sigma \in \mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[n]{\ell})/\mathbf{Q}_\ell^{\mathrm{nr}})$. Then since g acts as the ℓth powering map on roots of unity,

$$\frac{g\sigma g^{-1}(\sqrt[n]{\ell})}{\sqrt[n]{\ell}} = \frac{g\sigma(\zeta_{g^{-1}}\sqrt[n]{\ell})}{\sqrt[n]{\ell}} = \frac{g(\zeta_{g^{-1}}h\sqrt[n]{\ell})}{\sqrt[n]{\ell}} = \frac{g(h)\sqrt[n]{\ell}}{\sqrt[n]{\ell}} = h^\ell.$$

Applying the conjugation formula $ghg^{-1} = h^\ell$ to ρ^{ss} gives $\rho^{\mathrm{ss}}(ghg^{-1}) = \rho^{\mathrm{ss}}(h^\ell) = \rho^{\mathrm{ss}}(h)^\ell$. The two representations $h \mapsto \rho^{\mathrm{ss}}(h)^\ell$ and $h \mapsto \rho^{\mathrm{ss}}(h)$ of I_t are thus equivalent via conjugation by $\rho^{\mathrm{ss}}(g)$; we have $\rho^{\mathrm{ss}}(g)\rho^{\mathrm{ss}}(h)\rho^{\mathrm{ss}}(g^{-1}) = \rho^{\mathrm{ss}}(ghg^{-1}) = \rho^{\mathrm{ss}}(h)^\ell$. Consequently, the pair of characters $\{\alpha, \beta\}$ is stable under the ℓth power map, so as a set $\{\alpha, \beta\} = \{\alpha^\ell, \beta^\ell\}$. There are two possibilities:

— The *ordinary case*: $\alpha^\ell = \alpha$ and $\beta^\ell = \beta$.

— The *supersingular case*: $\alpha^\ell = \beta \neq \alpha$ and $\beta^\ell = \alpha \neq \beta$.

In the first case α and β take values in $\mathbf{F}_\ell^*$ and in the second case they take values in $\mathbf{F}_{\ell^2}^*$ but not in $\mathbf{F}_\ell^*$. By the results discussed in Section 2.1, this terminology is consistent with the terminology introduced above.

We first consider the supersingular case. Let Ψ denote one of the fundamental characters of level 2, and write $\alpha = \Psi^n$, $\beta = \Psi^{n\ell}$, with n an integer modulo ℓ^2-1. Next write the smallest non-negative representative for n in base ℓ: $n = a + \ell b$ with $0 \leq a, b \leq \ell - 1$. Then $\ell n \equiv b + \ell a \pmod{\ell^2-1}$. Switching α and β permutes a and b so, relabeling if necessary, we may assume that $a \leq b$. If $a = b$, then $\alpha = \Psi^a(\Psi')^a = \chi^a$, so α takes values in $\mathbf{F}_\ell^*$, which is not the supersingular case; thus we may assume that $0 \leq a < b \leq \ell - 1$. We now factor out by a power of the cyclotomic character:

$$\begin{aligned}\alpha &= \Psi^n = \Psi^a(\Psi')^b = \Psi^a(\Psi')^a(\Psi')^{b-a} = \chi^a(\Psi')^{b-a}\\ \beta &= \chi^a\Psi^{b-a}.\end{aligned}$$

Put another way,

$$\rho^{\mathrm{ss}} \sim \chi^a \otimes \begin{pmatrix}\Psi^{b-a} & 0\\ 0 & (\Psi')^{b-a}\end{pmatrix}.$$

The untwisted representation is $\left(\begin{smallmatrix}\Psi^{k-1} & 0\\ 0 & (\Psi')^{k-1}\end{smallmatrix}\right)$, where $k = 1 + b - a$. Since $2 \leq 1+b-a \leq \ell-1$, the weight of the untwisted representation is in the range considered above. Thus we are in good shape to define $k(\rho)$.

Before giving $k(\rho)$ it is necessary to understand how the weight changes upon twisting by a power of the cyclotomic character χ. This problem is addressed by the theory of mod ℓ modular forms, first developed by Serre [**95**] and Swinnerton-Dyer [**112**], then generalized by Katz [**59**]. A brief review of the geometric theory, which gives an excellent definition of mod ℓ modular forms, can be found in [**32**, §2], [**35**, §1], or [**46**, §2].

In [**61**], Katz defined spaces of mod ℓ modular forms, and a q-expansion map

$$\alpha : \bigoplus_{k\geq 0} M_k(\Gamma_1(N); \mathbf{F}_\ell) \to \mathbf{F}_\ell[[q]].$$

This map is not injective, because both the Hasse invariant of weight $\ell - 1$ and the constant 1 have the same q-expansion.

Definition 2.4. The *minimal weight filtration* $w(f) \in \mathbf{Z}$ of an element f of the *ring* of mod ℓ modular forms is the smallest integer k such that the q-expansion of f comes from a modular form of weight k; if no such k exists, do not define $w(f)$.

Definition 2.5. Define the operator $\theta = q\frac{d}{dq}$ on q-expansions by $\theta(\sum a_n q^n) = \sum n a_n q^n$.

For example, if f is an eigenform of weight k, then there is a mod ℓ eigenform θf of weight $k + \ell + 1$, still of level N, whose q-expansion is $\theta(\sum a_n q^n)$.

Theorem 2.6. *Let f be a mod ℓ modular form. Then $w(\theta f) = w(f) + \ell + 1$ if and only if $\ell \nmid w(f)$. In addition, if $\ell \mid w(f)$ then $w(\theta f) < w(f) + \ell + 1$.*

2.2.1. The supersingular case

We now give Serre's recipe for $k(\rho)$ in the supersingular case. The minimal weight before twisting is $1+b-a$, which is a positive integer that is not divisible by ℓ. Each twist by χ adds $\ell+1$ to the weight, so in the supersingular case we are motivated to define

$$k(\rho) := (1+b-a) + a(\ell+1) = 1 + \ell a + b.$$

We have to check that *at each step* the weight is prime to ℓ, so the minimal weight does not drop during any of the a twists by χ. Since $1 < 1+b-a < \ell$ and

$$(1+b-a) + a(\ell+1) \leq (\ell-1) + (\ell-2)(\ell+1) < \ell^2,$$

the weight can only drop if there exists c with $1 \leq c < a$ such that

$$(1+b-a) + c(\ell+1) \equiv 0 \pmod{\ell}.$$

If this occurred, then $c \equiv a-b-1 \pmod{\ell}$. But $1 \leq c < a \leq \ell-2$, so either $c = a-b-1$, which implies $c \leq 0$ since $a < b$, or $c = \ell+a-b-1 = a+\ell-1-b \geq a$, which would be a contradiction.

Assume that $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ arises from an eigenform f such that $a_\ell(f) = 0 \in \overline{\mathbf{F}}_\ell$. Now we sketch Edixhoven's proof that ρ arises from a mod ℓ eigenform of weight $k(\rho)$.

Let ρ^{ss} denote the semisimplification of the restriction of ρ to a decomposition group at ℓ. The restriction of ρ^{ss} to the inertia group at ℓ is

$$\rho^{\mathrm{ss}}|_{I_\ell} \sim \begin{pmatrix} \Psi^n & 0 \\ 0 & (\Psi')^n \end{pmatrix},$$

where Ψ and $\Psi' = \Psi^\ell$ are the two fundamental characters of level 2. If necessary, reorder Ψ and Ψ' so that $n = a + b\ell$ with $0 \leq a < b \leq \ell-1$. Then

$$\Psi^n = \Psi^{a+b\ell} = \Psi^a(\Psi')^b = \Psi^a(\Psi')^a(\Psi')^{b-a} = \chi^a(\Psi')^{b-a},$$

and

$$\rho^{\mathrm{ss}}|_{I_\ell} \sim \chi^a \otimes \begin{pmatrix} (\Psi')^{b-a} & 0 \\ 0 & \Psi^{b-a} \end{pmatrix}.$$

Recall that, motivated by Fontaine's theorem on Galois representations arising from supersingular eigenforms, we defined

$$k(\rho) = a(\ell+1) + (b-a+1) = 1 + \ell a + b.$$

The first step in showing that ρ arises from a form of weight $k(\rho)$, is to recall the well known result that, up to twist, all systems of mod ℓ eigenvalues occur in weight at most $\ell+1$. This is the subject of the next section.

2.2.2. Systems of mod ℓ eigenvalues

Theorem 2.7. *Suppose ρ is modular of level N and some weight k, and that $\ell \nmid N$. Then some twist $\rho \otimes \chi^{-\alpha}$ is modular of weight $\leq \ell+1$ and level N.*

This is a general theorem, applying to both the ordinary and supersingular cases. See Serre [**97**, Th. 3] when $N = 1$; significant further work was carried out by Jochnowitz [**55**] and Ash-Stevens [**1**, Thm. 3.5] when $\ell \geq 5$. Two proofs are given in [**32**, Thm. 3.4 and §7]. The original method of Serre, Tate, and Koike for treating questions like this is to use the Eichler-Selberg trace formula. As Serre has pointed out to us, the weight appears in that formula simply as an exponent; this makes more or less clear that a congruence modulo ℓ^2-1 gives information on modular forms mod ℓ.

As a digression, we pause to single out some of the tools involved in one possible proof of Theorem 2.7. Note that by twisting we may assume without loss of generality that $k \geq 2$. The group $\Gamma_1(N)$ acts by matrix multiplication on the real vector space $\mathbf{R}^2$. The Eichler-Shimura correspondence (see [**108**, §8.2]) is an isomorphism of real vector spaces

$$S_k(\Gamma_1(N)) \xrightarrow{\cong} H^1_P(\Gamma_1(N), \mathrm{Sym}^{k-2}(\mathbf{R}^2)).$$

The *parabolic* (or cuspidal) cohomology group H^1_P is the intersection, over all cusps $\alpha \in \mathbf{P}^1(\mathbf{Q})$, of the kernels of the restriction maps

$$\mathrm{res}_\alpha : H^1(\Gamma_1(N), \mathrm{Sym}^{k-2}(\mathbf{R}^2)) \to H^1(\Gamma_\alpha, \mathrm{Sym}^{k-2}(\mathbf{R}^2)),$$

where Γ_α denotes the stabilizer of α. For fixed z_0 in the upper half plane, the Eichler-Shimura isomorphism sends a cusp form f to the class of the cocycle $c : \Gamma_1(N) \to \mathrm{Sym}^{k-2}(\mathbf{R}^2)$ induced by

$$\gamma \mapsto \int_{z_0}^{\gamma(z_0)} \mathrm{Re}\left(f(z) \begin{pmatrix} z \\ 1 \end{pmatrix}^{k-2} dz \right),$$

where $\begin{pmatrix} z \\ 1 \end{pmatrix}^{k-2}$ denotes the image of $\begin{pmatrix} z \\ 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} z \\ 1 \end{pmatrix} \in \mathrm{Sym}^{k-2}(\mathbf{C}^2)$, and integration is coordinate wise. There is an action of the Hecke algebra $\mathbf{T}$ on

$$H^1_P(\Gamma_1(N), \mathrm{Sym}^{k-2}(\mathbf{R}^2)),$$

such that the Eichler-Shimura correspondence is an isomorphism of $\mathbf{T}$-modules.

The forms whose periods are integral form a lattice $H^1_P(\Gamma_1(N), \mathrm{Sym}^{k-2}(\mathbf{Z}^2))$ inside $H^1_P(\Gamma_1(N), \mathrm{Sym}^{k-2}(\mathbf{R}^2))$. Reducing this lattice modulo ℓ suggests that there is a relationship between mod ℓ modular forms and the cohomology group

$$H^1_P(\tilde{\Gamma}_1(N), \mathrm{Sym}^{k-2}(\mathbf{F}_\ell^2)),$$

where $\tilde{\Gamma}_1(N)$ is the image of $\Gamma_1(N)$ in $\mathrm{SL}(2, \mathbf{F}_\ell)$. Serre and Hida observed that for $k - 2 \geq \ell$ the $\tilde{\Gamma}_1(N)$ representations $\mathrm{Sym}^{k-2}(\mathbf{F}_\ell^2)$ are sums of representations arising in $\mathrm{Sym}^{k'-2}(\mathbf{F}_\ell^2)$ for $k' < k$. This essential idea is used in proving that all systems of eigenvalues occur in weight at most $\ell + 1$.

2.2.3. The supersingular case revisited

Let ρ be a supersingular mod ℓ representation that arises from some modular form. By Theorem 2.7 there is a form f of weight $k \leq \ell + 1$ such that $\chi^{-\alpha} \otimes \rho \sim \rho_f$. In fact, we may assume that $2 \leq k \leq \ell$; when $k = \ell + 1$ a theorem of Mazur (see [**32**, Thm. 2.8]) implies that there is a form of weight 2 giving rise to ρ_f, and when $k = 1$ we multiply f by the weight $\ell - 1$ Hasse invariant. To show that $w(\theta^\alpha f) = k(\rho)$ we investigate how application of the θ-operator changes the minimal weight. We have $(\rho_f \otimes \chi^\alpha)|_{I_\ell} \sim \left(\begin{smallmatrix} \Psi^n & 0 \\ 0 & (\Psi')^n \end{smallmatrix}\right)$ with $n = a + b\ell$ and $a < b$. Fontaine's theory (see Section 2.1) identifies the characters corresponding to $\rho_f|_{I_\ell}$ as powers Ψ^{k-1} and $(\Psi')^{k-1}$ of the fundamental characters. This gives an equality of unordered sets

$$\{\Psi^{k-1}\chi^\alpha, (\Psi')^{k-1}\chi^\alpha\} = \{\Psi^n, (\Psi')^n\}.$$

It is now possible to compute $w(\theta^\alpha f)$ by considering two cases, corresponding to the ways in which equality of unordered pairs can occur.

Case 1. Suppose that $\Psi^{k-1}\chi^{\alpha} = (\Psi')^n$. Since $\chi = \Psi^{\ell+1}$, we have

$$\Psi^{k-1+\alpha(\ell+1)} = \Psi^{k-1}\chi^{\alpha} = (\Psi')^n = (\Psi')^{a+b\ell} = \Psi^{b+a\ell}.$$

Comparing exponents of Ψ gives

$$k - 1 + \alpha(\ell+1) \equiv b + a\ell \pmod{\ell^2 - 1}, \tag{2.1}$$

which reduces modulo $\ell + 1$ to $k - 1 \equiv b - a \pmod{\ell+1}$; because $2 \leq k \leq \ell$, this implies that $k = 1 + b - a$. Reducing (2.1) modulo $\ell - 1$ and substituting $k = 1 + b - a$ gives $b - a + 2\alpha \equiv b + a \pmod{\ell - 1}$; we find the possible solutions $\alpha = a + m(\ell-1)/2$ with m an integer. No solution $\alpha = a + m(\ell-1)/2$, with m odd, satisfies (2.1), so $\alpha = a$ as an integer mod $\ell - 1$. Finally, we apply Theorem 2.6 and argue as in the end of Section 2.2, to show that

$$w(\theta^a f) = w(f) + a(\ell+1) = 1 + b - a + a\ell + a = 1 + b + a\ell = k(\rho).$$

Case 2. Suppose that $\Psi^{k-1}\chi^{\alpha} = \Psi^n$. Then

$$\Psi^{k-1+\alpha(\ell+1)} = \Psi^{k-1}\chi^{\alpha} = \Psi^n = \Psi^{a+b\ell}.$$

Comparing powers of Ψ, we obtain

$$k - 1 + \alpha(\ell+1) \equiv a + b\ell \pmod{\ell^2 - 1}, \tag{2.2}$$

which reduces modulo $\ell+1$ to $k - 1 \equiv a - b \pmod{\ell+1}$; thus $k = \ell + 2 - (b - a)$. The difference $b - a$ must be greater than 1; otherwise $k = \ell + 1$, contrary to our assumption that $2 \leq k \leq \ell$. Reducing (2.2) modulo $\ell - 1$ gives

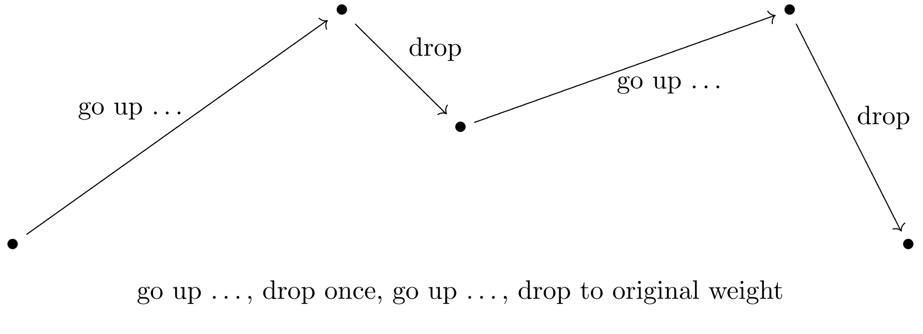
$$k - 1 + 2\alpha \equiv a + b \pmod{\ell - 1};$$

substituting $k = \ell + 2 - (b - a)$ we find that $\alpha = b - 1 + m(\ell - 1)/2$ with m an integer. If m is odd, then α does not satisfy (2.2), so $\alpha = b - 1$ as an integer modulo $\ell - 1$. It remains to verify the equality $w(\theta^{b-1}f) = w(\rho)$. Unfortunately, $k = \ell + 2 - (b - a)$ is not especially telling. The argument of Case 1 does not apply to compute $w(\theta^{\alpha}f)$; instead we use θ-cycles.

Because f is supersingular, Fermat's Little Theorem implies that $\theta^{\ell-1}f = f$. We use Tate's theory of θ-cycles (see [**32**, §7] and [**55**]) to compute $w(\theta^{b-1}f)$. The θ-cycle associated to f is the sequence of integers

$$w(f), w(\theta f), w(\theta^2 f), \ldots, w(\theta^{\ell-2}f), w(f).$$

The θ-cycle for any supersingular eigenform must behave as follows (see Theorem 2.6):



go up ..., drop once, go up ..., drop to original weight

Knowing this, we can deduce the exact θ-cycle. List ℓ numbers starting and ending with k:

$$\begin{aligned}&k,\ k+(\ell+1),\ k+2(\ell+1),\ \ldots,\ k+(\ell-k)(\ell+1),\\ &\ell+3-k,\ (\ell+3-k)+(\ell+1),\ \ldots,\ (\ell+3-k)+(k-3)(\ell+1),\\ &k\end{aligned}$$

The first and second lines contain $\ell+1-k$ and $k-2$ numbers, respectively. All told, ℓ numbers are listed; this must be the θ-cycle.

It is now possible to compute $w(\theta^{b-1}f)$. If

$$b-1 \le \ell-k = \ell-(\ell+2-b+a) = -2+b-a,$$

then $a \le -1$, a contradiction; thus $b-1 > \ell-k$. It follows that

$$w(\theta^{b-1}f) = \ell+3-k+(\ell+1)(b-2-(\ell-k)) = 1+b+a\ell = k(\rho),$$

verifying Serre's conjecture in this case.

2.2.4. The ordinary case

We next turn to the ordinary case, in which

$$\rho|_{I_\ell} \sim \begin{pmatrix}\alpha & * \\ 0 & \beta\end{pmatrix}$$

with $\alpha,\beta : I_\ell \to \mathbf{F}_\ell^*$ powers of the cyclotomic character. View $\rho|_{I_\ell}$ as the twist of a representation in which the lower right entry is 1:

$$\begin{pmatrix}\alpha & * \\ 0 & \beta\end{pmatrix} \sim \beta \otimes \begin{pmatrix}\alpha\beta^{-1} & * \\ 0 & 1\end{pmatrix}.$$

To determine the minimal weight of a form giving rise to $\rho|_{I_\ell}$, it is necessary to develop an ordinary version of θ-cycles. In general this is complicated, so we make the simplifying assumption that $\beta = 1$; then $\rho|_{I_\ell} \sim \begin{pmatrix}\chi^i & * \\ 0 & 1\end{pmatrix}$ with $1 \le i \le \ell-1$. Deligne showed that if f is of weight k and $\beta = 1$, then the associated representation is $\left(\begin{smallmatrix}\chi^{k-1} & * \\ 0 & 1\end{smallmatrix}\right)$ with $2 \le k \le \ell+1$. Motivated by this, our first reaction is to define $k(\rho)$ to be $i+1$. This definition does not distinguish between the extreme weights 2 and $\ell+1$ because they are congruent modulo $\ell-1$. Given a representation ρ arising from a form of weight either 2 or $\ell+1$, we cannot, in general, set $k(\rho) = 2$. For example, suppose $f = \Delta$ is the level 1 cusp form of weight 12 and ρ is the associated mod 11 representation. It would be wrong to set $k(\rho) = 2$, because there is no cusp form of weight 2 and level 1.

Warning: When $\ell = 2$ and our $k(\rho)$ is 3, Serre replaced $k(\rho)$ by 4 because there are no weight-3 modular forms whose character is of degree coprime to $\ell = 2$.

2.3. Distinguishing between weights 2 and $\ell+1$

We continue to motivate the definition of $k(\rho)$. Consider a representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ that arises from a newform f of the optimal level $N = N(\rho)$ and weight k satisfying $2 \le k \le \ell+1$. Assume that f is ordinary in the sense that $a_\ell(f) \ne 0 \in \overline{\mathbf{F}}_\ell$. Then, as discussed in Section 2.1,

$$\rho|_{I_\ell} \sim \left(\begin{smallmatrix}\chi^{k-1} & * \\ 0 & 1\end{smallmatrix}\right),$$

so $\rho|_{I_\ell}$ determines k modulo $\ell - 1$. This suggests a way to define $k(\rho)$ purely in terms of the Galois representation ρ, at least when $k \not\in \{2,\ \ell+1\}$.

The key to defining $k(\rho)$ when $k = 2$ or $k = \ell + 1$ is good reduction. To understand why this is so, we briefly summarize Shimura's geometric construction of Galois representations associated to newforms of weight 2.

2.3.1. Geometric construction of Galois representations

Shimura attached mod ℓ representations to a weight-2 newform $f = \sum a_n q^n$ of level N. Let E be the totally real or CM field $\mathbf{Q}(\ldots a_n \ldots)$. In [**108**, Thm. 7.14], Shimura described how to associate to f an abelian variety $A = A_f$ over $\mathbf{Q}$ of dimension $[E : \mathbf{Q}]$ furnished with an embedding $E \hookrightarrow \operatorname{End}_{\mathbf{Q}} A$ (see also Conrad's appendix). The mod ℓ representations attached to f are then found in the ℓ-torsion of A.

Over the complex numbers, the abelian variety A is found as a quotient of the Jacobian of the Riemann surface

$$X_1(N) := \overline{\Gamma_1(N)\backslash\mathfrak{h}} = \Gamma_1(N)\backslash\mathfrak{h} \cup \{\text{cusps}\}.$$

The Riemann surface $X_1(N)$ has a structure of algebraic curve over $\mathbf{Q}$; it is called the *modular curve* of level N. Its Jacobian $J_1(N)$ is an abelian variety over $\mathbf{Q}$ which, by work of Igusa, has good reduction at all primes $\ell \nmid N$. The dimension of $J_1(N)$ equals the genus of $X_1(N)$; for example, when $N = 1$, the curve $X_1(1)$ is isomorphic over $\mathbf{Q}$ to the projective line and $J_1(1) = 0$. There are (at least) two functorial actions of the Hecke algebra $\mathbf{T}$ on $J_1(N)$, and (at least) two definitions of $J_1(N)$. In the next section we will fix choices, and then construct A as the quotient of $J_1(N)$ by the image of the annihilator in $\mathbf{T}$ of f.

2.3.1.1. *Hecke operators on $J_1(N)$.* We pause to formulate a careful definition of $X_1(N)$ and of our preferred functorial action of the Hecke operators T_p on $J_1(N)$. For simplicity, we assume that $N > 4$ and $p \nmid N$. Following [**46**, Prop. 2.1] there is a smooth, proper, geometrically connected algebraic curve $X_1(N)$ over $\mathbf{Z}[1/N]$ that represents the functor assigning to each $\mathbf{Z}[1/N]$-scheme S the set of isomorphism classes of pairs (E, α), where E is a generalized elliptic curve over S and $\alpha : (\boldsymbol{\mu}_N)_S \hookrightarrow E^{\mathrm{sm}}[N]$ an embedding of group schemes over S whose image meets every irreducible component in each geometric fiber. Let $X_1(N, p)$ be the fine moduli scheme over $\mathbf{Z}[1/N]$ that represents the functor assigning to each $\mathbf{Z}[1/N]$-scheme S the set of isomorphism classes of triples (E, α, C), where E is a generalized elliptic curve over S, $\alpha : (\boldsymbol{\mu}_N)_S \hookrightarrow E^{\mathrm{sm}}[N]$ an embedding of group schemes over S, and C a locally free subgroup scheme of rank p in $E^{\mathrm{sm}}[p]$, such that $\mathrm{im}(\alpha) \times C$ meets every irreducible component in each geometric fiber of E. Let $\pi_1, \pi_2 : X_1(N, p) \to X_1(N)$ over $\mathbf{Z}[1/N]$ be the two standard degeneracy maps, which are defined on genuine elliptic curves by $\pi_1(E, \alpha, C) = (E, \alpha)$ and $\pi_2(E, \alpha, C) = (E', \alpha' = \varphi\alpha)$, where $E' = E/C$ and $\varphi : E \to E'$ is the associated p-isogeny. The Hecke operator $T_p = (T_p)^*$ acts on divisors D on $X_1(N)_{/\mathbf{Q}}$ by

$$T_p(D) = (\pi_1)_* \circ \pi_2^* D.$$

For example, if (E, α) is a non-cuspidal $\overline{\mathbf{Q}}$-point, then

$$T_p(E, \alpha) = \sum (E', \varphi \circ \alpha \circ [p]^{-1}),$$

The Hecke operatorThe Hecke operator $T_p = (T_p)^*$ acts on divisors D on $X_1(N)_{/\mathbf{Q}}$ by

$$T_p(D) = (\pi_1)_* \circ \pi_2^* D.$$

For example, if (E, α) is a non-cuspidal $\overline{\mathbf{Q}}$-point, then

$$T_p(E, \alpha) = \sum (E', \varphi \circ \alpha \circ [p]^{-1}),$$

where the sum is over all isogenies $\varphi : E \to E'$ of degree p, and $T_p = (T_p)^*$ acts on divisors D on $X_1(N)_{/\mathbf{Q}}$ by

$$T_p(D) = (\pi_1)_* \circ \pi_2^* D.$$

For example, if (E, α) is a non-cuspidal $\overline{\mathbf{Q}}$-point, then

$$T_p(E, \alpha) = \sum (E', \varphi \circ \alpha \circ [p]^{-1}),$$

where the sum is over all isogenies $\varphi : E \to E'$ of degree p, and where the sum is over all isogenies $\varphi : E \to E'$ of degree p, and $[p]^{-1}$ is the inverse of pth powering on $\boldsymbol{\mu}_N$. This map on divisors defines an endomorphism T_p of the Jacobian $J_1(N)$ associated to $X_1(N)$ via Picard functoriality.

For each prime p there is an involution $\langle p \rangle$ of $X_1(N)$ called a *diamond bracket operator*, defined functorially by

$$\langle p \rangle (E, \alpha) = (E, \alpha \circ [p]).$$

The diamond bracket operator defines a correspondence, such that the induced map $(\langle p \rangle)^*$ on $J_1(N)$ is

$$(\langle p \rangle)^* (E, \alpha) = (E, \alpha \circ [p^{-1}]).$$

If $(T_p)_*$ denotes the pth Hecke operator as defined in [**46**, §3], then

$$(T_p)_* = T_p \circ (\langle p^{-1} \rangle)^*,$$

Thus our T_p differs from Gross's $(T_p)_*$. Furthermore, upon embedding $X_1(N)$ into $J_1(N)$ and identifying weight-2 cusp forms with differentials on $J_1(N)$, Gross's $(T_p)_*$ induces, via Albanese functoriality, the usual Hecke action on cusp forms, whereas ours does not. In addition, we could have defined $X_1(N)$ by replacing the group scheme $\boldsymbol{\mu}_N$ by $(\mathbf{Z}/N\mathbf{Z})$. In this connection, see the discussion at the end of Section 5 of [**26**] and [**35**, §2.1].

2.3.1.2. *The representations attached to a newform.* Again let $\mathcal{O}$ be the ring of integers of $E = \mathbf{Q}(\ldots a_n \ldots)$, where $f = \sum a_n q^n$ is a weight-2 modular forms on $\Gamma_1(N)$. Recall that $A = A_f$ is the quotient of $J_1(N)$ by the image of the annihilator in $\mathbf{T}$ of f. In general, $\mathcal{O}$ need not be contained in $\operatorname{End} A$. However, by replacing A by an abelian variety $\mathbf{Q}$-isogenous to A, we may assume that $\mathcal{O}$ is contained in $\operatorname{End} A$ (see [**108**, pg. 199]). Let λ be a maximal ideal of $\mathcal{O}$ and set

$$A[\lambda] := \{P \in A(\overline{\mathbf{Q}}) : xP = 0 \text{ all } x \in \lambda\}.$$

By [**108**, Prop. 7.20, pg 190], $\dim_{\mathcal{O}/\lambda} A[\lambda] = 2$, so $A[\lambda]$ affords a 2-dimensional Galois representation, which is well-defined up to semisimplification. Let $\rho_{f,\lambda} : G_{\mathbf{Q}} \to A[\lambda]^{\mathrm{ss}}$ be the semisimplification of $A[\lambda]$.

2.3.1.3. *Good reduction.*

Definition 2.8. A finite group scheme G over $\mathbf{Q}_\ell^{\mathrm{nr}}$ is said to have *good reduction*, or to be *finite flat*, if it extends to a finite flat group scheme over the ring of integers $\mathcal{O}_{\mathbf{Q}_\ell^{\mathrm{nr}}}$ of $\mathbf{Q}_\ell^{\mathrm{nr}}$.

Proposition 2.9. *The representation $\rho_{f,\lambda}$ is finite flat at each prime $p \nmid N$.*

Proof. The finite flat group scheme extending $A[\lambda]$ is the scheme theoretic closure of $A[\lambda]$ in a good model $\mathcal{A}/\mathcal{O}_{\mathbf{Q}_\ell^{\mathrm{nr}}}$ of A. Such a model exists because A has good reduction at p. □

Consider again a Galois representation ρ as in the beginning of Section 2.3 such that $\rho|_{I_\ell} \sim \left(\begin{smallmatrix} \chi^{k-1} & * \\ 0 & 1 \end{smallmatrix}\right)$. If $k \not\equiv 2 \pmod{\ell - 1}$ then $k(\rho)$ is defined to equal k. If $k \equiv 2$ $(\mathrm{mod}\ \ell - 1)$, then

$$k(\rho) := \begin{cases} 2 & \text{if } \rho \text{ is finite flat,} \\ \ell + 1 & \text{otherwise.} \end{cases}$$

2.4. Representations arising from elliptic curves

Theorem 2.10. *Suppose $A/\mathbf{Q}$ is a semistable elliptic curve and that $\rho_{A,\ell}$ is irreducible. Let Δ_A denote the minimal discriminant of A. The representation $\rho_{A,\ell}$ is finite flat at ℓ if and only if $\ell \mid \mathrm{ord}_\ell \Delta_A$. If $p \neq \ell$, then $\rho_{A,\ell}$ is unramified at p if and only if $\ell \mid \mathrm{ord}_p \Delta_A$.*

Proof. The first statement is Proposition 5 of [**102**].

When A has good reduction at p, the second statement holds (see Exercise 15). Suppose A has multiplicative reduction at p. There is an unramified extension K of $\mathbf{Q}_p$ such that A has split multiplicative reduction at p. Consider the Tate curve $\mathbf{G}_m/q^{\mathbf{Z}}$ over K associated to A. Thus $\overline{\mathbf{Q}}_p^*/q^{\mathbf{Z}} \cong A(\overline{\mathbf{Q}}_p)$ as $\mathrm{Gal}(\overline{\mathbf{Q}}_p/K)$-modules. The ℓ-torsion points $A[\ell]$ correspond to the points $\{\zeta_\ell^a (q^{1/\ell})^b : 0 \leq a, b < \ell\}$ in the Tate curve. The extension $K(\zeta_\ell, q^{1/\ell})$ of K is unramified because $\ell \neq p$ and $\mathrm{ord}_p(q) = \mathrm{ord}_p(\Delta_A)$ is divisible by ℓ. Since an unramified extension of an unramified extension is unramified, the extension $K(\zeta_\ell, q^{1/\ell})$ of $\mathbf{Q}_p$ is unramified, which proves the second part of the theorem. □

2.4.1. Frey curves

Using Theorem 2.10 we see that the Shimura-Taniyama conjecture together with Serre's conjecture implies Fermat's Last Theorem. Suppose (a, b, c) is a solution to the Fermat equation $a^\ell + b^\ell = c^\ell$ with $\ell \geq 11$ and $abc \neq 0$. Consider the Frey curve A given by the equation $y^2 = x(x - a^\ell)(x + b^\ell)$; it is an elliptic curve with discriminant $\Delta_A = \frac{((abc)^2)^\ell}{2^8}$. By [**93**, §4.1, Prop. 6] the representation $A[\ell]$ is irreducible. Theorem 2.10 implies that $\rho_{A,\ell}$ is unramified, except possibly at 2 and ℓ. Thus $N(\rho) \mid 2$, and $k(\rho) = 2$ since $\ell \mid \mathrm{ord}_\ell(\Delta_A)$. But there are no cusp forms of level 2 and weight 2. The modularity of A (proved in [**114, 117**]), together with the weak conjecture of Serre (enough of which is proved in [**84**]), leads to a contradiction.

2.4.2. Examples

Using Theorem 2.10 we can frequently determine the Serre invariants $N(\rho)$ and $k(\rho)$ of a representation ρ attached to an elliptic curve. When $N(\rho) < N$, it is illustrative to verify directly that there is a newform of level $N(\rho)$ that also gives rise to ρ. For example, there is a unique weight-2 normalized newform

$$f = q + q^2 - q^3 - q^4 - 2q^5 - q^6 + 4q^7 - 3q^8 + q^9 + \cdots$$

on $\Gamma_0(33)$. One of the elliptic curves associated to f is the curve A given by the equation

$$y^2 + xy = x^3 + x^2 - 11x.$$

The discriminant of A is $\Delta = 3^6 \cdot 11^2$ and the conductor is $N = 3 \cdot 11$. Because A is semistable and there are no elliptic curves 3-isogenous to A, the associated mod 3 representation $\rho = \rho_{A,3} : G_{\mathbf{Q}} \to \mathrm{Aut}(A[3])$ is surjective (see Section 1.4). Since $3 \mid \mathrm{ord}_3 \Delta_A$, the Serre weight and level are $k(\rho) = 2$ and $N(\rho) = 11$. As predicted by Serre's conjecture, there is a weight-2 newform on $\Gamma_0(11)$ such that if B is one of the three elliptic curves of conductor 11 (it does not matter which), then $B[3] \approx A[3]$ as representations of $G_{\mathbf{Q}}$. Placing the eigenforms corresponding to A and B next to each other, we observe that their Fourier coefficients are congruent modulo 3:

$$\begin{array}{lcccccccccccc} f_A & = & q & +q^2 & -q^3 & -q^4 & -2q^5 & -q^6 & +4q^7 & -3q^8 & +q^9 & + & \cdots \\ f_B & = & q & -2q^2 & -q^3 & +2q^4 & +q^5 & +2q^6 & -2q^7 & & -2q^9 & + & \cdots . \end{array}$$

Next consider the elliptic curve A cut out by the equation

$$y^2 + y = x^3 + x^2 - 12x + 2.$$

It has conductor $N = 141 = 3 \cdot 47$ and discriminant $\Delta = 3^7 \cdot 47$. Since $\mathrm{ord}_3(\Delta)$ is divisible by 7, the mod 7 representation $\rho_{A,7}$ has Serre invariants $k(\rho_{A,7}) = 2$ and $N(\rho_{A,7}) = 47$. In confirmation of Serre's conjecture, we find a form $f \in S_2(\Gamma_0(47))$ that gives rise to $\rho_{A,7}$. The Fourier coefficients of f generate a quartic field.

Next consider $\rho_{A,3}$, whose Serre invariants are $N(\rho_{A,3}) = 47$ and, since 3 does not divide $\mathrm{ord}_3(\Delta)$, $k(\rho_{A,3}) = \ell + 1 = 4$. In $S_4(\Gamma_0(47))$ there are two conjugacy classes of eigenforms, which are defined over fields of degree 3 and 8, respectively. The one that gives rise to $\rho_{A,3}$ is

$$g = q + aq^2 + (-1/2a^2 - 5/2a - 1)q^3 + (a^2 - 8)q^4 + (a^2 + a - 10)q^5 + \cdots,$$

where $a^3 + 5a^2 - 2a - 12 = 0$.

2.5. Companion forms

Suppose f is a newform of weight k with $2 \leq k \leq \ell + 1$. Let ℓ be an ordinary prime, so $a_\ell(f)$ is not congruent to 0 modulo a prime λ lying over ℓ and

$$\rho_{f,\lambda}|_{I_\ell} \sim \begin{pmatrix} \chi^{k-1} & * \\ 0 & 1 \end{pmatrix}.$$

Is this representation split or not? Put another way, can $*$ be taken equal to 0, after an appropriate choice of basis? For how many ℓ do these representations split? We suspect that the ordinary split primes ℓ are in the minority, among all primes. How can we quantify the number of split primes?

If $* = 0$, then

$$\rho|_{I_\ell} \sim \begin{pmatrix} 1 & 0 \\ 0 & \chi^{k-1} \end{pmatrix},$$

so

$$\rho|_{I_\ell} \otimes \chi^{\ell-k} \sim \begin{pmatrix} \chi^{\ell-k} & 0 \\ 0 & 1 \end{pmatrix}.$$

Assume that $2 \leq 1+\ell-k \leq \ell+1$, so $k(\rho\otimes\chi^{\ell-k}) = 1+\ell-k$. Using the θ-operator we see that $\rho\otimes\chi^{\ell-k}$ is modular, of *some* weight and level. To say that it is modular of Serre's conjectured weight $k(\rho)$ is to make a much strong statement. If $\rho\otimes\chi^{\ell-k}$ is indeed modular of weight $1+\ell-k$, then by definition there exists an eigenform g of weight $1+\ell-k$ with $\rho_g \sim \rho_f \otimes \chi^{\ell-k}$. Such an eigenform g, if it exists, is called a *companion* of f. The existence of g is far from obvious.

We can extend the notion of companion form to the case when $k(\rho) = \ell$. In this case the companion has weight 1. If ρ is unramified at ℓ, then we expect ρ to also arise from a weight-1 eigenform.

The existence of a companion form was proved (assuming unchecked compatibilities) in most cases in which $k < \ell$ by Gross in [**46**] and in a few cases when $k = \ell$. Using new methods, Coleman and Voloch [**17**] proved all cases except $k = \ell = 2$. The arguments of Coleman and Voloch do not require verification of Gross's unchecked compatibilities.

CHAPTER 3

Optimizing the level

Consider an irreducible Galois representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ that arises from a newform of weight k and level N. Serre defined integers $k(\rho)$ and $N(\rho)$, and conjectured that ρ arises from a newform of weight $k(\rho)$ and level $N(\rho)$. In Chapter 2 we sketched Edixhoven's proof that if $\ell \nmid N$ then ρ arises from an newform of weight $k(\rho)$ and level N. In this chapter, we introduce some of the techniques used in proving that ρ arises from a newform level $N(\rho)$. For more details, see [**84, 87**].

In [**102**, §1.2] Serre defined the *optimal level* $N(\rho)$ as the prime-to-ℓ part of the Artin conductor of ρ. Recall that $N(\rho)$ is a product $\prod p^{n(p)}$ over prime numbers $p \neq \ell$. The integer $n(p)$ is defined by restricting ρ to a decomposition group D_p at p. Consider the sequence of ramification groups $G_0 \supset G_1 \supset \cdots \supset G_i \supset \cdots$ where G_0 is the inertia subgroup I_p of D_p. Let V be a vector space over $\overline{\mathbf{F}}_\ell$ affording the representation ρ, and for each $i \geq 0$ let V_i be the subspace of V consisting of those $v \in V$ that are fixed by G_i. Then

$$n(p) := \sum_{i=0}^{\infty} \frac{1}{(G_0 : G_i)} \dim V/V_i.$$

3.1. Reduction to weight 2

The optimal level $N(\rho)$ is not divisible by ℓ. The first step in level optimization is to strip the power of ℓ from N. When ℓ is odd, this is done explicitly in [**87**, §2]; for the case $\ell = 2$ see [**9**, §1]. Many of the arguments and key ideas are due to Serre [**94**]. This proof that ℓ can be stripped from the level uses concrete techniques of Serre [**95**, §3], [**98**, Thm. 5.4], and Queen [**78**, §3]; it involves multiplying f by suitable Eisenstein series and taking traces. Katz's theory of ℓ-adic modular forms suggests an alternative method. A classical form of weight 2 and level $M\ell^m$ is an ℓ-adic form of level M; the mod ℓ reduction of this form is classical of level M and some weight, and is congruent to f. See the appendices of [**60**] and the discussions in [**49**, §1] and [**50**, §1].

The next step is to replace f by a newform of weight between 2 and $\ell + 1$ that gives rise to a twist of ρ. Twisting ρ by the mod ℓ cyclotomic character χ preserves N; this is because $\rho \otimes \chi$ arises from $\theta(f) = q\frac{d}{dq}(f)$, which also has level N. Theorem 2.7 asserts that some twist $\rho \otimes \chi^i$ of ρ arises from a form g of weight between 2 and $\ell + 1$. If $\rho \otimes \chi^i$ arises from a newform of level N, then ρ also arises from a newform of the same level, so we can replace f by g and k by the weight of g. By results discussed in Chapter 2, we may assume that $k = k(\rho \otimes \chi^i)$. For the case $\ell = 2$ see [**9**, Prop. 1.3(a)].

We have reduced to considering a representation ρ that arises from a newform f of weight $k(\rho)$ and level N not divisible by ℓ. The weight satisfies $2 \leq k(\rho) \leq \ell+1$, but N need not equal $N(\rho)$. That N is a multiple of $N(\rho)$ is a theorem proved by both Carayol [**12**] and Livné [**70**, Prop. 0.1].

In order to lower N it is convenient to work systematically with form of weight 2. Paradoxically, even though we have just taken all powers of ℓ out of N, we are now going to allow one power of ℓ back into N. This allows us to reduce to weight 2 and realize ρ as a group of torsion points on an abelian variety. An alternative approach (see [**41, 57**]) is to avoid this crutch and work directly with representations coming from arbitrary weights between 2 and $\ell+1$; these are realized in étale cohomology groups. This later approach has the advantage that $X_0(N)$ has good reduction at ℓ.

Reduction to weight 2 is accomplished using a general relationship that originates with ideas of Koike and Shimura. In characteristic ℓ, eigenforms of level N whose weights satisfy $2 < k \leq \ell+1$ correspond to eigenforms of weight 2 and level ℓN (see [**87**, Thm. 2.2]):

$$\left\{ \ 2 < k \leq \ell+1, \text{ level } N \ \right\} \leftrightsquigarrow \left\{ \ k = 2, \text{ level } \ell N \ \right\}.$$

Thus we can and do work with weight 2 and level

$$N^* := \begin{cases} N & \text{if } k = 2, \\ N\ell & \text{if } k > 2. \end{cases}$$

3.2. Geometric realization of Galois representations

To understand representations arising from modular forms, it is helpful to realize these representations inside of geometric objects such as $J := J_1(N^*)$. These representations are constructed geometrically with the help of the Hecke algebra

$$\mathbf{T} := \mathbf{Z}[\ldots T_n \ldots],$$

which was defined in Section 2.3. Recall that $\mathbf{T}$ is a commutative subring of $\operatorname{End}_{\mathbf{Q}} J$ that is free as a module over $\mathbf{Z}$, and that its rank is equal to the dimension of J. When N is cube free, $\mathbf{T}$ is an order in a product of integer rings of number fields; this is a result of Coleman and Edixhoven (see [**16**, Thm. 4.1]). In contrast, the Hecke operators T_p, for $p^3 \mid N$, are usually not semisimple (see Exercise 3).

It is fruitful to view a newform f as a homomorphism

$$\mathbf{T} \to \mathcal{O} = \mathbf{Z}[\ldots a_n \ldots], \qquad T_n \mapsto a_n.$$

Letting $\varphi : \mathcal{O} \to \overline{\mathbf{F}}_\ell$ be the map sending a_p to $\operatorname{tr}(\rho(\operatorname{Frob}_p)) \in \overline{\mathbf{F}}_\ell$, we obtain an exact sequence $0 \to \mathfrak{m} \to \mathbf{T} \to \overline{\mathbf{F}}_\ell$ with $\mathfrak{m}$ a maximal ideal.

Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be an irreducible Galois representation that arises from a weight-2 newform f. The next step, after having attached a maximal ideal $\mathfrak{m}$ to f and φ, is to find a $\mathbf{T}/\mathfrak{m}$-vector space affording ρ inside of the group of ℓ-torsion points of J. Following [**71**, §II.7], we consider the $\mathbf{T}/\mathfrak{m}$-vector space

$$J[\mathfrak{m}] := \{P \in J(\overline{\mathbf{Q}}) : tP = 0 \text{ all } t \in \mathfrak{m}\} \subset J(\overline{\mathbf{Q}})[\ell] \approx (\mathbf{Z}/\ell\mathbf{Z})^{2g}.$$

Since the endomorphisms in $\mathbf{T}$ are $\mathbf{Q}$-rational, $J[\mathfrak{m}]$ comes equipped with a linear action of $G_{\mathbf{Q}}$.

That $\mathrm{tr}(\rho(\mathrm{Frob}_p))$ and $\det(\rho(\mathrm{Frob}_p))$ both lie in the subfield $\mathbf{T}/\mathfrak{m}$ of $\overline{\mathbf{F}}_\ell$ suggests that ρ has a model over $\mathbf{T}/\mathfrak{m}$, in the sense that ρ is equivalent to a representation taking values in $\mathrm{GL}(2, \mathbf{T}/\mathfrak{m}) \subset \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$.

Lemma 3.1. *The representation ρ has a model $\rho_\mathfrak{m}$ over the finite field $\mathbf{T}/\mathfrak{m}$.*

Proof. This is a classical result of I. Schur. Brauer groups of finite fields are trivial (see e.g., [**100**, X.7, Ex. a]), so the argument of [**99**, §12.2] proves the lemma.

Alternatively, when the residue characteristic ℓ of $\mathbf{T}/\mathfrak{m}$ is odd, the following more direct proof can be used. Complex conjugation acts through ρ as a matrix with distinct $\mathbf{F}_\ell$-rational eigenvalues; another well known theorem of Schur [**90**, IX a] (cf. [**116**, Lemme I.1]) then implies that ρ can be conjugated into a representation with values in $\mathrm{GL}(2, \mathbf{T}/\mathfrak{m})$. □

3.3. Multiplicity one

Let $V_\mathfrak{m}$ be a vector space affording $\rho_\mathfrak{m}$. Under the assumption that $\rho_\mathfrak{m}$ is absolutely irreducible, Boston, Lenstra, and Ribet (see [**6**]) proved that $J[\mathfrak{m}]$ is isomorphic as a $G_\mathbf{Q}$-module to a sum of copies of $V_\mathfrak{m}$:

$$J[\mathfrak{m}] \approx \bigoplus_{i=1}^{t} V_\mathfrak{m}.$$

The number of copies of $V_\mathfrak{m}$ is called the *multiplicity* of $\mathfrak{m}$. When ℓ is odd, the hypothesis of irreducibility of $\rho_\mathfrak{m}$ is equivalent to absolute irreducibility (see Exercise 3).

Proposition 3.2. *The multiplicity t is at least* 1.

Proof. Let $\mathbf{T} \subset \mathrm{End}(J)$ be the Hecke algebra associated to J. Because $\mathbf{T} \otimes \mathbf{Z}_\ell$ is an algebra of finite rank over the local ring $\mathbf{Z}_\ell$, we have a decomposition

$$\mathbf{T} \otimes \mathbf{Z}_\ell = \bigoplus_{\lambda | \ell} \mathbf{T}_\lambda,$$

where λ runs through the maximal ideals of $\mathbf{T}$ lying over ℓ, and $\mathbf{T}_\lambda$ denotes the completion of $\mathbf{T}$ at λ (see, e.g., [**37**, Cor. 7.6]). The Tate module

$$\mathrm{Tate}_\ell\, J := \mathrm{Hom}(\mathbf{Q}_\ell/\mathbf{Z}_\ell, \cup_{n\geq 1} J[\ell^n]) \cong \varprojlim J[\ell^n]$$

is a free $\mathbf{Z}_\ell$-module of rank equal to twice the dimension of J. For each maximal ideal λ of $\mathbf{T}$ lying over ℓ, let $e_\lambda \in \mathbf{T} \otimes \mathbf{Z}_\ell$ denote the corresponding idempotent; thus $e_\lambda^2 = e_\lambda$ and $\sum_{\lambda|\ell} e_\lambda = 1$. The map $x \mapsto \sum_\lambda e_\lambda x$ gives a decomposition

$$\mathrm{Tate}_\ell\, J \xrightarrow{\cong} \bigoplus_{\lambda|\ell} e_\lambda \,\mathrm{Tate}_\ell\, J.$$

The ring $\mathrm{End}(J) \otimes \mathbf{Z}_\ell$ operates faithfully on $\mathrm{Tate}_\ell\, J$ (see, e.g., [**74**, Lem. 12.2]), so each summand $e_\lambda \,\mathrm{Tate}_\ell\, J$ is nonzero. Set

$$\mathrm{Tate}_\lambda\, J := \mathrm{Hom}(\mathbf{Q}_\ell/\mathbf{Z}_\ell, \cup_{n\geq 1} J[\lambda^n]).$$

We claim that $\mathrm{Tate}_\lambda\, J$ is identified with $e_\lambda \,\mathrm{Tate}_\ell\, J$ under the natural inclusion $\mathrm{Tate}_\lambda\, J \subset \mathrm{Tate}_\ell\, J$. Denote by $\tilde{\lambda}$ the maximal ideal in $\mathbf{T} \otimes \mathbf{Z}_\ell$ generated by λ. Let n be a positive integer, and let I be the ideal in $\mathbf{T}_\lambda$ generated by ℓ^n. Because $\mathbf{T}_\lambda$ is a local ring with maximal ideal $\tilde{\lambda}$, there is an integer m such that $\tilde{\lambda}^m \subset I$. Since I is principal and generated by ℓ^n, and $\mathbf{T}$ acts on $e_\lambda J[\ell^n]$ through $\mathbf{T}_\lambda$, we have

$$e_\lambda J[\ell^n] = (e_\lambda J[\ell^n])[I] \subset (e_\lambda J[\ell^n])[\tilde{\lambda}^m] \subset (e_\lambda J[\ell^n])[\lambda^m] \subset J[\lambda^m].$$

This shows that $e_\lambda \operatorname{Tate}_\ell J \subset \operatorname{Tate}_\lambda J$. Next suppose $\lambda' \neq \lambda$ and let n be a positive integer. Since $\mathbf{T}_\lambda$ acts on $J[\lambda^n]$ through $\mathbf{T}/\lambda^n = \mathbf{T}_\lambda/\tilde{\lambda}^n$, we have $e_{\lambda'} J[\lambda^n] = 0$, so

$$J[\lambda^n] = \sum_{\text{all } \lambda'} e_{\lambda'} J[\lambda^n] = e_\lambda J[\lambda^n].$$

The other inclusion $\operatorname{Tate}_\lambda J = e_\lambda \operatorname{Tate}_\lambda J \subset e_\lambda \operatorname{Tate}_\ell J$, which we need to prove equality, then follows.

We apply the above conclusion with $\lambda = \mathfrak{m}$. Since $\operatorname{Tate}_\mathfrak{m} J \neq 0$, some $J[\mathfrak{m}^r]$ is nonzero; let r be the smallest such integer. Following [**71**, p. 112], observe that for each generating set of elements $a_1, \ldots, a_t$ of the $\mathbf{T}/\mathfrak{m}$-vector space $\mathfrak{m}^{r-1}/\mathfrak{m}^r$, the map $x \mapsto a_1 x \oplus \cdots \oplus a_t x$ is an injection of the module $J[\mathfrak{m}^r]/J[\mathfrak{m}^{r-1}]$ into the direct sum of t copies of $J[\mathfrak{m}]$. Thus $J[\mathfrak{m}]$ is nonzero. □

The special case $t = 1$, in which the multiplicity is one, plays a central role in the development of the theory. A detailed summary of multiplicity one results can be found in [**32**, §9], and some supplementary results are contained in [**117**, Thm. 2.1]. In general, the multiplicity can be greater than one (see [**72**, §13] and [**63**]).

3.3.1. Multiplicity one representations

Let $\rho : G_\mathbf{Q} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be an irreducible modular Galois representation such that

$$2 \leq k(\rho) \leq \ell + 1.$$

Consider pairs (N, α) where $N \geq 1$ is an integer with the property that $\ell \nmid N$ if $k(\rho) = 2$ and $\ell \mid\mid N$ if $k(\rho) > 2$, together with maps $\alpha : \mathbf{T}_N \to \overline{\mathbf{F}}_\ell$, such that $\alpha(T_p) = \operatorname{tr}(\rho(\operatorname{Frob}_p))$ and $\alpha(p\langle p\rangle) = \det(\rho(\operatorname{Frob}_p))$ for almost all p. Here $\mathbf{T}_N$ is the Hecke algebra associated to $S_2(\Gamma_1(N))$. Note that if (N, α) is such a pair and $\mathfrak{m} = \ker(\alpha)$, then

$$\rho \approx \rho_\mathfrak{m} \otimes_{\mathbf{T}/\mathfrak{m}} \overline{\mathbf{F}}_\ell,$$

where $\alpha : \mathbf{T}/\mathfrak{m} \hookrightarrow \overline{\mathbf{F}}_\ell$ and $\rho_\mathfrak{m}$ is the unique (up to isomorphism) semisimple representation over $\overline{\mathbf{F}}_\ell$ such that

$$\operatorname{tr}(\rho_\mathfrak{m}(\operatorname{Frob}_p)) = \alpha(T_p) \qquad \det(\rho_\mathfrak{m}(\operatorname{Frob}_p)) = \alpha(p\langle p\rangle)$$

for almost all p.

Definition 3.3. ρ is a multiplicity one representation if $J_1(N)[\ker\alpha]$ has dimension 2 for all pairs (N, α) as above.

Remark 3.4.
1. If $J_1(N)[\ker\alpha]$ has dimension 2 then $\rho_\mathfrak{m} = J_1(N)[\ker\alpha]$ by Eichler-Shimura, see [**6**].
2. The definition extends to arbitrary modular Galois representations ρ as follows. As explained in Section 2.2, every ρ has a twist $\rho \otimes \chi^i$ by some power of the cyclotomic character such that $k(\rho \otimes \chi^i) \leq \ell + 1$. We say that ρ is a *multiplicity one representation* if $\rho \otimes \chi^i$ is a multiplicity one representation.

3.3.2. Multiplicity one theorems

Techniques for proving multiplicity one results were pioneered by Mazur in [**71**] who considered $J_0(p)$ with p prime. Let f be an eigenform and fix a nonzero prime λ of the ring generated by the Fourier coefficients of f such that $\rho_{f,\lambda}$ is absolutely irreducible. View the Hecke algebra $\mathbf{T}$ as a subring of $\mathrm{End}(J_0(p))$, and let $\mathfrak{m}$ be the maximal ideal associated to f and λ. Let $V_{\mathfrak{m}}$ again be a two-dimensional $\mathbf{T}/\mathfrak{m}$-vector space that affords $\rho_{\mathfrak{m}} : G_{\mathbf{Q}} \to \mathrm{GL}(2, \mathbf{T}/\mathfrak{m})$. Mazur proved (see Prop. 14.2, ibid.) that $J[\mathfrak{m}] \approx V_{\mathfrak{m}}$, except perhaps when $\mathfrak{m}$ is ordinary of residue characteristic $\ell = 2$. The missing ordinary case can be treated under suitable hypothesis. If $\rho_{\mathfrak{m}}$ restricted to a decomposition group at 2 is not contained in the scalar matrices, then $J[\mathfrak{m}] \approx V_{\mathfrak{m}}$ (see, e.g., [**9**, Prop. 2.4]). The results of Mazur are extended in [**72**] and [**84**, §5].

Theorem 3.5. *An irreducible modular Galois representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is a multiplicity one representation, except perhaps when all of the following hypothesis on ρ are simultaneously satisfied:*

— *$k(\rho) = \ell$;*
— *ρ is unramified at ℓ;*
— *ρ is ordinary at ℓ;*
— *$\rho|_{D_\ell} \sim \left(\begin{smallmatrix} \alpha & * \\ 0 & \beta \end{smallmatrix}\right)$ with $\alpha = \beta$.*

Proof. See [**32**, §9], [**117**, Thm. 2.1], and [**9**, Prop. 2.4] for the case $\ell = 2$. □

In [**46**, §12] Gross proves multiplicity one when $\alpha \neq \beta$, $k(\rho) \leq \ell$, and ρ is ordinary; he uses this result in his proof of the existence of companion forms. In contrast, Coleman and Voloch [**17**] prove the existence of companion forms when $\alpha = \beta$ and $\ell > 2$ using a method that avoids the need for multiplicity one.

Remark 3.6. L. Kilford of London, England has recently discovered an example at prime level 503 in which multiplicity one fails. Let E_1, E_2, and E_3 be the three elliptic curves of conductor 503, and for each $i = 1, 2, 3$, let $\mathfrak{m}_i$ be the maximal ideal of $\mathbf{T} \subset \mathrm{End}(J_0(503))$ generated by 2 and all $T_p - a_p(E_i)$, with p prime. Each of the Galois representations $E_i[2]$ is irreducible, and one can check that $\mathfrak{m}_1 = \mathfrak{m}_2 = \mathfrak{m}_3$. If multiplicity one holds, then $E_1[2] = E_2[2] = E_3[2]$ inside of $J_0(503)$. However, this is not the case, as a modular symbols computation in the integral homology $H_1(X_0(N), \mathbf{Z})$ reveals that $E_1 \cap E_2 = \{0\}$.

3.3.3. Multiplicity one for mod 2 representations

For future reference, we now wish to consider multiplicity one in the following rather extreme situation. Suppose that $\ell = 2$, and let ρ be a mod ℓ representation arising from a form of weight either 2 or 3. If the weight is 3 then ρ is not finite at 2; this can be used to deduce multiplicity one by adapting the arguments of [**72**] (see the proof of [**9**, Prop. 2.4]). When the weight is 2, we have the following proposition.

Proposition 3.7. *Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_2)$ be an irreducible Galois representation that arises from a weight-2 form $f = \sum a_n q^n$ on $\Gamma = \Gamma_1(N) \cap \Gamma_0(2)$ with N odd, and let ε be the character of f. If $\overline{a}_2^2 \not\equiv \overline{\varepsilon}(2) \in \overline{\mathbf{F}}_2$, then ρ is a multiplicity one representation.*

Proof. Let $\mathfrak{m}$ be the maximal ideal associated to f in the Hecke algebra $\mathbf{T}$ attached to Γ. Because the weight of f is 2, the representation ρ is finite at 2. If ρ is supersingular then the inertia group I_2 operates through the two fundamental characters of level 2. These both have order $\ell^2 - 1 = 3 \neq 1$, so ρ is ramified and this can be used to deduce multiplicity one. If ρ is ordinary then $\rho|_{D_2} \sim \left(\begin{smallmatrix} \alpha & * \\ 0 & \beta \end{smallmatrix}\right)$ with β unramified and $\beta(\mathrm{Frob}_2) \equiv T_2 \bmod \mathfrak{m}$. The determinant $\alpha\beta$ of $\rho|_{D_2}$ is $\chi \cdot \varepsilon$ where χ is the mod 2 cyclotomic character and ε is unramified at 2. Since χ, ε, and β are unramified, α is also unramified. Since $\chi(\mathrm{Frob}_2) = 1$ and $\alpha\beta = \chi\varepsilon$, we have $\alpha(\mathrm{Frob}_2) = \beta^{-1}(\mathrm{Frob}_2)\varepsilon(2) = a_2^{-1}\varepsilon(2) \pmod{\mathfrak{m}}$. The further condition, under which we might not know multiplicity one, is $\alpha|_{D_2} = \beta|_{D_2}$; expressed in terms of the image of Frobenius, this becomes $a_2^{-1}\varepsilon(2) \equiv a_2 \pmod{\mathfrak{m}}$, or equivalently, $a_2^2 \equiv \varepsilon(2) \pmod{\mathfrak{m}}$. By hypothesis, this latter condition does not hold. □

3.4. The key case

We have set our problem up so that level optimization pertains to weight-2 forms of appropriate level, and takes place on Jacobians of modular curves. This level optimization problem was described, and partially treated, in a paper of Carayol [**12**]. In this paper, Carayol reduced the problem to the following key case.

Key case: Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be a Galois representation that arises from a weight-2 newform f of level pM, with $p \nmid \ell M$, and character $\varepsilon : (\mathbf{Z}/pM\mathbf{Z})^* \to \mathbf{C}^*$. Assume that ρ is unramified at p, and that ε factors through the natural map $(\mathbf{Z}/pM\mathbf{Z})^* \to (\mathbf{Z}/M\mathbf{Z})^*$. Show that ρ arises from a form of level M.

In the key case, the character ε of f is unramified at p. Thus f, a priori on $\Gamma_1(pM)$, is also on the bigger group $\Gamma_1(M) \cap \Gamma_0(p)$; that is, f lies in $S_2(\Gamma_1(M) \cap \Gamma_0(p))$.

Example 3.8. Consider the representation ρ arising from the 7-division points of the modular elliptic curve A of conductor $N_A = 3 \cdot 47$ and minimal discriminant $\Delta_A = 3^7 \cdot 47$. (The curve A is labeled **141A** in Cremona's notation [**20**].) The newform f corresponding to A is on $\Gamma_0(3{\cdot}47)$. As in Section 1.4, since $\mathrm{ord}_3(\Delta_A) = 7$, the representation ρ is unramified at 3 and $N(\rho) = 47$. To optimize the level means to find a form g on $\Gamma_0(47)$ that gives rise to ρ.

Example 3.9 (Frey curves)**.** The elliptic curves that Frey associated in [**42**] to hypothetical solutions of the Fermat equation $x^\ell + y^\ell = z^\ell$ give rise to mod ℓ Galois representations. According to Wiles's theorem [**117**], there is a weight-2 form f of level $2L$, with L big and square free, that gives rise to ρ. At the same time, $N(\rho) = 2$. Taking p to be any odd prime dividing L, we are put in the key case. If we can optimize the level, then we eventually reach a contradiction and thus deduce Fermat's Last Theorem.

The key case divides into two subcases; the more difficult one occurs when the following conditions are both satisfied:

— $p \equiv 1 \pmod{\ell}$;
— $\rho(\mathrm{Frob}_p)$ is a scalar matrix.

The second condition makes sense because $p \nmid N(\rho)$; since $\det(\rho(\mathrm{Frob}_p)) = \chi^{k-1}\varepsilon$, we know the scalar up to ± 1. The complementary case is easier; it can be treated using "Mazur's principle" (see Section 3.9). Though Example 3.8 falls into the easier case because $3 \not\equiv 1 \pmod 7$, the proof of Fermat's Last Theorem requires level optimization in both cases.

Consider a modular representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ that arises from a newform of level N and weight $k = k(\rho)$, and assume that $\ell \nmid N$. The goal of level optimization is to show that there is a newform of Serre's optimal level $N(\rho)$ that gives rise to ρ.

As discussed in Section 3.1, ρ arises from a newform $f = \sum a_n q^n$ on $\Gamma_1(N^*)$ of weight 2 and some character ε. Thus there is a homomorphism φ from $\mathcal{O} = \mathbf{Z}[\ldots a_n \ldots]$ to $\overline{\mathbf{F}}_\ell$ such that $\varphi(a_p) = \mathrm{tr}(\rho(\mathrm{Frob}_p))$ for all $p \nmid \ell N^*$. Let $\mathbf{T}$ be the Hecke algebra associated to $S_2(\Gamma_1(N^*))$. The maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ associated to ρ is the kernel of the map sending T_n to $\varphi(a_n)$. As was discussed in the previous chapter, the representation ρ is realized geometrically inside the subspace $J[\mathfrak{m}] \subset J[\ell]$ of the ℓ-torsion of the Jacobian J of $X_1(N^*)$.

Problem. Fix a divisor p of $N^*/N(\rho)$. Find a newform whose level is a divisor of N^*/p that also gives rise to ρ.

Lemma 3.10. *Let ρ be as above, and suppose p is a prime such that $p \mid N^*$ but $p \nmid \ell N(\rho)$, so ρ is unramified at p. Let ε_p denote the p part of ε. Then either $\varepsilon_p = 1$ or $p \equiv 1 \pmod \ell$.*

Proof. The character ε is initially defined as a homomorphism $(\mathbf{Z}/N^*\mathbf{Z})^* \to \mathcal{O}^*$; the reduction $\overline{\varepsilon}$ is obtained by composing ε with $\varphi : \mathcal{O} \to \overline{\mathbf{F}}_\ell$. Since ρ is unramified at p, the determinant $\det(\rho) = \chi_\ell^{k-1}\overline{\varepsilon} = \chi_\ell \overline{\varepsilon}$ is also unramified at p. Because χ_ℓ is ramified only at ℓ, the character $\overline{\varepsilon}$ is unramified at p. Let $M = N^*/p^r$ where $r = \mathrm{ord}_p(N^*)$, and write $(\mathbf{Z}/N^*\mathbf{Z})^* \cong (\mathbf{Z}/p^r\mathbf{Z})^* \times (\mathbf{Z}/M\mathbf{Z})^*$. By restricting ε to each factor, we write ε as a product of two characters: $\varepsilon = \varepsilon_p \cdot \varepsilon^{(p)}$ where ε_p is a character of $(\mathbf{Z}/p^r\mathbf{Z})^*$ and $\varepsilon^{(p)}$ is a character of $(\mathbf{Z}/M\mathbf{Z})^*$. The character $\varepsilon^{(p)}$ has conductor dividing M, so it is unramified at p. By class field theory, ε_p is totally ramified at p, so the reduction $\overline{\varepsilon}$ is unramified at p precisely when $\overline{\varepsilon}_p = 1$; equivalently, $\overline{\varepsilon}$ is unramified at p exactly when ε_p has order a power of ℓ. If ε_p is non-trivial, then, since the order of ε_p divides the order $p^{r-1}(p-1)$ of a generator of $(\mathbf{Z}/p^r\mathbf{Z})^*$, a power of ℓ divides $p^{r-1}(p-1)$, so $p \equiv 1 \pmod \ell$ since $\ell \neq p$. □

In addition to his conjectures about the optimal weight and level, Serre also made a conjecture about the optimal character of a form giving rise to ρ. Let p be a prime not dividing $\ell N(\rho)$. Serre's optimal character conjecture implies that ρ, which we know to arise from a form on $\Gamma_1(M) \cap \Gamma_1(p^r)$, arises from a form on $\Gamma_1(M) \cap \Gamma_0(p^r)$, and this has been proved in most cases.

3.5. Approaches to level optimization in the key case

As discussed in Section 3.4, results of Carayol and Livné (see [**12, 70**]) reduce the level optimization problem to the following key case. The weight-2 newform f, a priori on $\Gamma_1(N^*)$, is in fact on the bigger group $\Gamma_1(M) \cap \Gamma_0(p)$, where $Mp = N^*$, $p \nmid M$, and ρ is unramified at p. The goal is to show that ρ arises from a newform on $\Gamma_1(M)$. This has been achieved when ℓ is odd, and in many cases when $\ell = 2$, using several level optimization techniques.

I. **Mazur's principle**
If either $\rho(\mathrm{Frob}_p)$ is not a scalar matrix or $p \not\equiv 1 \pmod{\ell}$, then an argument of Mazur, explained in Section 3.9, can be used to optimize the level.

II. **Multiplicity one**
It is possible to optimize the level if ρ is a multiplicity one representation, as explained in [**84, 9**] and Section 3.11. The cases in which multiplicity one is known were reviewed in Section 3.3. In particular, we do not know multiplicity one in some cases when $k(\rho) = \ell$ and the eigenvalues of Frob_p are not distinct.

III. **Using a pivot**
Suppose that M can be written as a product $M = qK$ with q a prime not dividing pK, that ρ arises from a form on $\Gamma_1(K) \cap \Gamma_0(pq)$, and that ρ is ramified at q and unramified at p. Then q can be used as a "pivot" to remove p from the level. This approach grew out of [**83**], and was introduced in the short paper [**86**]. In Section 3.10 we describe the approach and discuss the terminology.

IV. **Without multiplicity one**
When ℓ is odd and $\varepsilon = 1$, the level optimization theorem was proved in [**87**] using an argument that does not require ρ to have multiplicity one. The hypothesis $\ell \neq 2$ is used in the proof of Proposition 7.8 of [**87**] to force splitting of a short exact sequence. In [**26**], Diamond extended the results of [**87**] to cover the case of arbitrary character, still under the assumption that ℓ is odd. One encounters seemingly insurmountable difficulties in trying to push this argument through when $\ell = 2$.

3.6. Some commutative algebra

In this section we set up some of the commutative algebra that is required in order to lower levels. There are two injective maps

$$\begin{tikzcd} S_2(\Gamma_1(M)) \arrow[r, hook, shift left] \arrow[r, hook, shift right] & S_2(\Gamma_1(M) \cap \Gamma_0(p)) . \end{tikzcd}$$

One is the inclusion $f(q) \mapsto f(q)$ and the other is $f(q) \mapsto f(q^p)$ (see Exercise 18). The *p-new subspace* $S_2(\Gamma_1(M) \cap \Gamma_0(p))^{p\text{-new}}$ is the complement, with respect to the Petersson inner product, of the subspace $\mathcal{S}$ generated by the two images of $S_2(\Gamma_1(M))$. The p-new subspace can also be defined algebraically as the kernel of the natural map from $S_2(\Gamma_1(M) \cap \Gamma_0(p))$ to the direct sum of two copies of $S_2(\Gamma_1(M))$.

Let $\mathbf{T}$ denote the Hecke algebra acting on $S_2(\Gamma_1(M) \cap \Gamma_0(p))$. If $p \nmid M$, then T_p acts on $\mathcal{S}$ as a direct sum of two copies of its action on $S_2(\Gamma_1(M))$; otherwise, T_p usually does not act diagonally (see Exercise 19). The image of $\mathbf{T}$ in $\mathrm{End}(\mathcal{S})$ is a quotient $\overline{\mathbf{T}}$ called the *p-new quotient.* A representation ρ associated to a maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ arises from level M if and only if $\mathfrak{m}$ arises by pullback from a maximal ideal of $\overline{\mathbf{T}}$. Because the map $\mathbf{T} \to \overline{\mathbf{T}}$ is surjective, $\mathfrak{m}$ arises from level M if and only if the image of $\mathfrak{m}$ in$\overline{\mathbf{T}}$ is not the unit ideal (see Exercise 21).

3.7. Aside: Examples in characteristic two

Sections 3.7 and 3.8 can be safely skipped on a first reading.

To orient the reader, we focus for the moment on mod 2 representations that arise from elliptic curves. We give examples in which one of the level optimizations methods applies but the others do not. We do not consider method **IV** because it is not applicable to mod 2 representations. The hypothesis of the "multiplicity one" method **II** when $\ell = 2$ are discussed after the statement of Theorem 3.19 in Section 3.11. We were unable to find an example in which none of the level optimization theorems applies.

We will repeatedly refer to the following theorem, which first appeared in [**85**].

Theorem 3.11. *Suppose ρ arises from a newform in $S_2(\Gamma_0(N))$. Let $p \nmid \ell N$ be a prime satisfying one or both of the identities*

$$\operatorname{tr}\rho(\mathrm{Frob}_p) = \pm(p+1) \pmod{\ell}.$$

Then ρ arises from a newform of level pN.

3.7.1. III applies but I and II do not

In this section we give a mod 2 representations in which the pivot hypothesis of **III** is satisfied, but the hypotheses of **I** and **II** are not. Our example is obtained by applying Theorem 3.11 to the mod 2 representation attached to a well-chosen elliptic curve.

We will find an elliptic curve E of conductor $M = qR$ such that $\rho = E[2]$ is absolutely irreducible, ramified at q, unramified at 2, and $\rho(\mathrm{Frob}_2) = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. Because of the last condition, [**9**, Prop. 2.4] does not imply that ρ is a multiplicity one representation, so **II** does not apply. (In fact, following Remark 3.6, one sees that ρ is not a multiplicity one representation.) Likewise, **I** does not apply because $\rho(\mathrm{Frob}_2)$ is a scalar and the p we will chose will satisfy $p \equiv 1 \pmod 2$. Next we choose a prime $p \nmid 2qR$ such that $\rho_{E,2}(\mathrm{Frob}_p) = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. Let f be the newform associated to E. By Theorem 3.11 there is a newform g of level pqR such that

$$\rho_{g,\lambda} \approx \rho_{E,2}.$$

In particular,

$$\rho_{g,\lambda}(\mathrm{Frob}_p) = \rho_{E,2}(\mathrm{Frob}_p) = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$$

is scalar and $p \equiv 1 \pmod 2$, so **I** does not apply. However, method **III** does apply with q used as a pivot.

For example, consider the elliptic curve E defined by the equation

$$y^2 + xy = x^3 - x^2 + 19x - 32.$$

The conductor of E is $N = 19 \cdot 109$, and the discriminant of the field $K = \mathbf{Q}(E[2])$ is $-19^3 \cdot 109^3$. We select $q = 19$ as our pivot. The prime $p = 73$ splits completely in K, so

$$\rho_{E,2}(\mathrm{Frob}_p) = \begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

By Theorem 3.11 there is a form g of level $109 \cdot 19 \cdot 73$ that is congruent to the newform f attached to E modulo a prime lying over 2. Method **III** can be used to optimize the level, but neither method **I** nor **II** applies.

3.7.2. II applies but I and III do not

We exhibit a mod 2 representation for which method **II** can be used to optimize

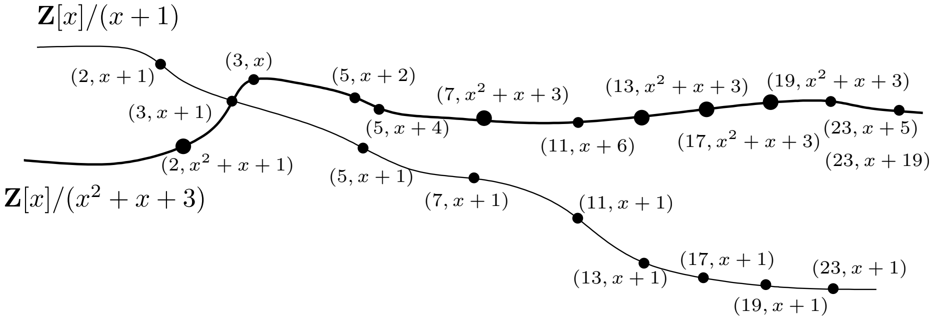


Figure 1. The spectrum of $\mathbf{T} \subset \mathrm{End}(S_2(\Gamma_0(33)))$, with $x = T_3$

the level, but neither method **I** nor **III** applies. Let K be the $\mathrm{GL}_2(\mathbf{F}_2)$-extension of $\mathbf{Q}$ obtained by adjoining all cube roots of 2. Then $K = \mathbf{Q}(E[2])$, where E is the elliptic curve $X_0(27)$ given by the equation $y^2 + y = x^3 - 7$. The prime $p = 31$ splits completely in K, so by Theorem 3.11 there is a newform f of level $31 \cdot 27$ and a maximal ideal λ of the appropriate Hecke algebra such that $\rho_{f,\lambda} \approx E[2]$. Neither method **I** nor **III** can be used to optimize the level of $\rho_{f,\lambda}$. Method **I** doesn't apply because 31 is odd and $\rho_{f,\lambda}(\mathrm{Frob}_{31}) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$; method **III** doesn't apply because the only odd prime that is ramified in K is 3, which does not exactly divide $31 \cdot 27$. If D_2 is a decomposition group at 2 then D_2 has image in $\mathrm{GL}_2(\mathbf{F}_2)$ of order 2, so it is not contained in the scalar matrices and **II** can be used to optimize the level of $\rho_{f,\lambda}$.

3.8. Aside: Sketching the spectrum of the Hecke algebra

It is helpful to understand the Hecke algebra geometrically using the language of schemes (see, e.g., [**38**]). The topological space underlying the scheme $\mathrm{Spec}(\mathbf{T})$ is the set of prime ideals of $\mathbf{T}$ endowed with the Zariski topology, in which the closed sets are the set of prime ideals containing a fixed ideal.

We can draw $\mathrm{Spec}(\mathbf{T})$ by sketching a diagram whose irreducible components correspond to the Galois conjugacy classes of eigenforms, and whose intersections correspond to congruences between eigenforms. When the level is not cube free, $\mathbf{T}$ can contain nilpotent elements, and then one might wish to include additional information. If $\sum a_n q^n$ is an eigenform, then the failure of $\mathbf{Z}[\ldots a_n \ldots]$ to be integrally closed can be illustrated by drawing singular points on the corresponding irreducible component; however, we do not do this below.

Example 3.12. The spectrum of the Hecke algebra associated to $\Gamma_0(33)$ is illustrated in Figure 1. The Hecke algebra $\mathbf{T} \subset S_2(\Gamma_0(33))$ has discriminant -99, as does the characteristic polynomial of T_3, so

$$\mathbf{T} = \mathbf{Z}[T_3]/((T_3+1)(T_3^2+T_3+3)) \cong \mathbf{Z}[x]/((x+1)(x^2+x+3)).$$

We sketch a curve corresponding to each of the two irreducible components. Some of the closed points (maximal) ideals are represented as dots. One component corresponds to the unique newform on $\Gamma_0(33)$, and the other corresponds to the two images of the newform on $\Gamma_0(11)$.

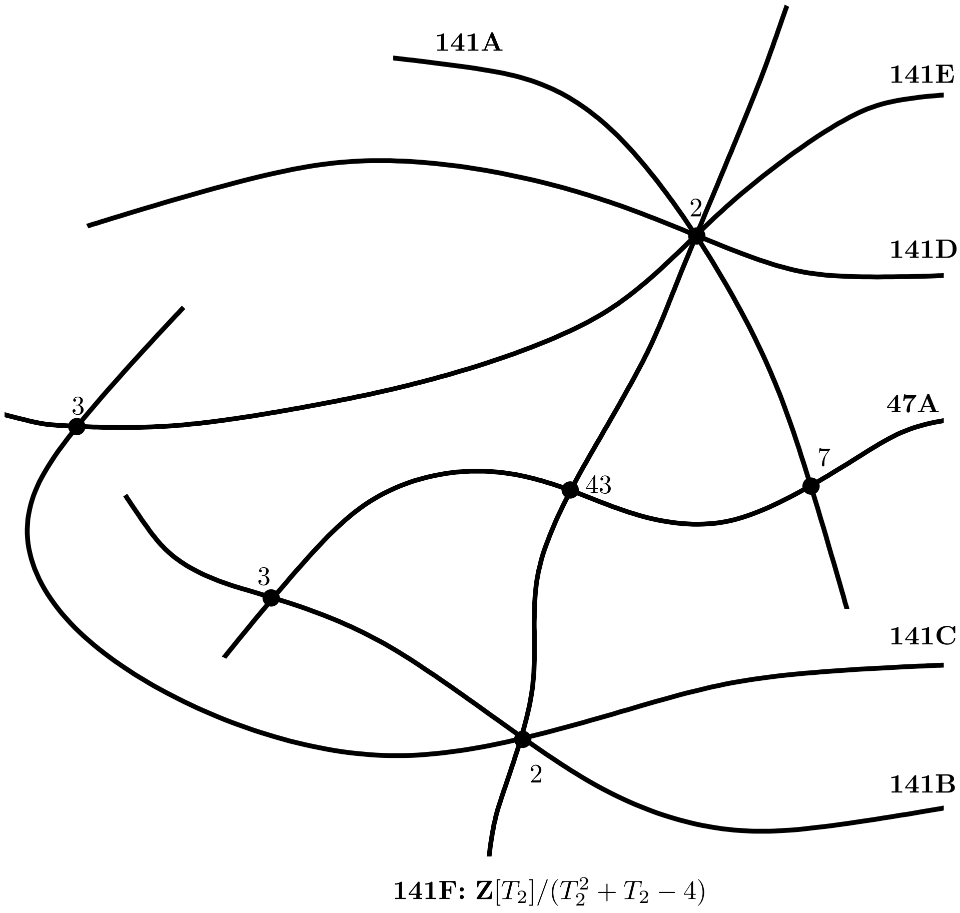


Figure 2. The spectrum of $\mathbf{T} \subset \mathrm{End}(S_2(\Gamma_0(141)))$

Example 3.13. Figure 2 is a diagram of the Hecke algebra associated to $S_2(\Gamma_0(3 \cdot 47))$. We have labeled fewer closed points than in Figure 1. The components are labeled by their isogeny class and the level at which they are new (the notation extends that of [**20**]). The component labeled **141F** corresponds to an eigenform whose Fourier coefficients generate a quadratic extension of $\mathbf{Q}$.

The newform corresponding to the elliptic curve A from Example 3.8 is labeled **141A**. Geometrically, the assertion that the level of $\rho_{A,7}$ can be optimized is represented by the characteristic-7 intersection between the component labeled **141A** and the old component **47A** coming from the unique Galois conjugacy class of newforms on $\Gamma_0(47)$.

3.9. Mazur's principle

A principle due to Mazur can be used to optimize the level in the key case, provided that a mild hypothesis is satisfied. The principle applies whenever $p \not\equiv 1 \pmod{\ell}$ and also in the case when $p \equiv 1 \pmod{\ell}$ but $\rho(\mathrm{Frob}_p)$ is not a scalar. This principle first appeared in [**84**, §6], then in [**26**, §4], and most recently when $\ell = 2$ in [**9**, pg. 7].

Theorem 3.14 (Mazur's Principle). *Suppose that $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ arises from a newform f of weight 2 and level Mp, with $p \nmid M$, and character ε of conductor dividing M. Assume that ρ is unramified at p and that either $\rho(\mathrm{Frob}_p)$ is not a scalar matrix or $p \not\equiv 1 \pmod{\ell}$. Then ρ arises from a modular of level dividing M.*

We will require the following basic fact later in the proof.

Lemma 3.15 (Li). *Let $f = \sum a_n q^n$ be a newform on $\Gamma_1(M) \cap \Gamma_0(p)$ of weight k. Then $a_p^2 = \varepsilon(p) p^{k-2}$.*

Proof. Li's proof is an easy application of her generalization to Γ_1 of the Atkin-Lehner theory of newforms [**69**, Thm. 3(iii)]. The newform f is an eigenvector for the operator W_p which is defined on $S_k(\Gamma_1(M) \cap \Gamma_0(p))$ by

$$W_p(f) = p^{k/2} f\left(\frac{apz + b}{Mpz + p}\right),$$

where a and b are integers such that $ap^2 - bMp = p$. By [**69**, Lem. 3],

$$g := T_p(f) + p^{k/2-1} W_p(f)$$

lies in $S_k(\Gamma_1(M))$. For all primes $q \nmid Mp$, the eigenvalue of T_q on the oldform g is the same as the eigenvalue of T_q on the newform f, so $g = 0$. By [**69**, Lem. 2] ${W_p}^2(f) = \varepsilon(p) f$, so $a_p^2 = \varepsilon(p) p^{k-2}$. □

Remark 3.16. The case of Lemma 3.15 that we will need can also be understood in terms of the local representation $\rho|_{G_p}$, which resembles the mod ℓ representation attached to a Tate curve, in the sense that $\rho|_{G_p} \sim \left(\begin{smallmatrix} \alpha\chi & * \\ 0 & \alpha \end{smallmatrix}\right)$. Our hypothesis include the assumption that ρ is unramified at p, so the two characters $\alpha\chi$ and α are unramified at p. Thus $\alpha(\mathrm{Frob}_p)$ makes sense; we have $\alpha(\mathrm{Frob}_p) = \overline{a}_p(f)$ and $\alpha\chi(\mathrm{Frob}_p) = \overline{a}_p(f)p$. Since $\det(\rho|_{G_p}) = \alpha^2\chi = \overline{\varepsilon}\chi$, we see that

$$\overline{a}_p^2 = \overline{\varepsilon}(p).$$

This local analysis of ρ was vastly generalized by Langlands in [**67**], which extends the analysis to include many ℓ-adic representations of possibly higher weight. See also [**13**].

Let $\mathbf{T}$ be the Hecke algebra associated to $\Gamma_1(M) \cap \Gamma_0(p)$, and let $\mathfrak{m}$ be the kernel of the following map $\mathbf{T} \to \overline{\mathbf{F}}_\ell$:

$$0 \longrightarrow \mathfrak{m} \longrightarrow \mathbf{T} \xrightarrow{T_n \mapsto \overline{a}_n,\ \langle d \rangle \mapsto \overline{\varepsilon}(d)} \overline{\mathbf{F}}_\ell.$$

As in Lemma 3.1, the determinants and traces of elements in the image of $\rho = \rho_{\mathfrak{m}}$ lie in $\mathbf{T}/\mathfrak{m} \subset \overline{\mathbf{F}}_\ell$, so there is a vector space $V \approx (\mathbf{T}/\mathfrak{m})^{\oplus 2}$ that affords $\rho_{\mathfrak{m}}$.

Next we realize $\rho_{\mathfrak{m}}$ as a group of division points in a Jacobian. The curve $X_1(Mp)$ corresponding to $\Gamma_1(Mp)$ covers the curve $X_1(M,p)$ corresponding to $\Gamma_1(M) \cap \Gamma_0(p)$. The induced map $J = \mathrm{Jac}(X_1(M,p)) \to J_1(Mp) = \mathrm{Jac}(X_1(Mp))$ has a finite kernel on which the Galois action is abelian.

Just as in Section 2.3.1.1, the Hecke algebra associated to $\Gamma_1(M) \cap \Gamma_0(p)$, can be constructed as a ring of correspondences on $X_1(M,p)$, then viewed as a subring $\mathbf{T} \subset \mathrm{End}_{\mathbf{Q}}(J)$. Inside of J we find the nonzero $G_{\mathbf{Q}}$-module $J[\mathfrak{m}] \approx \oplus_{i=1}^t V$. For the purposes of this discussion, we do not need to know that $J[\mathfrak{m}]$ is a direct sum

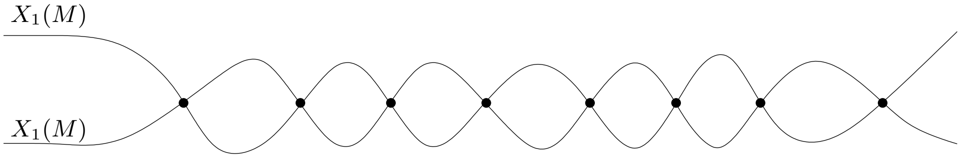


Figure 3. The reduction mod p of the Deligne-Rapoport model of $X_1(M,p)$

of copies of V. The following weaker assertion, known long ago to Mazur [**71**, §14, pg. 112], will suffice: $J[\mathfrak{m}]$ *is a successive extension of copies of* V. In particular, $V \subset J[\mathfrak{m}]$. A weaker conclusion, true since $\ell \in \mathfrak{m}$, is that $V \subset J[\ell]$,

Our hypothesis that ρ is unramified at p translates into the inclusion $V \subset J[\ell]^{I_p}$, where I_p is an inertia group at p. By [**104**, Lem. 2], if A is an abelian variety over $\mathbf{Q}$ with good reduction at p, then $A[\ell]^{I_p} \cong A_{\mathbf{F}_p}[\ell]$. However, the modular curve $X_1(M,p)$ has bad reduction at p, so J is likely to have bad reduction at p—in this case it does. We are led to consider the Néron model $\mathcal{J}$ of J (see, e.g., [**5**]), which is a smooth commutative group scheme over $\mathbf{Z}$ satisfying the following property: the restriction map $\mathrm{Hom}_{\mathbf{Z}}(\mathcal{S},\mathcal{J}) \longrightarrow \mathrm{Hom}_{\mathbf{Q}}(\mathcal{S}_{\mathbf{Q}},J)$ is bijective for all smooth schemes $\mathcal{S}$ over $\mathbf{Z}$. Passing to the scheme-theoretic closure, we have, inside of $\mathcal{J}$, a two-dimensional $\mathbf{T}/\mathfrak{m}$-vector space scheme $\mathcal{V}$.

In Section 2.3.1.1 we only defined $X_1(M,p)$ as a scheme over $\mathbf{Z}[1/Mp]$. Deligne and Rapoport [**25**] extended $X_1(M,p)$ to a scheme over $\mathbf{Z}[1/M]$ and computed the reduction modulo p. The introduction to [**62**] contains a beautiful historical discussion of the difficulties involved in extending modular curves over $\mathbf{Z}$.

We know a great deal about the reduction of $X_1(M,p)$ at p, which is frequently illustrated by the squiggly diagram in Figure 3. This reduction is the union of 2 copies of $X_1(M)_{\mathbf{F}_p}$ intersecting transversely at the supersingular points.

The subspace $S_2(\Gamma_1(M)) \oplus S_2(\Gamma_1(M))$ of $S_2(\Gamma_1(M) \cap \Gamma_0(p))$ is stable under the Hecke algebra $\mathbf{T}$, so there is a map $\mathbf{T} \to \mathrm{End}(S_2(\Gamma_1(M)) \oplus S_2(\Gamma_1(M)))$. The *p-old quotient* of $\mathbf{T}$ is the image $\overline{\mathbf{T}}$. Since the map $\mathbf{T} \to \overline{\mathbf{T}}$ is surjective, the image of $\mathfrak{m}$ in $\overline{\mathbf{T}}$ is an ideal $\overline{\mathfrak{m}}$. To optimize the level in the key case amounts to showing that $\overline{\mathfrak{m}}$ is not the unit ideal.

As is well known (cf. [**71**, Appendix, Prop 1.4]), the results of M. Raynaud [**82**] and Deligne-Rapoport [**25**] combine to produce an exact sequence

$$0 \longrightarrow \mathcal{T} \longrightarrow \mathcal{J}^0_{\mathbf{F}_p} \longrightarrow J_1(M)_{\mathbf{F}_p} \times J_1(M)_{\mathbf{F}_p} \longrightarrow 0, \tag{3.1}$$

where $\mathcal{T}$ is a torus, i.e., $\mathcal{T}_{\overline{\mathbf{F}}_p} \approx \mathbf{G}_m \times \cdots \times \mathbf{G}_m$, and $\mathcal{J}^0_{\mathbf{F}_p}$ is the identity component of $\mathcal{J}_{\mathbf{F}_p}$. There is a concrete description of $\mathcal{T}$ and of the maps in the exact sequence. Each object in the sequence is equipped with a functorial action of the Hecke algebra $\mathbf{T}$, and the sequence is $\mathbf{T}$-invariant. The p-old quotient $\overline{\mathbf{T}}$ can be viewed as coming from the action of $\mathbf{T}$ on $J_1(M)_{\mathbf{F}_p} \times J_1(M)_{\mathbf{F}_p}$.

By a generalization of [**104**, Lem. 2], the reduction map $J(\overline{\mathbf{Q}}_p)[\ell]^{I_p} \to \mathcal{J}_{\mathbf{F}_p}(\overline{\mathbf{F}}_p)$ is injective. Thus $V = \mathcal{V}_{\mathbf{F}_p}(\overline{\mathbf{F}}_p) \subset \mathcal{J}_{\mathbf{F}_p}(\overline{\mathbf{F}}_p)$. The component group $\Phi = \mathcal{J}_{\mathbf{F}_p}/\mathcal{J}^0_{\mathbf{F}_p}$ is *Eisenstein*, in the sense that it does not contain irreducible representations arising from eigenforms. Since V is irreducible, as a Galois module Φ does not contain an isomorphic copy of V, so $\mathcal{V}_{\mathbf{F}_p} \subset \mathcal{J}^0_{\mathbf{F}_p}$ and we have the following diagram:

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{T} \arrow[r] & \mathcal{J}^0_{\mathbf{F}_p} \arrow[r] & J_1(M)_{\mathbf{F}_p} \times J_1(M)_{\mathbf{F}_p} \arrow[r] & 0. \\
& & \mathcal{V}_{\mathbf{F}_p} \arrow[u, hook] \arrow[ul, dotted] \arrow[ur] & &
\end{tikzcd}$$

Since $\mathfrak{m}$ acts as 0 on V, the image $\overline{\mathfrak{m}}$ of $\mathfrak{m}$ acts as 0 on the image of V in $J_1(M)_{\mathbf{F}_p} \times J_1(M)_{\mathbf{F}_p}$. If $\overline{\mathfrak{m}} \neq (1)$ then we can optimize the level, so assume $\overline{\mathfrak{m}} = (1)$. Then the image of V in $J_1(M)_{\mathbf{F}_p} \times J_1(M)_{\mathbf{F}_p}$ is 0, so $V_{\mathbf{F}_p} \hookrightarrow \mathcal{T}$.

Let $\mathcal{X}_p(J) := \mathrm{Hom}(\mathcal{T}, \mathbf{G}_m)$ be the *character group* of $\mathcal{T}$. The action of $\mathbf{T}$ on $\mathcal{T}$ induces an action of $\mathbf{T}$ on $\mathcal{X}_p(J)$. Furthermore, $\mathcal{X}_p(J)$ supports an action of $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ which, because tori split over a quadratic extension, factors through the Galois group of $\mathbf{F}_{p^2}$. View the Galois action as an action of $\mathrm{Frob}_p \in D_p = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. With our conventions, the action of Frobenius on the torus is as follows (cf. [**26**, pg. 31]).

Lemma 3.17. *The Frobenius* Frob_p *acts as* pT_p *on* $\mathcal{T}(\overline{\mathbf{F}}_p)$.

Make the identification $\mathcal{T} \cong \mathrm{Hom}_{\mathbf{Z}}(\mathcal{X}_p(J), \mathbf{G}_m)$, so that

$$V \subset \mathcal{T}(\overline{\mathbf{F}}_p)[\ell] = \mathrm{Hom}_{\mathbf{Z}}(\mathcal{X}_p(J), \boldsymbol{\mu}_\ell).$$

By Lemma 3.17, Frob_p acts on $V \subset \mathcal{T}(\overline{\mathbf{F}}_p)$ as $pa_p \in \mathbf{T}/\mathfrak{m}$, i.e., as a *scalar*. The determinant of ρ is $\chi\varepsilon$, so we have simulatenously

$$\det(\rho(\mathrm{Frob}_p)) = \begin{cases} p\varepsilon(p) & \text{and} \\ (pa_p)^2. \end{cases}$$

By Lemma 3.15, $a_p^2 = \varepsilon(p)$, so $p^2 \equiv p \pmod{\ell}$. Since $p \neq \ell$, this can only happen if $p \equiv 1 \pmod{\ell}$, which completes the proof.

3.10. Level optimization using a pivot

In this section we discuss an approach to level optimization that does not rely on multiplicity one results. In this approach, we eliminate a prime p from the level by making use of the rational quaternion algebra that is ramified precisely at p and at a second prime q. The latter prime is, in the simplest case, an appropriate prime number at which ρ ramifies; in more complicated cases, it is an "auxiliary" prime at which ρ is unramified. The central role of q in the argument, and the fact that q stays fixed in the level while p is removed, leads us to refer to q as a "pivot."

The following theorem first appeared in [**86**].

Theorem 3.18. *Let* $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ *be an irreducible continuous representation that arises from an eigenform* f *on* $\Gamma_1(K) \cap \Gamma_0(pq)$ *with* p *and* q *distinct primes that do not divide* ℓK. *Make the* ***key assumption*** *that the representation* ρ *is ramified at* q *and unramified at* p. *Then* ρ *arises from a weight-*2 *eigenform on* $\Gamma_1(K) \cap \Gamma_0(q)$.

The case $\ell = 2$ is not excluded from consideration.

Before sketching the proof, we describe a famous application. Edixhoven suggested to the first author that such an approach might be possible in the context of

Fermat's Last Theorem. We associate to a (hypothetical) solution $a^\ell + b^\ell + c^\ell = 0$ of the Fermat equation with $\ell > 3$ a Galois representation $E[\ell]$ attached to an elliptic curve E. A theorem of Mazur implies that this representation is irreducible; a theorem of Wiles implies that it arises from a modular form. Using Tate's algorithm, we finds that the discriminant of E is $\Delta_E = \frac{(abc)^{2\ell}}{2^8}$, which is a perfect ℓth power away from 2, and that the conductor of E is $N_E = \mathrm{rad}(abc) = \prod_{p \mid abc} p$. Let $q = 2$; then $E[\ell]$ is ramified at q because $\ell \nmid \mathrm{ord}_2(\Delta_E) = -8$ (see Theorem 2.10), but $E[\ell]$ is unramified at all other primes p, again by Theorem 2.10. To complete the proof of Fermat Last Theorem, we use $q = 2$ as a pivot and inductively remove each odd factor from N. One complication that may arise (the second case of Fermat Last Theorem) is that $\ell \mid N$. Upon removing ℓ from the level (using Section 3.1), the weight may initially go up to $\ell + 1$. If this occurs, since $k(\rho) = 2$ we can use [**32**] to optimize the weight back to 2.

As demonstrated by the application to Fermat, in problems of genuine interest the setup of Theorem 3.18 occurs. There are, however, situations in which it does not apply such as the recent applications of level optimization as a key ingredient to a proof of Artin's conjecture for certain icosahedral Galois representations (see [**10**]).

3.10.1. Shimura curves

We cannot avoid considering Shimura curves. Denote by $X(K, pq)$ the modular curve associated to $\Gamma_1(K) \cap \Gamma_0(pq)$ and let $J := \mathrm{Jac}(X(K, pq))$ be its Jacobian. Likewise, denote by $X^{pq}(K)$ the Shimura curve associated to the quaternion algebra of discriminant pq. The curve $X^{pq}(K)$ is constructed as follows. Let B be an indefinite quaternion algebra over $\mathbf{Q}$ of discriminant pq. (Up to isomorphism, B is unique.) Let $\mathcal{O}$ be an Eichler order (i.e., intersection of two maximal orders) of level K (i.e., reduced discriminant Kpq) in B. Let Γ_∞ be the group of elements of $\mathcal{O}$ with (reduced) norm 1. After fixing an embedding $B \to M(2, \mathbf{R})$ (an embedding exists because B is indefinite), we obtain in particular an embedding $\Gamma_\infty \hookrightarrow \mathrm{SL}(2, \mathbf{R})$ and therefore an action of Γ_∞ on the upper half-plane $\mathfrak{h}$. Let $X^{pq}(K)$ be the standard canonical model, over $\mathbf{Q}$, of the compact Riemann surface $\Gamma_\infty \backslash \mathfrak{h}$, and let $J' = \mathrm{Jac}(X^{pq}(K))$ denote its Jacobian. The curve $X^{pq}(K)$ is furnished with Hecke correspondences T_n for $n \geq 1$. We write T_n for the endomorphism of J induced by the T_n on $X^{pq}(K)$ via Pic functoriality.

Set $J' := \mathrm{Jac}(X^{pq}(K))$ and $J := \mathrm{Jac}(X(K, pq))$. Work of Eichler, Jacquet-Langlands, and Shimura (see [**36, 51, 106**]) has uncovered a deep correspondence between certain automorphic forms and certain cusp forms. Combining their work with the isogeny theorem of Faltings [**40**], we find (noncanonically!) a map $J' \to J$ with finite kernel.

The pq-new part of J is $J_{pq\text{-new}} := \ker(J(K, pq) \longrightarrow J(K, p)^2 \oplus J(K, q)^2)$ where the map is induced by Albanese functoriality from the four maps

$$X(K, pq) \rightrightarrows X(K, p) \qquad \text{and} \qquad X(K, pq) \rightrightarrows X(K, q).$$

The image of $J' \to J$ is the pq-new part of J.

3.10.2. Character groups

Amazingly, there seems to be no canonical map $J' \to J$ between the Shimura and classical Jacobians described in the previous section. Surprisingly, there is a canonical relationship between the character groups of J' and J. The Čerednik-Drinfeld theory gives a description of $X^{pq}(K)$ in characteristic p (see [**14, 30**]). Using this we find a canonical $\mathbf{T}$-equivariant exact sequence

$$0 \to \mathcal{X}_p(J') \to \mathcal{X}_q(J) \to \mathcal{X}_q(J'') \to 0 \tag{3.2}$$

where $J'' = \mathrm{Jac}(X(K, q))^2$. This exact sequence relates a character group "in characteristic p" to two character groups "in characteristic q". We are now prepared to prove the theorem.

3.10.3. Proof

Proof of Theorem 3.18. By our key assumption, the representation ρ is ramified at q, so $\mathfrak{m} \subset \mathbf{T}$ is not q-old. We may as well suppose we are in a situation where we can not optimize the level, so we assume that $\mathfrak{m}$ is not p-old either and hope for a contradiction.

Localization is an exact functor, so the localization

$$0 \longrightarrow \mathcal{X}_p(J')_{\mathfrak{m}} \longrightarrow \mathcal{X}_q(J)_{\mathfrak{m}} \longrightarrow \mathcal{X}_q(J'')_{\mathfrak{m}} \longrightarrow 0 \tag{3.3}$$

of (3.2) is also exact. The Hecke algebra $\mathbf{T}$ acts on $\mathcal{X}_q(J'')$ through a quotient $\overline{\mathbf{T}}$. Since $\mathfrak{m}$ is not q-old, the image of $\mathfrak{m}$ in $\overline{\mathbf{T}}$ generates the unit ideal. Therefore $\mathcal{X}_q(J'')_{\mathfrak{m}} = 0$ and we obtain an isomorphism $\mathcal{X}_p(J')_{\mathfrak{m}} \approx \mathcal{X}_q(J)_{\mathfrak{m}}$. If R is a $\mathbf{T}$-module then $R/\mathfrak{m}R = R_{\mathfrak{m}}/\mathfrak{m}R_{\mathfrak{m}}$ so

$$\mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J) \approx \mathcal{X}_p(J')/\mathfrak{m}\mathcal{X}_p(J'). \tag{3.4}$$

Switching p and q and applying the same argument shows that

$$\mathcal{X}_p(J)/\mathfrak{m}\mathcal{X}_p(J) \approx \mathcal{X}_q(J')/\mathfrak{m}\mathcal{X}_q(J'). \tag{3.5}$$

Both (3.4) and (3.5) are isomorphisms of $\mathbf{T}/\mathfrak{m}$-vector spaces.

By [**6**] we have an isomorphism $J[\mathfrak{m}] \approx \bigoplus_{i=1}^{\lambda} V$, with $\lambda > 0$ and $J'[\mathfrak{m}] \approx \bigoplus_{i=1}^{\nu} V$. (It follows from [**51**] that $\nu > 0$, but we will not use this here.) Our hypothesis that V is unramified automatically propagates to all of $J[\mathfrak{m}] \approx \bigoplus_{i=1}^{\lambda} V$. Since V is irreducible and we are assuming that $\mathfrak{m}$ is not p-old, the same argument as in Section 3.9 shows that $J[\mathfrak{m}] \subset \mathcal{T}[\mathfrak{m}]$ where $\mathcal{T}$ is the toric part of $\mathcal{J}_{\mathbf{F}_p}$. This means that $\dim(\mathcal{X}_p(J)/\mathfrak{m}\mathcal{X}_p(J)) \geq 2\lambda$. Using the same argument with J replaced by J' gives that $\dim(\mathcal{X}_p(J')/\mathfrak{m}\mathcal{X}_p(J')) \geq 2\mu$.

As an I_q-module V is an extension of two copies of the trivial character. This follows from results of Langlands [**67**], since ρ is a mod ℓ representation of $G_{\mathbf{Q}}$ associated to some newform f whose level divides pqK and is divisible by q. (The admissible representation of $\mathrm{GL}(2, \mathbf{Q}_q)$ which is associated to f is a special representation.) Because V is ramified at q and there is an unramified line, we see that $\dim(V^{I_q}) = 1$. Thus $\dim J[\mathfrak{m}]^{I_q} = \lambda$; since $q \neq \ell$ and the action of inertia on character groups is trivial, we see that

$$\mathrm{Hom}(\mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J), \boldsymbol{\mu}_\ell) \subset J[\mathfrak{m}]^{I_q},$$

so $\dim \mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J) \leq \lambda$. A similar argument bounds $\dim \mathcal{X}_q(J')/\mathfrak{m}\mathcal{X}_q(J')$. We obtain the following quadruple of inequalities:

$$\begin{aligned}
\dim \mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J) &\leq \lambda,\\
\dim \mathcal{X}_q(J')/\mathfrak{m}\mathcal{X}_q(J') &\leq \mu,\\
\dim \mathcal{X}_p(J)/\mathfrak{m}\mathcal{X}_p(J) &\geq 2\lambda,\\
\dim \mathcal{X}_p(J')/\mathfrak{m}\mathcal{X}_p(J') &\geq 2\mu.
\end{aligned}$$

Combining these with (3.4, 3.5), we find that

$$\begin{aligned}
2\lambda &\leq \dim \mathcal{X}_p(J)/\mathfrak{m}\mathcal{X}_p(J)\\
&= \dim \mathcal{X}_q(J')/\mathfrak{m}\mathcal{X}_q(J')\\
&\leq \mu
\end{aligned}$$

and simulatenously that $2\mu \leq \lambda$. Together these imply that $4\lambda \leq \lambda$ so $\lambda = 0$. But Proposition 3.2 implies that the multiplicity of ρ in $J[\mathfrak{m}]$ is strictly positive. This contradiction implies that our assumption that $\mathfrak{m}$ is not p-old is false, hence $\mathfrak{m}$ is p-old and ρ arises from an eigenform on $\Gamma_1(K) \cap \Gamma_0(q)$. □

3.11. Level optimization with multiplicity one

Theorem 3.19. *Suppose $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is an irreducible multiplicity one representation that arises from a weight-2 newform f on $\Gamma_1(M) \cap \Gamma_0(p)$ and that p is unramified. Then there is a newform on $\Gamma_1(M)$ that also gives rise to ρ.*

We sketch a proof, under the assumption that $\ell > 2$. Buzzard [**9**] has given a proof when $\ell = 2$; his result has been combined with the results of [**28**] to prove a Wiles-like lifting theorem valid for many representations when $\ell = 2$, and hence (thanks to Taylor) to establish new examples of Artin's conjecture (see [**10**]).

The following diagram illustrates the multiplicity one argument:

$$\begin{tikzcd}
& Mpq \arrow[rr, rightsquigarrow, "\text{pivot}"] & & Mq \arrow[dr, rightsquigarrow, "\text{easy}"] & \\
Mp \arrow[ur, rightsquigarrow, "\text{level addition}"] & & & & M.
\end{tikzcd}$$

The pivot step is potentially the hardest; though it resembles the pivot step of Section 3.10, but the symmetry is broken. In Section 3.10 we knew that q could not be removed from the level, but here q can be.

We manufacture q as follows. Pick q to be one of the (infinitely many) primes not dividing $Mp\ell$ such that the following conditions both hold:

$$(1)\quad \begin{cases} \rho(\mathrm{Frob}_q) \text{ is not a scalar, } \textit{or} \\ q \not\equiv 1 \pmod{\ell}. \end{cases}$$

$$(2)\quad \begin{cases} \text{the ratio of the eigenvalues of} \\ \rho(\mathrm{Frob}_q) \text{ is either } q \text{ or } 1/q. \end{cases}$$

The second condition means that the characteristic polynomial of $\rho(\mathrm{Frob}_q)$ is of the form $(x - a)(x - qa)$ for some $a \in \overline{\mathbf{F}}_\ell^*$.

Lemma 3.20. *There are infinitely many primes q that simultaneously satisfy both of the two conditions listed above.*

Proof. First assume that $\ell > 2$. Using the Cebotarev density theorem, find infinitely many primes q such $\rho(\mathrm{Frob}_q) = \rho(c)$ where c denotes complex conjugation. The eigenvalues of $\rho(c)$ are 1 and -1 (Exercise 8), so their ratio is -1. This ratio is equal to q because

$$-1 = \chi(c) = \det(\rho(\mathrm{Frob}_q)) = \chi(\mathrm{Frob}_q) \equiv q \pmod{\ell},$$

and $q \not\equiv 1 \pmod{\ell}$ because ℓ is odd.

Next assume that $\ell = 2$. Because ρ is irreducible, the image $\rho(G_{\mathbf{Q}}) \subset \mathrm{GL}(2, \overline{\mathbf{F}}_2)$ has even order. After a possible change of basis we find $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \rho(G_{\mathbf{Q}})$. Using Cebotarev density, we find infinitely many q with $\rho(\mathrm{Frob}_q)$ conjugate to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. For such q, condition (1) is satisfied. Condition (2) is also satisfied because the ratio of the eigenvalues is 1 which, because q is an odd prime, is congruent to q modulo $\ell = 2$. □

Sketch of proof of Theorem 3.19. Choose q as in Lemma 3.20. With q thus chosen, we can raise the level. More precisely, there exists a pq-new form on $\Gamma_1(M) \cap \Gamma_0(pq)$. We illustrate this as follows.

$$\begin{tikzcd}
& (M,\underline{pq}) \arrow[dr, mapsto, "\text{goal}"] & & \\
(M,p) \arrow[ur, mapsto, "\text{raise}"] & & (M,q) \arrow[rr, mapsto, "\text{Mazur (condition 1)}"] & & M
\end{tikzcd}$$

We underline pq to emphasize that the situation at level (M, pq) is symmetrical in p and q.

Let $J = J(M, pq)$; there is a maximal ideal $\mathfrak{m}$ in $\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots] \subset \operatorname{End} J$ attached to the pq-newform f that gives rise to ρ. Applying the multiplicity one hypothesis at level Mpq, we have $J[\mathfrak{m}] = V$ where V is a $\mathbf{T}/\mathfrak{m}$-vector space that supports ρ. In everything so far, M can be divisible by 2; the distinction between whether or not 2 divides M arises mainly in verifying the multiplicity one hypothesis.

Let $J' = J^{pq}(M)$ be the Shimura curve analogue of $J_1(M)$. As described in Section 3.10, J' is constructed in a similar manner as $J_1(M)$, but with $M_2(\mathbf{Q})$ replaced by a quaternion algebra. Of primary importance is that $J'[\mathfrak{m}] \approx \bigoplus_{i=1}^{\nu} V$, for some $\nu \geq 1$. This follows *morally* because ρ arises from a pq-new form, though the actual argument is quite involved.

Assume that we cannot optimize the level. We have an exact sequence of character groups

$$0 \longrightarrow \mathcal{X}_p(J') \longrightarrow \mathcal{X}_q(J) \longrightarrow \mathcal{X}_q(J(M,q)^2) \longrightarrow 0.$$

After localizing at $\mathfrak{m}$ as in (3.3), we discover that

$$\dim_{\mathbf{T}/\mathfrak{m}} \mathcal{X}_p(J')/\mathfrak{m}\mathcal{X}_p(J') = \dim_{\mathbf{T}/\mathfrak{m}} \mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J). \tag{3.6}$$

Furthermore, since the component group of J' at p is a quotient of $\mathcal{X}_q(J(M,q)^2)$, we find that $V \hookrightarrow (J'[\mathfrak{m}]^{I_p})^{\mathrm{toric}}$. Thus $\dim \mathcal{X}_p(J')/\mathfrak{m}\mathcal{X}_p(J') \geq 2$, so (3.6) implies that $\dim \mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J) \geq 2$. The endomorphism Frob_q acts as a scalar (cf. Lemma 3.17) on

$$J[\mathfrak{m}]^{\mathrm{toric}} = \mathrm{Hom}(\mathcal{X}_q(J)/\mathfrak{m}\mathcal{X}_q(J), \mu_\ell).$$

Furthermore, $J[\mathfrak{m}]^{\mathrm{toric}} \subset J[\mathfrak{m}]$ and both $J[\mathfrak{m}]^{\mathrm{toric}}$ and $J[\mathfrak{m}]$ have dimension 2, so Frob_q acts as a scalar on $J[\mathfrak{m}]$. If $q \not\equiv 1 \pmod{\ell}$ then we could use Mazur's principle to optimize the level, so by condition 1 we may assume that $\rho(\mathrm{Frob}_q)$ is not a scalar. This contradiction completes the sketch of the proof. $\square$

CHAPTER 4

Exercises

The following exercises were used in the Park City problem sessions. D. Savitt, K. Kedlaya, and B. Conrad contributed some of the problems. In Section 4.2, we provide several solutions, many of which were suggested by students in the problem sessions. The solution of some of the problems in this section requires facts beyond those stated explicitly in this paper.

4.1. Exercises

Exercise 1. Suppose $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$ is a one-dimensional continuous odd Galois representation.

(1) Give an example to show that ρ need not be a power of the mod ℓ cyclotomic character.

(2) Assume that ρ is unramified outside ℓ. Deduce that ρ is a power of the mod ℓ cyclotomic character.

Exercise 2. The principal congruence subgroup $\Gamma(N)$ of level N is the kernel of the reduction map $\mathrm{SL}(2, \mathbf{Z}) \to \mathrm{SL}(2, \mathbf{Z}/N\mathbf{Z})$. The subgroup $\Gamma_1(N)$ consists of matrices of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ modulo N. Let $\Gamma \subset \mathrm{SL}(2, \mathbf{Z})$ be a subgroup that contains $\Gamma(N)$ for some N. Show that there exists $g \in \mathrm{GL}(2, \mathbf{Q})$ such that the conjugate $g^{-1}\Gamma g$, which is a subgroup of $\mathrm{GL}(2, \mathbf{Q})$, contains $\Gamma_1(N^2)$.

Exercise 3. Let k be a finite field of characteristic greater than 2, and consider an odd representation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, k)$. Prove that ρ is irreducible if and only if ρ is absolutely irreducible. (A representation is absolutely irreducible if it remains irreducible after composing with the embedding $\mathrm{GL}(2, \mathbf{F}_\ell) \hookrightarrow \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$.) Give an example to show that this assertion is false when k has characteristic 2.

Exercise 4. Let $A/\mathbf{Q}$ be an elliptic curve. Show that the group of $\mathbf{Q}$-rational endomorphisms $\mathrm{End}(A)$ of A is equal to $\mathbf{Z}$; that is, integer multiplications are the only $\mathbf{Q}$-rational endomorphisms of A. Assume further that A is isolated in its isogeny class, in the sense that if B is an elliptic curve that is isogenous to A over $\mathbf{Q}$, then A and B are isomorphic over $\mathbf{Q}$. Show that, for every prime number ℓ, the representation

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(A[\ell]) \approx \mathrm{GL}(2, \mathbf{F}_\ell)$$

is irreducible. Must ρ_ℓ be absolutely irreducible?

Exercise 5. Let $A/\mathbf{Q}$ be an elliptic curve and assume that for all ℓ the representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(A[\ell])$ is irreducible. Deduce that A is isolated in its isogeny class. This is the converse of Exercise 4.

Exercise 6. Suppose $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell)$ arises from the ℓ-torsion of an elliptic curve. Verify, using standard properties of the Weil pairing, that $\det(\rho)$ is the mod ℓ cyclotomic character.

Exercise 7. Let $f \in S_k(\Gamma_1(N))$ be a modular form that is an eigenform for all the Hecke operators T_p and for the diamond bracket operators $\langle d \rangle$. Let

$$\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$$

be the character of f, so $\langle d \rangle f = \varepsilon(d) f$ for all $d \in (\mathbf{Z}/N\mathbf{Z})^*$.

1. Show that f satisfies the following equation:
   for any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$,
   $$f(z) = \varepsilon(d)(cz+d)^{-k} f\Big(\frac{az+b}{cz+d}\Big).$$
2. Conclude that $\varepsilon(-1) = (-1)^k$.
3. Choose a prime ℓ and let ρ be one of the mod ℓ Galois representations associated to f. We have $\det(\rho) = \varepsilon \cdot \chi^{k-1}$ where χ is the mod ℓ cyclotomic character. Deduce that ρ is odd, in the sense that $\det(\rho(c)) = -1$ for c complex conjugation.

Exercise 8. Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \overline{\mathbf{F}}_\ell)$ be an odd Galois representation, and let $c \in G_{\mathbf{Q}}$ denote complex conjugation.

1. Prove that if $\ell \neq 2$ then $\rho(c)$ is conjugate over $\overline{\mathbf{F}}_\ell$ to the matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.
2. Give an example to show that when $\ell = 2$, the matrix $\rho(c)$ need not be conjugate to $\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Exercise 9. Show that there exists a *non-continuous* homomorphism

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \{\pm 1\}$$

where $\{\pm 1\}$ has the discrete topology; equivalently, that there is a non-closed subgroup of index two in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. To accomplish this, you must produce a map $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \{\pm 1\}$ such that

1. ρ is a homomorphism, and
2. ρ does not factor through $\mathrm{Gal}(K/\mathbf{Q})$ for any *finite* Galois extension $K/\mathbf{Q}$.

Exercise 10. A potential difficulty is that a representation ρ arising from a modular form sometimes takes values in a slightly smaller field than $\mathcal{O}/\lambda$. For example, let f be one of the two conjugate normalized eigenforms in $S_2(\Gamma_0(23))$. Then

$$f = q + \alpha q^2 + (-2\alpha - 1)q^3 + (-\alpha - 1)q^4 + 2\alpha q^5 + \cdots$$

with $\alpha^2 + \alpha - 1 = 0$. The coefficients of f lie in $\mathcal{O} = \mathbf{Z}[\alpha] = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$. Take λ to be the unique prime of $\mathcal{O}$ lying over 2; then $\mathcal{O}/\lambda \cong \mathbf{F}_4$, so $\overline{\rho}_{f,\lambda}$ is a homomorphism into $\mathrm{GL}(2, \mathbf{F}_4)$. Show that if $p \neq 2$ then $a_p \in \mathbf{Z}[\sqrt{5}]$, so that $\overline{\rho}_{f,\lambda}$ possesses a model over $\mathrm{GL}(2, \mathbf{F}_2)$.

Exercise 11. Let $A/\mathbf{Q}$ be an elliptic curve and $\ell \neq 2$ be a prime.

1. Prove that the field $\mathbf{Q}(A[\ell])$ generated by the coordinates of the points in $A[\ell]$ is strictly larger than $\mathbf{Q}$.
2. Given an example of an elliptic curve A such that $\mathbf{Q}(A[2]) = \mathbf{Q}$.

Exercise 12. Let A be an elliptic curve over $\mathbf{Q}$ defined by a Weierstrass equation $y^2 = x^3 + ax + b$ with $a, b \in \mathbf{Q}$.

(1) Describe the Galois representation

$$\rho = \rho_{A,2} : G_{\mathbf{Q}} \to \mathrm{GL}(2, \mathbf{F}_2).$$

(2) Give necessary and sufficient conditions for ρ to be reducible.

(3) Choose a prime p, and give an example in which ρ is ramified only at p.

Exercise 13. Let ε and ρ be a pair of continuous homomorphisms from $G_{\mathbf{Q}}$ to $\mathbf{F}_\ell^*$. Suppose that for all primes p at which both ε and ρ are unramified we have

$$\rho(\mathrm{Frob}_p) = \varepsilon(\mathrm{Frob}_p)p^i \in \mathbf{F}_\ell.$$

Deduce that $\rho = \varepsilon \cdot \chi^i$ where χ is the mod ℓ cyclotomic character.

Exercise 14. Let $A/\mathbf{Q}$ be an elliptic curve of conductor N, and let p be a prime number not dividing N. Denote by $\tilde{A}$ the mod p reduction of A. The Frobenius endomorphism $\Phi = \Phi_p : \tilde{A} \to \tilde{A}$ sends an affine point (x, y) to (x^p, y^p) and fixes ∞. The characteristic polynomial of the endomorphism induced by Φ on the Tate module of $\tilde{A}$ at some (any) prime $\ell \neq p$ is $X^2 - \mathrm{tr}(\Phi)X + \deg(\Phi)$.

(1) Show that $\deg(\Phi) = p$.

(2) Show that $\mathrm{tr}(\Phi) = p + 1 - \#A(\mathbf{F}_p)$, that is, "$\mathrm{tr}(\Phi) = a_p$."

(3) Choose a prime $\ell \nmid pN$. Then $\tilde{A}[\ell]$ is a vector space of dimension two over $\mathbf{F}_\ell$, and Φ induces a map $\tilde{A}[\ell] \to \tilde{A}[\ell]$. Show that this is the same as the map induced by some choice of $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

(4) Conclude that

$$\mathrm{tr}(\rho_{A,\ell}(\mathrm{Frob}_p)) = p + 1 - \#A(F_p) \pmod{\ell}.$$

Exercise 15. Let $A/\mathbf{Q}$ be an elliptic curve of conductor N, and let ℓ be a prime. Show that any prime p not dividing ℓN is unramified in $\mathbf{Q}(A[\ell])$. You may use the following fact which is proved using formal groups (see, e.g., [**109**, Prop. 3.1]): **Fact:** The map $A[\ell] \to \tilde{A}(\overline{\mathbf{F}}_p)$ is injective, where $\tilde{A}$ is the reduction of A modulo p.

Exercise 16. Show that the fundamental character of level 1 is the cyclotomic character $\chi|_{I_t}$. (Hint: This is trickier than it first appears, and requires Wilson's theorem from elementary number theory.)

Exercise 17. For each of the following semistable elliptic curves A, and each ℓ at which $\rho_{A,\ell}$ is *irreducible*, use Theorem 2.10 to compute Serre's minimal weight $k(\rho_{A,\ell})$ and level $N(\rho_{A,\ell})$.

N	$\lvert\Delta\rvert$	reducible ℓ	A
30	$2^4 \cdot 3^5 \cdot 5$	$2, 3$	$y^2 + xy + y = x^3 + x + 2$
210	$2^{12} \cdot 3^3 \cdot 5 \cdot 7$	$2, 3$	$y^2 + xy = x^3 - 41x - 39$
330	$2^4 \cdot 3^2 \cdot 5^4 \cdot 11^2$	2	$y^2 + xy = x^3 + x^2 - 102x + 324$
455	$5^3 \cdot 7^4 \cdot 13$	2	$y^2 + xy = x^3 - x^2 - 50x + 111$
2926	$2^8 \cdot 7^3 \cdot 11^4 \cdot 19^2$	2	$y^2 + xy + y = x^3 - x^2 + 1934x - 1935$

Attempt to verify Serre's conjecture directly in some of these cases.

Exercise 18. Let M be a positive integer and let p be a prime. Show that there is an injective linear map

$$S_2(\Gamma_1(M)) \hookrightarrow S_2(\Gamma_1(pM))$$

sending $f(q)$ to $f(q^p)$.

Exercise 19. Let M be an integer such that $S_2(\Gamma_1(M))$ has positive dimension, and let p be a prime (thus $M = 11$ or $M \geq 13$).

(1) Let $f \in S_2(\Gamma_1(M))$ be an eigenvector for T_p with eigenvalue λ. Show that T_p acting on $S_2(\Gamma_1(Mp))$ preserves the two-dimensional subspace generated by f and $f(pz)$ (see Section 1.5 for the definition of T_p when p divides the level). Show furthermore that if $\lambda^2 \neq 4p$ then T_p is diagonalizable on this 2-dimensional space. What are the eigenvalues of T_p on this space? In fact, one never has $\lambda^2 = 4p$; see [**16**] for more details.
(2) Show that for any $r > 2$, the Hecke operator T_p on $S_2(\Gamma_1(Mp^r))$ is not diagonalizable.
(3) Deduce that for $r > 2$ the Hecke algebra $\mathbf{T}$ associated to $S_2(\Gamma_1(Mp^r))$ has nilpotent elements, so it is not an order in a product of rings of integers of number fields.

Exercise 20. Let N be a positive integer. Show that the Hecke algebra $\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots] \subset \mathrm{End}(J_1(N))$ is of finite rank as a $\mathbf{Z}$-module.

Exercise 21. Suppose $N = pM$ with $(p, M) = 1$. There is an injection

$$S_2(\Gamma_1(M)) \oplus S_2(\Gamma_1(M)) \hookrightarrow S_2(\Gamma_1(M) \cap \Gamma_0(p))$$

given by $(f, g) \mapsto f(q) + g(q^p)$. The Hecke algebra $\mathbf{T} = \mathbf{T}_N$ acts through a quotient $\overline{\mathbf{T}}$ on the image of $S_2(\Gamma_1(M)) \oplus S_2(\Gamma_1(M))$. Suppose $\mathfrak{m} \subset \mathbf{T}$ is a maximal ideal that arises by pullback from a maximal ideal in $\overline{\mathbf{T}}$. Show that $\rho_{\mathfrak{m}}$ arises from a modular form of level M.

4.2. Solutions

Solution 1. 1. Let p be a prime different from ℓ and let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Gal}(\mathbf{Q}(\sqrt{p})/\mathbf{Q}) \approx \{\pm 1\} \hookrightarrow \mathbf{F}_\ell^*.$$

2. Let $K = \overline{\mathbf{Q}}^{\ker(\rho)}$. Then $K/\mathbf{Q}$ is abelian and ramified only at ℓ, so $K \subset \mathbf{Q}(\zeta_{\ell^\infty})$. But $[K : \mathbf{Q}] \mid \ell - 1$ so $K \subset \mathbf{Q}(\zeta_\ell)$.

Solution 2. Conjugate using $g = \left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Solution 3. If ρ is absolutely irreducible then it is irreducible, so assume that ρ is irreducible. If ρ is reducible over the algebraic closure $\overline{k}$ of k, then there is a vector $v \in \overline{k}^{\oplus 2}$ that generates a one-dimensional subspace stable under ρ. In particular, v is stable under complex conjugation, which has characteristic polynomial $x^2 - 1 = (x-1)(x+1)$. Since $-1 \neq 1$, this means that v must lie in one of the two 1-dimensional eigenspaces of complex conjugation, so v is a scalar multiple of an element w of $k^{\oplus 2}$. Then ρ leaves the subspace of $k^{\oplus 2}$ spanned by w invariant, so ρ is reducible, which contradicts our assumption.

Let $\rho : G_{\mathbf{Q}} \to \mathrm{GL}(2, \mathbf{F}_2)$ be any continuous representation whose image is the subgroup generated by $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Then ρ is irreducible because it has no one-dimensional invariant subspaces over $\mathbf{F}_2$. However, the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ is diagonalizable over $\mathbf{F}_4$.

Solution 4. Suppose $\varphi \in \mathrm{End}(E)$ is a nonzero endomorphism. The induced map $d\varphi$ on the differentials $H^0(A, \Omega) \approx \mathbf{Q}$ is multiplication by an integer n, so $d(\varphi - n) = 0$ which implies that $\varphi = n$.

Suppose that ρ_ℓ is reducible, so that there is a one-dimensional Galois stable subspace $V \subset A[\ell]$. The quotient $B = A/V$ is then an elliptic curve over $\mathbf{Q}$ and there is an isogeny $\pi : A \to B$ of degree ℓ. Because A is isolated in its isogeny class we have that $B = A$, so there is an endomorphism of A of degree ℓ. But all $\mathbf{Q}$-rational endomorphisms are multiplication by an integer, and multiplication by an integer has degree a perfect square.

The elliptic curve E given by the equation $y^2 = x^3 - 7x - 7$ has the property that $E[2]$ is irreducible but not absolutely irreducible. To see this, note that the splitting field of $x^3 - 7x - 7$ has Galois group cyclic of order 3.

Solution 5. Suppose all $\rho_{A,\ell}$ are irreducible, yet there exists an isogeny $\varphi : A \to B$ with $B \not\approx A$. Choose φ to have minimal possible degree and let $d = \deg(\varphi) > 1$. Let ℓ be the smallest prime divisor of d and choose a point $x \in \ker(\varphi)$ of exact order ℓ. If the order-ℓ cyclic subgroup generated by x is Galois stable, then $\rho_{A,\ell}$ is reducible, which is contrary to our assumption. Thus $\ker(\varphi)$ contains the full ℓ-torsion subgroup $A[\ell]$ of A. In particular, φ factors as illustrated below:

$$\begin{tikzcd}
A \arrow[rr, "\varphi"] \arrow[dr, "\ell"'] & & B. \\
& A/A[\ell] \arrow[ur] &
\end{tikzcd}$$

Since $A/A[\ell] \cong A$, there is an isogeny from A to B of degree equal to d/ℓ^2, which contradicts our assumption that d is minimal.

Solution 6. The Weil pairing $(\,,\,) : A[\ell] \times A[\ell] \to \boldsymbol{\mu}_\ell$ can be viewed as a map

$$\bigwedge^2 A[\ell] \xrightarrow{\cong} \boldsymbol{\mu}_\ell$$

sending $P \wedge Q$ to (P, Q). For any $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, we have $(P^\sigma, Q^\sigma) = (P, Q)^\sigma$. With the action $(P \wedge Q)^\sigma = P^\sigma \wedge Q^\sigma$, the map $\bigwedge^2 A[\ell] \to \boldsymbol{\mu}_\ell$ is a map of Galois modules. To compute $\det(\rho(\sigma))$ observe that if e_1, e_2 is a basis for $A[\ell]$, and $\rho(\sigma) = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$\begin{aligned}
\sigma(e_1 \wedge e_2) &= (ae_1 + ce_2) \wedge (be_1 + de_2) \\
&= (ad - bc) e_1 \wedge e_2 = \det(\rho(\sigma)) e_1 \wedge e_2
\end{aligned}$$

Thus $\bigwedge^2 A[\ell]$ gives the one-dimensional representation $\det(\rho)$. Since $\bigwedge^2 A[\ell]$ is isomorphic to $\boldsymbol{\mu}_\ell$ it follows that $\det(\rho) = \chi$.

Solution 7. The definition of $\langle d \rangle$ is as follows: choose any matrix $\sigma_d \in \Gamma_0(N)$ such that $\sigma_d \equiv \left(\begin{smallmatrix} d & 0 \\ 0 & d^{-1} \end{smallmatrix}\right) \pmod N$; then $\langle d \rangle f = f|_{\sigma_d}$. Observe that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and the matrices σ_d with $(d, N) = 1$ and $d < N$ are a system of coset representatives. Thus any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ can be written in the form $\sigma_d \cdot g$ for some $g \in \Gamma_1(N)$. We have

$$f = f|_{\sigma_d g} = (f|_{\sigma_d})|_g = (\varepsilon(d) f)|_g = \varepsilon(d) f|_g = \varepsilon(d)(cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right).$$

Solution 8.

(1) Since $c^2 = 1$, the minimal polynomial f of $\rho(c)$ divides $x^2 - 1$. Thus f is either $x + 1$, $x - 1$, or $x^2 - 1$. If $f = x + 1$ then $\rho(c) = -1 = \left(\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right)$. This implies that $\det(\rho(c)) = (-1)^2 = 1$, which is a contradiction since $\det(\rho(c)) = -1$ and the characteristic of the base field is odd. If $f = x - 1$, then $\rho(c) = 1$; again a contradiction. Thus the minimal polynomial of $\rho(c)$ is $x^2 - 1 = (x - 1)(x + 1)$. Since $-1 \neq 1$ there is a basis of eigenvectors for $\rho(c)$ such that the matrix of $\rho(c)$ with respect to this basis is $\left(\begin{smallmatrix}-1 & 0\\ 0 & 1\end{smallmatrix}\right)$.

(2) The following example shows that when $\ell = 2$ the matrix of $\rho_{A,\ell}$ need not be conjugate to $\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. Let A be the elliptic curve over $\mathbf{Q}$ defined by $y^2 = x(x^2 - a)$ with $a \in \mathbf{Q}$ not square. Then

$$A[2] = \{\infty, (0, 0), (\sqrt{a}, 0), (-\sqrt{a}, 0)\}.$$

The action of c on the basis $(0, 0), (-\sqrt{a}, 0)$ is represented by the matrix $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)$, since $c(-\sqrt{a}, 0) = (\sqrt{a}, 0) = (0, 0) + (-\sqrt{a}, 0)$.

Solution 9. The extension $\mathbf{Q}(\sqrt{d}, d \in \mathbf{Q}^*/(\mathbf{Q}^*)^2)$ is an extension of $\mathbf{Q}$ with Galois group $X \approx \prod \mathbf{F}_2$. The index-two open subgroups of X correspond to the quadratic extensions of $\mathbf{Q}$. However, Zorn's lemma implies that X contains many more index-two subgroups, which can be seen more precisely as follows.

(1) Choose a sequence $p_1, p_2, p_3, \ldots$ of distinct prime numbers. Define $\rho_1 : G_{\mathbf{Q}} \to \prod \mathbf{F}_2$ by

$$\rho_1(\sigma)_i = \begin{cases} 0 & \text{if } \sigma \text{ acts trivially on } \mathbf{Q}(\sqrt{p_i}),\\ 1 & \text{otherwise}\end{cases}$$

Thus ρ_1 is just

$$G_{\mathbf{Q}} \to \mathrm{Gal}(\mathbf{Q}(\sqrt{p_1}, \sqrt{p_2}, \ldots)/\mathbf{Q}) \approx \prod \mathbf{F}_2.$$

(2) Let $\oplus\mathbf{F}_2 \subset \prod \mathbf{F}_2$ be the subgroup of elements having only finitely many nonzero coordinates. Then $\prod \mathbf{F}_2/\oplus\mathbf{F}_2$ is a vector space over $\mathbf{F}_2$ of dimension > 0. By Zorn's lemma, there is a basis $\mathcal{B}$ of $\prod \mathbf{F}_2/\oplus\mathbf{F}_2$. Let $b \in \mathcal{B}$ and let W be the subspace spanned by $\mathcal{B} - \{b\}$. Then $V = (\prod \mathbf{F}_2/\oplus\mathbf{F}_2)/W$ is an $\mathbf{F}_2$-vector space of dimensional 1.

(3) Let ρ be the composite map

$$\begin{tikzcd}
\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \arrow[d] \arrow[drr, "\rho"] & & \\
\prod \mathbf{F}_2 \arrow[r] & V \arrow[r, "\cong"'] & \{\pm 1\}
\end{tikzcd}$$

(4) Let $H = \ker(\rho) \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. If $\sigma(\sqrt{p_i}) = -\sqrt{p_i}$ and $\sigma(\sqrt{p_j}) = \sqrt{p_j}$ for $i \neq j$, then $\sigma \in H$. Thus H does not fix any $\mathbf{Q}(\sqrt{p_i})$, so the fixed field of H equals $\mathbf{Q}$. The largest finite Galois group quotient through which ρ factors is then $\mathrm{Gal}(\mathbf{Q}/\mathbf{Q}) = \{1\}$. Since $\rho \neq 1$, we conclude that ρ does not factor through any finite Galois group quotient, which proves that ρ is not continuous.

Solution 10. We have $f = f_1 + \alpha f_2$ with $f_1, f_2 \in S_2(\Gamma_0(23))$ and

$$\begin{aligned} f_1 &= q - q^3 - q^4 + \cdots \\ f_2 &= q^2 - 2q^3 - q^4 + 2q^5 + \cdots . \end{aligned}$$

Because $S_2(\Gamma_0(23))$ has dimension 2, it is spanned by f_1 and f_2. Let $\eta(q) = q^{\frac{1}{24}} \prod_{n \geq 1}(1 - q^n)$. Then $g = (\eta(q)\eta(q^{23}))^2 \in S_2(\Gamma_0(23))$. Expanding we find that $g = q^2 - 2q^3 + \cdots$, so $g = f_2$. Next observe that g is a power series in q^2 modulo 2:

$$\begin{aligned} g &= q^2 \prod (1 - q^n)^2 (1 - q^{23n})^2 \\ &\equiv q^2 \prod (1 - q^{2n})(1 - q^{46n}) \pmod 2 \\ &\equiv q^2 \prod (1 + q^{2n} + q^{46n} + q^{48n}) \pmod 2 \end{aligned}$$

Thus the coefficient in f_2 of q^p with $p \neq 2$ prime is even, and the proposition follows.

Solution 11.

(1) Let $\zeta \in \boldsymbol{\mu}_\ell$ be a primitive ℓth root of unity. Since $\bigwedge^2 A[\ell] \cong \boldsymbol{\mu}_\ell$, there exists $P, Q \in A[\ell]$ such that $P \wedge Q = \zeta$. Since $\ell > 2$ there exists σ such that $\zeta^\sigma \neq \zeta$, hence $P^\sigma \wedge Q^\sigma \neq P \wedge Q$. This is impossible if all ℓ-torsion is rational, since then $P^\sigma = P$ and $Q^\sigma = Q$.

(2) Consider the elliptic curve defined by $y^2 = (x - a)(x - b)(x - c)$ where a, b, c are distinct rational numbers.

Solution 12.

(1) Let K be the splitting field of $x^3 + ax + b$. Then ρ embeds $\text{Gal}(K/\mathbf{Q})$ in $\text{GL}(2, \mathbf{F}_2)$:

$$\begin{tikzcd}
\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \arrow[rr, "\rho"] \arrow[dr] & & \text{GL}(2, \mathbf{F}_2) \\
& \text{Gal}(K/\mathbf{Q}) \arrow[ur, hook] &
\end{tikzcd}$$

(2) The representation ρ is reducible exactly when the polynomial $x^3 + ax + b$ has a rational root.

(3) Examples: $y^2 = x(x^2 - 23)$, $y^2 = x^3 + x - 1$.

Solution 13. Consider the character $\tau = \varepsilon\chi/\rho$. By assumption, $\tau(\text{Frob}_p) = 1$ for all unramified p. Let K be an extension of $\mathbf{Q}$ such that τ factors through $\text{Gal}(K/\mathbf{Q})$. For any $\sigma \in \text{Gal}(K/\mathbf{Q})$, the Cebotarev density theorem implies that there are infinitely many primes p such that $\text{Frob}_p = \sigma$. Thus for any σ, $\tau(\sigma) = \tau(\text{Frob}_p) = 1$, so $\tau = 1$ and hence $\rho = \varepsilon\chi$.

Solution 14.

(1) See, e.g., [**109**, 2.11].

(2) By [**109**, 5.5], $\Phi - 1$ is separable, so $\#A(\mathbf{F}_p) = \deg(\Phi - 1)$. Since Φ has degree p, there exists an isogeny $\overline{\Phi}$ (the dual isogeny, see [**109**, III.6]),

such that $\Phi\overline{\Phi} = p$. Letting bars denote the dual isogeny, we have

$$\begin{aligned}\#A(\mathbf{F}_p) = \deg(\Phi - 1) &= (\Phi - 1)\overline{(\Phi - 1)}\\ &= \Phi\overline{\Phi} - \Phi - \overline{\Phi} + 1\\ &= p - \mathrm{tr}(\Phi) + 1\end{aligned}$$

(3) Both maps are pth powering on coordinates.

Solution 15. Since $\ell \neq p$ and A has good reduction at p, the natural map $A[\ell] \to \tilde{A}[\ell]$ is an isomorphism. We have the following commutative diagram

$$\begin{tikzcd}
\mathrm{Gal}(\mathbf{Q}_p(A[\ell])/\mathbf{Q}_p) \arrow[r, hook] \arrow[d] & \mathrm{Aut}(A[\ell]) \arrow[d, "\cong"] \\
\mathrm{Gal}((\mathcal{O}/\lambda)_{\mathbf{F}_p}) \arrow[r] & \mathrm{Aut}(\tilde{A}[\ell])
\end{tikzcd}$$

It follows that the first vertical map must be injective, which is the same as $\mathbf{Q}_p(A[\ell])$ being unramified over $\mathbf{Q}_p$.

Solution 16. The fundamental character Ψ of level one is the composition

$$\mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{nr}}(\sqrt[\ell-1]{\ell})/\mathbf{Q}_\ell^{\mathrm{nr}}) \to \boldsymbol{\mu}_{\ell-1}(\overline{\mathbf{Q}}_\ell) \to \boldsymbol{\mu}_{\ell-1}(\overline{\mathbf{F}}_\ell^*) = \mathbf{F}_\ell^*.$$

Let π be such that $\pi^{\ell-1} = \ell$. Then $\Psi(\sigma) = \frac{\sigma(\pi)}{\pi} \pmod{\pi}$. Let $\zeta \in \overline{\mathbf{Q}}_\ell$ be a primitive ℓth root of unity. Now

$$\prod_{a=1}^{\ell-1}(\zeta^a - 1) = \ell,$$

so

$$(\zeta - 1)^{\ell-1}\prod_{a=1}^{\ell-1},\frac{\zeta^a - 1}{\zeta - 1} = \ell$$

and (this is where Wilson's theorem is used),

$$\prod_{a=1}^{\ell-1}\frac{\zeta_\ell^a - 1}{\zeta - 1} \equiv 1 \pmod{\zeta - 1}.$$

Since the polynomial $x^{\ell-1} - 1$ has roots over $\mathbf{F}_\ell$, by Hensel's lemma there is a unit $u \in \mathbf{Q}_\ell(\pi)$ such that

$$u^{\ell-1} = \prod_{a=1}^{\ell-1}\frac{\zeta^a - 1}{\zeta - 1}.$$

We can take $\pi = (\zeta - 1)u$. Then

$$\begin{aligned}\frac{\sigma(\pi)}{\pi} &= \frac{(\zeta^{\chi(\sigma)} - 1)\sigma(u)}{(\zeta - 1)u}\\ &= \frac{(\zeta - 1)(\zeta^{\chi(\sigma)-1} + \cdots + 1)\sigma(u)}{(\zeta - 1)u}\\ &= (\zeta^{\chi(\sigma)-1} + \cdots + 1)\sigma(u)/u\\ &\equiv \chi(\sigma) \pmod{\zeta - 1}.\end{aligned}$$

Solution 17. We write $N = N(\rho)$ and $k = k(\rho)$ to save space. The essential tool is Theorem 2.10.

(1) $\ell = 5$: $N = 6$, $k = 6$, $\ell > 5$, $N = 30$, $k = 2$.
(2) $\ell = 5$: $N = 2 \cdot 3 \cdot 7$, $k = 6$; $\ell = 7$: $N = 2 \cdot 3 \cdot 5$, $k = 8$; $\ell > 7$: $N = 2 \cdot 3 \cdot 5 \cdot 7$, $k = 2$.
(3) $\ell = 3$: $N = 2 \cdot 5 \cdot 11$, $k = 4$; $\ell = 5$: $N = 2 \cdot 3 \cdot 11$, $k = 6$; $\ell = 7$: $N = 2 \cdot 3 \cdot 5 \cdot 11$, $k = 2$; $\ell = 11$: $N = 2 \cdot 3 \cdot 5$, $k = 12$; $\ell > 11$: $N = 2 \cdot 3 \cdot 5 \cdot 11$, $k = 2$.
(4) $\ell = 3$: $N = 7 \cdot 13$, $k = 2$; $\ell = 5$: $N = 7 \cdot 13$, $k = 6$; $\ell = 7$: $N = 5 \cdot 13$, $k = 8$; $\ell = 13$: $N = 5 \cdot 7$, $k = 14$; $\ell = 11, \ell > 13$: $N = 5 \cdot 7 \cdot 13$, $k = 2$.
(5) $\ell = 3$: $N = 2 \cdot 11 \cdot 19$, $k = 2$; $\ell = 7$: $N = 2 \cdot 11 \cdot 19$, $k = 8$; $\ell = 11$: $N = 2 \cdot 7 \cdot 19$, $k = 12$; $\ell = 19$: $N = 2 \cdot 7 \cdot 11$, $k = 20$; ℓ other: $N = 2 \cdot 7 \cdot 11 \cdot 19$, $k = 2$.

Solution 20. One approach is to view $J_1(N)$ as a complex torus, and note that the endomorphism ring is the set of automorphism of a complex vector space that fix a lattice. Another approach is to use the deeper finiteness theorems that are valid in arbitrary characteristic, see, e.g., [**74**, Thm. 12.5].

CHAPTER 5

Appendix by Brian Conrad: The Shimura construction in weight 2

The purpose of this appendix is to explain the ideas of Eichler-Shimura for constructing the two-dimensional ℓ-adic representations attached to classical weight-2 Hecke eigenforms. We assume familiarity with the theory of schemes and the theory of newforms, but the essential arithmetic ideas are due to Eichler and Shimura. We warn the reader that a complete proof along the lines indicated below requires the verification of a number of compatibilities between algebraic geometry, algebraic topology, and the classical theory of modular forms. As the aim of this appendix is to explain the key arithmetic ideas of the proof, we must pass over in silence the verification of many such compatibilities. However, we at least make explicit what compatibilities we need. To prove them all here would require a serious digression from our expository goal; see [**18**, Ch. 3] for details. It is also worth noting that the form of the arguments we present is *exactly* the weight-2 version of Deligne's more general proof of related results in weight > 1, up to the canonical isomorphism

$$\mathbf{Q}_\ell \otimes_{\mathbf{Z}_\ell} \varprojlim \mathrm{Pic}^0_{X/k}[\ell^n](k) \cong H^1_{\text{ét}}(X, \mathbf{Q}_\ell(1)) \cong H^1_{\text{ét,c}}(Y, \mathbf{Q}_\ell(1))$$

for a proper smooth connected curve X over a separably closed field k of characteristic prime to ℓ, and Y a dense open in X. Using ℓ-adic Tate modules allows us to bypass the general theory of étale cohomology which arises in the case of higher weight.

5.1. Analytic preparations

Fix $i = \sqrt{-1} \in \mathbf{C}$ for all time. Fix an integer $N \geq 5$ and let $X_1(N)^{\mathrm{an}}$ denote the classical analytic modular curve, the "canonical" compactification of $Y_1(N)^{\mathrm{an}} = \Gamma_1(N)\backslash\mathfrak{h}$, where $\mathfrak{h} = \{z \in \mathbf{C} : \mathrm{Im}\, z > 0\}$ and $\Gamma_1(N) \subset \mathrm{SL}_2(\mathbf{Z})$ acts on the left via linear fractional transformations. The classical theory identifies the $\mathbf{C}$-vector space $H^0(X_1(N)^{\mathrm{an}}, \Omega^1_{X_1(N)^{\mathrm{an}}})$ with $S_2(\Gamma_1(N), \mathbf{C})$, the space of weight-2 cusp forms. Note that the classical Riemann surface $X_1(N)^{\mathrm{an}}$ has genus 0 if we consider $N < 5$, while $S_2(\Gamma_1(N), \mathbf{C}) = 0$ if $N < 5$. Thus, assuming $N \geq 5$ is harmless for what we will do.

The Hodge decomposition for the compact Riemann surface $X_1(N)^{\mathrm{an}}$ supplies us with an isomorphism of $\mathbf{C}$-vector spaces

$$\begin{aligned}
&S_2(\Gamma_1(N),\mathbf{C})\oplus\overline{S_2(\Gamma_1(N),\mathbf{C})}\\
&\qquad\qquad\cong H^0(X_1(N)^{\mathrm{an}},\Omega^1_{X_1(N)^{\mathrm{an}}})\oplus H^0(X_1(N)^{\mathrm{an}},\overline{\Omega}^1_{X_1(N)^{\mathrm{an}}})\\
&\qquad\qquad\xrightarrow{\sim} H^1(X_1(N)^{\mathrm{an}},\underline{\mathbf{C}})\\
&\qquad\qquad\cong H^1(X_1(N)^{\mathrm{an}},\underline{\mathbf{Z}})\otimes_{\mathbf{Z}}\mathbf{C}
\end{aligned}$$

(where $\underline{A}$ denotes the constant sheaf attached to an abelian group A). This will be called the (weight-2) *Shimura isomorphism.* We want to define "geometric" operations on $H^1(X_1(N)^{\mathrm{an}},\underline{\mathbf{Z}})$ which recover the classical Hecke operators on $S_2(\Gamma_1(N),\mathbf{C})$ via the above isomorphism.

The "geometric" (or rather, cohomological) operations we wish to define can be described in two ways. First, we can use explicit matrices and explicit "upper-half plane" models of modular curves. This has the advantage of being concrete, but it provides little conceptual insight and encourages messy matrix calculations. The other point of view is to identify the classical modular curves as the base of certain universal analytic families of (generalized) elliptic curves with level structure. A proper discussion of this latter point of view would take us too far afield, so we will have to settle for only some brief indications along these two lines (though this is how to best verify compatibility with the algebraic theory via schemes).

Choose a matrix $\gamma_n\in\mathrm{SL}_2(\mathbf{Z})$ with $\gamma_n\equiv\left(\begin{smallmatrix}n^{-1}&*\\0&n\end{smallmatrix}\right)\pmod N$, for $n\in(\mathbf{Z}/N\mathbf{Z})^*$. The action of γ_n on $\mathfrak{h}$ induces an action on $Y_1(N)^{\mathrm{an}}$ and even on $X_1(N)^{\mathrm{an}}$. Associating to each $z\in\mathfrak{h}$ the data of the elliptic curve $\mathbf{C}/[1,z]=\mathbf{C}/(\mathbf{Z}+\mathbf{Z}z)$ and the point $1/N$ of exact order N, we may identify $Y_1(N)^{\mathrm{an}}$ as a *set* with the set of isomorphism classes of pairs (E,P) consisting of an elliptic curve E over $\mathbf{C}$ and a point $P\in E$ of exact order N. The map $Y_1(N)^{\mathrm{an}}\to Y_1(N)^{\mathrm{an}}$ induced by γ_n can then described on the underlying set by $(E,P)\mapsto(E,nP)$, so it is "intrinsic", depending only on $n\in(\mathbf{Z}/N\mathbf{Z})^*$. We denote by $I_n : X_1(N)^{\mathrm{an}}\to X_1(N)^{\mathrm{an}}$ the induced map on $X_1(N)^{\mathrm{an}}$. Once this data (E,P) is formulated in a relative context over an analytic base, we could define the analytic map I_n conceptually, without using the matrix γ_n. We ignore this point here.

The map $z\mapsto\frac{-1}{Nz}$ on $\mathfrak{h}$ induces a map $Y_1(N)^{\mathrm{an}}\to Y_1(N)^{\mathrm{an}}$ which extends to $w_N : X_1(N)^{\mathrm{an}}\to X_1(N)^{\mathrm{an}}$. More conceptually and more generally, if $\zeta\in\boldsymbol{\mu}_N(\mathbf{C})$ is a primitive Nth root of unity, consider the rule w_ζ that sends $(E,P)\in Y_1(N)^{\mathrm{an}}$ to $(E/P,P'\mod P)$, where $P'\in E$ has exact order N and $\langle P,P'\rangle_N=\zeta$, with $\langle\,,\,\rangle_N$ the Weil pairing on N-torsion points (following the sign conventions of [**62, 77**]; opposite the convention of [**109**]). More specifically, on $\mathbf{C}/[1,z]$ we have $\langle\frac{1}{N},\frac{z}{N}\rangle_N=e^{2\pi i/N}$. The map w_ζ extends to an analytic map $X_1(N)^{\mathrm{an}}\to X_1(N)^{\mathrm{an}}$. When $\zeta=e^{2\pi i/N}$, we have $w_\zeta=w_N$ due to the above sign convention.

We have induced pullback maps

$$w_\zeta^*,I_n^*:H^1(X_1(N)^{\mathrm{an}},\underline{\mathbf{Z}})\to H^1(X_1(N)^{\mathrm{an}},\underline{\mathbf{Z}}).$$

We write $\langle n\rangle^*$ rather than I_n^*.

Finally, choose a prime p. Define $\Gamma_1(N,p)\subset\mathrm{SL}_2(\mathbf{Z})$ to be $\Gamma_1(N,p)=\Gamma_1(N)\cap\Gamma_0(p)$ when $p\nmid N$ and $\Gamma_1(N,p)=\Gamma_1(N)\cap\Gamma_0(p)^t$ when $p\mid N$, where the group $\Gamma_0(p)^t$ is the transpose of $\Gamma_0(p)$. Define $Y_1(N,p)^{\mathrm{an}}=\Gamma_1(N,p)\backslash\mathfrak{h}$ and let $X_1(N,p)^{\mathrm{an}}$

be its “canonical” compactification. Using the assignment

$$z \mapsto (\mathbf{C}/[1,z], \frac{1}{N}, \langle\frac{1}{p}\rangle)$$

when $p \nmid N$ and

$$z \mapsto (\mathbf{C}/[1,z], \frac{1}{N}, \langle\frac{z}{p}\rangle)$$

when $p \mid N$, we may identify the *set* $Y_1(N,p)^{\mathrm{an}}$ with the set of isomorphism classes of triples (E,P,C) where $P \in E$ has exact order N and $C \subset E$ is a cyclic subgroup of order p, meeting $\langle P\rangle$ trivially (a constraint if $p \mid N$). Here and below, we denote by $\langle P\rangle$ the (cyclic) subgroup generated by P.

There are unique analytic maps

$$\pi_1^{(p)}, \pi_2^{(p)} : X_1(N,p)^{\mathrm{an}} \to X_1(N)^{\mathrm{an}}$$

determined on $Y_1(N,p)^{\mathrm{an}}$ by

$$\pi_1^{(p)}(E,P,C) = (E,P)$$

and

$$\pi_2^{(p)}(E,P,C) = (E/C, P \mod C).$$

For example, $\pi_1^{(p)}$ is induced by $z \mapsto z$ on $\mathfrak{h}$, in terms of the above upper half plane uniformization of $Y_1(N)^{\mathrm{an}}$ and $Y_1(N,p)^{\mathrm{an}}$.

We define

$$T_p^* = (\pi_1^{(p)})_* \circ (\pi_2^{(p)})^* : H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \to H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$$

where $(\pi_1^{(p)})_* : H^1(X_1(N,p)^{\mathrm{an}}, \underline{\mathbf{Z}}) \to H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$ is the canonical trace map associated to the finite map $\pi_1^{(p)}$ of compact Riemann surfaces. More specifically, we have a canonical isomorphism

$$H^1(X_1(N,p)^{\mathrm{an}}, \underline{\mathbf{Z}}) \cong H^1(X_1(N)^{\mathrm{an}}, (\pi_1^{(p)})_*\underline{\mathbf{Z}})$$

since $(\pi_1^{(p)})_*$ is exact on abelian sheaves, and there is a unique trace map of sheaves $(\pi_1^{(p)})_*\underline{\mathbf{Z}} \to \underline{\mathbf{Z}}$ determined on stalks at $x \in X_1(N)^{\mathrm{an}}$ by

$$\begin{aligned} \prod_{\pi_1^{(p)}(y)=x} \mathbf{Z} &\to \mathbf{Z} \\ (a_y) &\mapsto \Sigma_y e_y a_y \end{aligned} \tag{5.1}$$

where e_y is the ramification degree of y over x via $\pi_1^{(p)}$.

A fundamental compatibility, whose proof we omit for reasons of space, is:

Theorem 5.1. *The weight-*2 *Shimura isomorphism*

$$\mathrm{Sh}_{\Gamma_1(N)} : S_2(\Gamma_1(N), \mathbf{C}) \oplus \overline{S_2(\Gamma_1(N), \mathbf{C})} \cong H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \otimes_{\mathbf{Z}} \mathbf{C}$$

from (5.1) *identifies* $\langle n\rangle \oplus \overline{\langle n\rangle}$ *with* $\langle n\rangle^* \otimes 1$, $T_p \oplus \overline{T}_p$ *with* $T_p^* \otimes 1$, *and* $w_N \oplus \overline{w}_N$ *with* $w^*_{e^{2\pi i/N}} \otimes 1$.

Let $\mathbf{T}_1(N) \subset \mathrm{End}_{\mathbf{Z}}(H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}))$ be the subring generated by the T_p^*'s and $\langle n\rangle^*$'s. By Theorem 5.1, this is identified via the Shimura isomorphism with the classical (weight-2) Hecke ring at level N. In particular, this ring is commutative (which can be seen directly via cohomological considerations as well). It is clearly a finite flat $\mathbf{Z}$-algebra.

The natural map

$$\mathbf{T}_1(N) \otimes_{\mathbf{Z}} \mathbf{C} \hookrightarrow \mathrm{End}_{\mathbf{C}}(H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \otimes_{\mathbf{Z}} \mathbf{C}) \tag{5.2}$$

induces an *injection* $\mathbf{T}_1(N) \otimes \mathbf{C} \hookrightarrow \mathrm{End}_{\mathbf{C}}(S_2(\Gamma_1(N), \mathbf{C}))$, by Theorem 5.1. This is the classical realization of Hecke operators in weight 2.

Another compatibility we need is between the cup product on $H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$ and the (non-normalized) Petersson product on $S_2(\Gamma_1(N), \mathbf{C})$. To be precise, we define an isomorphism $H^2(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \cong \mathbf{Z}$ using the i-orientation of the complex manifold $X_1(N)^{\mathrm{an}}$ (i.e., the "$i\mathrm{d}z \wedge \mathrm{d}\overline{z}$" orientation), so we get via cup product a (perfect) pairing

$$(\ ,\ )_{\Gamma_1(N)} : H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \otimes_{\mathbf{Z}} H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \to H^2(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \cong \mathbf{Z}.$$

This induces an analogous pairing after applying $\otimes_{\mathbf{Z}}\mathbf{C}$. For $f, g \in S_2(\Gamma_1(N), \mathbf{C})$ we define

$$\langle f, g\rangle_{\Gamma_1(N)} = \int_{\Gamma_1(N)\backslash\mathfrak{h}} f(z)\overline{g}(z)\mathrm{d}x\mathrm{d}y$$

where this integral is absolutely convergent since f and g have exponential decay near the cusps. This is a perfect Hermitian pairing.

Theorem 5.2. *Under the weight-*2 *Shimura isomorphism* $\mathrm{Sh}_{\Gamma_1(N)}$,

$$\big(\mathrm{Sh}_{\Gamma_1(N)}(f_1 + \overline{g}_1), \mathrm{Sh}_{\Gamma_1(N)}(f_2 + \overline{g}_2)\big)_{\Gamma_1(N)} = 4\pi \cdot (\langle f_1, g_2\rangle_{\Gamma_1(N)} - \langle f_2, g_1\rangle_{\Gamma_1(N)}).$$

Note that *both* sides are antilinear in g_1, g_2 and alternating with respect to interchanging the pair (f_1, g_1) and (f_2, g_2). The extra factor of 4π is harmless for our purposes since it does not affect formation of adjoints. What is important is that in the classical theory, conjugation by the involution w_N takes each $T \in \mathbf{T}_1(N)$ to its adjoint with respect to the Petersson product. The most subtle case of this is $T = T_p^*$ for $p \mid N$. For $p \nmid N$ the adjoint of T_p^* is $\langle p^{-1}\rangle^* T_p^*$ and the adjoint of $\langle n\rangle^*$ is $\langle n^{-1}\rangle^*$. These classical facts (especially for T_p^* with $p \mid N$) yield the following important corollary of Theorem 5.2.

Corollary 5.3. *With respect to the pairing* $[x, y]_{\Gamma_1(N)} = (x, w_\zeta^* y)_{\Gamma_1(N)}$ *with* $\zeta = e^{2\pi i/N}$*, the action of* $\mathbf{T}_1(N)$ *on* $H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$ *is equivariant. That is,*

$$[x, Ty]_{\Gamma_1(N)} = [Tx, y]_{\Gamma_1(N)}$$

for all $T \in \mathbf{T}_1(N)$*. With respect to* $(\ ,\ )_{\Gamma_1(N)}$*, the adjoint of* T_p^* *for* $p \nmid N$ *is* $\langle p^{-1}\rangle^* T_p^*$ *and the adjoint of* $\langle n\rangle^*$ *is* $\langle n^{-1}\rangle^*$ *for* $n \in (\mathbf{Z}/N\mathbf{Z})^*$.

Looking back at the "conceptual" definition of w_ζ^* for an arbitrary primitive Nth root of unity $\zeta \in \boldsymbol{\mu}_N(\mathbf{C})$, which gives an analytic involution of $X_1(N)^{\mathrm{an}}$, one can check that $w_{\zeta^j}^* \circ w_\zeta^* = \langle j\rangle^*$ for $j \in (\mathbf{Z}/N\mathbf{Z})^*$. Since $\langle j\rangle^*$ is a unit in $\mathbf{T}_1(N)$ and $\mathbf{T}_1(N)$ is *commutative*, we conclude that Corollary 5.3 is true with $\zeta \in \boldsymbol{\mu}_N(\mathbf{C})$ *any* primitive Nth root of unity (by reduction to the case $\zeta = e^{2\pi i/N}$).

Our final step on the analytic side is to reformulate everything above in terms of Jacobians. For any compact Riemann surface X, there is an isomorphism of complex Lie groups

$$\mathrm{Pic}^0_X \cong H^1(X, \mathscr{O}_X)/H^1(X, \underline{\mathbf{Z}}) \tag{5.3}$$

via the exponential sequence

$$0 \to \underline{\mathbf{Z}} \to \mathscr{O}_X \xrightarrow{e^{2\pi i(\cdot)}} \mathscr{O}_X^* \to 1$$

and the identification of the underlying group of Pic^0_X with

$$H^1(X, \mathscr{O}_X^*) \cong \check{H}^1(X, \mathscr{O}_X^*)\,,$$

where the line bundle $\mathscr{L}$ with trivializations $\varphi_i : \mathscr{O}_{U_i} \cong \mathscr{L}|U_i$ corresponds to the class of the Čech 1-cocycle

$$\{\varphi_j^{-1} \circ \varphi_i : \mathscr{O}_{U_i \cap U_j} \cong \mathscr{O}_{U_i \cap U_j}\} \in \prod_{i<j} H^0(U_i \cap U_j, \mathscr{O}_X^*)$$

for an ordered open cover $\{U_i\}$. Beware that the tangent space isomorphism

$$T_0(\mathrm{Pic}^0_X) \cong H^1(X, \mathscr{O}_X)$$

coming from (5.3) is $-2\pi i$ times the "algebraic" isomorphism arising from

$$0 \to \mathscr{O}_X \to \mathscr{O}_{X[\varepsilon]}^* \to \mathscr{O}_X^* \to 1,$$

where $X[\varepsilon] = (X, \mathscr{O}_X[\varepsilon]/\varepsilon^2)$ is the non-reduced space of "dual numbers over X". This extra factor of $-2\pi i$ will not cause problems. We will use (5.3) to "compute" with Jacobians.

Let $f : X \to Y$ be a finite map between compact Riemann surfaces. Since f is finite flat, there is a natural trace map $f_*\mathscr{O}_X \to \mathscr{O}_Y$, and it is not difficult to check that this is compatible with the trace map $f_*\underline{\mathbf{Z}} \to \underline{\mathbf{Z}}$ as defined in (5.1). In particular, we have a trace map

$$f_* : H^1(X, \mathscr{O}_X) \cong H^1(Y, f_*\mathscr{O}_X) \to H^1(Y, \mathscr{O}_Y).$$

Likewise, we have compatible pullback maps $f^*\mathscr{O}_Y \cong \mathscr{O}_X$ and $f^*\underline{\mathbf{Z}} \cong \underline{\mathbf{Z}}$.

Thus, any such f gives rise to *commutative* diagrams

$$\begin{tikzcd}
H^1(Y, \mathscr{O}_Y) \arrow[r, "f^*"] & H^1(X, \mathscr{O}_X) & H^1(X, \mathscr{O}_X) \arrow[r, "f_*"] & H^1(Y, \mathscr{O}_Y) \\
H^1(Y, \underline{\mathbf{Z}}) \arrow[r, "f^*"] \arrow[u] & H^1(X, \underline{\mathbf{Z}}) \arrow[u] & H^1(X, \underline{\mathbf{Z}}) \arrow[r, "f_*"] \arrow[u] & H^1(Y, \underline{\mathbf{Z}})\,, \arrow[u]
\end{tikzcd}$$

where the columns are induced by the canonical maps $\underline{\mathbf{Z}} \to \mathscr{O}_Y$ and $\underline{\mathbf{Z}} \to \mathscr{O}_X$. Passing to quotients on the columns therefore gives rise to maps

$$f^* : \mathrm{Pic}^0_Y \to \mathrm{Pic}^0_X, \qquad f_* : \mathrm{Pic}^0_X \to \mathrm{Pic}^0_Y$$

of analytic Lie groups. These maps are "computed" by

Lemma 5.4. *In the above situation, $f^* = \mathrm{Pic}^0(f)$ is the map induced by Pic^0 functoriality and $f_* = \mathrm{Alb}(f)$ is the map induced by Albanese functoriality. These are dual with respect to the canonical autodualities of Pic^0_X, Pic^0_Y.*

The significance of the theory of Jacobians is that by (5.3) we have a canonical isomorphism

$$\begin{aligned} T_\ell(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}) &\cong H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}_\ell) \\ &\cong H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}}) \otimes_{\mathbf{Z}} \mathbf{Z}_\ell, \end{aligned} \tag{5.4}$$

connecting the ℓ-adic Tate module of $\mathrm{Pic}^0_{X_1(N)}$ with the $\mathbf{Z}$-module $H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$ that "encodes" $S_2(\Gamma_1(N), \mathbf{C})$ via the Shimura isomorphism. Note that this isomorphism is defined in terms of the analytic construction (5.3) which depends upon the choice of i. The intrinsic isomorphism (compatible with étale cohomology) has $\mathbf{Z}$ above replaced by $2\pi i\mathbf{Z} = -2\pi i\mathbf{Z}$.

Definition 5.5. We define endomorphisms of $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$ via

$$T_p^* = \mathrm{Alb}(\pi_1^{(p)}) \circ \mathrm{Pic}^0(\pi_2^{(p)}), \quad \langle n\rangle^* = \mathrm{Pic}^0(I_n), \quad w_\zeta^* = \mathrm{Pic}^0(w_\zeta).$$

By Lemma 5.4, it follows that the above isomorphism (5.4) carries the operators on $T_\ell(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}})$ over to the ones *previously defined* on $H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Z}})$ (which are, in turn, compatible with the classical operations via the Shimura isomorphism). By the faithfulness of the "Tate module" functor on complex tori, we conclude that $\mathbf{T}_1(N)$ *acts* on $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$ in a unique manner compatible with the above definition, and (5.4) is an isomorphism of $\mathbf{T}_1(N) \otimes_{\mathbf{Z}} \mathbf{Z}_\ell$-modules. We call this the $(\,)^*$-*action* of $\mathbf{T}_1(N)$ on $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$.

We must warn the reader that under the canonical isomorphism of $\mathbf{C}$-vector spaces

$$\begin{aligned} S_2(\Gamma_1(N), \mathbf{C}) &\cong H^0(X_1(N)^{\mathrm{an}}, \Omega^1_{X_1(N)^{\mathrm{an}}}) \\ &\cong H^0(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}, \Omega^1_{\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}}) \\ &\cong \mathrm{Cot}_0(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}), \end{aligned}$$

the $(\,)^*$-action of $T \in \mathbf{T}_1(N)$ on $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$ does *not* go over to the classical action of T on $S_2(\Gamma_1(N), \mathbf{C})$, but rather the adjoint of T with respect to the Petersson pairing. To clear up this matter, we make the following definition:

Definition 5.6.

$$(T_p)_* = \mathrm{Alb}(\pi_2^{(p)}) \circ \mathrm{Pic}^0(\pi_1^{(p)}), \quad \langle n\rangle_* = \mathrm{Alb}(I_n), \quad (w_\zeta)_* = \mathrm{Alb}(w_\zeta).$$

Since $I_n^{-1} = I_{n^{-1}}$ and $w_\zeta^{-1} = w_\zeta$ on $X_1(N)^{\mathrm{an}}$, we have $(w_\zeta)_* = w_\zeta^*$ and $\langle n\rangle_* = \langle n^{-1}\rangle^*$. We claim that the above $(\,)_*$ operators are the *dual* morphisms (with respect to the canonical principal polarization of $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$) of the $(\,)^*$ operators and induce exactly the *classical* action of T_p and $\langle n\rangle$ on $S_2(\Gamma_1(N), \mathbf{C})$, so we also have a well-defined $(\,)_*$-action of $\mathbf{T}_1(N)$ on $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$, dual to the $(\,)^*$-action. By Theorem 5.2, Corollary 5.3, and Lemma 5.4, this follows from the following general fact about compact Riemann surfaces. The proof is non-trivial.

Lemma 5.7. *Let X be a compact Riemann surface, and use the i-orientation to define $H^2(X, \underline{\mathbf{Z}}) \cong \mathbf{Z}$. Use $1 \mapsto e^{2\pi i/\ell^n}$ to define $\mathbf{Z}/\ell^n \cong \mu_{\ell^n}(\mathbf{C})$ for all n. The* *diagram*

$$\begin{tikzcd}
H^1(X, \underline{\mathbf{Z}}_\ell) \otimes_{\mathbf{Z}_\ell} H^1(X, \underline{\mathbf{Z}}_\ell) \arrow[r, "\cup"] \arrow[d, "\cong"] & \mathbf{Z}_\ell \arrow[d, "\cong"] \\
T_\ell(\mathrm{Pic}^0_X) \otimes_{\mathbf{Z}_\ell} T_\ell(\mathrm{Pic}^0_X) \arrow[r] & \varprojlim \mu_{\ell^n}(\mathbf{C})
\end{tikzcd}$$

anticommutes (*i.e., going around from upper left to lower right in the two possible ways gives results that are negatives of each other*), *where the bottom row is the ℓ-adic Weil pairing* (*with respect to the canonical principal polarization* $\mathrm{Pic}^0_X \cong \widehat{\mathrm{Pic}^0_X}$ *for the "second"* Pic^0_X *in the lower left.*)

Note that the sign doesn't affect formation of adjoints. It ultimately comes from the sign on the bottom of [**77**, pg. 237] since our Weil pairing sign convention agrees with [**77**].

We now summarize our findings in terms of $V_\ell(N) = \mathbf{Q}_\ell \otimes_{\mathbf{Z}_\ell} T_\ell(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}})$, which has a perfect alternating Weil pairing

$$(\ ,\ )_\ell : V_\ell(N) \otimes V_\ell(N) \to \mathbf{Q}_\ell(1)$$

and has two $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$-actions, via the $(\,)^*$-actions and the $(\,)_*$-actions. Since $(w_\zeta)_* = w_\zeta^*$, we simply write w_ζ for this operator on $V_\ell(N)$.

Theorem 5.8. *Let $\mathbf{T}_1(N)$ act on $V_\ell(N)$ with respect to the $(\,)^*$-action or with respect to the $(\,)_*$-action. With respect to $(\ ,\ )_\ell$, the adjoint of T_p for $p \nmid N$ is $\langle p\rangle^{-1}T_p$ and the adjoint of $\langle n\rangle$ is $\langle n\rangle^{-1}$ for $n \in (\mathbf{Z}/N\mathbf{Z})^*$. With respect to*

$$[x, y]_\ell = (x, w_\zeta(y))_\ell$$

for $\zeta \in \boldsymbol{\mu}_N(\mathbf{C})$ a primitive Nth root of unity, the action of $\mathbf{T}_1(N)$ on $V_\ell(N)$ is self-adjoint. In general, adjointness with respect to $(\ ,\ )_\ell$ interchanges the $(\,)_*$-action and $(\,)^*$-action.

It should be noted that when making the translation to étale cohomology, the $(\,)^*$-action plays a more prominent role (since this is what makes (5.4) a $\mathbf{T}_1(N)$-equivariant map). However, when working directly with Tate modules and arithmetic Frobenius elements, it is the $(\,)_*$-action which gives the cleaner formulation of Shimura's results.

An important consequence of Theorem 5.8 is

Corollary 5.9. *The $\mathbf{Q}_\ell \otimes_{\mathbf{Z}} \mathbf{T}_1(N)$-module $V_\ell(N)$ is free of rank 2 for either action, and* $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N), \mathbf{Q})$ *is free of rank 1 over* $\mathbf{Q} \otimes \mathbf{T}_1(N)$ (*hence likewise with* $\mathbf{Q}$ *replaced by any field of characteristic* 0).

Remark 5.10. The assertion about $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N), \mathbf{Q})$ is equivalent to the intrinsic condition that $\mathbf{Q} \otimes \mathbf{T}_1(N)$ is *Gorenstein*. Also, this freeness clearly makes the two assertions about $V_\ell(N)$ for the $(\,)_*$- and $(\,)^*$-actions equivalent. *For the proof*, the $(\,)^*$-action is what we use. But in what follows, it is the case of the $(\,)_*$-action that we need!

Proof. Using (5.4) and the choice of $(\,)^*$-action on $V_\ell(N)$, it suffices to prove

- $H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Q}})$ is free of rank 2 over $\mathbf{Q} \otimes \mathbf{T}_1(N)$,
- $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N), \mathbf{Q})$ is free of rank 1 over $\mathbf{Q} \otimes \mathbf{T}_1(N)$.

Using $[\,,\,]_{\Gamma_1(N)}$, we have

$$H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Q}}) \cong \mathrm{Hom}_{\mathbf{Q}}(H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Q}}), \mathbf{Q}) \tag{5.5}$$

as $\mathbf{Q} \otimes \mathbf{T}_1(N)$-modules, so we may study this $\mathbf{Q}$-dual instead. Since $\mathbf{Q} \otimes \mathbf{T}_1(N)$ is semilocal, a finite module over this ring is locally free of constant rank if and only if it is *free* of that rank. But local freeness of constant rank can be checked after faithfully flat base change. Applying this with the base change $\mathbf{Q} \to \mathbf{C}$, and noting that $\mathbf{C} \otimes \mathbf{T}_1(N)$ is semilocal, it suffices to replace $\mathbf{Q}$ by $\mathbf{C}$ above.

Note that *if* the right hand side of (5.5) is free of rank 2, so is the left side, so choosing a basis of the left side and feeding it into the right hand side shows that $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N)^{\oplus 2}, \mathbf{Q})$ is free of rank 2. In particular, the direct summand $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N), \mathbf{Q})$ is flat over $\mathbf{Q} \otimes \mathbf{T}_1(N)$ with full support over $\mathrm{Spec}(\mathbf{Q} \otimes \mathbf{T}_1(N))$, so it must be locally free with local rank at least 1 at all points of $\mathrm{Spec}(\mathbf{Q} \otimes \mathbf{T}_1(N))$. Consideration of $\mathbf{Q}$-dimensions then forces $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q} \otimes \mathbf{T}_1(N), \mathbf{Q})$ to be locally free of rank 1, hence free of rank 1. In other words, it suffices to show that $\mathrm{Hom}_{\mathbf{Q}}(H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{Q}}), \mathbf{Q})$ is free of rank 2 over $\mathbf{T}_1(N) \otimes \mathbf{Q}$, or equivalently that $\mathrm{Hom}_{\mathbf{C}}(H^1(X_1(N)^{\mathrm{an}}, \underline{\mathbf{C}}), \mathbf{C})$ is free of rank 2 over $\mathbf{T}_1(N) \otimes \mathbf{C}$.

Via the Shimura isomorphism (in weight 2), which is compatible with the Hecke actions, we are reduced to showing that $\mathrm{Hom}(S_2(\Gamma_1(N), \mathbf{C}), \mathbf{C})$ is free of rank 1 over $\mathbf{C} \otimes \mathbf{T}_1(N)$. For this purpose, we will study the $\mathbf{C} \otimes \mathbf{T}_1(N)$-equivariant $\mathbf{C}$-bilinear pairing

$$S_2(\Gamma_1(N), \mathbf{C}) \otimes_{\mathbf{C}} (\mathbf{C} \otimes \mathbf{T}_1(N)) \to \mathbf{C}$$
$$(f, T) \mapsto a_1(Tf)$$

were $a_1(\cdot)$ is the "Fourier coefficient of q". This is $\mathbf{C} \otimes \mathbf{T}_1(N)$-equivariant, since $\mathbf{T}_1(N)$ is commutative. It suffices to check that there's no nonzero kernel on either side of this pairing. Since

$$\mathbf{C} \otimes \mathbf{T}_1(N) \to \mathrm{End}_{\mathbf{C}}(S_2(\Gamma_1(N), \mathbf{C}))$$

is *injective* (as noted in (5.2)) and $a_1(TT_n f) = a_n(Tf)$ for $T \in \mathbf{T}_1(N)$, the kernel on the right is trivial. Since $a_1(T_n f) = a_n(f)$, the kernel on the left is also trivial. □

5.2. Algebraic preliminaries

Let S be a scheme. An *elliptic curve* $E \to S$ is a proper smooth group scheme with geometrically connected fibers of dimension 1 (necessarily of genus 1). It follows from [**62**, Ch.2] that the group structure is commutative and uniquely determined by the identity section. Fix $N \geq 1$ and assume $N \in H^0(S, \mathcal{O}_S^*)$ (i.e., S is a $\mathbf{Z}[\frac{1}{N}]$-scheme). Thus, the map $N : E \to E$ is finite *étale* of degree N^2 as can be checked on geometric fibers. A *point of exact order* N on E is a section $P : S \to E$ which is killed by N (i.e., factors through the finite étale group scheme $E[N]$) and induces a point of exact order N on geometric fibers.

It follows from the stack-theoretic methods in [**25**] or the more explicit descent arguments in [**62**] that for $N \geq 5$ there is a proper smooth $\mathbf{Z}[\frac{1}{N}]$-scheme $X_1(N)$ equipped with a finite flat map to $\mathbf{P}^1_{\mathbf{Z}[\frac{1}{N}]}$, such that the open subscheme $Y_1(N)$ lying over $\mathbf{P}^1_{\mathbf{Z}[\frac{1}{N}]} - \{\infty\} = \mathbf{A}^1_{\mathbf{Z}[\frac{1}{N}]}$ is the base of a universal object $(E_1(N), P) \to Y_1(N)$ for elliptic curves with a point of exact order N over variable $\mathbf{Z}[\frac{1}{N}]$-schemes.

Moreover, the fibers of $X_1(N) \to \mathrm{Spec}\,\mathbf{Z}[\frac{1}{N}]$ are *geometrically connected*, as this can be checked on a single geometric fiber and by choosing the complex fiber we may appeal to the fact (whose proof requires some care) that there is an isomorphism $(X_1(N) \times_{\mathbf{Z}[\frac{1}{N}]} \mathbf{C})^{\mathrm{an}} \cong X_1(N)^{\mathrm{an}}$ identifying the "algebraic" data $(\mathbf{C}/[1,z], \frac{1}{N})$ in $Y_1(N)(\mathbf{C}) \subset X_1(N)(\mathbf{C})$ with the class of $z \in \mathfrak{h}$ in $\Gamma_1(N)\backslash\mathfrak{h} = Y_1(N)^{\mathrm{an}} \subset X_1(N)^{\mathrm{an}}$ (and $X_1(N)^{\mathrm{an}}$ *is* connected, as $\mathfrak{h}$ is). These kinds of compatibilities are somewhat painful to check unless one develops a full-blown relative theory of elliptic curves in the analytic world (in which case the verifications become quite mechanical and natural).

Again fixing $N \geq 5$, but now also a prime p, we want an algebraic analogue of $X_1(N,p)^{\mathrm{an}}$ over $\mathbf{Z}[\frac{1}{Np}]$. Let $(E,P) \to S$ be an elliptic curve with a point of exact order N over a $\mathbf{Z}[\frac{1}{Np}]$-scheme S. We're interested in studying triples $(E,P,C) \to S$ where $C \subset E$ is an order-p finite locally free S-subgroup-scheme which is not contained in the subgroup generated by P on geometric fibers (if $p \mid N$). Methods in [**25**] and [**62**] ensure the existence of a universal such object $(E_1(N,p),P,C) \to Y_1(N,p)$ for a smooth affine $\mathbf{Z}[\frac{1}{Np}]$-scheme which naturally sits as the complement of a relative Cartier divisor in a proper smooth $\mathbf{Z}[\frac{1}{Np}]$-scheme $X_1(N,p)$ which is finite flat over $\mathbf{P}^1_{\mathbf{Z}[\frac{1}{Np}]}$ (with $Y_1(N,p)$ the preimage of $\mathbf{A}^1_{\mathbf{Z}[\frac{1}{Np}]}$). Base change to $\mathbf{C}$ and analytification recovers $X_1(N,p)^{\mathrm{an}}$ as before, so $X_1(N,p) \to \operatorname{Spec} \mathbf{Z}[\frac{1}{Np}]$ has geometrically connected fibers.

There are maps of $\mathbf{Z}[\frac{1}{Np}]$-schemes (respectively, $\mathbf{Z}[\frac{1}{N}]$-schemes)

$$\begin{tikzcd} & Y_1(N,p) \arrow[dl, "\pi_1^{(p)}"'] \arrow[dr, "\pi_2^{(p)}"] & & & Y_1(N) \arrow[r, "I_n"] & Y_1(N) \\ Y_1(N)[\frac{1}{p}] & & Y_1(N)[\frac{1}{p}] & & & \end{tikzcd}$$

determined by $(E,P,C) \xrightarrow{\pi_1^{(p)}} (E,P)$ and $(E,P,C) \xrightarrow{\pi_2^{(p)}} (E/C,P)$ (which makes sense in $Y_1(N)$ if $p \mid N$ by the "disjointness" condition on C and P) and $I_n(E,P) = (E,nP)$. Although $\pi_2^{(p)}$ is *not* a map over $\mathbf{A}^1_{\mathbf{Z}[\frac{1}{Np}]}$, it can be shown that these all uniquely extend to (necessarily finite *flat*) maps, again denoted $\pi_1^{(p)}$, $\pi_2^{(p)}$, I_n between $X_1(N,p)$, $X_1(N)[\frac{1}{p}]$, $X_1(N)$. A proof of this fact requires the theory of minimal regular proper models of curves over a Dedekind base; the analogous fact over $\mathbf{Q}$ is an immediate consequence of basic facts about proper smooth curves over a field, but in order to most easily do some later calculations in characteristic $p \nmid N$ it is convenient to know that we have the map I_p defined on $X_1(N)$ over $\mathbf{Z}[1/N]$ (though this could be bypassed by using liftings to characteristic 0 in a manner similar to our later calculations of T_p in characteristic p).

Likewise, over $\mathbf{Z}[\frac{1}{N}, \zeta_N]$ we can define, for any primitive Nth root of unity $\zeta = \zeta_N^i$ ($i \in (\mathbf{Z}/N\mathbf{Z})^*$), an operator $w_\zeta : Y_1(N)_{/\mathbf{Z}[\frac{1}{N},\zeta_N]} \to Y_1(N)_{/\mathbf{Z}[\frac{1}{N},\zeta_N]}$ via $w_\zeta(E,P) = (E/\langle P\rangle, P')$ where $\langle P\rangle$ is the order-N étale subgroup-scheme generated by P and $P' \in (E[N]/\langle P\rangle)(S)$ is uniquely determined by the relative Weil pairing condition $\langle P, P'\rangle_N = \zeta$ (with $P' \in E[N](S)$ here). This really does extend to $X_1(N)_{/\mathbf{Z}[\frac{1}{N},\zeta_N]}$, and one checks that $w_{\zeta^j} w_\zeta = I_j$ for $j \in (\mathbf{Z}/N\mathbf{Z})^*$. In particular, $w_\zeta^2 = 1$.

Since $X_1(N) \to \operatorname{Spec} \mathbf{Z}[\frac{1}{N}]$ is a proper smooth scheme with geometrically connected fibers of dimension 1, $\operatorname{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$ is an abelian scheme over $\mathbf{Z}[\frac{1}{N}]$ and hence is the Néron model of its generic fiber. We have scheme-theoretic Albanese and Pic^0 functoriality for finite (flat) maps between proper smooth curves (with geometrically connected fibers) over any base at all, and analytification of such a situation over $\mathbf{C}$ recovers the classical theory of Pic^0 as used in Section 5.1.

For example, we have endomorphisms

$$\langle n\rangle^* = \mathrm{Pic}^0(I_n), \quad \langle n\rangle_* = \mathrm{Alb}(I_n)$$

on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$,

$$w_\zeta^* = \mathrm{Pic}^0(w_\zeta) = \mathrm{Alb}(w_\zeta) = (w_\zeta)_*$$

on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N},\zeta_N]}}$, and

$$\begin{aligned} T_p^* &= \mathrm{Alb}(\pi_1^{(p)}) \circ \mathrm{Pic}^0(\pi_2^{(p)}) \\ (T_p)_* &= \mathrm{Alb}(\pi_2^{(p)}) \circ \mathrm{Pic}^0(\pi_1^{(p)}) \end{aligned}$$

on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{Np}]}}$. A key point is that by the *Néronian property*, T_p^* and $(T_p)_*$ uniquely extend to endomorphisms of $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$, even though the $\pi_i^{(p)}$ do *not* make sense over $\mathbf{Z}[\frac{1}{N}]$ from what has gone before. In particular, it makes sense to study T_p^* and $(T_p)_*$ on the abelian variety $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{F}_p}}$ over $\mathbf{F}_p$ for $p \nmid N$. This will be rather crucial later, but note it requires the Néronian property in the definition.

Passing to the analytifications, the above constructions recover the operators defined on $\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}$ in Section 5.1. The resulting subring of

$$\mathrm{End}(\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}) \subset \mathrm{End}(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}})$$

generated by T_p^*, $\langle n\rangle^*$ (respectively, by $(T_p)_*$, $\langle n\rangle_*$) is identified with $\mathbf{T}_1(N)$ via its $(\,)^*$-action (respectively, via its $(\,)_*$-action) and using

$$\varprojlim \mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}[\ell^n](\overline{\mathbf{Q}}) \cong T_\ell(\mathrm{Pic}^0_{X_1(N)^{\mathrm{an}}}) \tag{5.6}$$

(using $\overline{\mathbf{Q}} \subset \mathbf{C}$) endows our "analytic" $V_\ell(N)$ with a canonical *continuous* action of $G_{\mathbf{Q}} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ unramified at all $p \nmid N\ell$ (via Néron-Ogg-Shafarevich) and *commuting* with the action of $\mathbf{T}_1(N)$ (via either the $(\,)^*$-action or the $(\,)_*$-action). We also have an endomorphism $w_\zeta = w_\zeta^* = (w_\zeta)_*$ on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N},\zeta_N]}}$ and it is easy to see that

$$(g^{-1})^* w_{g(\zeta)} g^* = w_\zeta$$

on $\overline{\mathbf{Q}}$-points, where $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and g^* denotes the natural action of g on $\overline{\mathbf{Q}}$-points (corresponding to base change of degree 0 line bundles on $X_1(N)_{/\overline{\mathbf{Q}}}$). Since $w_\zeta = w_{\zeta^{-1}}$ (as $(E,P) \cong (E,-P)$ via -1), we see that w_ζ is defined over the real subfield $\mathbf{Q}(\zeta_N)^+$. By étale descent, the operator w_ζ is defined over $\mathbf{Z}[\frac{1}{N},\zeta_N]^+$.

In any case, w_ζ acts on $V_\ell(N)$, recovering the operator in Section 5.1, and so this conjugates the $(\,)^*$-action to the $(\,)_*$-action, taking each $T \in \mathbf{T}_1(N)$ (for either action on $V_\ell(N)$) to its Weil pairing adjoint, via the canonical principal polarization of the abelian scheme $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$. Using Corollary 5.3 and (5.6) we obtain

Lemma 5.11. *Let $\mathbf{T}_1(N)$ act on $V_\ell(N)$ through either the $(\,)^*$-action or the $(\,)_*$-action. Then $\rho_{N,\ell} : G_{\mathbf{Q}} \to \mathrm{Aut}(V_\ell(N)) \cong \mathrm{GL}(2, \mathbf{Q}_\ell \otimes \mathbf{T}_1(N))$ is a continuous representation, unramified at $p \nmid N\ell$.*

The main result we are after is

Theorem 5.12. *Let $\mathbf{T}_1(N)$ act on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$ via the $(\,)_*$-action. For any $p \nmid N\ell$, the characteristic polynomial of $\rho_{N,\ell}(\mathrm{Frob}_p)$ is*

$$X^2 - (T_p)_* X + p\langle p\rangle_*$$

relative to the $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$-module structure on $V_\ell(N)$, where Frob_p denotes an arithmetic Frobenius element at p.

The proof of Theorem 5.12 will make essential use of the w_ζ operator. For the remainder of this section, we admit Theorem 5.12 and deduce its consequences. Let $f \in S_2(\Gamma_1(N), \mathbf{C})$ be a *newform* of level N. Let $K_f \subset \mathbf{C}$ be the number field generated by $a_p(f)$ for all $p \nmid N$, where $f = \sum a_n(f)q^n$, so by weak multiplicity one $a_n(f) \in K_f$ for all $n \geq 1$ and the Nebentypus character χ_f has values in K_f. Let $\mathfrak{p}_f \subset \mathbf{T}_1(N)$ be the minimal prime corresponding to f (i.e., the kernel of the map $\mathbf{T}_1(N) \to K_f$ sending each $T \in \mathbf{T}_1(N)$ to its eigenvalue on f).

We now require $\mathbf{T}_1(N)$ to act on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}$ via its $(\,)_*$-action.

Definition 5.13. A_f is the quotient of $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Q}}}$ by $\mathfrak{p}_f \subset \mathbf{T}_1(N)$.

By construction, A_f has good reduction over $\mathbf{Z}[\frac{1}{N}]$ and the action of $\mathbf{T}_1(N)$ on $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Q}}}$ induces an action of $\mathbf{T}_1(N)/\mathfrak{p}$ on A_f, hence an action of $K_f \cong (\mathbf{T}_1(N)/\mathfrak{p}) \otimes_{\mathbf{Z}} \mathbf{Q}$ on A_f in the "up-to-isogeny" category.

Theorem 5.14 (Shimura)**.** *We have $\dim A_f = [K_f : \mathbf{Q}]$ and $V_\ell(A_f)$ is free of rank 2 over $\mathbf{Q}_\ell \otimes_{\mathbf{Q}} K_f$, with Frob_p having characteristic polynomial*

$$X^2 - (1 \otimes a_p(f))X + 1 \otimes p\chi_f(p)$$

for all $p \nmid N\ell$.

Proof. By Lemma 5.11 and Theorem 5.12, we just have to check that the $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$-linear map

$$V_\ell(\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Q}}}) \to V_\ell(A_f)$$

identifies the right hand side with the quotient of the left hand side by $\mathfrak{p}_f$. More generally, for any exact sequence

$$B' \to B \to A \to 0$$

of abelian varieties over a field of characteristic prime to ℓ, we claim

$$V_\ell(B') \to V_\ell(B) \to V_\ell(A) \to 0$$

is exact. We may assume the base field is algebraically closed, and then may appeal to Poincaré reducibility (see [**77**, pg. 173]). □

Choosing a place λ of K_f over ℓ and using the natural realization of $K_{f,\lambda}$ as a factor of $\mathbf{Q}_\ell \otimes K_f$, we deduce from Theorem 5.14:

Corollary 5.15. *Let $f \in S_2(\Gamma_1(N), \mathbf{C})$ be a newform and λ a place of K_f over ℓ. There exists a continuous representation*

$$\rho_{f,\lambda} : G_{\mathbf{Q}} \to GL(2, K_{f,\lambda})$$

unramified at all $p \nmid N\ell$, with Frob_p *having characteristic polynomial*

$$X^2 - a_p(f)X + p\chi_f(p) \in K_{f,\lambda}[X].$$

5.3. Proof of Theorem 5.12

Fix $p \nmid N$ and let

$$J_p = \mathrm{Pic}^0_{X_1(N)_{/\mathbf{F}_p}} \cong \mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}} \times_{\mathbf{Z}[\frac{1}{N}]} \mathbf{F}_p$$

with $\mathbf{T}_1(N)$ acting through the $(\,)_*$-action. Fix a choice of Frob_p, or more specifically fix a choice of place in $\overline{\mathbf{Q}}$ over p. Note that this determines a preferred algebraic closure $\overline{\mathbf{F}}_p$ as a quotient of the ring of algebraic integers, and in particular a map $\mathbf{Z}[1/N, \zeta_N] \to \overline{\mathbf{F}}_p$. Thus, we may view w_ζ as inducing an endomorphism of the abelian variety $J_p \times_{\mathbf{F}_p} \overline{\mathbf{F}}_p$ over $\overline{\mathbf{F}}_p$ (whereas the elements in $\mathbf{T}_1(N)$ induce endomorphisms of J_p over $\mathbf{F}_p$). The canonical isomorphism

$$V_\ell(\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Q}}}) \cong V_\ell(\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}) \cong V_\ell(J_p)$$

identifies the Frob_p-action on $\overline{\mathbf{Q}}$-points on the left hand side with the (arithmetic) Frobenius action on $\overline{\mathbf{F}}_p$-points on the right hand side. Obviously $V_\ell(J_p)$ is a module over the ring $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$ and is free of rank 2 as such. For *any* $\mathbf{F}_p$-schemes Z, Z' and any $\mathbf{F}_p$-map $f : Z \to Z'$ the diagram

(5.7)
$$\begin{tikzcd} Z \arrow[r, "f"] \arrow[d, "F_Z"'] & Z' \arrow[d, "F_{Z'}"] \\ Z \arrow[r, "f"] & Z' \end{tikzcd}$$

commutes, where columns are absolute Frobenius. Taking $Z = \mathrm{Spec}\,\overline{\mathbf{F}}_p$, $Z' = J_p$, we see that the Frob_p action of $V_\ell(J_p)$ through $\overline{\mathbf{F}}_p$-points is *identical* to the action induced by the intrinsic absolute Frobenius morphism $F : J_p \to J_p$ over $\mathbf{F}_p$. Here is the essential input, to be proven later.

Theorem 5.16 (Eichler-Shimura)**.** *In* $\mathrm{End}_{\overline{\mathbf{F}}_p}(J_p)$,

$$(T_p)_* = F + \langle p\rangle_* F^\vee, \qquad w_\zeta^{-1} F w_\zeta = \langle p\rangle_*^{-1} F$$

where $F^\vee$ denotes the dual morphism.

The extra relation involving w_ζ is crucial. The interested reader should compare this with [**108**, Cor. 7.10].

Let us admit Theorem 5.16 and use it to prove Theorem 5.12. We will then prove Theorem 5.16. Using an $\mathbf{F}_p$-rational base point P (e.g., the cusp 0), we get a commutative diagram

$$\begin{tikzcd} X_1(N)_{/\mathbf{F}_p} \arrow[r, hook] \arrow[d, "F_{X_1(N)}"'] & J_p \arrow[d, "F"] \\ X_1(N)_{/\mathbf{F}_p} \arrow[r, hook] & J_p \end{tikzcd}$$

where $F_{X_1(N)}$ denotes the absolute Frobenius morphism of $X_1(N)_{/\mathbf{F}_p}$, so by Albanese functoriality $F = \mathrm{Alb}(F_{X_1(N)})$. Thus

$$\begin{aligned} FF^\vee &= \mathrm{Alb}(F_{X_1(N)}) \circ \mathrm{Pic}^0(F_{X_1(N)}) \\ &= \deg(F_{X_1(N)}) = p \end{aligned}$$

as $X_1(N)_{/\mathbf{F}_p}$ is a smooth *curve*. We conclude from $(T_p)_* = F + \langle p\rangle_* F^\vee$ that

$$F^2 - (T_p)_* F + p\langle p\rangle_* = 0$$

on J_p, hence in $V_\ell(J_p)$. Thus, $\rho_{N,\ell}(\mathrm{Frob}_p)$ satisfies the expected quadratic polynomial

$$X^2 - (T_p)_* X + p\langle p\rangle_* = 0.$$

Let $X^2 - aX + b$ be the *true* characteristic polynomial, which $\rho_{N,\ell}(\mathrm{Frob}_p)$ must also satisfy, by Cayley-Hamilton. We must *prove* that $a = (T_p)_*$, and then $b = p\langle p\rangle_*$ is forced. It is this matter which requires the second relation.

We want $\mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(\rho_{N,\ell}(\mathrm{Frob}_p)) = (T_p)_*$ or equivalently

$$\mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(V_\ell(F)) = (T_p)_*.$$

Using the modified Weil pairing

$$[x, y]_\ell = (x, w_\zeta y)_\ell$$

and using the fact that $V_\ell(J_p) \cong V_\ell(\mathrm{Pic}^0_{X_1(N)/\mathbf{Q}})$ respects Weil pairings (by invoking the relativization of this concept, here over $\mathbf{Z}[\frac{1}{N}]$) we may identify (via Theorem 5.8 and a choice $\mathbf{Q}_\ell(1) \cong \mathbf{Q}_\ell$ as $\mathbf{Q}_\ell$-vector spaces)

$$V_\ell(J_p) \cong \mathrm{Hom}_{\mathbf{Q}_\ell}(V_\ell(J_p), \mathbf{Q}_\ell) := V_\ell(J_p)^*$$

as $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$-modules, but taking the F-action over to the $\langle p\rangle_* F^\vee$-action, since adjoints with respect to Weil pairings are dual morphisms and $w_\zeta^{-1} F^\vee w_\zeta$ is dual to $w_\zeta^{-1} F w_\zeta = \langle p\rangle_*^{-1} F = F\langle p\rangle_*^{-1}$ (absolute Frobenius commutes with all morphisms of $\mathbf{F}_p$-schemes!)

Since $V_\ell(J_p)$ is free of rank 2 over $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$ and $\mathrm{Hom}_{\mathbf{Q}_\ell}(\mathbf{Q}_\ell \otimes \mathbf{T}_1(N), \mathbf{Q}_\ell)$ is free of rank 1 over $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$, by Corollary 5.9, we conclude

$$\mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(F | V_\ell(J_p)) = \mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(\langle p\rangle_* F^\vee | V_\ell(J_p)^*).$$

We wish to invoke the following applied to the $\mathbf{Q}_\ell$-algebra $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$ and the $\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)$-module $V_\ell(J_p)$:

Lemma 5.17. *Let $\mathcal{O}$ be a commutative ring, A a finite locally free $\mathcal{O}$-algebra with* $\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$ *a locally free A-module (necessarily of rank 1). Let M be a finite locally free A-module, $M^* = \mathrm{Hom}_{\mathcal{O}}(M, \mathcal{O})$, so M^* is finite and locally free over A with the same rank as M. For any A-linear map $f : M \to M$ with $\mathcal{O}$-dual $f^* : M^* \to M^*$, automatically A-linear,*

$$\mathrm{char}(f) = \mathrm{char}(f^*)$$

in $A[T]$ (these are the characteristic polynomials).

Proof. Without loss of generality $\mathcal{O}$ is local, so A is semilocal. Making faithfully flat base change to the henselization of $\mathcal{O}$ (or the completion if $\mathcal{O}$ is noetherian or if we first reduce to the noetherian case), we may assume that A is a product of local rings. Without loss of generality, A is then local, so

$$M = \oplus A e_i$$

if free, and $\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$ is free of rank 1 over A. Choose an isomorphism

$$h : A \cong \mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$$

as A-modules, so the projections

$$\pi_i : M \to Ae_i \cong A$$

satisfy $e_i^* = h(i) \circ \pi_i$ in M^*. These e_i^* are an A-basis of M^* and we compute matrices over A:

$$\mathrm{Mat}_{\{e_i\}}(f) = \mathrm{Mat}_{\{e_i^*\}}(f^*)^t.$$

□

We conclude that

$$\mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(F|V_\ell(J_p)) = \mathrm{tr}_{\mathbf{Q}_\ell \otimes \mathbf{T}_1(N)}(\langle p\rangle_* f^\vee|V_\ell(J_p)).$$

By Theorem 5.16, we have

$$\begin{aligned} 2(T_p)_* &= \mathrm{tr}((T_p)_*|V_\ell(J_p)) \\ &= \mathrm{tr}(F + \langle p\rangle_* F^\vee|V_\ell(J_p)) \\ &= 2\,\mathrm{tr}(F|V_\ell(J_p)). \end{aligned}$$

This proves that $\mathrm{tr}(F|V_\ell(J_p)) = (T_p)_*$, so indeed $X^2 - (T_p)_*X + p\langle p\rangle_*$ *is* the characteristic polynomial. Finally, there remains

Proof of Theorem 5.16. It suffices to check the maps coincide on a Zariski dense subset of $J_p(\overline{\mathbf{F}}_p) = \mathrm{Pic}^0(X_1(N)_{/\overline{\mathbf{F}}_p})$. If g is the genus of $X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}$ and we fix an $\overline{\mathbf{F}}_p$-rational base point, we get an induced surjective map

$$X_1(N)^g_{/\overline{\mathbf{F}}_p} \to J_{p/\overline{\mathbf{F}}_p},$$

so for any dense open $U \subset X_1(N)_{\overline{\mathbf{F}}_p}$, $U^g \to (J_p)_{/\overline{\mathbf{F}}_p}$ hits a Zariski dense subset of $\overline{\mathbf{F}}_p$-points. Taking U to be the ordinary locus of $Y_1(N)_{/\overline{\mathbf{F}}_p}$, it suffices to study what happens to a difference $(x) - (x')$ for $x, x' \in Y_1(N)(\overline{\mathbf{F}}_p)$ corresponding to (E, P), (E', P') over $\overline{\mathbf{F}}_p$ with E and E' *ordinary* elliptic curves.

By the commutative diagram (5.7), the map

$$J_p(\overline{\mathbf{F}}_p) \to J_p(\overline{\mathbf{F}}_p)$$

induced by F is the same as the map induced by the pth power map in $\overline{\mathbf{F}}_p$. By *definition* of Pic^0 functoriality, this corresponds to base change of an invertible sheaf on $X_1(N)_{/\overline{\mathbf{F}}_p}$ by the absolute Frobenius on $\overline{\mathbf{F}}_p$. By *definition* of $Y_1(N)_{/\overline{\mathbf{F}}_p}$ as a universal object, such base change induces on $Y_1(N)(\overline{\mathbf{F}}_p)$ *exactly* "base change by absolute Frobenius" on elliptic curves with a point of exact order N over $\overline{\mathbf{F}}_p$. We conclude

$$F((x) - (x')) = (E^{(p)}, P^{(p)}) - ((E')^{(p)}, P^{(p)})$$

where $(\,)^{(p)}$ denotes base change by absolute Frobenius on $\overline{\mathbf{F}}_p$.

Since $p = FF^\vee = F^\vee F$ and F is bijective on $\overline{\mathbf{F}}_p$-points, we have

$$\begin{aligned} F^\vee((x) - (x')) &= pF^{-1}((x) - (x')) \\ &= p((E^{(p^{-1})}, P^{(p^{-1})}) - ((E')^{(p^{-1})}, (P')^{(p^{-1})})). \end{aligned}$$

Thus,

$$\langle p\rangle_* F^\vee((x) - (x')) = p(E^{(p^{-1})}, pP^{(p^{-1})}) - p((E')^{(p^{-1})}, p(P')^{(p^{-1})})$$

so

$$\begin{aligned}(F + \langle p\rangle_* F^\vee)((x) - (x')) &= (E^{(p)}, P^{(p)}) + p(E^{(p^{-1})}, pP^{(p^{-1})}) \\ &\quad - ((E')^{(p)}, (P')^{(p)}) + p((E')^{(p^{-1})}, p(P')^{(p^{-1})}).\end{aligned}$$

Computing $(T_p)_*$ on $J_p = \mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}} \times_{\mathbf{Z}[\frac{1}{N}]} \mathbf{F}_p$ is more subtle because $(T_p)_*$ was defined over $\mathbf{Z}[\frac{1}{Np}]$ (or over $\mathbf{Q}$) as $(\pi_2)_*\pi_1^*$ and was *extended* over $\mathbf{Z}[\frac{1}{N}]$ by the Néronian property. That is, we do *not* have a direct definition of $(T_p)_*$ in characteristic p, so we will need to lift to characteristic 0 to compute. It is *here* that the ordinariness assumption is crucial, for we shall see that, in some sense,

$$(T_p)_*((x) - (x')) = (F + \langle p\rangle_* F^\vee)((x) - (x'))$$

as *divisors* for ordinary points x, x'. This is, of course, much stronger than the mere linear equivalence that we need to prove.

Before we dive into the somewhat subtle calculation of $(T_p)_*((x) - (x'))$, let's quickly take care of the relation $w_\zeta^{-1} F w_\zeta = \langle p\rangle_*^{-1} F$, or equivalently,

$$F w_\zeta = w_\zeta \langle p^{-1}\rangle_* F.$$

All maps here are induced by maps on $X_1(N)_{/\overline{\mathbf{F}}_p}$, with $F = \mathrm{Alb}(F_{X_1(N)})$, $w_\zeta = \mathrm{Alb}(w_{\zeta|_{X_1(N)}})$, $\langle p^{-1}\rangle_* = \mathrm{Alb}(I_{p^{-1}})$. Thus, it suffices to show

$$F_{X_1(N)} \circ w_\zeta = w_\zeta I_{p^{-1}} F_{X_1(N)}$$

on $X_1(N)_{/\overline{\mathbf{F}}_p}$, and we can check by studying $x = (E, P) \in Y_1(N)(\overline{\mathbf{F}}_p)$:

$$F_{X_1(N)} w_\zeta(x) = F_{X_1(N)}(E/P, P') = (E^{(p)}/P^{(p)}, (P')^{(p)})$$

where $\langle P, P'\rangle_N = \zeta$, so $\langle P^{(p)}, (P')^{(p)}\rangle_N = \zeta^p$ by compatibility of the (relative) Weil pairing with respect to base change. Meanwhile,

$$w_\zeta I_{p^{-1}} F_{X_1(N)}(x) = w_\zeta(E^{(p)}, p^{-1}P^{(p)}) = (E^{(p)}/(p^{-1}P^{(p)}), Q)$$

where $\langle p^{-1}P^{(p)}, Q\rangle_N = \zeta$, or equivalently $\langle P^{(p)}, Q\rangle = \zeta^p$. Since $Q = (P')^{(p)}$ is such a point, this second relation is established.

Now we turn to the problem of computing

$$(T_p)_*((x) - (x'))$$

for "ordinary points" $x = (E, P)$, $x' = (E', P')$ as above. Let $R = \mathbf{Z}_p^{\mathrm{un}}$, $W(\overline{\mathbf{F}}_p)$, or more generally any henselian (e.g., complete) discrete valuation ring with residue field $\overline{\mathbf{F}}_p$ and fraction field K of characteristic 0. Since $p \nmid N$, R is a $\mathbf{Z}[\frac{1}{N}]$-algebra. Since $Y_1(N)$ is *smooth* over $\mathbf{Z}[\frac{1}{N}]$, we conclude from the (strict) henselian property that $Y_1(N)(R) \to Y_1(N)(\overline{\mathbf{F}}_p)$ is surjective. Of course, this can be seen "by hand": if (E, P) is given over $\overline{\mathbf{F}}_p$, choose a Weierstrass model $\mathcal{E} \hookrightarrow \mathbf{P}^2_R$ lifting E (this is canonically an elliptic curve, by [**62**, Ch 2]). The finite *étale* group scheme $\mathcal{E}[N]$ is *constant* since R is strictly henselian. Thus there exists a unique closed immersion of group schemes $\mathbf{Z}/N\mathbf{Z} \hookrightarrow \mathcal{E}[N]$ lifting $P : \mathbf{Z}/N\mathbf{Z} \hookrightarrow E[N]$.

Let $(\mathcal{E}, \mathcal{P})$, $(\mathcal{E}', \mathcal{P}')$ over R lift x, x' respectively. We view these sections to $X_1(N)_{/R} \to \operatorname{Spec} R$ as relative effective Cartier divisors of degree 1. Using the reduction map

$$\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}(R) \to J_p(\overline{\mathbf{F}}_p)$$

and the *definition* of $(T_p)_*$, we see that $(T_p)_*((x)-(x'))$ is the image of $(T_p)_*((\mathcal{E},\mathcal{P})-(\mathcal{E}',\mathcal{P}'))$. Now R is *NOT* a $\mathbf{Z}[\frac{1}{Np}]$-algebra but K *is*, and we have an injection (even bijection)

$$\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}(R) \hookrightarrow \mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}(K),$$

as $\mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}} \to \operatorname{Spec}\mathbf{Z}[\frac{1}{N}]$ is separated (even proper).

Thus, we will first compute $(T_p)_*((x)-(x'))$ by working with $\overline{K}$-points, where $\overline{K}$ is an algebraic closure of K. Since $p \nmid N$, we have

$$(\pi_2)_*\pi_1^*((\mathcal{E},\mathcal{P})_{/\overline{K}}) = \sum_C (\mathcal{E}_{\overline{K}}/C, \mathcal{P}_{\overline{K}} \mod C)$$

where C runs through all $p+1$ order-p subgroups of $\mathcal{E}_{/\overline{K}}$. Since $\mathcal{E} \to \operatorname{Spec} R$ has *ordinary* reduction, and R is strictly henselian, the connected-étale sequence of $\mathcal{E}[p]$ is the short exact sequence of finite flat R-group schemes

$$0 \to \mu_p \to \mathcal{E}[p] \to \underline{\mathbf{Z}/p\mathbf{Z}} \to 0.$$

Enlarging R to a finite extension does not change the residue field $\overline{\mathbf{F}}_p$, so we may assume that

$$\mathcal{E}[p]_{/K} \cong \underline{\mathbf{Z}/p\mathbf{Z}} \times \underline{\mathbf{Z}/p\mathbf{Z}}.$$

Taking the scheme-theoretic closure in $\mathcal{E}[p]$ of the $p+1$ distinct subgroups of $\mathcal{E}[p]_{/K}$ gives $p+1$ *distinct* finite flat subgroup schemes $\mathcal{C} \subset \mathcal{E}$ realizing the $p+1$ distinct C's over $\overline{K}$.

Exactly one of these $\mathcal{C}$'s is killed by $\mathcal{E}[p] \to \underline{\mathbf{Z}/p\mathbf{Z}}$ over R, as this can be checked on the generic fiber, so it must be $\mu_p \hookrightarrow \mathcal{E}[p]$. For the remaining $\mathcal{C}$'s, the map $\mathcal{C} \to \underline{\mathbf{Z}/p\mathbf{Z}}$ is an isomorphism on the generic fiber. We claim these maps

$$\mathcal{C} \to \underline{\mathbf{Z}/p\mathbf{Z}}$$

over R are isomorphisms. Indeed, if $\mathcal{C}$ is *étale* this is clear, yet $\mathcal{C} \hookrightarrow \mathcal{E}[p]$ is a finite flat closed subgroup-scheme of order p, so a consideration of the closed fiber shows that if $\mathcal{C}$ is *not* étale then it is multiplicative. But $\mathcal{E}[p]$ has a *unique* multiplicative subgroup-scheme since

$$\mathcal{E}[p]^\vee \cong \mathcal{E}[p]$$

by Cartier-Nishi duality and $\mathcal{E}[p]$ has a *unique* order-p *étale* quotient (as any such quotient must kill the μ_p we have inside $\mathcal{E}[p]$.)

Thus,

$$\begin{aligned}(\pi_2)_*\pi_1^*((\mathcal{E},\mathcal{P})_{/\overline{K}}) &= \sum_{\mathcal{C}} (\mathcal{E}/\mathcal{C}, \mathcal{P} \mod \mathcal{C}) - \sum_{\mathcal{C}'} (\mathcal{E}'/\mathcal{C}', \mathcal{P}' \mod \mathcal{C}') \\ &\in \mathrm{Pic}^0_{X_1(N)_{/\mathbf{Z}[\frac{1}{N}]}}(R)\end{aligned}$$

coincides with $(T_p)_*((\mathcal{E},\mathcal{P})-(\mathcal{E}',\mathcal{P}'))$ as both induce the same $\overline{K}$-point. Passing to closed fibers,

$$\begin{aligned}(T_p)_*((x)-(x')) = {} & (E/\mu_p, P \mod \mu_p) + p(E/\underline{\mathbf{Z}/p\mathbf{Z}'}, P \mod \mathbf{Z}/p\mathbf{Z}) \\ & - (E'/\mu_p, P' \mod \mu_p) + p(E'/\underline{\mathbf{Z}/p\mathbf{Z}}, P' \mod \mathbf{Z}/p\mathbf{Z})\end{aligned}$$

where $E[p] \cong \mu_p \times \underline{\mathbf{Z}/p\mathbf{Z}}$ and $E'[p] \cong \mu_p \times \underline{\mathbf{Z}/p\mathbf{Z}}$ are the *canonical* splittings of the connected-étale sequence over the perfect field $\overline{\mathbf{F}}_p$.

Now consider the relative Frobenius morphism

$$F_{E/\overline{\mathbf{F}}_p} : E \to E^{(p)},$$

which sends O to O (and P to $P^{(p)}$) and so is a map of *elliptic curves* over $\overline{\mathbf{F}}_p$. Recall that in characteristic p, for any map of schemes $X \to S$ we define the relative Frobenius map $F_{X/S} : X \to X^{(p)}$ to be the unique S-map fitting into the diagram

$$\begin{tikzcd} X \arrow[r, "F_{X/S}"'] \arrow[rr, bend left, "F_X"] \arrow[dr] & X^{(p)} \arrow[r] \arrow[d] & X \arrow[d] \\ & S \arrow[r, "F_S"'] & S \end{tikzcd}$$

where F_S, F_X are the absolute Frobenius maps. Since $E \to \operatorname{Spec} \overline{\mathbf{F}}_p$ is smooth of pure relative dimension 1, $F_{E/\overline{\mathbf{F}}_p}$ is finite flat of degree $p^1 = p$. It is bijective on points, so $\ker(F_{E/\overline{\mathbf{F}}_p})$ must be connected of order p.

The *only* such subgroup-scheme of E is $\mu_p \hookrightarrow E[p]$ by the *ordinariness*. Thus

$$E/\mu_p \cong E^{(p)}$$

is easily seen to take P mod μ_p to $P^{(p)}$.

Similarly, we have

$$\begin{tikzcd} E \arrow[r, "F_{E/\overline{\mathbf{F}}_p}"'] \arrow[rr, bend left, "p"] & E^{(p)} \arrow[r, "F^\vee_{E/\mathbf{F}_p}"'] & E \end{tikzcd}$$

so $F^\vee_{E/\mathbf{F}_p}$ is étale of degree p and base extension by $\mathrm{Frob}^{-1} : \overline{\mathbf{F}}_p \to \overline{\mathbf{F}}_p$ gives

$$\begin{tikzcd} E^{(p^{-1})} \arrow[r] \arrow[rr, bend left, "p"] & E \arrow[r] & E^{(p^{-1})} \\ P^{(p^{-1})} \arrow[r, mapsto] & P \arrow[r, mapsto] & p \cdot P^{(p^{-1})}. \end{tikzcd}$$

As the second map in this composite is étale of degree p, we conclude

$$(E_{/\underline{\mathbf{Z}/p\mathbf{Z}}}, P \bmod \mathbf{Z}/p\mathbf{Z}) \cong (E^{(p^{-1})}, pP^{(p^{-1})}).$$

Thus, in $\mathrm{Pic}^0_{X_1(N)}(\overline{\mathbf{F}}_p)$,

$$\begin{aligned} (T_p)_*((x) - (x')) = {} & (E^{(p)}, P^{(p)}) + p \cdot (E^{(p^{-1})}, p \cdot P^{(p^{-1})}) \\ & - ((E')^{(p)}, (P')^{(p)}) - p \cdot ((E')^{(p^{-1})}, p \cdot (P')^{(p^{-1})}) \end{aligned}$$

which we have seen is equal to $(F + \langle p \rangle_* F^\vee)((x) - (x'))$.

□

CHAPTER 6

Appendix by Kevin Buzzard: A mod ℓ multiplicity one result

In this appendix, we explain how the ideas of [**46**] can be used to prove the following mild strengthening of the multiplicity one results in §9 of [**32**].

The setup is as follows. Let f be a normalised cuspidal eigenform of level N, and weight k, defined over $\overline{\mathbf{F}}_\ell$, with $\ell \nmid N$ and $2 \leq k \leq \ell + 1$. Let N^* denote N if $k = 2$, and $N\ell$ if $k > 2$. Let $J_{\mathbf{Q}}$ be the Jacobian of the curve $X_1(N^*)_{\mathbf{Q}}$, and let H denote the Hecke algebra in $\mathrm{End}(J_{\mathbf{Q}})$ generated over $\mathbf{Z}$ by T_p for all primes p, and all the Diamond operators of level N^*. It is well-known (for example by Proposition 9.3 of [**46**]) that there is a characteristic 0 normalised eigenform F in $S_2(\Gamma_1(N^*))$ lifting f. Let $\mathfrak{m}$ denote the maximal ideal of H associated to F (note that $\mathfrak{m}$ depends only on f and not on the choice of F), and let $\mathbf{F} = H/\mathfrak{m}$, which embeds naturally into $\overline{\mathbf{F}}_\ell$. Suppose the representation $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ associated to f is absolutely irreducible, and furthermore assume that if $k = \ell + 1$ then ρ_f is not isomorphic to a representation coming from a form of weight 2 and level N.

Theorem 6.1. *If ρ_f is ramified at ℓ, or if ρ_f is unramified at ℓ but $\rho_f(\mathrm{Frob}_\ell)$ is not a scalar matrix, then $J_{\mathbf{Q}}(\overline{\mathbf{Q}})[\mathfrak{m}]$ has $H/\mathfrak{m}$-dimension two, and hence is a model for (precisely one copy of) ρ_f.*

The motivation for putting ourselves in the setup above is that every absolutely irreducible modular mod ℓ representation has a twist coming from a modular form of level prime to ℓ and weight at most $\ell + 1$. In particular, every modular mod ℓ representation has a twist coming from a form satisfying the conditions of our setup. Furthermore, if f is as in our setup, then Theorems 2.5 and 2.6 of [**32**] tell us the precise structure of the restriction of ρ_f to D_ℓ, a decomposition group at ℓ. These results are explained in Section 2.2. Using them, it is easy to deduce

Corollary 6.2. *Let ρ be an absolutely irreducible modular mod ℓ representation, such that $\rho_f(D_\ell)$ is not contained within the scalars. Then some twist of ρ comes from a modular form satisfying the conditions of the theorem, and hence ρ is a multiplicity one representation in the sense of Remark 3.4.2.*

The theorem, commonly referred to as a "multiplicity one theorem", is a mild extension of results of Mazur, Ribet, Gross and Edixhoven. It was announced for $\ell = 2$ as Proposition 2.4 of [**9**] but the proof given there is not quite complete—in fact, the last line of the proof there is a little optimistic. I would hence like to thank Ribet and Stein for the opportunity to correct this oversight in [**9**].

Proof of Theorem. Firstly we observe that the only case not dealt with by Theorem 9.2 of [**32**] is the case when $k = \ell$ and ρ_f is unramified at ℓ, with $\rho_f(\mathrm{Frob}_\ell)$ a non-scalar matrix whose eigenvalues are equal. Moreover, using Theorems 2.5 and 2.6 of [**32**] we see that in this case f must be ordinary at ℓ. We are hence in a position to use the detailed construction of ρ_f given in §§11–12 of [**46**]. We will follow the conventions set up in the present paper for normalisations of Hecke operators, and so in particular the formulae below differ from the ones in [**46**] by a twist.

The maximal ideal $\mathfrak{m}$ of H associated to f gives rise as in (12.5) of [**46**] to an idempotent $e \in H_\ell := H \otimes_{\mathbf{Z}} \mathbf{Z}_l$, such that the completion $H_\mathfrak{m}$ of H at $\mathfrak{m}$ is just eH_ℓ. Let G denote $e(J_{\mathbf{Q}_\ell}[\ell^\infty])$, the part of the ℓ-divisible group of J which is associated to $\mathfrak{m}$. Then $H_\mathfrak{m}$ acts on G, and it is proved in Propositions 12.8 and 12.9 of [**46**] that there is an exact sequence of ℓ-divisible groups

$$0 \to G^0 \to G \to G^e \to 0$$

over $\mathbf{Q}_\ell$, which is $H_\mathfrak{m}$-stable. Let

$$0 \to T^0 \to T \to T^e \to 0$$

be the exact sequence of Tate modules of these groups. We now explain explicitly, following [**46**], how the group D_ℓ acts on these Tate modules.

If $k > 2$ then there is a Hecke operator U_ℓ in $H_\mathfrak{m}$, and we define $u = U_\ell$. If $k = 2$ then there is a Hecke operator T_ℓ in $H_\mathfrak{m}$ and because we are in the ordinary case we know that T_ℓ is a unit in $H_\mathfrak{m}$. We define u to be the unique root of the polynomial $X^2 - T_\ell X + \ell\langle\ell\rangle$ in $H_\mathfrak{m}$ which is a unit (u exists by an appropriate analogue of Hensel's lemma).

The calculations of Propositions 12.8 and 12.9 of [**46**] show that, under our conventions, the absolute Galois group D_ℓ of $\mathbf{Q}_\ell$ acts on T^e as $\lambda(u)$, where $\lambda(x)$ denotes the unramified character taking Frob_ℓ to x. Moreover, these propositions also tell us that D_ℓ acts on T^0 via the character $\chi_\ell\lambda(u^{-1}\langle\ell\rangle_N)\chi^{\ell-2}$, where χ_ℓ is the cyclotomic character and χ is the Teichmüller character. The key point is that this character takes values in $H^\times$.

The next key observation is that a standard argument on differentials, again contained in the proof of Propositions 12.8 and 12.9 of [**46**], shows that $G^e[\mathfrak{m}] = \mathfrak{m}^{-1}\ell T^e/\ell T^e$ has $H_\mathfrak{m}/\mathfrak{m}$-dimension 1 and that $G^0[\mathfrak{m}]$ has dimension $d^0 \geq 1$. (Note that the fact that $G^e[\mathfrak{m}]$ has dimension 1 implies, via some simple linear algebra, that the sequence $0 \to G^0[\mathfrak{m}] \to G[\mathfrak{m}] \to G^e[\mathfrak{m}] \to 0$ is exact, as asserted by Gross.) Furthermore, because we can identify $G^0[\mathfrak{m}]$ with $\mathfrak{m}^{-1}\ell T^0/\ell T^0$, we see that the action of D_ℓ on $G^0[\mathfrak{m}]$ is via a character which takes values in $(H/\mathfrak{m})^\times$. In particular, D_ℓ acts as scalars on $G^0[\mathfrak{m}]$.

Let us now assume that ρ_f is unramified at ℓ, and that $\rho_f(\mathrm{Frob}_\ell)$ is a non-diagonalisable matrix with eigenvalue $\alpha \in H/\mathfrak{m}$. Choose a model ρ for ρ_f defined over $\mathrm{GL}_2(H/\mathfrak{m})$. By the theorem of Boston, Lenstra and Ribet, we know that $G[\mathfrak{m}]$ is isomorphic to a direct sum of d copies of ρ, or more precisely, d copies of the restriction of ρ to D_ℓ. Here d is an integer satisfying $2d = d^0 + d^e$. Hence, if $G[\mathfrak{m}]^\alpha$ denotes the subspace of $G[\mathfrak{m}]$ where Frob_ℓ acts as α, then the $H/\mathfrak{m}$-dimension of $G[\mathfrak{m}]^\alpha$ is at most d. On the other hand, Frob_ℓ acts on $G[\mathfrak{m}]^0$ as a scalar, and hence this scalar must be α, and so we see $G[\mathfrak{m}]^0 \subseteq G[\mathfrak{m}]^\alpha$. Hence $d^0 \leq d = (d^0+1)/2$. We deduce that $d^0 \leq 1$ and hence $d^0 = d = 1$ and the theorem is proved. □

We remark that L. Kilford has found examples of mod 2 forms f of weight 2, such that ρ_f is unramified at 2 and $\rho_f(\mathrm{Frob}_2)$ is the identity, and where $J_{\mathbf{Q}}(\overline{\mathbf{Q}})[\mathfrak{m}]$ has $H/\mathfrak{m}$-dimension 4, and so one cannot hope to extend the theorem to this case. See Remark 3.6 for more details, or [**64**]. A detailed analysis of what is happening in this case, at least in the analogous setting of forms of weight 2 on $J_0(p)$, with p prime, has been undertaken by Emerton in [**39**]. In particular, Emerton proves that multiplicity one fails if and only if the analogue of the exact sequence $0 \to T^0 \to T \to T^e \to 0$ fails to split as a sequence of $H_{\mathfrak{m}}$-modules.

BIBLIOGRAPHY

1. A. Ash and G. Stevens, *Cohomology of arithmetic groups and congruences between systems of Hecke eigenvalues*, J. Reine Angew. Math. **365** (1986), 192–220.
2. A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.
3. B. J. Birch, *Cyclotomic fields and Kummer extensions*, Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965), Thompson, Washington, D.C., 1967, pp. 85–93.
4. B. J. Birch and W. Kuyk (eds.), *Modular functions of one variable. IV*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 476.
5. S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.
6. N. Boston, H.W. Lenstra, Jr., and K.A. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sci. Paris Sér. I Math. **312** (1991), no. 4, 323–328.
7. C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*, or Wild* 3*-adic exercises*, http://www.math.harvard.edu/HTML/Individuals/Richard_Taylor.html
8. S. Brueggeman, *The non-existence of certain Galois extensions unramified outside 5*, Journal of Number Theory **75** (1999), 47–52.
9. K. Buzzard, *On level-lowering for mod 2 representations*, to appear in Mathematics Research Letters.
10. K. Buzzard, M. Dickinson, N. Shepherd-Barron, and R. Taylor, *On icosahedral Artin representations*, in preparation.
11. H. Carayol, *Sur les représentations* ℓ*-adiques associées aux formes modulaires de Hilbert*, Ann. scient. Éc. Norm. Sup., 4^{eb} série **19** (1986), 409–468.
12. ______, *Sur les représentations galoisiennes modulo* ℓ *attachées aux formes modulaires*, Duke Math. J. **59** (1989), 785–801.
13. W. Casselman, *On representations of* GL_2 *and the arithmetic of modular curves*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 107–141. Lecture Notes in Math., Vol. 349.
14. I. V. Čerednik, *Uniformization of algebraic curves by discrete arithmetic subgroups of* $\mathrm{PGL}_2(k_w)$ *with compact quotient spaces*, Mat. Sb. (N.S.) **100(142)**

(1976), no. 1, 59–88, 165.

15. R. F. Coleman, *Serre's conjecture: The Jugentraum of the 20th century*, Mat. Contemp. **6** (1994), 13–18, XII School of Algebra, Part I (Portuguese) (Diamantina, 1992).
16. R. F. Coleman and B. Edixhoven, *On the semi-simplicity of the U_p-operator on modular forms*, Math. Ann. **310** (1998), no. 1, 119–127.
17. R. F. Coleman and J. F. Voloch, *Companion forms and Kodaira-Spencer theory*, Invent. Math. **110** (1992), no. 2, 263–281.
18. B. Conrad, *Modular forms, cohomology, and the Ramanujan conjecture*, in preparation.
19. B. Conrad, F. Diamond, and R. Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), no. 2, 521–567.
20. J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.
21. C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1962, Pure and Applied Mathematics, Vol. XI.
22. H. Darmon, *Serre's conjectures*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993–1994), Amer. Math. Soc., Providence, RI, 1995, pp. 135–153.
23. H. Darmon, F. Diamond, and R. Taylor, *Fermat's last theorem*, Current developments in mathematics, 1995 (Cambridge, MA), Internat. Press, Cambridge, MA, 1994, pp. 1–154.
24. P. Deligne, *Formes modulaires et représentations ℓ-adiques.*, Sém. Bourbaki no. 355, 1968/69 (Berlin and New York), Springer-Verlag, 1971, Lecture Notes in Mathematics, Vol. 179, pp. 139–172.
25. P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349.
26. F. Diamond, *The refined conjecture of Serre*, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993) (Cambridge, MA), Internat. Press, 1995, pp. 22–37.
27. F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem (Providence, RI), Amer. Math. Soc., 1995, pp. 39–133.
28. M. Dickinson, *On the modularity of certain* 2*-adic galois representations*, Harvard Ph.D. thesis (2000).
29. D. Doud, *S_4 and $\tilde{S}_4$ extensions of $\mathbf{Q}$ ramified at only one prime*, J. Number Theory **75** (1999), no. 2, 185–197.
30. V. G. Drinfeld, *Coverings of p-adic symmetric domains*, Funkcional. Anal. i Prilov zen. **10** (1976), no. 2, 29–40.
31. B. Edixhoven, *L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein"*, Astérisque (1991), no. 196–197, 7–8, 159–170 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).
32. ———, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594.

33. B. Edixhoven, *Le rôle de la conjecture de Serre dans la démonstration du théorème de Fermat*, Gaz. Math. (1995), no. 66, 25–41.
34. ______, *Erratum and addendum: "The role of Serre's conjecture in the proof of Fermat's theorem"*, Gaz. Math. (1996), no. 67, 19.
35. ______, *Serre's conjecture*, Modular forms and Fermat's last theorem (Boston, MA, 1995) (New York), Springer, 1997, pp. 209–242.
36. M. Eichler, *Quadratische Formen und Modulfunktionen*, Acta Arith. **4** (1958), 217–239.
37. D. Eisenbud, *Commutative algebra with a view toward algebraic geometry*, Springer-Verlag, New York, 1995.
38. D. Eisenbud and J. Harris, *Schemes, The language of modern algebraic geometry*, Springer-Verlag, Berlin, Graduate Texts in Mathematics, Vol. 197.
39. M. Emerton, *Supersingular elliptic curves, theta series and weight two modular forms*, preprint.
40. G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366.
41. G. Faltings and B. W. Jordan, *Crystalline cohomology and* $\mathrm{GL}(2,\mathbf{Q})$, Israel J. Math. **90** (1995), no. 1-3, 1–66.
42. G. Frey, *Links between stable elliptic curves and certain Diophantine equations*, Ann. Univ. Sarav. Ser. Math. **1** (1986), no. 1, iv+40.
43. ______, *Links between solutions of* $A-B=C$ *and elliptic curves*, Number theory (Ulm, 1987), Springer, New York, 1989, pp. 31–62.
44. A. Fröhlich, *Local fields*, Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965), Thompson, Washington, D.C., 1967, pp. 1–41.
45. K. Fujiwara, *Level optimization in the totally real case*, in preparation (1999).
46. B. H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. **61** (1990), no. 2, 445–517.
47. R. Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.
48. Y. Hellegouarch, *Invitation aux mathématiques de Fermat-Wiles*, Masson, Paris, 1997.
49. H. Hida, *Galois representations into* $\mathrm{GL}_2(\mathbf{Z}_p[[X]])$ *attached to ordinary cusp forms*, Invent. Math. **85** (1986), no. 3, 545–613.
50. ______, *Iwasawa modules attached to congruences of cusp forms*, Ann. Sci. École Norm. Sup. (4) **19** (1986), no. 2, 231–273.
51. H. Jacquet and R. P. Langlands, *Automorphic forms on* $\mathrm{GL}(2)$, Springer-Verlag, Berlin, 1970, Lecture Notes in Mathematics, Vol. 114.
52. F. Jarvis, *On Galois representations associated to Hilbert modular forms*, J. Reine Angew. Math. **491** (1997), 199–216.
53. ______, *Level lowering for modular mod* ℓ *representations over totally real fields*, Math. Ann. **313** (1999), no. 1, 141–160.
54. ______, *Mazur's principle for totally real fields of odd degree*, Compositio Math. **116** (1999), no. 1, 39–79.
55. N. Jochnowitz, *A study of the local components of the Hecke algebra mod* ℓ, Trans. Amer. Math. Soc. **270** (1982), no. 1, 253–267.
56. ______, *The index of the Hecke ring,* T_k*, in the ring of integers of* $T_k \otimes \mathbf{Q}$, Duke Math. J. **46** (1979), no. 4, 861–869.

57. B. W. Jordan and R. Livné, *Conjecture "epsilon" for weight $k > 2$*, Bull. Amer. Math. Soc. (N.S.) **21** (1989), no. 1, 51–56.

58. K. Joshi, *Remarks on methods of Fontaine and Faltings*, Internat. Math. Res. Notices **1999**, no. 22, 1199–1209.

59. N. M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 69–190. Lecture Notes in Mathematics, Vol. 350.

60. ______, *Higher congruences between modular forms*, Ann. of Math. (2) **101** (1975), 332–367.

61. ______, *A result on modular forms in characteristic p*, Modular functions of one variable, V (Proc. Second Internat. Conf., Univ. Bonn, Bonn, 1976) (Berlin), Springer, 1977, pp. 53–61. Lecture Notes in Math., Vol. 601.

62. N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

63. C. Khare, *Multiplicities of mod p Galois representations*, Manuscripta Math. **95** (1998), no. 2, 181–188.

64. L. J. P. Kilford, *Some examples of non-Gorenstein Hecke algebras associated to modular forms*, in preparation.

65. A. W. Knapp, *Elliptic curves*, Princeton University Press, Princeton, NJ, 1992.

66. S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and Walter Feit, Corrected reprint of the 1976 original.

67. R. P. Langlands, *Modular forms and ℓ-adic representations*, Proceedings of the International Summer School, University of Antwerp, RUCA, July 17–August 3, 1972 (Berlin) (P. Deligne and W. Kuyk, eds.), Springer, 1973, pp. 361–500. Lecture Notes in Math., Vol. 349.

68. ______, *Base change for* GL(2), Princeton University Press, Princeton, N.J., 1980.

69. W-C. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.

70. R. Livné, *On the conductors of mod ℓ Galois representations coming from modular forms*, J. Number Theory **31** (1989), no. 2, 133–141.

71. B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

72. B. Mazur and K. A. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque (1991), no. 196-197, 6, 215–255 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

73. L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations (Berlin), Springer, 1994, pp. 59–94.

74. J. S. Milne, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

75. T. Miyake, *Modular forms*, Springer-Verlag, Berlin, 1989, Translated from the Japanese by Yoshitaka Maeda.

76. H. Moon, *Finiteness results on certain mod p Galois representations*, to appear in J. Number Theory.

77. D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.
78. C. Queen, *The existence of p-adic Abelian L-functions*, Number theory and algebra (New York), Academic Press, 1977, pp. 263–288.
79. A. Raji, *On the levels of modular mod ℓ Galois representations of totally real fields*, Princeton University Ph.D. thesis, 1998.
80. R. Ramakrishna, *Lifting Galois representations*, Invent. Math. **138** (1999), no. 3, 537–562.
81. ______, *Deforming Galois representations and the conjectures of Serre and Fontaine-Mazur*, preprint, ftp://math.cornell.edu/pub/ravi (2000).
82. M. Raynaud, *Spécialisation du foncteur de Picard*, Inst. Hautes Études Sci. Publ. Math. No. **38** (1970), 27–76.
83. K. A. Ribet, *From the Taniyama-Shimura conjecture to Fermat's last theorem*, Ann. Fac. Sci. Toulouse Math. (5) **11** (1990), no. 1, 116–139.
84. ______, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), no. 2, 431–476.
85. ______, *Raising the levels of modular representations*, Séminaire de Théorie des Nombres, Paris 1987–88, Birkhäuser Boston, Boston, MA, 1990, pp. 259–271.
86. ______, *Lowering the levels of modular representations without multiplicity one*, International Mathematics Research Notices (1991), 15–19.
87. ______, *Report on mod ℓ representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Motives (Seattle, WA, 1991), Amer. Math. Soc., Providence, RI, 1994, pp. 639–676.
88. A. Robert, *Elliptic curves*, Springer-Verlag, Berlin, 1973, Notes from postgraduate lectures given in Lausanne 1971/72, Lecture Notes in Mathematics, Vol. 326.
89. T. Saito, *Modular forms and p-adic Hodge theory*, Invent. Math. **129** (1997), 607–620.
90. I. Schur, *Arithmetische Untersuchungen über endliche Gruppen linearer Substitutionen*, Sitz. Pr. Akad. Wiss. (1906), 164–184, Gesam. Abhl., **I**, 177–197, Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1973.
91. J-P. Serre, *Groupes de Lie l-adiques attachés aux courbes elliptiques*, Les Tendances Géom. en Algébre et Théorie des Nombres, Éditions du Centre National de la Recherche Scientifique, Paris, 1966, pp. 239–256 (= Collected Papers **70**).
92. ______, *Une interprétation des congruences relatives à la fonction τ de Ramanujan*, Séminaire Delange-Pisot-Poitou *n^o* **14** (1967–68) (= C.P. **80**).
93. ______, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331 (= C.P. **94**).
94. ______, *Congruences et formes modulaires [d'après H. P. F. Swinnerton-Dyer]*, Séminaire Bourbaki, 24e année (1971/1972), Exp. No. 416 (Berlin), Springer, 1973, pp. 319–338. Lecture Notes in Math., Vol. 317 (= C.P. **95**).
95. ______, *Formes modulaires et fonctions zêta p-adiques*, Proceedings of the International Summer School, University of Antwerp, RUCA, July 17–August 3, 1972 (Berlin), Springer, 1973, pp. 191–268. Lecture Notes in Math., Vol. 350 (= C.P. **97**).

96. ______, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

97. ______, *Valeurs propres des opérateurs de Hecke modulo ℓ*, Astérisque **24–25** (1975), 109–117 (= C.P. **104**).

98. ______, *Divisibilité de certaines fonctions arithmétiques*, Enseign. Math. (2) **22** (1976), no. 3-4, 227–260 (= C.P. **108**).

99. ______, *Linear representations of finite groups*, Springer-Verlag, New York, 1977, Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.

100. ______, *Local fields*, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

101. ______, *Lettre à J.-F. Mestre*, Current trends in arithmetical algebraic geometry (Arcata, Calif., 1985), Amer. Math. Soc., Providence, RI, 1987, pp. 263–268 (= C.P. **142**).

102. ______, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. **54** (1987), no. 1, 179–230 (= C.P. **143**).

103. ______, *Travaux de Wiles (et Taylor, ...)*, Partie I, Séminaire Bourbaki, **803** (1995) (= C.P. **168**).

104. J-P. Serre and J. T. Tate, *Good reduction of abelian varieties*, Ann. of Math. (2) **88** (1968), 492–517 (= C.P. **79**).

105. N. I. Shepherd-Barron and R. Taylor, *Mod 2 and mod 5 icosahedral representations*, J. Amer. Math. Soc. **10** (1997), no. 2, 283–298.

106. H. Shimizu, *On zeta functions of quaternion algebras*, Ann. of Math. (2) **81** (1965), 166–193.

107. G. Shimura, *A reciprocity law in non-solvable extensions*, J. Reine Angew. Math. **221** (1966), 209–220.

108. ______, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

109. J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

110. ______, *Advanced topics in the arithmetic of elliptic curves*, Springer-Verlag, New York, 1994.

111. C. M. Skinner and A. J. Wiles, *Ordinary representations and modular forms*, Proc. Nat. Acad. Sci. U.S.A. **94** (1997), no. 20, 10520–10527.

112. H. P. F. Swinnerton-Dyer, *On ℓ-adic representations and congruences for coefficients of modular forms*, Proceedings of the International Summer School, University of Antwerp, RUCA, July 17–August 3, 1972 (Berlin), Springer, 1973, pp. 1–55. Lecture Notes in Math., Vol. 350.

113. J. T. Tate, *The non-existence of certain Galois extensions of* $\mathbf{Q}$ *unramified outside* 2, Contemporary Math. **174** (1994), 153–156.

114. R. Taylor and A. J. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

115. J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. (N.S.) **5** (1981), no. 2, 173–175.

116. J.-L. Waldspurger, *Quelques propriétés arithmétiques de certaines formes automorphes sur* GL(2), Compositio Math. **54** (1985), no. 2, 121–171.

117. A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

INDEX

Italic page numbers are used to indicate pages with important information about the entry, while page numbers in normal type indicate a textual reference.

8 An introduction to computing modular forms using modular symbols

An introduction to computing modular forms using modular symbols

William A. Stein

Abstract

We explain how weight-two modular forms on $\Gamma_0(N)$ are related to modular symbols, and how to use this to explicitly compute spaces of modular forms.

Introduction

The definition of the spaces of modular forms as functions on the upper half plane satisfying a certain equation is very abstract. The definition of the Hecke operators even more so. Nevertheless, one wishes to carry out explicit investigations involving these objects.

We are fortunate that we now have methods available that allow us to transform the vector space of cusp forms of given weight and level into a concrete object, which can be explicitly computed. We have the work of Atkin-Lehner, Birch, Swinnerton-Dyer, Manin, Mazur, Merel, and many others to thank for this (see, e.g., [3, 6, 15, 16]). For example, we can use the Eichler-Selberg trace formula, as extended in [11], to compute characteristic polynomials of Hecke operators. Then the method described in [25] gives a basis for certain spaces of modular forms. Alternatively, we can compute Θ-series using Brandt matrices and quaternion algebras as in [12, 18], or we can use a closely related geometric method that involves the module of enhanced supersingular elliptic curves [17]. Another related method of Birch [2] is very fast, but gives only a piece of the full space of modular forms. The power of the modular symbols approach was demonstrated by Cremona in his book [6] in which he systematically computes a large table of invariants of all elliptic curves of conductor up to 1000 (his online tables [7] go well beyond $100,000$).

Though the above methods are each beautiful and well suited to certain applications, we will only discuss the modular symbols method further, as it has many advantages. We will primarily discuss the theory in this summary paper, leaving an explicit description of the objects involved for other papers. Nonetheless, there is a definite gap between the theory on the one hand, and an efficient running machine implementation on the other. To implement the algorithms hinted at below requires making absolutely everything completely explicit, then finding intelligent and efficient ways of performing the necessary manipulations. This is a nontrivial and tedious task, with room for error at every step. Fortunately, Sage [24] has extensive capabilities for computing with modular forms and includes Cremona's programs; we will give a few examples below. See also the author's Magma [4] package for computing with modular forms and modular symbols.

In this paper we will focus exclusively on the case of weight-2 modular forms for $\Gamma_0(N)$. The methods explained here extend to modular forms of integer weight greater than 2; for more details see the author's book [23] and Merel's paper [16].

Section 1 contains a brief summary of basic facts about modular forms, Hecke operators, and integral homology. Section 2 introduces modular symbols, and describes how to compute with them. Section 3 outlines an algorithm for constructing cusp forms using modular symbols in conjunction with Atkin-Lehner theory.

This paper assumes some familiarity with algebraic curves, Riemann surfaces, and homology groups of compact surfaces. A few basic facts about modular forms are recalled, but only briefly. In particular, only a roundabout attempt is made to motivate why one might be interested in modular forms; for this, see many of the references in the bibliography. No prior exposure to modular symbols is assumed.

1 Modular forms and Hecke operators

All of the objects we will consider arise from the modular group $\mathrm{SL}_2(\mathbf{Z})$ of two-by-two integer matrices with determinant equal to one. This group acts via linear fractional transformations on the complex upper half plane $\mathfrak{h}$, and also on the extended upper half plane

$$\mathfrak{h}^* = \mathfrak{h} \cup \mathbf{P}^1(\mathbf{Q}) = \mathfrak{h} \cup \mathbf{Q} \cup \{\infty\}.$$

See [21, §1.3–1.5] for a careful description of the topology on $\mathfrak{h}^*$. A basis of neighborhoods for $\alpha \in \mathbf{Q}$ is given by the sets $\{\alpha\} \cup D$, where D is a disc in $\mathfrak{h}$ that is tangent to the real line at α. Let N be a positive integer and consider the group $\Gamma_0(N)$ of matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ such that $N \mid c$. This group acts on $\mathfrak{h}^*$ by linear fractional transformations, and the quotient $\Gamma_0(N)\backslash\mathfrak{h}^*$ is a Riemann surface, which we denote by $X_0(N)$. Shimura showed in [21, §6.7] that $X_0(N)$ has a canonical structure of algebraic curve over $\mathbf{Q}$.

A *cusp form* is a function f on $\mathfrak{h}$ such that $f(z)dz$ is a holomorphic differential on $X_0(N)$. Equivalently, a cusp form is a holomorphic function f on $\mathfrak{h}$ such that

(a) the expression $f(z)dz$ is invariant under replacing z by $\gamma(z)$ for each $\gamma \in \Gamma_0(N)$, and

(b) $f(z)$ is holomorphic at each element of $\mathbf{P}^1(\mathbf{Q})$, and moreover $f(z)$ tends to 0 as z tends to any element of $\mathbf{P}^1(\mathbf{Q})$.

The space of cusp forms on $\Gamma_0(N)$ is a finite dimensional complex vector space, of dimension equal to the genus g of $X_0(N)$. Viewed topologically, as a 2-dimensional real manifold, $X_0(N)(\mathbf{C})$ is a g-holed torus.

Condition (b) in the definition of $f(z)$ means that $f(z)$ has a Fourier expansion about each element of $\mathbf{P}^1(\mathbf{Q})$. Thus, at ∞ we have

$$\begin{aligned} f(z) &= a_1 e^{2\pi i z} + a_2 e^{2\pi i 2z} + a_3 e^{2\pi i 3z} + \cdots \\ &= a_1 q + a_2 q^2 + a_3 q^3 + \cdots , \end{aligned}$$

where, for brevity, we write $q = q(z) = e^{2\pi i z}$.

Example 1.1. Let E be the elliptic curve defined by the equation $y^2 + xy = x^3 + x^2 - 4x - 5$. For $p \neq 3, 13$, let $a_p = p + 1 - \#\tilde{E}(\mathbf{F}_p)$, where $\tilde{E}$ is the reduction of E mod p, and let $a_3 = -11$, $a_{13} = 1$. For n composite, define a_n using the relations at the end of Section 3. Then

$$f = q + a_2q^2 + a_3q^3 + a_4q^4 + a_5q^5 + \cdots = q + q^2 - 11q^3 + 2q^5 + \cdots$$

is the q-expansion of a modular form on $\Gamma_0(39)$. The Shimura-Taniyama conjecture, which is now a theorem (see [5]) asserts that any q-expansion constructed as above from an elliptic curve over $\mathbf{Q}$ is a modular form. We define the above elliptic curve and compute the associated modular form f using Sage as follows:

```
sage: E = EllipticCurve([1,1,0,-4,-5]); E
Elliptic Curve defined by y^2 + x*y  = x^3 + x^2 - 4*x - 5
over Rational Field
sage: E.q_eigenform(10)
q + q^2 - q^3 - q^4 + 2*q^5 - q^6 - 4*q^7 - 3*q^8 + q^9 + O(q^10)
```

The Hecke operators are a family of *commuting* endomorphisms of $S_2(N)$, which are defined as follows. The complex points of the open subcurve $Y_0(N) = \Gamma_0(N)\backslash\mathfrak{h}$ are in bijection with pairs (E, C), where E is an elliptic curve over $\mathbf{C}$ and C is a cyclic subgroup of $E(\mathbf{C})$ of order N. If $p \nmid N$ then there are two natural maps π_1 and π_2 from $Y_0(pN)$ to $Y_0(N)$; the first, π_1, sends (E, C) to (E, C'), where C' is the unique cyclic subgroup of C of order N, and the second, π_2, sends a point $(E, C) \in Y_0(N)(\mathbf{C})$ to $(E/D, C/D)$, where D is the unique cyclic subgroup of C of order p. These maps extend in a unique way to maps from $X_0(pN)$ to $X_0(N)$:

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\pi_2"'] \arrow[dr, "\pi_1"] & \\ X_0(N) & & X_0(N). \end{tikzcd}$$

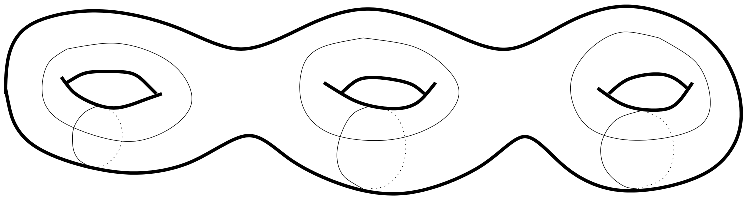
The pth *Hecke operator* T_p is $(\pi_1)_* \circ (\pi_2)^*$; it acts on most objects attached to $X_0(N)$, such as divisors and cusp forms. There is a Hecke operator T_n for every positive integer n, but we will not need to consider those with n composite.

Example 1.2. There is a basis of $S_2(39)$ so that

$$T_2 = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -2 & 1 \\ 0 & -1 & 1 \end{pmatrix} \text{ and } T_5 = \begin{pmatrix} 1 & -1 & -1 \\ -2 & 2 & -2 \\ -3 & -1 & -1 \end{pmatrix}.$$

Notice that these matrices commute, and that 1 is an eigenvalue of T_2, and 2 is an eigenvalue of T_5. We compute each of the above matrices and verify that they commute using Sage as follows:

```
sage: S = CuspForms(39)
sage: T2 = S.hecke_matrix(2); T2
```



$$H_1(X_0(39), \mathbf{Z}) \cong \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$$

Figure 1: The homology of $X_0(39)$.

```
[ 0  2 -1]
[ 1 -2  1]
[ 0 -1  1]
sage: T5 = S.hecke_matrix(5); T5
[ 1 -1 -1]
[-2  2 -2]
[-3 -1 -1]
sage: T2*T5 == T5*T2
True
```

The first homology group $H_1(X_0(N), \mathbf{Z})$ is the group of singular 1-cycles modulo homology relations. Recall that topologically $X_0(N)$ is a g-holed torus, where g is the genus of $X_0(N)$. The group $H_1(X_0(N), \mathbf{Z})$ is thus a free abelian group of rank $2g$ (see, e.g., [10, Ex. 19.30]), with two generators corresponding to each hole, as illustrated in the case $N = 39$ in Diagram 1.

The Hecke operators T_p act on $H_1(X_0(N), \mathbf{Z})$, and integration defines a nondegenerate Hecke-equivariant pairing

$$\langle\, , \rangle : S_2(N) \times H_1(X_0(N), \mathbf{Z}) \to \mathbf{C}.$$

Explicitly, for a path x,

$$\langle f, x\rangle = 2\pi i \int_x f(z)dz,$$

where the integral may be viewed as a complex line integral along an appropriate piece of the preimage of x in the upper half plane. The pairing is Hecke equivariant in the sense that for every prime p, we have $\langle fT_p, x\rangle = \langle f, T_p x\rangle$. As we will see, modular symbols allow us to make explicit the action of the Hecke operators on $H_1(X_0(N), \mathbf{Z})$; the above pairing then translates this into a wealth of information about cusp forms.

For a more detailed survey of the basic facts about modular curves and modular forms, we urge the reader to consult the book [9] by Diamond and Shurman along with Diamond and Im's survey paper [8]. For a discussion of how to draw a picture of the ring generated by the Hecke operators, see [19, §3.8].

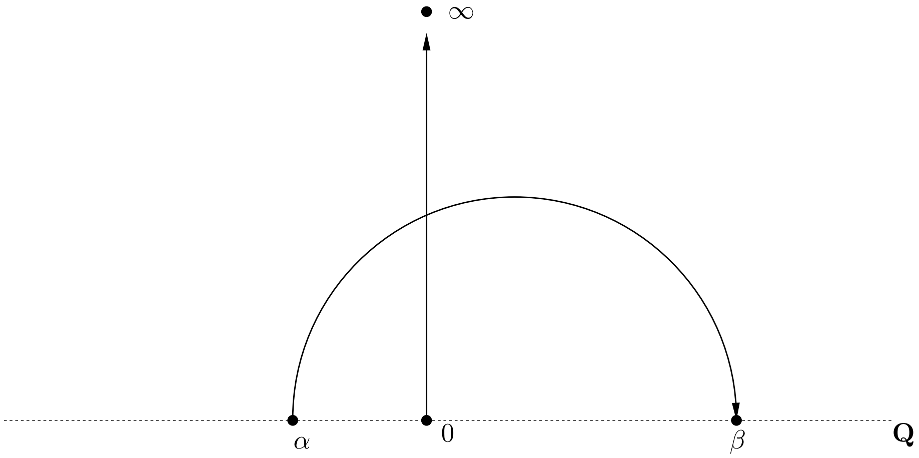


Figure 2: The modular symbols $\{\alpha, \beta\}$ and $\{0, \infty\}$.

2 Modular symbols

The modular symbols formalism provides a presentation of $H_1(X_0(N), \mathbf{Z})$ in terms of paths between elements of $\mathbf{P}^1(\mathbf{Q})$. Furthermore, a trick due to Manin gives an explicit finite list of generators and relations for the space of modular symbols.

The *modular symbol* defined by a pair $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$ is denoted $\{\alpha, \beta\}$. As illustrated in Figure 2, this modular symbol should be viewed as the homology class, relative to the cusps, of a geodesic path from α to β in $\mathfrak{h}^*$. The homology group relative to the cusps is a slight enlargement of the usual homology group, in that we allow paths with endpoints in $\mathbf{P}^1(\mathbf{Q})$ instead of restricting to closed loops.

Motivated by this picture, we declare that modular symbols satisfy the following homology relations: if $\alpha, \beta, \gamma \in \mathbf{Q} \cup \{\infty\}$, then

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0.$$

Furthermore, we quotient out by any torsion, so, e.g., $\{\alpha, \alpha\} = 0$ and $\{\alpha, \beta\} = -\{\beta, \alpha\}$.

Denote by $\boldsymbol{\mathcal{M}}_2$ the free abelian group with basis the set of symbols $\{\alpha, \beta\}$ modulo the three-term homology relations above and modulo any torsion. There is a left action of $\mathrm{GL}_2(\mathbf{Q})$ on $\boldsymbol{\mathcal{M}}_2$, whereby a matrix g acts by

$$g\{\alpha, \beta\} = \{g(\alpha), g(\beta)\},$$

and g acts on α and β by a linear fractional transformation. The space $\boldsymbol{\mathcal{M}}_2(N)$ of *modular symbols for* $\Gamma_0(N)$ is the quotient of $\boldsymbol{\mathcal{M}}_2$ by the submodule generated by the infinitely many elements of the form $x - g(x)$, for x in $\boldsymbol{\mathcal{M}}_2$ and g in $\Gamma_0(N)$, and modulo any torsion. A *modular symbol for* $\Gamma_0(N)$ is an element of this space. We frequently denote the equivalence class that defines a modular symbol by giving a representative element.

In [14], Manin proved that there is a natural injection $H_1(X_0(N),\mathbf{Z}) \to \boldsymbol{\mathcal{M}}_2(N)$. The image of $H_1(X_0(N),\mathbf{Z})$ in $\boldsymbol{\mathcal{M}}_2(N)$ can be identified as follows. Let $\boldsymbol{\mathcal{B}}_2(N)$ denote the free abelian group whose basis is the finite set $\Gamma_0(N)\backslash\mathbf{P}^1(\mathbf{Q})$. The *boundary map* $\delta : \boldsymbol{\mathcal{M}}_2(N) \to \boldsymbol{\mathcal{B}}_2(N)$ sends $\{\alpha,\beta\}$ to $[\beta]-[\alpha]$, where $[\beta]$ denotes the basis element of $\boldsymbol{\mathcal{B}}_2(N)$ corresponding to $\beta \in \mathbf{P}^1(\mathbf{Q})$. The kernel $\boldsymbol{\mathcal{S}}_2(N)$ of δ is the subspace of *cuspidal* modular symbols. An element of $\boldsymbol{\mathcal{S}}_2(N)$ can be thought of as a linear combination of paths in $\mathfrak{h}^*$ whose endpoints are cusps, and whose images in $X_0(N)$ are a linear combination of loops. We thus obtain a map $\varphi : \boldsymbol{\mathcal{S}}_2(N) \to H_1(X_0(N),\mathbf{Z})$.

Theorem 2.1. *The map φ given above defines a canonical isomorphism*

$$\boldsymbol{\mathcal{S}}_2(N) \cong H_1(X_0(N),\mathbf{Z}).$$

2.1 Manin's trick

In this section, we describe a trick of Manin that shows that the space of modular symbols can be computed.

By reducing modulo N, one sees that the group $\Gamma_0(N)$ has finite index in $\mathrm{SL}_2(\mathbf{Z})$. Let $r_0, r_1, \ldots, r_m$ be distinct right coset representatives for $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbf{Z})$, so that

$$\mathrm{SL}_2(\mathbf{Z}) = \Gamma_0(N)r_o \cup \Gamma_0(N)r_1 \cup \cdots \cup \Gamma_0(N)r_m,$$

where the union is disjoint. For example, when N is prime, a list of coset representatives is

$$\begin{pmatrix}1&0\\0&1\end{pmatrix}, \begin{pmatrix}1&0\\1&1\end{pmatrix}, \begin{pmatrix}1&0\\2&1\end{pmatrix}, \begin{pmatrix}1&0\\3&1\end{pmatrix}, \ldots, \begin{pmatrix}1&0\\N-1&1\end{pmatrix}, \begin{pmatrix}0&-1\\1&0\end{pmatrix}.$$

In general, the right cosets of $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbf{Z})$ are in bijection with the elements of $\mathbf{P}^1(\mathbf{Z}/N\mathbf{Z})$ (see [6, §2.2] for complete details).

The following trick of Manin (see [14, §1.5] and [6, §2.1.6]) allows us to write every modular symbol as a $\mathbf{Z}$-linear combination of symbols of the form $r_i\{0,\infty\}$. In particular, the finitely many symbols $r_i\{0,\infty\}$ generate $\boldsymbol{\mathcal{M}}_2(N)$.

Because of the relation $\{\alpha,\beta\} = \{0,\beta\} - \{0,\alpha\}$, it suffices to consider modular symbols of the form $\{0, b/a\}$, where the rational number b/a is in lowest terms. Expand b/a as a continued fraction and consider the successive convergents in lowest terms:

$$\frac{b_{-2}}{a_{-2}} = \frac{0}{1}, \quad \frac{b_{-1}}{a_{-1}} = \frac{1}{0}, \quad \frac{b_0}{a_0} = \frac{b_0}{1}, \ldots, \quad \frac{b_{n-1}}{a_{n-1}}, \quad \frac{b_n}{a_n} = \frac{b}{a}$$

where the first two are added formally. Then

$$b_k a_{k-1} - b_{k-1}a_k = (-1)^{k-1},$$

so that

$$g_k = \begin{pmatrix} b_k & (-1)^{k-1}b_{k-1} \\ a_k & (-1)^{k-1}a_{k-1}\end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}).$$

Hence

$$\left\{\frac{b_{k-1}}{a_{k-1}}, \frac{b_k}{a_k}\right\} = g_k\{0,\infty\} = r_i\{0,\infty\},$$

for some i, is of the required special form.

Example 2.2. Let $N = 11$, and consider the modular symbol $\{0, 4/7\}$. We have

$$\frac{4}{7} = 0 + \frac{1}{1+\frac{1}{1+\frac{1}{3}}},$$

so the partial convergents are

$$\frac{b_{-2}}{a_{-2}} = \frac{0}{1}, \quad \frac{b_{-1}}{a_{-1}} = \frac{1}{0}, \quad \frac{b_0}{a_0} = \frac{0}{1}, \quad \frac{b_1}{a_1} = \frac{1}{1}, \quad \frac{b_2}{a_2} = \frac{1}{2}, \quad \frac{b_3}{a_3} = \frac{4}{7}.$$

Thus

$$\begin{aligned}\{0,4/7\} &= \{0,\infty\} + \{\infty,0\} + \{0,1\} + \{1,1/2\} + \{1/2,4/7\} \\ &= \begin{pmatrix}1 & -1\\ 2 & -1\end{pmatrix}\{0,\infty\} + \begin{pmatrix}4 & 1\\ 7 & 2\end{pmatrix}\{0,\infty\} \\ &= 2\cdot\left[\begin{pmatrix}1 & 4\\ 1 & 5\end{pmatrix}\{0,\infty\}\right]\end{aligned}$$

2.2 Manin symbols

As above, fix coset representatives $r_0, \ldots, r_m$ for $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbf{Z})$. Denote the modular symbol $r_i\{0,\infty\}$ by $[r_i]$. The symbols $[r_0], \ldots, [r_m]$ are called *Manin symbols*, and they are equipped with a right action of $\mathrm{SL}_2(\mathbf{Z})$, which is given by $[r_i]g = [r_j]$, where $\Gamma_0(N)r_j = \Gamma_0(N)r_ig$. Recall that $\mathrm{SL}_2(\mathbf{Z})$ is generated by the two matrices $\sigma = \left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right)$ and $\tau = \left(\begin{smallmatrix}1 & -1\\ 1 & 0\end{smallmatrix}\right)$ (see Theorem 2 of [20, VII.1.2]).

Theorem 2.3 (Manin)**.** *The Manin symbols* $[r_0], \ldots, [r_m]$ *satisfy the following relations:*

$$[r_i] + [r_i]\sigma = 0$$
$$[r_i] + [r_i]\tau + [r_i]\tau^2 = 0.$$

Furthermore, these relations generate all relations (modulo torsion relations).

This theorem, which is proved in [14, §1.7], provides a finite presentation for the space of modular symbols.

2.3 Hecke operators on modular symbols

When p is a prime not dividing N, define

$$T_p\{\alpha,\beta\} = \begin{pmatrix}p & 0\\ 0 & 1\end{pmatrix}\{\alpha,\beta\} + \sum_{r \bmod p}\begin{pmatrix}1 & r\\ 0 & p\end{pmatrix}\{\alpha,\beta\}.$$

As mentioned before, this definition is compatible with the integration pairing $\langle\, ,\, \rangle$ of Section 1, in the sense that $\langle fT_p, x\rangle = \langle f, T_p x\rangle$. When $p \mid N$, the definition is the same, except that the matrix $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is dropped.

For example, when $N = 11$ we have

$$\begin{aligned} T_2\{0, 1/5\} &= \{0, 2/5\} + \{0, 1/10\} + \{1/2, 3/5\} \\ &= -2\{0, 1/5\}. \end{aligned}$$

In [16], L. Merel gives a description of the action of T_p directly on Manin symbols $[r_i]$ (see also, [6, §2.4]). For example, when $p = 2$ and N is odd, we have

$$T_2([r_i]) = [r_i]\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + [r_i]\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + [r_i]\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} + [r_i]\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}.$$

3 Computing the space of modular forms

In this section we describe how to use modular symbols to construct a basis of $S_2(N)$ consisting of modular forms that are eigenvectors for every element of the ring $\mathbf{T}'$ generated by the Hecke operator T_p, with $p \nmid N$. Such eigenvectors are called *eigenforms*.

Suppose M is a positive integer that divides N. As explained in [13, VIII.1–2], for each divisor d of N/M there is a natural *degeneracy map* $\beta_{M,d} : S_2(M) \to S_2(N)$ given by $\beta_{M,d}(f(q)) = f(q^d)$. The *new subspace* of $S_2(N)$, denoted $S_2(N)^{\text{new}}$, is the orthogonal complement with respect to the Petersson inner product of the images of all maps $\beta_{M,d}$, with M and d as above.

The theory of Atkin and Lehner [1] asserts that, as a $\mathbf{T}'$-module, $S_2(N)$ is built up as follows:

$$S_2(N) \quad = \bigoplus_{M|N,\ d|N/M} \beta_{M,d}(S_2(M)^{\text{new}}).$$

To compute $S_2(N)$ it thus suffices to compute $S_2(M)^{\text{new}}$ for each positive divisor M of N.

We now turn to the problem of computing $S_2(N)^{\text{new}}$. Atkin and Lehner [1] also proved that $S_2(N)^{\text{new}}$ is spanned by eigenforms, each of which occurs with multiplicity one in $S_2(N)^{\text{new}}$. Moreover, if $f \in S_2(N)^{\text{new}}$ is an eigenform then the coefficient of q in the q-expansion of f is nonzero, so it is possible to normalize f so that coefficient of q is 1. With f so normalized, if $T_p(f) = a_p f$, then the pth Fourier coefficient of f is a_p. If $f = \sum_{n=1}^{\infty} a_n q^n$ is a normalized eigenvector for all T_p, then the a_n, with n composite, are determined by the a_p, with p prime, by the following formulas: $a_{nm} = a_n a_m$ when n and m are relatively prime, and $a_{p^r} = a_{p^{r-1}} a_p - p a_{p^{r-2}}$ for $p \nmid N$ prime. When $p \mid N$, $a_{p^r} = a_p^r$. We conclude that in order to compute $S_2(N)^{\text{new}}$, it suffices to compute all systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$ of the Hecke operators $T_2, T_3, T_5, \ldots$ acting on $S_2(N)^{\text{new}}$. Given a system of eigenvalues, the corresponding eigenform is $f = \sum_{n=1}^{\infty} a_n q^n$, where the a_n, for n composite, are determined by the recurrence given above.

In light of the pairing $\langle\, ,\, \rangle$ introduced in Section 1, computing the above systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$ amounts to computing the systems of eigenvalues of the Hecke operators T_p on the subspace V of $\boldsymbol{\mathcal{S}}_2(N)$ that corresponds to the new subspace of $S_2(N)$. For each proper divisor M of N and each divisors d of N/M, let $\phi_{M,d} : \boldsymbol{\mathcal{S}}_2(N) \to \boldsymbol{\mathcal{S}}_2(M)$ be the map sending x to $\left(\begin{smallmatrix} t & 0 \\ 0 & 1 \end{smallmatrix}\right) x$. Then V is the intersection of the kernels of all maps $\phi_{M,d}$.

The computation of the systems of eigenvalues of a collection of commuting diagonalizable endomorphisms involves standard linear algebra techniques, such as computation of characteristic polynomials and kernels of matrices. There are, however, several tricks that greatly speed up this process, some of which are described in [22, §3.5.4].

Example 3.1. All forms in $S_2(39)$ are new. Up to Galois conjugacy, the eigenvalues of the Hecke operators T_2, T_3, T_5, and T_7 on $\boldsymbol{\mathcal{S}}_2(39)$ are $\{1, -1, 2, -4\}$ and $\{a, 1, -2a - 2, 2a+2\}$, where $a^2+2a-1 = 0$. (Note that these eigenvalues occur with multiplicity two.) Thus $S_2(39)$ has dimension 3, and is spanned by

$$f_1 = q + q^2 - q^3 - q^4 + 2q^5 - q^6 - 4q^7 + \cdots ,$$

$$f_2 = q + aq^2 + q^3 + (-2a - 1)q^4 + (-2a - 2)q^5 + aq^6 + (2a + 2)q^7 + \cdots ,$$

and the Galois conjugate of f_2. We compute f_1 and f_2 using Sage as follows:

```
sage: CuspForms(39).newforms('a')
[q + q^2 - q^3 - q^4 + 2*q^5 + O(q^6),
 q + a1*q^2 + q^3 + (-2*a1 - 1)*q^4 + (-2*a1 - 2)*q^5 + O(q^6)]
```

3.1 Summary

To compute the q-expansion, to some precision, of each eigenforms in $S_2(N)$, we use the degeneracy maps so that we only have to solve the problem for $S_2(N)^{\mathrm{new}}$. Here, using modular symbols we compute the systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$, then write down each of the corresponding eigenforms $q + a_2q^2 + a_3q^3 + \cdots$.

References

[1] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[2] B. J. Birch, *Hecke actions on classes of ternary quadratic forms*, Computational number theory (Debrecen, 1989), de Gruyter, Berlin, 1991, pp. 191–212.

[3] B. J. Birch and W. Kuyk (eds.), *Modular functions of one variable. IV*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 476.

[4] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[5] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***, or Wild* 3*-adic exercises*, (2000), http://www.math.harvard.edu/HTML/Individuals/Richard_Taylor.html.

[6] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[7] J. E. Cremona, *Elliptic Curves Data*, `http://www.warwick.ac.uk/~masgaj/ftp/data/`, 2008.

[8] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[9] Fred Diamond and Jerry Shurman, *A first course in modular forms*, Graduate Texts in Mathematics, vol. 228, Springer-Verlag, New York, 2005.

[10] M. J. Greenberg and J. R. Harper, *Algebraic topology*, Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., 1981, A first course.

[11] H. Hijikata, *Explicit formula of the traces of Hecke operators for* $\Gamma_0(N)$, J. Math. Soc. Japan **26** (1974), no. 1, 56–82.

[12] D. R. Kohel, *Hecke module structure of quaternions*, preprint (1998).

[13] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[14] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66.

[15] B. Mazur, *Courbes elliptiques et symboles modulaires*, Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 414, Springer, Berlin, 1973, pp. 277–294. Lecture Notes in Math., Vol. 317.

[16] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations (Berlin), Springer, 1994, pp. 59–94.

[17] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242.

[18] A. Pizer, *An algorithm for computing modular forms on* $\Gamma_0(N)$, J. Algebra **64** (1980), no. 2, 340–390.

[19] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, IAS/Park City Mathematics Institute 1999.

[20] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[21] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

[22] W. A. Stein, *Explicit approaches to modular abelian varieties*, U. C. Berkeley Ph.D. thesis (2000).

[23] William Stein, *Modular forms, a computational approach*, Graduate Studies in Mathematics, vol. 79, American Mathematical Society, Providence, RI, 2007, With an appendix by Paul E. Gunnells.

[24] William Stein, *Sage: Open Source Mathematical Software*, The Sage Group, 2008, `http://www.sagemath.org`.

[25] H. Wada, *Tables of Hecke operations. I*, Seminar on Modern Methods in Number Theory (Tokyo), Inst. Statist. Math., 1971, p. 10.

9 The field generated by the points of small prime order on an elliptic curve, with L. Merel

The field generated by the points of small prime order on an elliptic curve

Loïc Merel and William A. Stein

Introduction

Let $\bar{\mathbf{Q}}$ be an algebraic closure of $\mathbf{Q}$, and for any prime number p, denote by $\mathbf{Q}(\mu_p)$ the cyclotomic subfield of $\bar{\mathbf{Q}}$ generated by the pth roots of unity.

Theorem . — *Let p be a prime. If there exists an elliptic curve E over $\mathbf{Q}(\mu_p)$ such that the points of order p of $E(\bar{\mathbf{Q}})$ are all $\mathbf{Q}(\mu_p)$-rational, then $p = 2, 3, 5, 13$ or $p > 1000$.*

The case $p = 7$ was treated by Emmanuel Halberstadt. The part of the theorem that concerns the case $p \equiv 3 \pmod 4$ is given in [3]. In this paper, we give the details that permit our treating the more difficult case in which $p \equiv 1 \pmod 4$. We treat this last case with the aid of Proposition 2 below, which is not present in *loc. cit.*. The case $p = 13$ is currently under investigation by Marusia Rebolledo, as part of her Ph.D. thesis.

1. Counterexamples define points on $X_0(p)(\mathbf{Q}(\sqrt{p}))$

First we recall some of the results and notation of [3]. Let $S_2(\Gamma_0(p))$ denote the space of cusp forms of weight 2 for the congruence subgroup $\Gamma_0(p)$. Denote by $\mathbf{T}$ the subring of $\operatorname{End} S_2(\Gamma_0(p))$ generated by the Hecke operators T_n for all integers n. Let $f \in S_2(\Gamma_0(p))$ have q-expansion $\sum_{n=1}^{\infty} a_n q^n$. When χ is a Dirichlet character, denote by $L(f, \chi, s)$ the entire function which extends the Dirichlet series $\sum_{n=1}^{\infty} a_n \chi(n)/n^s$.

Let S be the set of isomorphism classes of supersingular elliptic curves in characteristic p. Denote by Δ_S the group formed by the divisors of degree 0 with support on S. It is equipped with a structure of $\mathbf{T}$-module (induced, for example, from the action of the Hecke correspondences on the fiber at p of the regular minimal model of $X_0(p)$ over $\mathbf{Z}$).

Let $j \in \bar{\mathbf{F}}_p - J_S$, where J_S denotes the set of supersingular modular invariants. We denote by ι_j the homomorphism of groups $\Delta_S \longrightarrow \bar{\mathbf{F}}_p$ that associates to $\sum_E n_E[E]$ the quantity $\sum_E n_E/(j - j(E))$, where $j(E)$ denotes the modular invariant of E.

One says that an element $j \in \mathbf{F}_p$ is *anomalous* if there exists an elliptic curve over $\mathbf{F}_p$ with modular invariant j that possesses an $\mathbf{F}_p$-rational point of order p (then necessarily $j \notin J_S$).

Let p be a prime that is congruent to 1 modulo 4. In the following proposition we prove, under a hypothesis on p, that if E is an elliptic curve over $\mathbf{Q}(\mu_p)$ all of whose torsion is $\mathbf{Q}(\mu_p)$-rational, then for each subgroup $C \subset E(\bar{\mathbf{Q}})$ of order p, the point (E, C) on $X_0(p)$ is defined over $\mathbf{Q}(\sqrt{p})$. As we will see in Proposition 2, this $\mathbf{Q}(\sqrt{p})$-rationality conclusion is contrary to fact, from which we conclude that such elliptic curves E do not exist when the hypothesis on p is satisfied. In Section 3 we verify this hypothesis for $p = 11$ and $13 < p < 1000$.

Proposition 1. — *Suppose that p is congruent to* 1 *modulo* 4*. Suppose that for all anomalous $j \in \mathbf{F}_p$ and all non-quadratic Dirichlet characters $\chi\colon (\mathbf{Z}/p\mathbf{Z})^* \longrightarrow \mathbf{C}^*$, there exists $t_\chi \in \mathbf{T}$ and $\delta \in \Delta_S$ such that $L(f, \chi, 1) \neq 0$ for every newform $f \in t_\chi S_2(\Gamma_0(p))$ and $\iota_j(t_\chi \delta) \neq 0$.*

Let E be an elliptic curve over $\mathbf{Q}(\mu_p)$, such that the points of order p of $E(\bar{\mathbf{Q}})$ are all $\mathbf{Q}(\mu_p)$-rational. Then for all subgroups C of order p of $E(\bar{\mathbf{Q}})$, there exists an elliptic curve E_C over $\mathbf{Q}(\sqrt{p})$ equipped with a $\mathbf{Q}(\sqrt{p})$-rational subgroup D_C of order p, and the pairs (E, C) and (E_C, D_C) are $\bar{\mathbf{Q}}$-isomorphic.

Proof. — We prove the proposition using the results of [3]. The hypothesis $\iota_j(t_\chi \delta) \neq 0$ forces $t_\chi \notin p\mathbf{T}$ and, *a fortiori*, $t_\chi \neq 0$; in addition, the non-vanishing hypothesis on the L-series forces the hypothesis $H_p(\chi)$ of *loc. cit.*, introduction.

By assumption, hypothesis $H_p(\chi)$ is satisfied for all non-quadratic Dirichlet characters χ of conductor p. Thus Corollary 3 of Proposition 6 of *loc. cit.* implies that E has potentially good reduction at the prime ideal $\mathcal{P}$ of $\mathbf{Z}[\mu_p]$ that lies above p.

Denote by j the modular invariant of the fiber at $\mathcal{P}$ of the Néron model of E. According to the corollary of Proposition 15 of *loc. cit.*, j is anomalous.

Let C be a subgroup of $E(\bar{\mathbf{Q}})$ of order p. By assumption E is an elliptic curve over $\mathbf{Q}(\mu_p)$ whose points of order p are all $\mathbf{Q}(\mu_p)$-rational, so the pair (E, C) defines a $\mathbf{Q}(\mu_p)$-rational point P of the modular curve $X_0(p)$.

Consider the morphism $\phi_\chi = \phi_{t_\chi} : X_0(p) \to J_0(p)$ obtained by composing the standard embedding of $X_0(p)$ into $J_0(p)$ with t_χ. As in section 1.3 of *loc. cit.*, ϕ_χ extends to a map from the minimal regular model of $X_0(p)$ to the Néron model of $J_0(p)$. When $\iota_j(t_\chi \delta) \neq 0$, this map is a formal immersion at the point $P_{/\mathbf{F}_p}$, according to *loc. cit.*, Proposition 4. The hypothesis that $L(f, \chi, 1) \neq 0$ for every newform $f \in t_\chi S_2(\Gamma_0(p))$, translates into $L(t_\chi J_0(p), \chi, 1) \neq 0$, which in turn implies that the χ-isotypical component of $t_\chi J_0(p)(\mathbf{Q}(\mu_p))$ is finite (this is Kato's theorem, see the discussion in section 1.5 of *loc. cit.*). We can then apply Corollary 1 of Proposition 6 of *loc. cit.*. This proves that P is $\mathbf{Q}(\sqrt{p})$-rational, which translates into the conclusion of Proposition 1.

Remark 1: Proposition 1 is true even under the weaker hypothesis that t_χ lies in $\mathbf{T} \otimes \mathbf{Z}[\chi]$, which acts $\mathbf{Z}[\chi]$-linearly on modular forms.

2. Elliptic curves and quadratic fields

PROPOSITION 2. — *Let p be a prime number > 5 and congruent to 1 modulo 4. Let E be an elliptic curve over $\bar{\mathbf{Q}}$. There exists a subgroup $C \subset E(\bar{\mathbf{Q}})$ of order p such that (E, C) can not be defined over $\mathbf{Q}(\sqrt{p})$.*

Proof. — We procede by contradiction, i.e., we assume that for all cyclic subgroups C of order p of $E(\bar{\mathbf{Q}})$, the pair (E, C) can be defined over $\mathbf{Q}(\sqrt{p})$. We choose such a pair (E_0, C_0) over $\mathbf{Q}(\sqrt{p})$.

Assume first that all twists of E are quadratic, i.e. that $j(E)$ is neither 0 nor 1728. We show that the group $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$ acts by scalars on the $\mathbf{F}_p$-vector space $E_0(\bar{\mathbf{Q}})[p]$. For this it suffices to show that all subgroups of order p of $E_0(\bar{\mathbf{Q}})[p]$ are stable by $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$.

Suppose C_1 is a cyclic subgroup of order p of $E_0(\bar{\mathbf{Q}})[p]$. By assumption, there exists a quadratic twist E_1 of E_0 and a cyclic subgroup C_1' of $E_1(\bar{\mathbf{Q}})[p]$ that is defined over $\mathbf{Q}(\sqrt{p})$, such that the image of C_1 by the isomorphism $E_0 \simeq E_1$ is C_1'. Since $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$ leaves C_1' stable and the action of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$ on $E_0(\bar{\mathbf{Q}})[p]$ is a quadratic twist of the action on $E_1(\bar{\mathbf{Q}})[p]$, we see that $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$ leaves C_1 stable. Thus $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$ fixes all lines in $E_0(\bar{\mathbf{Q}})[p]$, and hence acts by scalars. Denote by α the corresponding character of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p}))$.

Because of the Weil pairing, α^2 coincides with the cyclotomic character modulo p, and it factors through $\mathrm{Gal}(\mathbf{Q}(\mu_p)/\mathbf{Q}(\sqrt{p}))$. But, when $p \equiv 1 \pmod 4$, the group $\mathrm{Gal}(\mathbf{Q}(\mu_p)/\mathbf{Q}(\sqrt{p}))$ is of even order, and the characters modulo p form a group generated by the reduction modulo p of the cyclotomic character, which, therefore, can not be a square.

Next suppose that $j(E) = 0$ or $j(E) = 1728$. Indeed, in these two cases E has complex multiplication by an order of $K = \mathbf{Q}[\sqrt{-3}]$ or $\mathbf{Q}[\sqrt{-1}]$. Let $d_K = 3$ or $d_K = 2$ in these two cases respectively. Let C be a subgroup of order p of $E(\bar{\mathbf{Q}})$. Consider the map $\rho_0 : \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{p})) \longrightarrow \mathrm{Aut}\, E_0(\bar{\mathbf{Q}})[p]$. Since E has complex multiplication, the image of ρ_0 has no element of order p. Therefore, there are at least two subgroups, including C_0, of order p of $E(\bar{\mathbf{Q}})$ stable under the image of ρ_0. Call the other subgroup C_1. Let C_2 be a subgroup of order p of $E(\bar{\mathbf{Q}})$ which is distinct from C_0 and C_1. The pair (E, C_2) can be defined over $\mathbf{Q}(\sqrt{p})$. Therefore, there exists an extension field K_2 of $\mathbf{Q}(\sqrt{p})$, whose degree d_2 divides $2d_K$, such that the image of the restriction of ρ_0 to $\mathrm{Gal}(\bar{\mathbf{Q}}/K_2)$ leaves stable three distinct subgroups of order p of $E_0(\bar{\mathbf{Q}})$, and therefore consists only of scalars. If $d_2 \leq 2$, one concludes as in the cases where $j(E) \neq 0$ and $j(E) \neq 1728$. We suppose now that $d_2 > 2$. The projective image of ρ_0 has order d_K.

Since E is an elliptic curve over $\bar{\mathbf{Q}}$ with complex multiplication by a field of class number one, there is a model for E that is defined over $\mathbf{Q}$. Consider the map $\rho : \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{Aut}\, E(\bar{\mathbf{Q}})[p]$. By the theory of complex multiplication, the projective image of ρ has order $2(p+1)$ or $2(p-1)$. There exists a field extension L of degree dividing d_K of $\mathbf{Q}(\sqrt{p})$ such that the restrictions to $\mathrm{Gal}(\bar{\mathbf{Q}}/L)$ of the projective images of ρ and ρ_0 coincide. Therefore one has $(p-1)|d_K^2$ or $(p+1)|d_K^2$. This imposes $p = 5$ and $d_K = 2$.

3. Verification of the hypothesis of Proposition 1 Let p be a prime number. In this section we explain how we used a computer to verify that the second hypothesis of Proposition 1 are satisfied for $p = 11$ and $13 < p < 1000$. (In the present paper, this verification is only required for p that are congruent to 1 modulo 4.)

We first list the anomalous j-invariants $j \in \mathbf{F}_p$. Since p is fairly small in the range of our computations, we created this list by simply enumerating all of the elliptic curves over $\mathbf{F}_p$ and counting the number of points on each curve. For example, when $p = 31$ the anomalous j-invariants are $j = 10, 14$.

Let $\chi : \mathbf{Z}/p\mathbf{Z} \longrightarrow \mathbf{C}$ be a non-quadratic Dirichlet character, and denote by $\mathbf{Z}[\chi]$ the subring of $\mathbf{Q}(\zeta_{p-1})$ generated by the image of χ. Denote by $S_2(\Gamma_0(p); \mathbf{Z})$ the set of modular forms $f \in S_2(\Gamma_0(p))$ whose Fourier expansion at the cusp ∞ lies in $\mathbf{Z}[[q]]$.

We study the $\mathbf{T}$-modules $\mathbf{T}$, Δ_S, and $S_2(\Gamma_0(p); \mathbf{Z})$. After extension of scalars to $\mathbf{Q}$, these are $\mathbf{T} \otimes \mathbf{Q}$-modules that are free of rank 1, of which the irreducible sub-$\mathbf{T} \otimes \mathbf{Q}$ modules are the annihilators of the minimal prime ideals of $\mathbf{T}$. We compute a list of the minimal prime ideals of $\mathbf{T}$ by computing appropriate kernels and characteristic polynomials of Hecke operators of small index on Δ_S, which we find using the graph method of Mestre and Oesterlé [4].

Having computed the minimal prime ideals of $\mathbf{T}$, we verify that some nontrivial ideal $\mathcal{I}$ of $\mathbf{T}$ (always a minimal prime ideal in the range of our computations) simultaneously satisfies the following three conditions:

1) For each anomalous j-invariant, there exists $x \in \Delta_S$ such that $\mathcal{I}x = 0$ and $\iota_j(x) \neq 0$.

2) Each of the newforms $f \in S_2(\Gamma_0(p))$ with $\mathcal{I}f = 0$ satisfies $L(f, \chi, 1) \neq 0$.

3) The image of $\mathcal{I}$ in the $\mathbf{T}$-module $\mathbf{T}/p\mathbf{T}$ is a direct factor.

Let $\mathcal{I}$ be an ideal of $\mathbf{T}$. Here is how we verify these conditions for $\mathcal{I}$.

Verification of condition 1.

We verified that $\mathcal{I}$ satisfies the first condition by finding a $\mathbf{T}$-eigenvector v of $\Delta_S \otimes \bar{\mathbf{Z}}$ that is annihilated by $\mathcal{I}$ and satisfies $\iota_j(v) \neq 0$ for all anomalous j-invariants. Because ι_j is a homomorphism, this implies the existence of x as in condition 1.

Verification of condition 2.

We verified the second condition using modular symbols. Our method is purely algebraic, so we do not perform any approximate computation of integrals. Using the algorithm described in [2], we compute the action of the Hecke algebra $\mathbf{T}$ on the space $\mathrm{Hom}_{\mathbf{Q}[\chi]}(H_1(X_0(p); \mathbf{Q}[\chi]), \mathbf{Q}[\chi])$. By intersecting the kernels of appropriate elements of $\mathbf{T}$, we find a basis $\varphi_1, \ldots, \varphi_n$ for the subspace of $\mathrm{Hom}_{\mathbf{Q}[\chi]}(H_1(X_0(p); \mathbf{Q}[\chi]), \mathbf{Q}[\chi])$ that is annihilated by $\mathcal{I}$. Let $\Phi_{\mathcal{I}} = \varphi_1 \times \cdots \times \varphi_n$ denote the linear map $H_1(X_0(p); \mathbf{Q}[\chi]) \longrightarrow \mathbf{Q}[\chi]^n$ defined by the φ_i.

Let $\mathbf{T}_{\mathbf{Q}[\chi]} = \mathbf{T} \otimes \mathbf{Q}[\chi]$, where $\mathbf{Q}[\chi]$ is the number field generated the image of χ. The *χ-twisted winding element* (denoted θ_χ in [3])

$$\mathbf{e}_\chi = \sum_{a \in (\mathbf{Z}/p\mathbf{Z})^*} \bar{\chi}(a) \left\{ \infty, \frac{a}{p} \right\}$$

generates the χ-*twisted winding submodule* $\mathbf{T}_{\mathbf{Q}[\chi]} \cdot \mathbf{e}_\chi$. To compute this submodule, we use that $\mathbf{T}$ is generated, even as a $\mathbf{Z}$-module, by $T_1, T_2, \ldots, T_b$, for any $b \geq (p+1)/6$ (see [1]).

Lemma 3. — *Let $\mathcal{I}$ be a minimal prime ideal of $\mathbf{T}$, and let $\chi : (\mathbf{Z}/N\mathbf{Z})^* \longrightarrow \mathbf{C}^*$ be a nontrivial Dirichlet character. Then the dimension of the $\mathbf{Q}[\chi]$-vector space $\Phi_{\mathcal{I}}(\mathbf{T}_{\mathbf{Q}[\chi]} \cdot \mathbf{e}_\chi)$ is equal to the cardinality of the set of newforms f such that $\mathcal{I}f = 0$ and $L(f, \chi, 1) \neq 0$.*
Proof. — We have

$$\dim_{\mathbf{Q}[\chi]} \Phi_{\mathcal{I}}(\mathbf{T}_{\mathbf{Q}[\chi]} \cdot \mathbf{e}_\chi) = \dim_{\mathbf{C}} \Phi_{\mathcal{I}}(\mathbf{T}_{\mathbf{C}} \cdot \mathbf{e}_\chi).$$

This dimension is invariant upon changing the basis $\varphi_1, \ldots, \varphi_n$ used to define $\Phi_{\mathcal{I}}$. In particular, over $\mathbf{C}$ there is a basis $\varphi'_1, \ldots, \varphi'_n$ so that the resulting map $\Phi'_{\mathcal{I}}$ satisfies

$$\Phi'_{\mathcal{I}}(x) = \Big(\mathrm{Re}(\int_x f^{(1)}), \mathrm{Im}(\int_x f^{(1)}), \ldots, \mathrm{Re}(\int_x f^{(d)}), \mathrm{Im}(\int_x f^{(d)})\Big),$$

where $f^{(1)}, \ldots, f^{(d)}$ are the Galois conjugates of a newform $f^{(1)} = \sum a_n^{(1)} q^n$ such that $\mathcal{I}f^{(1)} = 0$. Furthermore, $\Phi'_{\mathcal{I}}$ is a $\mathbf{T}_{\mathbf{C}}$-module homomorphism if we declare that $\mathbf{T}_{\mathbf{C}}$ acts on $\mathbf{R}^{2d} = \mathbf{C}^d$ via

$$T_n(x_1, y_1, \ldots, x_d, y_d) = T_n(z_1, \ldots, z_d) = (a_n^{(1)} z_1, \ldots, a_n^{(d)} z_d),$$

where $z_j = x_j + iy_j$ and the $a_n^{(j)}$ are Fourier coefficients of the $f^{(j)}$.

As explained in Section 2.2 of [3], $\int_{\mathbf{e}_\chi} f = * \cdot L(f, \chi, 1)$, where $*$ is some nonzero real or pure-imaginary complex number, according to whether $\chi(-1)$ equals 1 or -1, respectively. Combining this observation with the equality

$$\dim_{\mathbf{C}} \Phi_{\mathcal{I}}(\mathbf{T}_{\mathbf{C}} \cdot \mathbf{e}_\chi) = \dim_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}} \cdot \Phi_{\mathcal{I}}(\mathbf{e}_\chi)),$$

and that the image of $\mathbf{T}_{\mathbf{C}}$ in $\mathrm{End}(\mathbf{C}^d)$ is equal to the diagonal matrices, proves the asserted equality.

Remark 2: The dimension of $\Phi_{\mathcal{I}}(\mathbf{T}_{\mathbf{Q}[\chi]} \cdot \mathbf{e}_\chi)$ is unchanged if χ is replaced by a Galois-conjugate character.

In practice, computations over the cyclotomic field $\mathbf{Q}[\chi]$ are extremely expensive. Fortunately, for our application it suffices to give a lower bound on the dimension appearing in the lemma. Such a bound can be efficiently obtained by instead computing the reductions of Φ, χ, and the χ-twisted winding submodule modulo a suitable maximal ideal of the ring of integers of $\mathbf{Q}[\chi]$ that splits completely; this amounts to performing the above linear algebra over a relatively small finite field $\mathbf{F}_\ell$ where ℓ is congruent to 1 modulo $p-1$.

Remark 3: For every newform f in $S_2(\Gamma_0(p))$, with $p \leq 1000$, and every mod p Dirichlet character χ, we found that $L(f, \chi, 1) \neq 0$ if and only if $L(f^\sigma, \chi, 1) \neq 0$ for all conjugates f^σ of f. More generally, for any f and χ, this equivalence holds if $\mathbf{Q}[\chi]$ is linearly disjoint from the field $K_f = (\mathbf{T}/\mathcal{I}) \otimes \mathbf{Q}$. The first few primes for which there is a form f and a mod p character χ such that the linear disjointness hypothesis fails are $p = 31, 113, 127$, and 191. The analogue of this nonvanishing observation is false if we instead consider newforms on $\Gamma_1(p)$ and allow χ to be arbitrary. For example, let f be one of the two Galois-conjugate newforms in $S_2(\Gamma_1(13))$. Then there is a character $\chi : (\mathbf{Z}/7\mathbf{Z})^* \longrightarrow \mathbf{C}^*$ of order 3 such that $L(f, \chi, 1) = 0$ and $L(f^\sigma, \chi, 1) \neq 0$.

Verification of condition 3.

The third condition is satisfied for all $p < 10000$, except possibly $p = 389$, because we have verified that the discriminant of $\mathbf{T}$ is prime to p for all such $p \neq 389$, so the ring $\mathbf{T}/p\mathbf{T}$ is semisimple. The discriminant computation was carried out by the second author as follows. Using the method of [4], we computed discrimininants of characteristic polynomials mod p of the Hecke operators T_2, T_3, T_5, and T_7. In the few cases when all four of these characteristic polynomials had discriminant equal to 0 mod p, we resorted to modular symbols to compute several more characteristic polynomials until we found one having nonzero discriminant modulo p.

We consider the remaining case $p = 389$ in detail. There are exactly five minimal prime ideals of $\mathbf{T}$, which we denote $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_6$, and $\mathcal{P}_{20}$, where the quotient field of $\mathbf{T}/\mathcal{P}_i$ has dimension i. The discriminant of the characteristic polynomial of T_2 is exactly divisible by 389. Since the field of fractions of $\mathbf{T}/\mathcal{P}_{20}$ has discriminant divisible by 389, we see that 389 is not the residue characteristic of any congruence prime. Let $\mathcal{O}_i = \mathbf{T}/\mathcal{P}_i$. The natural map $\mathbf{T} \to \prod \mathcal{O}_i$ has finite kernel and cokernel each of order coprime to 389, so $\mathbf{T}/389\mathbf{T} \cong \prod \mathcal{O}_i/389\mathcal{O}_i$. The nonquadratic characters $\chi : (\mathbf{Z}/p\mathbf{Z})^* \to \mathbf{C}^*$ have orders $1, 4, 97, 193, 388$. We must verify that for each of these degrees, one of the ideals $\mathcal{P}_i$ satisfies conditions 1–3. We check as above that conditions 1–3 for χ of order 4 are satisfied by $\mathcal{P}_2$ and conditions 1–3 for χ of order greater than 4 are satisfied by $\mathcal{P}_1$. When χ is the trivial character, conditions 1–3 are satisfied only by $\mathcal{P}_{20}$.

Summary.

For each prime $p < 1000$ different than $2, 3, 5, 7, 13$, we verified the existence of an ideal that satisfies the three conditions given above, as follows. We consider each Galois conjugacy class of non-quadratic characters χ. We find a single newform f such that $L(f, \chi, 1) \neq 0$ for all conjugates of f and of χ. Then we let $\mathcal{I}$ be the annihilator of f, and try to verify condition 1 for *all* of the anamolous j-invariants in $\mathbf{F}_p$. When the three conditions are satisfied for an ideal $\mathcal{I}$ of $\mathbf{T}$, there exists $t_\chi \in \mathbf{T}$ that is annihilated by $\mathcal{I}$ and is the inverse image of a projector of $\mathbf{T}/p\mathbf{T}$ on the complement of $\mathcal{I} + p\mathbf{T}$. Putting $\delta = x$, one has $\iota_j(t_\chi \delta) = \iota_j(\delta) \neq 0$ (because ι_j takes its values in characteristic p, it follows that δ is annihilated by $\mathcal{I}$ and $t_\chi \in 1 + p\mathbf{T} + \mathcal{P}$). Every newform $f \in t_\chi S_2(\Gamma_0(p))$ satisfies $\mathcal{I}f = 0$, and therefore, by our second condition, $L(f, \chi, 1) \neq 0$. The pair (t_χ, δ) then satisfies the conditions required by Proposition 1.

Acknowledgment: We would like to thank Barry Mazur for providing us several useful comments.

Bibliography

[1] A. Agashe, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Ser. I Math. 328 (1999), no. 5, 369–374.

[2] J. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, (1997).

[3] L. Merel, *Sur la nature non cyclotomique des points d'ordre fini des courbes elliptiques*, To appear in Duke Math. Journal.

[4] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata), 217–242, (1986).

10 Appendex on Generating the Hecke algebra, with A. Agashe

Appendix by A. Agashe and W. Stein.

In this appendix, we apply a result of J. Sturm* to obtain a bound on the number of Hecke operators needed to generate the Hecke algebra as an abelian group. This bound was suggested to the authors of this appendix by Loïc Merel and Ken Ribet.

Theorem. *The ring* $\mathbf{T}$ *of Hecke operators acting on the space of cusp forms of weight* k *and level* N *is generated as an abelian group by the Hecke operators* T_n *with*

$$n \leq \frac{kN}{12} \cdot \prod_{p|N} \left(1 + \frac{1}{p}\right).$$

Proof. For any ring R, let $S_k(N; R) = S_k(N; \mathbf{Z}) \otimes R$, where $S_k(N; \mathbf{Z})$ is the subgroup of cusp forms with integer Fourier expansion at the cusp ∞, and let $\mathbf{T}_R = \mathbf{T} \otimes_{\mathbf{Z}} R$. There is a perfect pairing $S_k(N; R) \otimes_R \mathbf{T}_R \to R$ given by $\langle f, T\rangle \mapsto a_1(T(f))$.

Let M be the submodule of $\mathbf{T}$ generated by $T_1, T_2, \ldots, T_r$, where r is the largest integer $\leq \frac{kN}{12} \cdot \prod_{p|N} \left(1 + \frac{1}{p}\right)$. Consider the exact sequence of additive abelian groups

$$0 \to M \xrightarrow{i} \mathbf{T} \to \mathbf{T}/M \to 0.$$

Let p be a prime and use that tensor product is right exact to obtain an exact sequence

$$M \otimes \mathbf{F}_p \xrightarrow{\bar{i}} \mathbf{T} \otimes \mathbf{F}_p \to (\mathbf{T}/M) \otimes \mathbf{F}_p \to 0.$$

Suppose that $f \in S_k(N; \mathbf{F}_p)$ pairs to 0 with each of $T_1, \ldots, T_r$. Then $a_m(f) = a_1(T_m f) = \langle f, T_m\rangle = 0$ in $\mathbf{F}_p$ for each $m \leq r$. By Theorem 1 of Sturm's paper, it follows that $f = 0$. Thus the pairing restricted to the image of $M \otimes \mathbf{F}_p$ in $\mathbf{T} \otimes \mathbf{F}_p$ is nondegenerate, so

$$\dim_{\mathbf{F}_p} \bar{i}(M \otimes \mathbf{F}_p) = \dim_{\mathbf{F}_p} S_k(N, \mathbf{F}_p) = \dim_{\mathbf{F}_p} \mathbf{T} \otimes \mathbf{F}_p.$$

It follows that $(\mathbf{T}/M) \otimes \mathbf{F}_p = 0$; repeating the argument for all primes p shows that $\mathbf{T}/M = 0$, as claimed.

Remark. In general, the theorem is not true if one considers only T_n where n runs over the *primes* less than the bound. Consider, for example, $S_2(11)$, where the bound is 2 and T_2 is the 1×1 matrix $[2]$, which does not generate the full Hecke algebra as a $\mathbf{Z}$-submodule of $\mathrm{End}(S_2(\Gamma_0(N), \mathbf{Z}))$. One needs, in addition, the matrix $[1]$.

* J. Sturm, *On the Congruence of Modular Forms.* Number theory (New York, 1984–1985), 275–280, Lecture Notes in Math., 1240, Springer, Berlin-New York, 1987.

11 Component Groups of Purely Toric Quotients, with B. Conrad

Component Groups of Purely Toric Quotients

William A. Stein Brian Conrad

April 21, 2003

Abstract

Suppose $\pi : J \to A$ is an optimal quotient of abelian varieties over a p-adic field, optimal in the sense that $\ker(\pi)$ is connected. Assume that J is equipped with a symmetric principal polarization θ (e.g., any Jacobian of a curve has such a polarization), that J has semistable reduction, and that A has purely toric reduction. In this paper, we express the group of connected components of the Néron model of A in terms of the monodromy pairing on the character group of the torus associated to J. We apply our results in the case when A is an optimal quotient of the modular Jacobian $J_0(N)$. For each prime p that exactly divides N, we obtain an algorithm to compute the component group of A at p.

1 Introduction

Let A be an abelian variety over the rational numbers $\mathbf{Q}$. Birch and Swinnerton-Dyer found a conjectural formula for the order of the Shafarevich-Tate group of A. The Tamagawa numbers c_p of A are among the quantities that appear in this formula. We now recall the definition of the Tamagawa numbers of an abelian variety (the definition of Néron model and component groups is given in Section 2).

Definition 1.1 (Tamagawa number). Let p be a prime, let $\mathcal{A}$ be the Néron model of A over the p-adic integers $\mathbf{Z}_p$, and let $\Phi_{A,p}$ be the component group of $\mathcal{A}$ at p. Then the *Tamagawa number* c_p of A at p is the order of the subgroup $\Phi_{A,p}(\mathbf{F}_p)$ of $\mathbf{F}_p$-rational points in $\Phi_{A,p}(\overline{\mathbf{F}}_p)$.

Remark 1.2. The Tamagawa number is defined in a different way in some other papers, but the definitions are equivalent.

When A has dimension one, A is called an elliptic curve, and A can be defined by a Weierstrass equation $y^2 = x^3 + ax + b$. Using that elliptic curves (and their related integral models) can be described by simple equations, Tate found an efficient algorithm to compute all of the Tamagawa numbers of A (see [18]). In the case when A is the Jacobian of a genus 2 curve, [7] discusses a method for computing the Tamagawa numbers of A. In this paper, we consider the situation in which A has purely toric reduction at p, with no constraint on the dimension of A. For such A we give an explicit description of the order of the group of connected components of the closed fiber of the Néron model of A. In the case when $A = A_f$ is a quotient of $J_0(N)$ attached to a newform $f \in S_2(\Gamma_0(N))$ and $p \mid\mid N$, our method is completely explicit, and yields an algorithm to compute the Tamagawa number c_p of A (up to a bounded power of 2).

This paper is structured as follows. In Sections 2–6 we state and prove an explicit formula involving component groups of fairly general abelian varieties. Then in Section 7 we turn to quotients of modular Jacobians $J_0(N)$. We give some tables and discussed the arithmetic of quotients of $J_0(N)$ when N is prime. In Section 8 we prove a couple of facts about toric reduction that are used in the proof of Theorem 6.1.

Acknowledgement: This paper was inspired by lectures of R. Coleman and K. Ribet, and a letter from Ribet to Mestre (see [17]), which contains some of the results of the present paper in the case when A has dimension 1. The second author would like to thank A. Abbes, A. Agashe, D. Kohel, and D. Lorenzini, for helpful conversations. Both authors were partially supported by the NSF and the Clay Mathematics Institute during work on this paper.

2 The Main Results

In this section, we summarize the main contributions of this paper. First we recall the precise definition of the component group of an abelian variety, then we state our main theorem.

Let R be a discrete valuation ring with field of fractions K and maximal ideal $\mathfrak{m}$, and let $k = R/\mathfrak{m}$ be the residue class field. Let A be an abelian variety over K.

Definition 2.1 (Néron model). A *Néron model* of A is a smooth commutative group scheme $\mathcal{A}$ over R such that A is its generic fiber and $\mathcal{A}$ satisfies the Néron mapping property: the restriction map

$$\mathrm{Hom}_R(S, \mathcal{A}) \longrightarrow \mathrm{Hom}_K(S_K, A)$$

is bijective for all smooth schemes S over R.

The Néron mapping property implies that $\mathcal{A}$ is unique up to a unique isomorphism, so we will refer without hesitation to "the" Néron model of A. Néron models are separated and of finite type as opposed to just locally of finite type, even though their universal property is on the category of arbitrary smooth R-schemes. For more about Néron models see [2].

The closed fiber $\mathcal{A}_k$ of $\mathcal{A}$ is a group scheme over k, which need not be connected. Denote by $\mathcal{A}_k^0$ the connected component of $\mathcal{A}_k$ that contains the identity. We have an exact sequence

$$0 \longrightarrow \mathcal{A}_k^0 \longrightarrow \mathcal{A}_k \longrightarrow \Phi_A \longrightarrow 0,$$

where Φ_A is a finite étale group scheme over k. Equivalently, Φ_A is a commutative finite group equipped with a continuous action of $\mathrm{Gal}(\overline{k}/k)$.

Definition 2.2 (Component group). The *component group* of an abelian variety A over K is the group scheme $\Phi_A = \mathcal{A}_k/\mathcal{A}_k^0$.

2.1 Statement of the Theorem

We now state our main result, supressing some of the definitions of the terms used until later (see Section 6 below for a more complete statement and the proof). Let K be as above, and suppose $\pi : J \to A$ is an optimal quotient. Assume that J is equipped with a symmetric principal polarization λ, in the sense of Definition 5.1. For example, the θ polarization of the Jacobian of a curve is a symmetric principal polarization. Also assume that J has semistable reduction, and that A has purely toric reduction.

We express the component group of A in terms of the monodromy pairing associated to J. Let $m_A = \sqrt{\deg(\theta_A)}$, where $\theta_A : A^\vee \to A$ is induced by the canonical principal polarization of J arising from the θ-divisor. Let X_J be the character group of the toric part of the closed fiber of the Néron model of J. Let $\mathcal{L}$ be the saturation of the image of X_A in

X_J. The monodromy pairing induces a map $\alpha : X_J \to \mathrm{Hom}(\mathcal{L}, \mathbf{Z})$. Let Φ_X be the cokernel of α and $m_X = [\alpha(X_J) : \alpha(\mathcal{L})]$ be the order of the finite group $\alpha(X_J)/\alpha(\mathcal{L})$. The main result of this paper is that

$$\frac{\#\Phi_A}{m_A} = \frac{\#\Phi_X}{m_X},$$

and this is recorded as Theorem 6.1 below.

Using the snake lemma, one sees that Φ_X is isomorphic to the image of the natural map $\Phi_J \to \Phi_A$, and the above formula implies that the cokernel of the map $\Phi_J \to \Phi_A$ has order m_A/m_X. A non-obvious consequence of this is that $m_X \mid m_A$.

In the context of modular forms, if the optimal quotient $J \to A$ arises from a newform on $\Gamma_0(N)$, then the quantities m_A, m_X and Φ_X can be explicitly computed, hence we can compute $\#\Phi_A$. Note that the authors have not computed the structure of Φ_A as a group.

3 Optimal Quotients

Let K be as in Section 2, let J be an abelian variety equipped with a symmetric principal polarization θ_J (see Definition 5.1). For example, J could be the Jacobian of a curve equipped with the canonical principal polarization arising from the θ-divisor.

Definition 3.1 (Optimal quotient). An *optimal quotient* of J is an abelian variety A and a smooth surjective morphism $\pi : J \to A$ whose kernel is connected (i.e., an abelian variety).

Remark 3.2. Any connected scheme of finite type over a field is geometrically connected if it contains a rational point (e.g., if it is a group scheme). See [8, IV$_2$, §4.5.13].

Let $\pi : J \to A$ be an optimal quotient. Denote by $J^\vee$ and $A^\vee$ the abelian varieties dual to J and A, respectively. Upon composing the dual of π with $\theta_J^\vee = \theta_J$, we obtain a map

$$A^\vee \xrightarrow{\pi^\vee} J^\vee \xrightarrow{\theta_J} J.$$

Proposition 3.3. *The map $\theta_J \circ \pi^\vee : A^\vee \to J$ is a closed immersion.*

Proof. Since θ_J is an isomorphism, we want to prove that $\pi^\vee$ is a closed immersion. It is a general fact that duals to surjections of abelian varieties with abelian variety kernel are closed immersions, but for lack of an adequate reference we recall the proof. Since a monomorphism between smooth finite type group schemes over a field is necessarily a closed immersion, it suffices to show that the commutative proper group scheme $\ker(\pi^\vee)$ vanishes. Since a non-zero commutative proper group scheme G over a field F necessarily has a non-zero finite subgroup scheme $G[n]$ for some n (since either $(G_{/\overline{F}})^0_{\mathrm{red}}$ is an abelian variety or else G is finite and non-zero), it suffices to show that $\ker(\pi^\vee)[n]$ vanishes for all positive integers n. In other words, it suffices to show that the induced map $A^\vee[n] \to J^\vee[n]$ is a closed immersion for all n.

Since Cartier duality interchanges faithfully flat maps and closed immersions, and the scheme-theoretic Weil pairing identifies the Cartier dual of the map induced by $\pi^\vee$ on n-torsion with $\pi : J[n] \to A[n]$, we just have to show that these latter maps are faithfully flat for all integers n. Using the short exact sequence

$$0 \to \ker(\pi) \to J \to A \to 0$$

in the abelian category of fppf abelian sheaves over $\mathrm{Spec}(K)$, the snake lemma gives an exact sequence

$$0 \to \ker(\pi)[n] \to J[n] \to A[n] \to 0$$

because $n : \ker(\pi) \to \ker(\pi)$ is a faithfully flat map (hence fppf surjective), as $\ker(\pi)$ is an abelian variety. This gives an isomorphism of group schemes

$$J[n]/\ker(\pi)[n] \simeq A[n]$$

compatible with the maps from $J[n]$, whence $J[n] \to A[n]$ is faithfully flat.

□

Henceforth we will abuse notation and denote the injection $A^\vee \to J$ by $\pi^\vee$. We define θ_A to be the composite $\pi \circ \pi^\vee$, so the kernel of θ_A equals the scheme-theoretic intersection of $A^\vee$ and $B = \ker(\pi)$, as depicted in the following diagram:

$$\begin{tikzcd}
A^\vee \cap B \arrow[r] \arrow[d] & B \arrow[d] \\
A^\vee \arrow[r, hook, "\pi^\vee"] \arrow[rd, "\theta_A"'] & J \arrow[d, "\pi"] \\
& A.
\end{tikzcd}$$

Since θ_A is a polarization (due to how its definition uses the polarization θ_J) the degree of θ_A is a perfect square (see [16, §16, p. 150]).

Definition 3.4 (Degree). Define the *degree* of A as a quotient of J to be the integer

$$m_A = \sqrt{\# \ker(\theta_A)}.$$

4 The Closed Fiber of the Néron Model

In this section we recall some terminology associated with closed fibers of Néron models. Let K, R, and k be as in Section 2, and let $\Phi_A = \mathcal{A}_k/\mathcal{A}_k^0$ be the group scheme of connected components of the closed fiber $\mathcal{A}_k$. By Chevalley's structure theorem (see [3], or [4] for a modern account), if K is a perfect extension field of k (e.g., $K = \overline{k}$) then there is a unique short exact sequence

$$0 \to \mathcal{C} \to \mathcal{A}_K^0 \to \mathcal{B} \to 0$$

with $\mathcal{C}$ a smooth affine algebraic K-group and $\mathcal{B}$ an abelian variety. Moreover, there is a unique exact sequence

$$0 \to \mathcal{T} \to \mathcal{C} \to \mathcal{U} \to 0$$

with $\mathcal{T}$ a torus and $\mathcal{U}$ unipotent.

Using the rigidity of tori, one can show that $\mathcal{T}$ is induced by a unique torus in $\mathcal{A}_k^0$. In particular, the condition that $\mathcal{B} = \mathcal{U} = 0$ is equivalent to the condition that $\mathcal{A}_k^0$ be a torus, and the condition that $\mathcal{U} = 0$ is equivalent to the condition that $\mathcal{A}_k^0$ be the extension of an abelian variety by a torus (i.e., be a semi-abelian variety). These conditions can be checked on a geometric closed fiber.

Definition 4.1. The abelian variety A is said to have *purely toric reduction* if $\mathcal{A}_k^0$ is torus, and to have *semistable reduction* if $\mathcal{A}_k^0$ is a semi-abelian variety (i.e., $\mathcal{A}_{\overline{k}}^0$ has vanishing unipotent part).

4.1 The Monodromy Pairing on the Character Group

Definition 4.2 (Character group of torus). The *character group*

$$X_A = \mathrm{Hom}_{\overline{k}}(\mathcal{T}_{\overline{k}}, \mathbf{G}_{m\overline{k}})$$

is a free abelian group of rank t contravariantly associated to A.

As discussed in [9], if A is semistable there is a *monodromy pairing* $X_A \times X_{A^\vee} \to \mathbf{Z}$ and an exact sequence

$$0 \to X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

Also, the canonical isomorphism $(A^\vee)^\vee \cong A$ induces an isomorphism

$$X_{A^\vee} \times X_{(A^\vee)^\vee} \cong X_A \times X_{A^\vee},$$

which identifies the monodromy pairing associated to $A^\vee$ with that associated to A.

Example 4.3 *(Tate curve).* Suppose $E = \mathbf{G}_m/q^{\mathbf{Z}}$ is a Tate curve over $\mathbf{Q}_p^{\mathrm{ur}}$. The monodromy pairing on $X_E = q^{\mathbf{Z}}$ is

$$\langle q, q \rangle = \mathrm{ord}_p(q) = -\,\mathrm{ord}_p(j).$$

Thus Φ_E is cyclic of order $-\,\mathrm{ord}_p(j)$.

Suppose J is an abelian variety equipped with a symmetric principal polarization. Since J is self dual via the given symmetric principal polarization, we can view the monodromy pairing on J as a pairing $X_J \times X_J \to \mathbf{Z}$. Because the principal polarization on J is symmetric the resulting pairing $X_J \times X_J \to \mathbf{Z}$ is symmetric, so there is no ambiguity about left versus right definitions of $X_J \to \mathrm{Hom}(X_J, \mathbf{Z})$. The above exact sequence then becomes

$$0 \to X_J \to \mathrm{Hom}(X_J, \mathbf{Z}) \to \Phi_J \to 0.$$

5 The Degree of a Symmetric Isogeny

We next relate the degree of the isogeny $A^\vee \to A$ defined at the end of Section 3 to the order of the cokernel of the induced map on the character groups of tori defined in Section 4.1. Let K be as in Section 2, and let A be an abelian variety over K.

Definition 5.1 (Symmetric isogeny). A *symmetric isogeny* $\varphi : A^\vee \to A$ is an isogeny such that the map

$$\varphi^\vee : A^\vee \to (A^\vee)^\vee = A$$

is equal to φ.

If J and A are as in Section 3 then the principal polarization θ_J of J is symmetric, so the natural map $A^\vee \to A$ is a symmetric isogeny.

Lemma 5.2. *Suppose that A is a purely toric abelian variety over K and that $\varphi : A^\vee \to A$ is a symmetric isogeny. Let $\varphi_a : X_A \to X_{A^\vee}$ denote the induced map on character groups. Then*

$$\deg(\varphi) = \#\,\mathrm{coker}(\varphi_a)^2.$$

Proof. By Corollary 8.7 applied to our isogeny φ (so what we are presently calling $A^\vee$ and A are respectively called A and B in the discussion surrounding Theorem 8.6), we deduce that

$$\deg(\varphi) = \#\ker(\varphi) = \#\ker(\varphi_t) \cdot \#\ker(\varphi_t^\vee)$$

where φ_t and $\varphi_t^\vee$ are the maps induced by φ and $\varphi^\vee$ on closed fiber tori.

Since the character group X_A is, by definition, $\mathrm{Hom}_{\overline{k}}(\mathcal{T}_{\overline{k}}, \mathbf{G}_{m\overline{k}})$, where $\mathcal{T}$ is the toric part of the closed fiber of A, it follows that $\#\ker(\varphi_t) = \#\,\mathrm{coker}(\varphi_a)$. Since $\varphi = \varphi^\vee$, this proves the lemma. □

6 Statement and Proof of the Main Theorem

Let K be as in Section 2, and let $\pi : J \to A$ be an optimal quotient. Assume that J is equipped with a symmetric principal polarization λ, that J has semistable reduction, and that A has purely toric reduction. Let X_A, $X_{A^\vee}$, and X_J denote the character groups of the toric parts of the closed fibers of the abelian varieties A, $A^\vee$, and J, respectively.

Let $\pi : J \to A$ be an optimal quotient, and let $\theta : A^\vee \to A$ denote the induced polarization. Let π_*, π^*, θ_*, and θ^* be the maps induced on character groups by the various functorialities, as indicated in the following two key diagrams:

$$\begin{tikzcd}
A^\vee \arrow[r, hook, "\pi^\vee"] \arrow[dr, "\theta"'] & J \arrow[d, "\pi"] \\
& A
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
X_A \arrow[r, hook, "\pi^*"] \arrow[dr, "\theta^*"] & X_J \arrow[d, "\pi_*"] \\
& X_{A^\vee}. \arrow[ul, bend left, "\theta_*"]
\end{tikzcd}$$

The surjectivity of π_* is proved in Theorem 8.2. The injectivity of π^* follows because

$$\theta_*\pi_*\pi^* = \theta_*\theta^* = \deg(\theta) \neq 0,$$

and multiplication by a nonzero integer on a free abelian group is injective.

Let $\mathcal{L}$ be the *saturation* of π^*X_A in X_J; thus π^*X_A is a finite-index subgroup of $\mathcal{L}$ and the quotient $X_J/\mathcal{L}$ is torsion free. Let

$$\alpha : X_J \to \mathrm{Hom}(\pi^*X_A, \mathbf{Z})$$

be the map defined by the monodromy pairing restricted to $X_J \times \pi^*X_A$. For L of finite index in $\mathcal{L}$, define the *degree* of L to be

$$m_L = [\alpha(X_J) : \alpha(L)],$$

and the *component group* of L to be

$$\Phi_L = \mathrm{coker}(X_J \to \mathrm{Hom}(L, \mathbf{Z})).$$

When $L = \mathcal{L}$ and A is fixed, for simplicity we write $m_X = m_{\mathcal{L}}$ and $\Phi_X = \Phi_{\mathcal{L}}$.

Recall that Φ_A is the component group of A and m_A is the square root of the degree of the induced map $A^\vee \to A$.

Theorem 6.1. *For any subgroup L of finite index in $\mathcal{L}$, the following relation holds:*

$$\frac{\#\Phi_A}{m_A} = \frac{\#\Phi_L}{m_L}.$$

6.1 Proof of the Main Theorem

The notation in this section is as in previous section.

Lemma 6.2. *Let $\pi_* : X_J \to X_{A^\vee}$ and $\alpha : X_J \to \mathrm{Hom}(\pi^*X_A, \mathbf{Z})$ be as in previous section. Then*

$$\ker(\pi_*) = \ker(\alpha).$$

Proof. Suppose $x \in \ker(\pi_*)$, and let $y = \pi^* z$ with $z \in X_A$. Then

$$\langle x, y\rangle = \langle x, \pi^* z\rangle = \langle \pi_* x, z\rangle = 0,$$

so $x \in \ker(\alpha)$. Next let $x \in \ker(\alpha)$. Then for all $z \in X_A$,

$$0 = \langle x, \pi^* z\rangle = \langle \pi_* x, z\rangle,$$

so $\pi_* x$ is in the kernel of the monodromy map

$$X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}).$$

Since $X_{A^\vee}$ and $\mathrm{Hom}(X_A, \mathbf{Z})$ are free of the same finite rank and the cokernel is torsion, the monodromy map is injective. Thus $\pi_* x = 0$ and $x \in \ker(\pi_*)$. □

Let $\pi^* : X_A \to X_J$ be as in previous section.

Lemma 6.3. *The monodromy-pairing map* $X_J \to \mathrm{Hom}(X_J, \mathbf{Z})$ *composed with restriction* $\mathrm{Hom}(X_J, \mathbf{Z}) \to \mathrm{Hom}(\pi^* X_A, \mathbf{Z})$ *gives rise to an exact sequence*

$$X_J \to \mathrm{Hom}(\pi^* X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

Proof. Lemma 6.2 gives the following commutative diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & X_J/\ker(\alpha) \arrow[r] \arrow[d, "\cong"] & \mathrm{Hom}(\pi^* X_A, \mathbf{Z}) \arrow[r] \arrow[d, "\cong"] & \mathrm{coker}(\alpha) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & X_{A^\vee} \arrow[r] & \mathrm{Hom}(X_A, \mathbf{Z}) \arrow[r] & \Phi_A \arrow[r] & 0.
\end{tikzcd}$$

By Lemma 6.2, the first vertical map is an isomorphism. The second is an isomorphism because it is induced by the isomorphism $\pi^* : X_A \to \pi^* X_A$. It follows that $\mathrm{coker}(\alpha) \cong \Phi_A$, as claimed. □

Recall that $\mathcal{L}$ denotes the saturation of $\pi^* X_A$ in X_J, and that $L \subset \mathcal{L}$ denotes a subgroup of finite index.

Lemma 6.4. *The rational number* $\dfrac{\#\Phi_L}{m_L}$ *is independent of the choice of* L.

Proof. Suppose L' is another finite index subgroup of $\mathcal{L}$, and let $n = [L : L']$. Here n is a rational number, the lattice index of L' in L. Since α is injective when restricted to $\mathcal{L}$, it follows that

$$m_{L'} = [\alpha(X_J) : \alpha(L')] = [\alpha(X_J) : \alpha(L)] \cdot [\alpha(L) : \alpha(L')] = m_L \cdot n.$$

Similarly, $\#\Phi_{L'} = \#\Phi_L \cdot n$. □

Recall that $m_A = \sqrt{\deg(\theta)}$ and

$$\Phi_A \cong \mathrm{coker}(X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z})),$$

where m_A is the degree of A and Φ_A is the component group of A.

Proof of Theorem 6.1. By Lemma 6.4 we may assume that $L = \pi^* X_A$. With this choice of L, Lemma 6.3 asserts that $\Phi_L \cong \Phi_A$. By Lemma 6.2, properties of the index, and Lemma 5.2 we have

$$\begin{aligned} m_L &= [\alpha(X_J) : \alpha(L)] \\ &= [\pi_*(X_J) : \pi_*(L)] \\ &= [X_{A^\vee} : \pi_*(\pi^* X_A)] \\ &= [X_{A^\vee} : \theta^* X_A] \\ &= \#\operatorname{coker}(\theta^*) \\ &= \sqrt{\deg(\theta)} = m_A. \end{aligned}$$

□

Recall that $\Phi_{\mathcal{L}}$ denotes the cokernel of the natural map $X_J \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})$ induced by composing the monodromy map $X_J \to \operatorname{Hom}(X_J, \mathbf{Z})$ with the natural restriction map $\operatorname{Hom}(X_J, \mathbf{Z}) \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})$.

Proposition 6.5. *The group $\Phi_{\mathcal{L}}$ is canonically isomorphic to the image of the map from Φ_J to Φ_A induced by $\pi : J \to A$. Thus*

$$\operatorname{image}(\Phi_J \to \Phi_A) \cong \Phi_{\mathcal{L}}.$$

Proof. Since $\pi^* X_A \subset \mathcal{L} \subset X_J$, an application of Lemma 6.3 gives the following commutative diagram with exact rows:

$$\begin{tikzcd}
X_J \arrow[r] \arrow[d, equal] & \operatorname{Hom}(X_J, \mathbf{Z}) \arrow[r] \arrow[d] & \Phi_J \arrow[r] \arrow[d] & 0 \\
X_J \arrow[r] \arrow[d, equal] & \operatorname{Hom}(\mathcal{L}, \mathbf{Z}) \arrow[r] \arrow[d] & \Phi_{\mathcal{L}} \arrow[r] \arrow[d] & 0 \\
X_J \arrow[r] & \operatorname{Hom}(\pi^* X_A, \mathbf{Z}) \arrow[r] & \Phi_A \arrow[r] & 0.
\end{tikzcd}$$

The map $\operatorname{Hom}(\mathcal{L}, \mathbf{Z}) \to \operatorname{Hom}(\pi^* X_A, \mathbf{Z})$ is an isomorphism, so the map $\Phi_{\mathcal{L}} \to \Phi_A$ is injective. Thus

$$\operatorname{image}(\Phi_J \to \Phi_A) \cong \operatorname{image}(\Phi_J \to \Phi_{\mathcal{L}}).$$

The cokernel of $\operatorname{Hom}(X_J, \mathbf{Z}) \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})$ surjects onto the cokernel of $\Phi_J \to \Phi_{\mathcal{L}}$. Using the exact sequence

$$0 \to \mathcal{L} \to X_J \to X_J/\mathcal{L} \to 0,$$

we find that

$$\operatorname{coker}(\operatorname{Hom}(X_J, \mathbf{Z}) \to \operatorname{Hom}(\mathcal{L}, \mathbf{Z})) \subset \operatorname{Ext}^1(X_J/\mathcal{L}, \mathbf{Z}).$$

Because $\mathcal{L}$ is saturated, the quotient $X_J/\mathcal{L}$ is torsion free, so the indicated Ext^1 group vanishes. Thus the map $\Phi_J \to \Phi_{\mathcal{L}}$ is surjective, from which the proposition follows. □

Corollary 6.6. *The cokernel of the map from Φ_J to Φ_A induced by $\pi : J \to A$ has order $m_A/m_{\mathcal{L}}$. Thus*

$$\#\operatorname{coker}(\Phi_J \to \Phi_A) = \frac{m_A}{m_{\mathcal{L}}}.$$

Proof. Combine Theorem 6.1 and Proposition 6.5. □

7 Optimal Quotients of $J_0(N)$

In this section we specialize the general results of the rest of this paper to the concrete case in which $J = J_0(N)$ is the Jacobian of a modular curve, and $A = A_f$ is an optimal quotient of J attached to a modular forms. The paper [12] contains more computations like these.

7.1 Modular Curves and Semistable Reduction

Let $X_0(N)$ be the modular curve associated to the subgroup $\Gamma_0(N)$ of $\mathrm{SL}_2(\mathbf{Z})$ that consists of those matrices which are upper triangular modulo N. The algebraic curve $X_0(N)_{\mathbf{C}}$ can be constructed as a Riemann surface as the quotient

$$\Gamma_0(N)\backslash\left(\{z : z \in \mathbf{C}, \operatorname{Im}(z) > 0\} \cup \mathbf{P}^1(\mathbf{Q})\right),$$

and $X_0(N)$ has a canonical structure of algebraic curve over $\mathbf{Q}$.

It is well known that the p-new part of the Jacobian $J_0(N)$ of $X_0(N)$ has purely toric reduction at p when $p \mid\mid N$. Let us briefly recall the reason, writing $N = Mp$. Using the description of closed fibers of modular curves [10, Ch. 13] and Raynaud's result relating Néron models and Picard functors (as summarized in [2, Ch. 9]), the standard finite *flat* degeneracy maps $X_0(Mp) \to X_0(M)$ over $\mathbf{Z}_{(p)}$ induce a *"pushfoward"* map on Néron model connected components

$$\mathrm{Pic}^0_{X_0(Mp)/\mathbf{Z}_{(p)}} \longrightarrow \mathrm{Pic}^0_{X_0(M)/\mathbf{Z}_{(p)}} \times \mathrm{Pic}^0_{X_0(M)/\mathbf{Z}_{(p)}}$$

which on the closed fiber is the map induced by *pullback* to the two components $X_0(M)_{/\mathbf{F}_p}$ in $X_0(Mp)_{/\mathbf{F}_p}$. The kernel of this latter map is a torus [2, Ex. 9.2.8], yet this kernel is visibly isogenous to the semistable mod p fiber of the dual of $J_0(Mp)^{\mathrm{new}}$, whence the purely toric conclusion.

7.2 Newforms and Optimal Quotients

The Hecke algebra

$$\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots] \subset \mathrm{End}(J_0(N))$$

is a commutative ring of endomorphisms of $J_0(N)$ of $\mathbf{Z}$-rank equal to the dimension of $J_0(N)$. The character group $X_{J,p}$ of $J_0(N)$ at p is equipped with a functorial action of $\mathbf{T}$. The Hecke algebra $\mathbf{T}$ also acts on the complex vector space $S = S_2(\Gamma_0(N), \mathbf{C})$ of cusp forms.

Let f be a newform, and associate to f the ideal I_f of the Hecke algebra $\mathbf{T}$ of elements which annihilate f. Then $\mathcal{O}_f = \mathbf{T}/I_f$ is an order in the ring of integers of the totally real number field K_f obtained by adjoining the Fourier coefficients of f to $\mathbf{Q}$. The quotient

$$A_f = J_0(N)/I_f J_0(N)$$

is an optimal quotient of $J_0(N)$ of dimension equal to $[K_f : \mathbf{Q}]$. As discussed in the previous section, A_f is purely toric at p.

7.3 Tamagawa Numbers

Let $\mathrm{Frob}_p : X_J \to X_J$ denote the map induced by the Frobenius automorphism. We have $\mathrm{Frob}_p = -W_p$, where W_p is the map induced by the Atkin-Lehner involution on $J_0(p)$. Let f be a newform, $A = A_f$ the corresponding optimal quotient, and w_p the sign of the eigenvalue of W_p on f.

Proposition 7.1.

$$\Phi_A(\mathbf{F}_p) = \begin{cases} \Phi_A(\overline{\mathbf{F}}_p) & \textit{if } w_p = -1, \\ \Phi_A(\overline{\mathbf{F}}_p)[2] & \textit{if } w_p = 1. \end{cases}$$

Proof. If $w_p = -1$, then $\mathrm{Frob}_p = 1$ and the $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$-action of $\Phi_A(\overline{\mathbf{F}}_p)$ is trivial. In this case $\Phi(\mathbf{F}_p) = \Phi(\overline{\mathbf{F}}_p)$. Next suppose $w_p = 1$. Recall that we have an exact sequence

$$0 \to X_{A^\vee} \to \mathrm{Hom}(X_A, \mathbf{Z}) \to \Phi_A \to 0.$$

Since W_p acts as $+1$ on f, it also acts as $+1$ on each of the modules A, X_A, $\mathrm{Hom}(X_A, \mathbf{Z})$, and Φ_A. Thus $\mathrm{Frob}_p = -W_p$ acts as -1 on Φ_A. Since the subgroup of 2-torsion elements of a finite abelian group equals the subgroup of elements fixed under -1, it follows that $\Phi_A(\mathbf{F}_p) = \Phi_A(\overline{\mathbf{F}}_p)[2]$. □

Warning: When extending this result to the whole of $J_0(N)$, be careful. The action of $\mathrm{Frob}_p = T_p$ need not be by ± 1, even though it must be by an involution of order 2. For example, the component group of $J_0(65)$ at 5 is cyclic of order 42. The action of Frob_5 is by multiplication by -13. Note that $(-13)^2 = 1 \pmod{42}$. The fixed points of multiplication by -13 is the order 14 subgroup of $\mathbf{Z}/42\mathbf{Z}$.

7.4 Computing Component Groups

Using modular symbols, we can enumerate the optimal quotients A_f of $J_0(N)$ (see, e.g., [1]) and compute the degree m_A (see [12, §3.1]). Suppose p is a prime that exactly divides N. As explained in [12], the method of graphs (see [14]) or the ideal theory of quaternion algebras (see [11]) can be used to compute $X = X_{J_0(N),p}$ with its $\mathbf{T}$-action and the monodromy pairing. We can then compute the following three modules:

1. the saturated submodule $\mathcal{L} = \bigcap_{t \in I_f} \ker(t)$ of X,
2. the character group degree $m_X = m_{\mathcal{L}}$, and
3. $\Phi_X = \Phi_{\mathcal{L}}$.

By Theorem 6.1 we obtain

$$\#\Phi_{A,p} = \#\Phi_X \cdot \frac{m_A}{m_X}.$$

7.5 The Eisenstein Nature of Component Groups

The theorem below, which generalizes some of the results of [13] and [15], was conjectured by the second author after computing many component groups of quotients of $J_0(p)$ using the results of this paper. M. Emerton read an early version of this paper and subsequently announced a proof of the theorem below (see [6]).

Theorem 7.2 (Emerton). *Let p be a prime and let $f_1, \ldots, f_n$ be a set of representatives for the Galois-conjugacy classes of newforms in $S_2(\Gamma_0(p))$. Let $A_1, \ldots, A_n$ be the optimal quotients associated to $f_1, \ldots, f_n$, respectively. Then for each i, $i = 1, \ldots, n$, we have*

$$\#A_i(\mathbf{Q})_{\mathrm{tor}} = \#\Phi_{A_i}(\overline{\mathbf{F}}_p) = \#\Phi_{A_i}(\mathbf{F}_p).$$

Furthermore,

$$\#\Phi_{J_0(p)}(\overline{\mathbf{F}}_p) = \prod_{i=1}^{n} \#\Phi_{A_i}(\overline{\mathbf{F}}_p).$$

Before Emerton proved the above assertion, the second author verified it using the algorithm of this paper for all $p \leq 757$, and, up to a power of 2, for all $p < 2000$.

Remark 7.3. It is tempting to guess that, e.g., the natural map

$$\Phi_{J_0(113)}(\overline{\mathbf{F}}_{113}) \to \prod_{i=1}^{4} \Phi_{A_i}(\overline{\mathbf{F}}_{113})$$

is an isomorphism, but this is incorrect. Two of the $\Phi_{A_i}(\overline{\mathbf{F}}_{113})$ have order 2, so the product $\prod_{i=1}^{4} \Phi_{A_i}(\overline{\mathbf{F}}_{113})$ is not a cyclic group. However, Mazur proved that the groups $\Phi_{J_0(p)}(\overline{\mathbf{F}}_p)$ are cyclic for all primes p.

7.6 Examples

In this section we give some examples of the numbers involved in computing component groups of quotients of $J_0(N)$. For more examples, see [12]. We use the notation for abelian varieties that is described in [1]. For example **65A** is the "first" abelian variety quotient of $J_0(65)$ attached to a newform.

7.6.1 Quotients of $J_0(N)$

Table 1 contains many of the quantities involved in the computation of component groups for each of the newform optimal quotients for $N \in \{65, 66, 68, 69\}$.

7.6.2 Quotients of $J_0(p)^-$

We computed the quantities m_A, m_X, and Φ_X for each abelian variety A_f associated to a newform of prime level p with $p \leq 631$. Table 2 lists those A_f for which $w_p = -1$, along with the order of the corresponding component group. The first column, which is labeled "A" contains a description of A_f, the second column, labeled "d", contains the dimension of A_f, and the third column, labeled "$\#\Phi_A$", contains the order $\#\Phi_{A_f,p}(\overline{\mathbf{F}}_p)$ of the component group.

Remark 7.4. Theorem 7.2 together with [13, Prop. II.17.10] imply that the component groups of the A_f for which $w_p = +1$ are trivial, so we do not list them. An optimal quotient A_f of $J_0(p)$ with nonzero component group has nonzero rational torsion (by Theorem 7.2), so it factors through the Eisenstein quotient of $J_0(p)$. Also w_p acts as -1 on the Eisenstein quotient of $J_0(p)$, which is [13, Prop. II.17.10], and which is a deep result because of subtleties at the prime 2 (see the discussion in [13, III.1]).

Table 1: Component groups of quotients of $J_0(N)$

A	dim	p	w_p	$\#\Phi_X$	m_X	m_A	$\#\Phi_A$
65A	1	5	$+$	1	2	2	1
		13	$+$	1	2		1
65B	2	5	$+$	3	2^2	2^2	3
		13	$-$	3	2^2		3
65C	2	5	$-$	7	2^2	2^2	7
		13	$+$	1	2^2		1
66A	1	2	$+$	1	2	2^2	2
		3	$-$	3	2^2		3
		11	$+$	1	2^2		1
66B	1	2	$-$	2	2	2^2	2^2
		3	$+$	1	2^2		1
		11	$+$	1	2^2		1
66C	1	2	$-$	1	2	$2^2 \cdot 5$	$2 \cdot 5$
		3	$-$	1	2^2		5
		11	$-$	1	$2^2 \cdot 5$		1
68A	2	17	$+$	2	$2 \cdot 3$	$2 \cdot 3$	2
69A	1	3	$-$	2	2	2	2
		23	$+$	1	2		1
69B	2	3	$+$	2	2	$2 \cdot 11$	$2 \cdot 11$
		23	$-$	2	$2 \cdot 11$		2

Table 2: Component groups of quotients of $J_0(p)^-$

A	d	$\#\Phi_A$	A	d	$\#\Phi_A$	A	d	$\#\Phi_A$	A	d	$\#\Phi_A$
11A	1	5	**157B**	7	13	**313A**	2	1	**487B**	2	3
17A	1	2^2	**163C**	7	3^3	**313C**	12	$2 \cdot 13$	**487C**	3	1
19A	1	3	**167B**	12	83	**317B**	15	79	**487D**	16	3^3
23A	2	11	**173B**	10	43	**331D**	16	$5 \cdot 11$	**491C**	29	$5 \cdot 7^2$
29A	2	7	**179A**	1	1	**337B**	15	$2^2 \cdot 7$	**499C**	23	83
31A	2	5	**179C**	11	89	**347D**	19	173	**503B**	1	1
37B	1	3	**181B**	9	$3 \cdot 5$	**349B**	17	29	**503C**	1	1
41A	3	$2 \cdot 5$	**191B**	14	$5 \cdot 19$	**353A**	1	2	**503D**	3	1
43B	2	7	**193C**	8	2^4	**353B**	3	2	**503F**	26	251
47A	4	23	**197C**	10	7^2	**353D**	14	$2 \cdot 11$	**509B**	28	127
53B	3	13	**199A**	2	1	**359D**	24	179	**521B**	29	$2 \cdot 5 \cdot 13$
59A	5	29	**199C**	10	$3 \cdot 11$	**367B**	19	61	**523C**	26	$3 \cdot 29$
61B	3	5	**211A**	2	5	**373C**	17	31	**541B**	24	$3^2 \cdot 5$
67A	1	1	**211D**	9	7	**379B**	18	$3^2 \cdot 7$	**547C**	25	$7 \cdot 13$
67C	2	11	**223C**	12	37	**383C**	24	191	**557B**	1	1
71A	3	5	**227B**	2	1	**389A**	1	1	**557D**	26	139
71B	3	7	**227C**	2	1	**389E**	20	97	**563A**	1	1
73A	1	2	**227E**	10	113	**397B**	2	1	**563E**	31	281
73C	2	3	**229C**	11	19	**397C**	5	11	**569B**	31	$2 \cdot 71$
79B	5	13	**233A**	1	2	**397D**	10	3	**571A**	1	1
83B	6	41	**233C**	11	29	**401B**	21	$2^2 \cdot 5^2$	**571B**	1	1
89B	1	2	**239B**	17	$7 \cdot 17$	**409B**	20	$2 \cdot 17$	**571C**	2	1
89C	5	11	**241B**	12	$2^2 \cdot 5$	**419B**	26	$11 \cdot 19$	**571D**	2	1
97B	4	2^3	**251B**	17	5^3	**421B**	19	$5 \cdot 7$	**571F**	4	1
101B	7	5^2	**257B**	14	2^6	**431B**	1	1	**571I**	18	$5 \cdot 19$
103B	6	17	**263B**	17	131	**431D**	3	1	**577A**	2	3
107B	7	53	**269C**	16	67	**431F**	24	$5 \cdot 43$	**577B**	2	1
109A	1	1	**271B**	16	$3^2 \cdot 5$	**433A**	1	1	**577C**	3	1
109C	4	3^2	**277B**	3	1	**433B**	3	1	**577D**	18	2^4
113A	1	2	**277D**	9	23	**433D**	16	$2^2 \cdot 3^2$	**587C**	31	293
113B	2	2	**281B**	16	$2 \cdot 5 \cdot 7$	**439C**	25	73	**593B**	1	2
113D	3	7	**283B**	14	47	**443C**	1	1	**593C**	2	1
127B	7	$3 \cdot 7$	**293B**	16	73	**443E**	22	$13 \cdot 17$	**593E**	27	$2 \cdot 37$
131B	10	$5 \cdot 13$	**307A**	1	1	**449B**	23	$2^4 \cdot 7$	**599C**	37	$13 \cdot 23$
137B	7	$2 \cdot 17$	**307B**	1	1	**457C**	20	$2 \cdot 19$	**601B**	29	$2 \cdot 5^2$
139A	1	1	**307C**	1	1	**461D**	26	$5 \cdot 23$	**607D**	31	101
139C	7	23	**307D**	1	1	**463B**	22	$7 \cdot 11$	**613C**	27	$3 \cdot 17$
149B	9	37	**307E**	2	3	**467C**	26	233	**617B**	28	$2 \cdot 7 \cdot 11$
151B	3	1	**307F**	9	17	**479B**	32	239	**619B**	30	103
151C	6	5^2	**311B**	22	$5 \cdot 31$	**487A**	2	1	**631B**	32	$3 \cdot 5 \cdot 7$

8 Appendix: Some Facts Concerning Toric Reduction

Let R be a discrete valuation ring with fraction field K and residue field k. For any abelian variety A over K, with Néron model $\mathcal{A}$ over R, we denote by X_A the character group of the toric part of $\mathcal{A}_k^0$ (the connected component of the closed fiber of $\mathcal{A}$). All group schemes below are understood to be commutative.

Our aim in this appendix is to prove a couple of facts (Theorem 8.2 and Theorem 8.6) which are no doubt well-known to experts but for which published proofs do not appear to be readily available. We begin with a simple and basic lemma.

Lemma 8.1. *Let $f : G \to G'$ be a map between multiplicative (resp. étale) finite flat group schemes over R. The map f is a closed immersion (resp. faithfully flat) if and only if the generic fiber map f_K is a closed immersion (resp. faithfully flat).*

Proof. Cartier duality interchanges étaleness and multiplicativeness, as well as closed immersions and faithfully flat maps (as the latter two properties may be checked on the closed fiber, for which one is reduced to the standard case of finite commutative group schemes over a field). Thus, it suffices to consider the étale case. By faithfully flat base change to a strict henselization of R, we are reduced to the case where our finite étale group schemes are constant. Since faithful flatness is equivalent to surjectivity (for maps between étale schemes over a base), the lemma is now physically clear.

□

Now we turn to the first of the two main results we want to prove. Let $\pi : J \to A$ be an optimal quotient of abelian varieties over K (i.e., we assume that $\ker \pi$ is an abelian variety over K), and assume that J has semistable reduction over R (so A does too). We do not yet make any hypotheses of purely toric reduction. The dual abelian varieties $A^\vee$ and $J^\vee$ again have semistable reduction, as they are isogenous to A and J respectively.

Theorem 8.2. *With notation as above, the map $X_{J^\vee} \to X_{A^\vee}$ induced by π is surjective.*

Proof. The underlying idea comes down to two facts: Lemma 8.1 and the fact that we can lift tori on the level of ℓ-divisible groups for any prime ℓ. More precisely, we argue as follows. By Proposition 3.3, the map $\pi^\vee : A^\vee \to J^\vee$ is a closed immersion of abelian varieties. We will use this to prove that the induced map $\pi_t^\vee$ on closed fiber tori of Néron models is a closed immersion. Since the "character group" functor sets up an anti-equivalence of categories between tori over a field F and finite free $\mathbf{Z}$-modules with continuous action of $\mathrm{Gal}(F_s/F)$, identifying closed immersions of tori with surjections of character groups and surjections of tori with "saturated injections" of character groups (i.e., injections with torsion-free cokernel), the closed immersion property for $\pi_t^\vee$ on the closed fiber tori will yield the desired surjection of character groups.

In general the "Néron model" functor doesn't behave well for closed immersions. That is, just because $\pi^\vee$ is a closed immersion, it does not follow purely formally that $\pi^\vee$ induces a closed immersion on Néron models. Nevertheless, we claim quite generally that if $B \to B'$ is a closed immersion of abelian varieties over K with semistable Néron models, then the induced map $T \to T'$ on closed fiber tori is a closed immersion. For this it is sufficient to prove that the induced map on ℓ-divisible groups $T[\ell^\infty] \to T'[\ell^\infty]$ is a closed immersion for *all* primes ℓ (i.e., all maps $T[\ell^n] \to T'[\ell^n]$ are closed immersions). Indeed, suppose we verify this closed immersion property on torsion, and let H be the kernel of $T \to T'$, so $H[\ell^n] = 0$ for all primes ℓ and positive integers n. The torus $(H^0_{/\overline{k}})_{\mathrm{red}}$ must vanish (as it has no non-trivial torsion) and hence H is finite. If N is the order of H, then $H = H[N] = 0$. The map $T \to T'$ is then a monomorphism between algebraic groups over a field and hence is a closed immersion, as desired.

In order to verify that the ℓ-divisible group maps $T[\ell^\infty] \to T'[\ell^\infty]$ are closed immersions for all ℓ, we can make the faithfully flat base change to the henselization of R (which commutes with formation of Néron models) to reduce to the case where R is henselian. Now we recall the following basic result of Grothendieck:

Lemma 8.3. *Let R be a henselian local ring, G a quasi-finite separated R-scheme of finite presentation. There is a unique decomposition*

$$G = G_f \coprod G'$$

into disjoint clopen pieces with G_f finite over R (called the "finite part"of G) and G' having empty closed fiber. The formation of G_f is functorial in G and is compatible with henselian local base change and formation of fiber products over R.

If moreover G is a group scheme over R, then G_f is a clopen subgroup scheme and there exists a unique multiplicative closed R-subgroup scheme G_μ inside of G whose closed fiber is the multiplicative part of the closed fiber of G (G_μ is called the "multiplicative part" of G). The formation of G_μ is functorial in G.

Proof. For the first part, see [8, IV_4, 18.5.11(c)] (aside from the functorial properties, which are obvious). The second part, concerning group schemes, is a mechanical consequence of the first part (including the functoriality of the finite part). For example, the existence of G_μ follows from considering the connected-étale sequence of the Cartier dual of G_f over the henselian local base R, and the uniqueness and functoriality follows from the functoriality of $G \rightsquigarrow G_f$ and the functoriality of the connected-étale sequence.

□

Remark 8.4. Assuming R in Lemma 8.3 is a discrete valuation ring (with fraction field K and residue field k), let us make some observations concerning the behavior of Lemma 8.3 with respect to primary components, as this will be useful later. Let's suppose that N and M are relatively prime integers with NM divisible by the order of G_K, and hence killing G. Thus, by functoriality we have $G = G[N] \times_R G[M]$ where $G[N]$ and $G[M]$ are quasi-finite separated R-group schemes. We claim that $G[N]$ and $G[M]$ are also *flat* over R, whence it follows that the formation of G_f and G_μ is compatible with passage to "primary components".

In other words, if ℓ is a prime and ℓ^n is divisible by the ℓ-part of the order of G_K, then we claim that $G[\ell^n]$ is R-flat. From the clopen decomposition $G = G_f \coprod G'$, it is easy to see that $G[\ell^n] = G_f[\ell^n] \coprod X_n$ for some finite K-scheme X_n, so for the issue of R-flatness we can replace G with G_f. We are thereby reduced to the finite flat case, so we can use the proof of [10, 1.7.2].

The significance of Lemma 8.3 for our purposes is the following standard consequence.

Corollary 8.5. *Let A be an abelian variety over the fraction field K of a henselian discrete valuation ring R with residue field k. Let $\mathcal{A}$ be the Néron model of A, and assume that $\mathcal{A}$ has semistable reduction. For every prime ℓ, there exists a unique multiplicative ℓ-divisible group Γ_ℓ inside of $\mathcal{A}$ whose closed fiber is the ℓ-divisible group of the torus T of $\mathcal{A}_k^0$. The formation of Γ_ℓ is functorial in A.*

Proof. Fix ℓ. By the semistability hypothesis, the multiplication maps $\ell^n : \mathcal{A} \to \mathcal{A}$ are quasi-finite flat, so $\mathcal{A}[\ell^n]$ is a quasi-finite flat separated R-group scheme. Let $\mathcal{A}[\ell^n]_\mu$ denote its multiplicative part (as in Lemma 8.3), so the multiplicative $T[\ell^n] \hookrightarrow \mathcal{A}[\ell^n]_k$ lies inside of $(\mathcal{A}[\ell^n]_\mu)_k$. The "closed fiber" functor is an equivalence of categories between finite flat multiplicative group schemes over R and k (since Cartier duality reduces this to the étale case, and the "closed fiber" functor is an equivalence of categories between finite étale R-schemes and finite étale k-schemes [8, IV$_4$, 18.5.12]). Thus, there exists a unique multiplicative closed R-subgroup scheme $\Gamma^{(n)} \hookrightarrow \mathcal{A}[\ell^n]_\mu$ whose closed fiber is $T[\ell^n]$.

Moreover, using the equivalence of categories just mentioned, $\mathcal{A}[\ell^n]_\mu$ lies inside of $\mathcal{A}[\ell^{n+1}]_\mu$ and $\Gamma^{(n)}$ lies inside of $\Gamma^{(n+1)}$. The resulting system $\Gamma_\ell = \{\Gamma^{(n)}\}$ over R forms an ℓ-divisible group on the closed fiber and hence is an ℓ-divisible group over R. This settles the desired existence, as well as the desired uniqueness. The functoriality of Γ_ℓ in A follows from the functoriality of toric parts on the closed fiber of Néron models.

□

Returning to the proof of Theorem 8.2, recall that we were studying the map of toric parts $j_t : T \to T'$ induced by a closed immersion $j : B \hookrightarrow B'$ of semistable abelian varieties over K, with R henselian. We wanted the map

$$j_t[\ell^\infty] : T[\ell^\infty] \to T'[\ell^\infty]$$

to be a closed immersion for all primes ℓ (as we have seen that this forces $T \to T'$ to be a closed immersion, which is what we really want to show). Fix ℓ. By Corollary 8.5 there exist unique multiplicative ℓ-divisible groups Γ and Γ' over R in the respective Néron models $\mathcal{B}$ and $\mathcal{B}'$ such that Γ and Γ' respectively lift the ℓ-divisible groups of the tori of the closed fibers. Hence, it suffices to show that the R-map $\gamma : \Gamma \to \Gamma'$ induced by the Néron functoriality map $\mathrm{N}(j)$ is a closed immersion. The generic fiber map γ_K is a closed immersion since it "sits inside" the generic fiber ℓ-divisible groups of B and B', the map between which is a closed immersion since $j : B \to B'$ is a closed immersion. Now we use Lemma 8.1 (applied at all finite torsion levels) to conclude that γ is a closed immersion. This completes the proof of Theorem 8.2.

□

We now turn to a result which requires a stronger hypothesis on the closed fiber. Note that we retain the hypothesis that R is henselian (this hypothesis arose in the proof of Theorem 8.2, even though it wasn't needed for the statement). Let A and B be abelian varieties over K with purely toric reduction (i.e., their Néron models have closed fiber connected components which are tori). Let $\varphi : A \to B$ be an isogeny, and let $\varphi_t : T_A \to T_B$ be the induced map on the closed fiber toric parts (i.e., connected components) of the Néron models. We denote by $\varphi_t^\vee : T_{B^\vee} \to T_{A^\vee}$ the analogous map induced by the dual isogeny $\varphi^\vee$. Since the map φ_t is an isogeny (by functoriality), the kernel $\ker(\varphi_t)$ is a finite multiplicative k-group scheme.

For any finite multiplicative k-group scheme G, we let $\widetilde{G}$ denote the (unique) multiplicative finite flat R-group scheme with closed fiber G. For example, $\widetilde{\ker(\varphi_t)}$ is a multiplicative R-group scheme which lies inside of

$$\ker(\mathrm{N}(\varphi))_\mu$$

(where $\mathrm{N}(\varphi)$ is the map induced by Néron functoriality). Thus, we have a natural closed immersion

$$(\widetilde{\ker\varphi_t})_K \hookrightarrow \ker\varphi$$

and likewise we have a natural quotient map

$$\ker(\varphi^\vee)^\vee \to \widetilde{\ker(\varphi_t^\vee)}^\vee_K$$

dual to the natural closed immersion using the isogeny $\varphi^\vee$.

By the duality theory for abelian varieties (particularly the adjointness of φ and $\varphi^\vee$ with respect to the scheme-theoretic Weil pairing over K), there is a canonical perfect duality K-group scheme duality between $\ker(\varphi)$ and $\ker(\varphi^\vee)$ over K, whence there is a natural quotient map of K-group schemes

$$\ker(\varphi) \simeq \ker(\varphi^\vee)^\vee \to \widetilde{\ker(\varphi_t^\vee)}_K^\vee.$$

Theorem 8.6. *The diagram of K-group schemes*

$$0 \to \widetilde{\ker(\varphi_t)}_K \to \ker(\varphi) \to \widetilde{\ker(\varphi_t^\vee)}_K^\vee \to 0$$

is exact.

The content of the proof is the Grothendieck Orthogonality Theorem. Moreover, Theorem 8.6 is implicit in Grothendieck's construction of the monodromy pairing for semiabelian varieties.

Proof. The exact sequence of the theorem says that the finite flat K-group schemes p

$$\ker(\varphi)/\widetilde{\ker(\varphi_t)}_K \quad \text{and} \quad \widetilde{\ker(\varphi_t^\vee)}_K$$

are canonically Cartier dual to each other compatibly with the perfect duality between $\ker(\varphi)$ and $\ker(\varphi^\vee)$. More precisely, let $\mathcal{A}$ and $\mathcal{B}$ denote the Néron models of A and B, respectively, let

$$G = \ker(\mathcal{A} \to \mathcal{B}), \quad G^\vee = \ker(\mathcal{B}^\vee \to \mathcal{A}^\vee),$$

so G and $G^\vee$ are both quasi-finite flat separated R-group schemes whose generic fibers are the $\ker(\varphi)$ and $\ker(\varphi^\vee)$ in the theorem (the R-flatness of G and $G^\vee$ arises from the semiabelian condition, since any quasi-finite morphism between semi-abelian schemes is necessarily flat, as can be checked on geometric fibers). Being quasi-finite flat and separated, the R-group schemes G and $G^\vee$ have canonical respective "finite parts" G_f and $G_f^\vee$ and "multiplicative parts" G_μ and $G_\mu^\vee$ (as in Lemma 8.3). Beware that we do *not* claim $G_f^\vee$ (resp. $G_\mu^\vee$) is the Cartier dual to G_f (resp. G_μ); usually such duality does not hold.

Since G_μ and $G_\mu^\vee$ are finite flat R-group schemes, the quotients G/G_μ and $G^\vee/G_\mu^\vee$ make sense as quasi-finite flat separated R-group schemes. The theorem almost says that there is a canonical duality between $(G/G_\mu)_K = G_K/(G_\mu)_K$ and $(G_\mu^\vee)_K$, induced by the duality between $G_K = \ker(\varphi)$ and $G_K^\vee = \ker(\varphi^\vee)$, except for the mild problem that G_μ might be larger than $\widetilde{\ker(\varphi_t)}$ (i.e., possibly $(G_\mu)_k$ is not entirely inside of $\mathcal{A}_k^0$) and likewise $G_\mu^\vee$ might be larger than $\widetilde{\ker(\varphi_t^\vee)}$.

We will work on ℓ-primary components for each prime ℓ individually. In order to permit this, we use Remark 8.4. We will first treat the more subtle case when ℓ is the residue characteristic, and then we'll handle the case when it isn't. The advantage of working with the case in which ℓ is the residue characteristic is that multiplicative finite k-group schemes are automatically *connected*. Thus, in this case $(G_\mu)_\ell = \widetilde{\ker(\varphi_t)}_\ell$ and $(G_\mu^\vee)_\ell = \widetilde{\ker(\varphi_t^\vee)}_\ell$.

Since

$$0 \to G_K \to A \to B \to 0$$

is an exact sequence of abelian sheaves on the fppf site over $\mathrm{Spec}(K)$, by the usual snake lemma argument (and the fact that the ℓ-part $(G_K)_\ell$ of G_K is killed by a big power of ℓ) we obtain an exact sequence

$$0 \to (G_K)_\ell \to A[\ell^\infty] \to B[\ell^\infty] \to 0.$$

Arguing as in Corollary 8.5, there is an exact sequence over $\mathrm{Spec}(R)$

$$0 \to (G_\mu)_\ell \to \mathcal{A}[\ell^\infty]_t \to \mathcal{B}[\ell^\infty]_t \to 0$$

which lifts the exact sequence involving ℓ-divisible groups of tori on the closed fiber (as $(G_\mu)_\ell$ *must* be in the relative connected component of $\mathcal{A}$). Passing to the generic fiber over K gives us a commutative diagram with exact rows and closed immersions along columns

$$\begin{tikzcd}
0 \arrow[r] & ((G_\mu)_K)_\ell \arrow[r] \arrow[d] & A[\ell^\infty]_t \arrow[r] \arrow[d] & B[\ell^\infty]_t \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & (G_K)_\ell \arrow[r] & A[\ell^\infty] \arrow[r] & B[\ell^\infty] \arrow[r] & 0
\end{tikzcd}$$

where $A[\ell^\infty]_t$ denotes the K-fiber of the ℓ-divisible group $\{\mathcal{A}[\ell^n]_t\}$, and likewise for $B[\ell^\infty]_t$.

Using the snake lemma in the category of fppf abelian sheaves over $\mathrm{Spec}(K)$, we get a short exact sequence of cokernels

$$0 \to ((G/G_\mu)_K)_\ell \to A[\ell^\infty]/(\text{toric}) \to B[\ell^\infty]/(\text{toric}) \to 0$$

where all maps are the natural ones, and right two terms are ℓ-divisible groups over K.

The Grothendieck Orthogonality Theorem (see [9, Exp. IX, Prop 5.6]) asserts that the perfect scheme-theoretic Weil pairing between $A[\ell^n]$ and $A^\vee[\ell^n]$ makes $A[\ell^n]_t$ and $A^\vee[\ell^n]_f$ exact annhilators, where $A[\ell^n]_f$ denotes the K-fiber of the finite part of the $\mathcal{A}^0[\ell^n]$ and $A[\ell^n]_t$ denotes the K-fiber of the unique R-subgroup scheme in $\mathcal{A}^0[\ell^n]$ lifting the ℓ^n-torsion on the closed fiber torus. By the purely toric condition applied to $A^\vee$, we see $A^\vee[\ell^n]_f = A^\vee[\ell^n]_t$. Thus, the orthogonality theorem says that $A[\ell^n]/A[\ell^n]_t$ and $A^\vee[\ell^n]_t$ are in perfect duality via the scheme-theoretic Weil pairing over K.

Passing to the limit, we get a canonical isomorphism of ℓ-divisible groups

$$A[\ell^\infty]/(\text{toric}) = (A^\vee[\ell^\infty]_t)^\vee.$$

But φ and $\varphi^\vee$ are adjoint with respect to Weil pairing, so we conclude that the diagram

$$\begin{tikzcd}
A[\ell^\infty]/(\text{toric}) \arrow[r, equal] \arrow[d, "\varphi"] & (A^\vee[\ell^\infty]_t)^\vee \arrow[d, "(\varphi_t^\vee)^\vee"] \\
B[\ell^\infty]/(\text{toric}) \arrow[r, equal] & (B^\vee[\ell^\infty]_t)^\vee
\end{tikzcd}$$

commutes. Thus, we get an isomorphism between the kernels of these vertical isogenies. The kernel of the left column is $((G/G_\mu)_K)_\ell$, as we saw above. Meanwhile, the kernel of the right is (by duality theory of ℓ-divisible groups) exactly the dual of $\ker(\varphi_t^\vee) = (G_\mu^\vee)_K$. This gives the desired perfect duality between $(G/G_\mu)_K$ and $(G_\mu^\vee)_K$ on ℓ-primary parts for ℓ equal to the residue characteristic.

Now we consider the case when ℓ is not equal to the residue characteristic. There is no loss of generality in passing to the case of a strictly henselian base R. Thus, the closed fiber tori have *constant* ℓ-divisible groups. Also, we can work with $\mathbf{Z}_\ell$-modules of geometric points (over K) via Tate's construction. The "toric" part of the ℓ-adic Tate module $T_\ell(A)$ is exactly the (saturated) maximal submodule with trivial Galois action, since a compatible system of ℓ-power torsion points in $A(K) = \mathcal{A}(R)$ must lie entirely inside of $\mathcal{A}^0(R)$ (thanks to the finiteness of the component group) and we can identify $\mathcal{A}^0(R)[\ell^n]$ with the (constant) ℓ^n-torsion on the split torus $\mathcal{A}_k^0$ over the separably closed k.

Using inverse limits, we see that $T_\ell(A) \to T_\ell(B)$ is injective with cokernel $(G_K)_\ell$ (= geometric points of ℓ-part), and this cokernel is exactly $\ker(\varphi)_\ell$. Likewise, the cokernel of the map

$$T_\ell(A)_t \to T_\ell(B)_t$$

on "toric" parts (i.e., ℓ-adic Tate module generic fibers of the lifts of the ℓ-divisible groups of closed fiber tori) is $(\widetilde{\ker(\varphi_t)}_K)_\ell$.

Thus, we get a commutative diagram with horizontal exact sequences

$$\begin{tikzcd}
0 \arrow[r] & T_\ell(A)_t \arrow[r] \arrow[d] & T_\ell(B)_t \arrow[r] \arrow[d] & (\widetilde{\ker(\varphi_t)}_K)_\ell \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & T_\ell(A) \arrow[r] & T_\ell(B) \arrow[r] & \ker(\varphi)_\ell \arrow[r] & 0
\end{tikzcd}$$

with columns given by the natural maps. These vertical maps are all injective, so by the snake lemma we get a short exact sequence of cokernels. We can now use the exact same Weil pairing arguments with the Grothendieck orthogonality theorem (now in the easier "$\ell \neq p$" form of the orthogonality theorem [9, Exp IX, 2.4]), essentially just as we argued in the previous case. One distinction is that the use of inverse limit Tate modules rather than direct limits causes some maps to switch direction.

More specifically we have a commutative square with horizontal isomporphisms (thanks to the orthogonality theorem)

$$\begin{tikzcd}
T_\ell(A)/T_\ell(A)_t \arrow[r, equal] \arrow[d, "\varphi"'] & T_\ell(A^\vee)_t^\vee \arrow[d, "T_\ell(\varphi_t^\vee)^\vee"] \\
T_\ell(B)/T_\ell(B)_t \arrow[r, equal] & T_\ell(B^\vee)_t^\vee
\end{tikzcd}$$

This induces an isomorphism

$$((\ker\varphi)/\widetilde{\ker(\varphi_t)}_K)_\ell \simeq ((\widetilde{\ker\varphi_t^\vee})_K)_\ell^\vee$$

between the vertical cokernels, and by construction this isomorphism is compatible with Weil pairings, whence the desired perfect pairing has been shown.

□

Let R be an arbitrary dvr (not necessarily henselian), let $\varphi : A \to B$ be an isogeny, and let $\varphi_t : T_A \to T_B$ be the induced map on the closed fiber toric parts, as above.

Corollary 8.7. *The order of* $\ker(\varphi)$ *is the product of the orders of* $\ker\varphi_t$ *and* $\ker\varphi_t^\vee$*.*

Proof. Pass to the henselization of R and use Theorem 8.6. □

References

[1] A. Agashe and W. A. Stein, *Visibility of Shafarevich-Tate groups of abelian varieties: Evidence for the Birch and Swinnerton-Dyer conjecture*, (2001).

[2] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.

[3] C. Chevalley, *Une démonstration d'un théorème sur les groupes algébriques*, J. Math. Pures Appl. (9) **39** (1960), 307–317.

[4] B. Conrad, *A modern proof of Chevalley's theorem on algebraic groups*, `http://www-math.mit.edu/~dejong/papers/chev.dvi`

[5] B. Edixhoven, *L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein"*, Astérisque (1991), no. 196–197, 7–8, 159–170 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[6] M. Emerton, *Optimal quotients of modular Jacobians*, (2001), preprint.

[7] E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. Comp. **70** (2001), no. 236, 1675–1697 (electronic).

[8] A. Grothendieck, *Éléments de géométrie algébrique*, Publications Mathématiques IHES, **4,8,11,17,20,24,28,32**, 1960–7.

[9] A. Grothendieck, *Groupes de monodromie en géométrie algébrique*, Lecture Notes in Math **288**, Springer-Verlag, Heidelberg (1972).

[10] N. Katz, B. Mazur, *Arithmetic moduli of elliptic curves*, Princetion University Press, Princeton, New Jersey, 1985.

[11] D. R. Kohel, *Hecke module structure of quaternions*, In K. Miyake, ed., *Class Field Theory – Its Centenary and Prospect*, The Advanced Studies in Pure Mathematics Series, Math Soc. Japan.

[12] D. R. Kohel and W. A. Stein, *Component Groups of Quotients of $J_0(N)$*, Proceedings of the 4th International Symposium (ANTS-IV), Leiden, Netherlands, July 2–7, 2000 (Berlin), Springer, 2000.

[13] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[14] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242.

[15] J.-F. Mestre and J. Oesterlé, *Courbes de Weil semi-stables de discriminant une puissance m-ième*, J. Reine Angew. Math. **400** (1989), 173–184.

[16] D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.

[17] K. A. Ribet, *Letter about component groups of elliptic curves*, `arXiv:math.AG/0105124v1` (2001).

[18] J. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1975, pp. 33–52. Lecture Notes in Math., Vol. 476.

12 Visibility of ShafarevichTate Groups of Abelian-Varieties, with A. Agashe

Journal of Number Theory **97**, 171–185 (2002)
doi:10.1006/jnth.2002.2810

Visibility of Shafarevich–Tate Groups of Abelian Varieties

Amod Agashe

Department of Mathematics, University of Texas, Austin, Texas 78712
E-mail: amod@math.utexas.edu

and

William Stein

Department of Mathematics, Harvard University, Cambridge, Massachusetts
E-mail: was@math.harvard.edu

Communicated by K. Ribet

Received November 24, 2001; revised February 27, 2002

We investigate Mazur's notion of visibility of elements of Shafarevich–Tate groups of abelian varieties. We give a proof that every cohomology class is visible in a suitable abelian variety, discuss the visibility dimension, and describe a construction of visible elements of certain Shafarevich–Tate groups. This construction can be used to give some of the first evidence for the Birch and Swinnerton–Dyer conjecture for abelian varieties of large dimension. We then give examples of visible and invisible Shafarevich–Tate groups. © 2002 Elsevier Science (USA)

Key Words: visibility; Shafarevich–Tate group; Birch and Swinnerton–Dyer conjecture; modular abelian variety.

INTRODUCTION

If a genus 0 curve X over $\mathbf{Q}$ has a point in every local field $\mathbf{Q}_p$ and in $\mathbf{R}$, then it has a global point over $\mathbf{Q}$. For genus 1 curves, this "local-to-global principle" frequently fails. For example, the nonsingular projective curve defined by the equation $3x^3 + 4y^3 + 5z^3 = 0$ has a point over each local field and $\mathbf{R}$, but has no $\mathbf{Q}$-point. The Shafarevich–Tate group of an elliptic curve E, denoted $Ш(E)$, is a group that measures the extent to which a local-to-global principle fails for the genus one curves with Jacobian E. More generally, if A is an abelian variety over a number field K, then the elements of the Shafarevich–Tate group $Ш(A)$ of A correspond to the torsors for A that have a point everywhere locally, but not globally. In this paper, we study a geometric way of realizing (or "visualizing") torsors corresponding to elements of $Ш(A)$.

0022-314X/02 $35.00
© 2002 Elsevier Science (USA)
All rights reserved.

Let A be an abelian variety over a field K. If $\iota : A \hookrightarrow J$ is a closed immersion of abelian varieties, then the subgroup of $H^1(K, A)$ *visible in* J (with respect to ι) is $\ker(H^1(K, A) \to H^1(K, J))$. We prove that every element of $H^1(K, A)$ is visible in some abelian variety, and give bounds on the smallest size of an abelian variety in which an element of $H^1(K, A)$ is visible. Next assume that K is a number field. We give a construction of visible elements of $Ш(A)$, which we demonstrate by giving evidence for the Birch and Swinnerton–Dyer conjecture for a certain 20-dimensional abelian variety. We also give an example of an elliptic curve E over $\mathbf{Q}$ of conductor N whose Shafarevich–Tate group is not visible in $J_0(N)$ but is visible in $J_0(Np)$ for some prime p.

This paper is organized as follows. Section 1 contains the definition of visibility for cohomology classes and elements of Shafarevich–Tate groups. Then in Section 1.3, we use a restriction of scalars construction to prove that every cohomology class is visible in some abelian variety. Next, in Section 2, we investigate the visibility dimension of cohomology classes. Section 3 contains a theorem that can be used to construct visible elements of Shafarevich–Tate groups. Finally, Section 4, contains examples and applications of our visibility results in the context of modular abelian varieties.

1. VISIBILITY

In Section 1.1 we introduce visible cohomology classes, then in Section 1.2 we discuss visible elements of Shafarevich–Tate groups. In Section 1.3, we use restriction of scalars to deduce that every cohomology class is visible somewhere.

For a field K and a smooth commutative K-group scheme G, we write $H^i(K, G)$ to denote the group cohomology $H^i(\mathrm{Gal}(K_s/K), G(K_s))$ where K_s is a fixed separable closure of K; equivalently, $H^i(K, G)$ denotes the ith étale cohomology of G viewed as an étale sheaf on $\mathrm{Spec}(K)_{\text{ét}}$.

1.1. *Visible elements of* $H^1(K, A)$.

In [Maz99], Mazur introduced the following definition. Let A be an abelian variety over an arbitrary field K.

Definition 1.1. Let $\iota : A \hookrightarrow J$ be an embedding, of A into an abelian variety J over K. Then the *visible subgroup of* $H^1(K, A)$ *with respect to the embedding* ι is

$$\mathrm{Vis}_J(H^1(K, A)) = \mathrm{Ker}(H^1(K, A) \to H^1(K, J)).$$

The visible subgroup $\mathrm{Vis}_J(H^1(K, A))$ depends on the choice of embedding ι, but we do not include ι in the notation, as it is usually clear from context.

The Galois cohomology group $H^1(K, A)$ has a geometric interpretation as the group of classes of torsors X for A (see [LT58]). To a cohomology class $c \in H^1(K, A)$, there is a corresponding variety X over K and a map $A \times X \to X$ that satisfies axioms similar to those for a simply transitive group action. The set of equivalence classes of such X forms a group, the Weil–Chatelet group of A, which is canonically isomorphic to $H^1(K, A)$.

There is a close relationship between visibility and the geometric interpretation of Galois cohomology. Suppose $\iota : A \to J$ is an embedding and $c \in \mathrm{Vis}_J(H^1(K, A))$. We have an exact sequence of abelian varieties $0 \to A \to J \to C \to 0$, where $C = J/A$. A piece of the associated long exact sequence of Galois cohomology is

$$0 \to A(K) \to J(K) \to C(K) \to H^1(K, A) \to H^1(K, J) \to \cdots,$$

so there is an exact sequence

$$0 \to J(K)/A(K) \to C(K) \to \mathrm{Vis}_J(H^1(K, A)) \to 0. \tag{1.1}$$

Thus there is a point $x \in C(K)$ that maps to c. The fiber X over x is a subvariety of J, which, when equipped with its natural action of A, lies in the class of torsors corresponding to c. This is the origin of the terminology "visible". Also, we remark that when K is a number field, $\mathrm{Vis}_J(H^1(K, A))$ is finite because it is torsion and is the surjective image of the finitely generated group $C(K)$.

1.2. *Visible Elements of* Ш(A). Let A be an abelian variety over a number field K. The Shafarevich–Tate group of A, which is defined below, measures the failure of the local-to-global principle for certain torsors. The *Shafarevich–Tate group* of A is

$$Ш(A) := \mathrm{Ker}\left(H^1(K, A) \to \prod_v H^1(K_v, A)\right),$$

where the product is over all places of K.

Definition 1.2. If $\iota : A \hookrightarrow J$ is an embedding then the *visible subgroup of* Ш(A) *with respect to* ι is

$$\mathrm{Vis}_J(Ш(A)) := Ш(A) \cap \mathrm{Vis}_J(H^1(K, A)) = \mathrm{Ker}(Ш(A) \to Ш(J)).$$

1.3. *Every Element is Visible Somewhere.*

Proposition 1.3. *Every element of $H^1(K, A)$ is visible in some abelian variety J.*

Proof. Fix $c \in H^1(K, A)$. There is a finite separable extension L of K such that $\mathrm{res}_L(c) = 0 \in H^1(K, A)$. Let $J = \mathrm{Res}_{L/K}(A_L)$ be the Weil restriction of scalars from L to K of the abelian variety A_L (see [BLR90, Sect. 7.6]). Thus J is an abelian variety over K of dimension $[L : K] \cdot \dim(A)$, and for any scheme S over K, we have a natural (functorial) group isomorphism $A_L(S_L) \cong J(S)$. The functorial injection $A(S) \hookrightarrow A_L(S_L) \cong J(S)$ corresponds via Yoneda's Lemma to a natural K-group scheme map $\imath : A \to J$, and by construction $\imath$, is a monomorphism. But $\imath$ is proper and thus is a closed immersion (see [Gro66, Sect. 8.11.5]). Using the Shapiro lemma one finds, after a tedious computation, that there is a canonical isomorphism $H^1(K, J) \cong H^1(L, A)$ which identifies $\imath_*(c)$ with $\mathrm{res}_L(c) = 0$. ∎

Remark 1.4. 1. In [CM00], de Jong gave a totally different proof of the above proposition in the case when A is an elliptic curve over a number field. His argument actually displays A as visible inside the Jacobian of a curve.

2. L. Clozel has remarked that the method of proof above is a standard technique in the theory of algebraic groups.

2. THE VISIBILITY DIMENSION

Let A be an abelian variety over a field K and fix $c \in H^1(K, A)$.

Definition 2.1. The *visibility dimension* of c is the minimum of the dimensions of the abelian varieties J such that c is visible in J.

In Section 2.1 we prove an elementary lemma which, when combined with the proof of Proposition 1.3, gives an upper bound on the visibility dimension of c in terms of the order of c and the dimension of A. Then, in Section 2.2, we consider the visibility dimension in the case when $A = E$ is an elliptic curve. After summarizing the results of Mazur and Klenke on the visibility dimension, we apply a theorem of Cassels to deduce that the visibility dimension of $c \in \text{Ш}(E)$ is at most the order of c.

2.1. *A Simple Bound.*

The following elementary lemma, which the second author learned from Hendrik Lenstra, will be used to give a bound on the visibility dimension in terms of the order of c and the dimension of A.

Lemma 2.2. *Let G be a group, M be a finite (discrete) G-module, and $c \in H^1(G, M)$. Then there is a subgroup H of G such that* $\mathrm{res}_H(c) = 0$ *and* $\#(G/H) \leqslant \#M$.

Proof. Let $f : G \to M$ be a cocycle corresponding to c, so $f(\tau\sigma) = f(\tau) + \tau f(\sigma)$ for all $\tau, \sigma \in G$. Let $H = \ker(f) = \{\sigma \in G : f(\sigma) = 0\}$. The map $\tau H \mapsto f(\tau)$ is a well-defined injection from the coset space G/H to M. ∎

The following is a general bound on the visibility dimension.

PROPOSITION 2.3. *The visibility dimension of any $c \in H^1(K, A)$ is at most $d \cdot n^{2d}$ where n is the order of c and d is the dimension of A.*

Proof. The map $H^1(K, A[n]) \to H^1(K, A)[n]$ is surjective and $A[n]$ has order n^{2d}, so Lemma 2.2 implies that there is an extension L of K of degree at most n^{2d} such that $\mathrm{res}_L(c) = 0$. The proof of Proposition 1.3 implies that c is visible in an abelian variety of dimension $[L : K] \cdot \dim A \leqslant dn^{2d}$. ∎

2.2. *The Visibility Dimension for Elliptic Curves.* We now consider the case when $A = E$ is an elliptic curve over a number field K. Mazur proved in [Maz99] that every nonzero $c \in \text{Ш}(E)[3]$ has visibility dimension 2 (note that Proposition 2.3 only implies that the visibility dimension is $\leqslant 3$). Mazur's result is particularly nice because it shows that c is visible in an abelian variety that is isogenous to the product of two elliptic curves. Using similar techniques, Klenke proved in [Kle01] that every nonzero $c \in H^1(K, E)[2]$ has visibility dimension 2 (note that Proposition 2.3 only implies that the visibility dimension of any $c \in H^1(K, E)[2]$ is $\leqslant 4$). It is unknown whether the visibility dimension of every nonzero element of $H^1(K, E)[3]$ is 2, and it is not known whether elements of $\text{Ш}(E)[5]$ must have visibility dimension 2.

When c lies in $\text{Ш}(E)$ we use a classical result of Cassels to strengthen the conclusion of Proposition 2.3.

PROPOSITION 2.4. *Let E be an elliptic curve over a number field K and let $c \in \text{Ш}(E)$. Then the visibility dimension of c is at most the order of c.*

Proof. Let n be the order of c. In view of the restriction of scalars construction in the proof of Proposition 1.3, it suffices to show that there is an extension L of K of degree n such that $\mathrm{res}_L(c) = 0$. Without the hypothesis that c lies in $\text{Ш}(E)$, such an extension L might not exist, as Cassels observed in [Cas63]. However, in that same paper, Cassels proved that such an L exists when $c \in \text{Ш}(E)$ (see also [O'N01] for another proof). ∎

Remark 2.5. In contrast to the case of dimension 1, it seems to be an open problem to determine whether or not elements of $\text{Ш}(A)[n]$ split over an extension of degree n.

3. CONSTRUCTION OF VISIBLE ELEMENTS

The goal of this section is to state and prove the main result of this paper, which we use to construct visible elements of Shafarevich–Tate groups and sometimes give a nontrivial lower bound for the order of the Shafarevich–Tate group of an abelian variety thus providing new evidence for the conjecture of Birch and Swinnerton–Dyer (see Section 4.1 and [AS02]). The Tamagawa numbers $c_{A,v}$ and $c_{B,v}$ will be defined in Section 3.1.

Theorem 3.1. *Let A and B be abelian subvarieties of an abelian variety J over a number field K such that $A \cap B$ is finite. Let N be an integer divisible by the residue characteristics of primes of bad reduction for B. Suppose n is an integer such that for each prime $p|n$, we have $e_p < p-1$ where e_p is the largest ramification of any prime of K lying over p, and that*

$$\gcd\left(n, N \cdot \#(J/B)(K)_{\text{tor}} \cdot \#B(K)_{\text{tor}} \cdot \prod_{\text{all places } v} (c_{A,v} \cdot c_{B,v})\right) = 1,$$

where $c_{A,v} = \#\Phi_{A,v}(\mathbf{F}_\ell)$ (resp., $c_{B,\ell}$) is the Tamagawa number of A (resp., B) at v (see Section 3.1 for the definition of $\Phi_{A,v}$). Suppose furthermore that $B[n] \subset A$ as subgroup schemes of J. Then there is a natural map

$$\varphi : B(K)/nB(K) \to \mathrm{Vis}_J(\text{Ш}(A)),$$

such that $\ker(\varphi) \subset J(K)/(B(K) + A(K))$. *If $A(K)$ has rank* 0, *then* $\ker(\varphi) = 0$ (*more generally*, $\ker(\varphi)$ *has order at most n^r where r is the rank of $A(K)$*).

Remark 3.2. Mazur has proved similar results for elliptic curves using flat cohomology (unpublished), and discussions with him motivated this theorem.

In Section 3.1 we recall a definition of the Tamagawa numbers of an abelian variety. In Section 3.2 we prove a lemma, which gives a condition under which there is an unramified nth root of an unramified point. In Section 3.3, we use the snake lemma to produce a map

$$B(K)/nB(K) \hookrightarrow \mathrm{Vis}_J(H^1(K, A))$$

with bounded kernel. Finally, in Section 3.4, we use a local analysis at each place of K to show that the image of the above map lies in Ш(A).

3.1. *Tamagawa Numbers.* Let A be an abelian variety over a local field K with residue class field k, and let $\mathscr{A}$ be the Néron model of A over the ring of integers of K. The closed fiber $\mathscr{A}_k$ of $\mathscr{A}$ need not be connected. Let $\mathscr{A}_k^0$ denote the geometric component of $\mathscr{A}$ that contains the identity. The group

$\Phi_{\mathscr{A}} = \mathscr{A}_k/\mathscr{A}_k^0$ of connected components is a finite group scheme over k. This group scheme is called the *component group* of $\mathscr{A}$, and the *Tamagawa number* of A is $c_A = \#\Phi_{\mathscr{A}}(k)$.

Now suppose that A is an abelian variety over a global field K. For every place v of K, the *Tamagawa number* of A at v, denoted $c_{A,v}$ or just c_v, is the Tamagawa number of A_{K_v}, where K_v is the completion of K at v.

3.2. *Smoothness and Surjectivity.* In this section, we recall some well-known lemmas that we will use in Section 3.4 to produce unramified cohomology classes. The authors are grateful to B. Conrad for explaining the proofs of these lemmas.

LEMMA 3.3. *If G is a finite-type smooth commutative group scheme over a strictly henselian local ring R and the fibers of G over R are (geometrically) connected, then the multiplication map*

$$n_G : G(R) \to G(R)$$

is surjective when $n \in R^\times$.

Proof. Pick an element $g \in G(R)$ and form the cartesian diagram

$$\begin{tikzcd} Y_g \arrow[r, "\psi"] \arrow[d] & \mathrm{Spec}(R) \arrow[d, "g"] \\ G \arrow[r, "n_G"] & G \end{tikzcd}$$

We want to prove that ψ has a section. Since R is strictly henselian, by [Gro67, 18.8.1] it suffices to show that Y_g is étale over R with nonempty closed fiber, or more generally that n_G is étale and surjective.

By Lemma 2(b) of [BLR90, Sect. 7.3], n_G is étale. The image of the étale n_G must be an open subgroup scheme, and on fibers over $\mathrm{Spec}(R)$ we get surjectivity since an open subgroup scheme of a smooth connected (hence irreducible) group scheme over a field must fill up the whole space [Gro70, $\mathrm{VI_A}$, 0.5]. ∎

LEMMA 3.4. *Let A be an abelian variety over the fraction field K of a strictly henselian dvr (e.g., K could be the maximal unramified extension a local field). Let n be an integer not divisible by the residue characteristic of K. Suppose that x is a point of $A(K)$ whose reduction lands in the identity* *component of the closed fiber of the Néron model of A. Then there exists $z \in A(K)$ such that $nz = x$.*

Proof. Let $\mathscr{A}$ denote the Néron model of A over the valuation ring R of K, and let $\mathscr{A}^0$ denote the "identity component" (i.e., the open subgroup scheme obtained by removing the nonidentity components of the closed fiber of $\mathscr{A}$). The hypothesis on the reduction of $x \in A(K) = \mathscr{A}^0(R)$ says exactly that $x \in \mathscr{A}^0(R)$. Since connected schemes over a field are geometrically connected when there is a rational point [Gro65, Proposition 4.5.13], the fibers of $\mathscr{A}^0$ over $\operatorname{Spec}(R)$ are geometrically connected. The lemma now follows from Lemma 3.3 with $G = \mathscr{A}^0$. ∎

Remark 3.5. M. Baker noted that this argument can also be formulated in terms of formal groups when R is the strict henselization of a *complete* dvr.

Lemma 3.6. *Let $\mathscr{J} \xrightarrow{\phi} \mathscr{C}$ be a smooth surjective morphism of schemes over a strictly Henselian local ring R. Then the induced map $\mathscr{J}(R) \to \mathscr{C}(R)$ is surjective.*

Proof. The argument is similar to that of the proof of Lemma 3.3. Pick an element $g \in \mathscr{C}(R)$ and form the cartesian diagram

$$\begin{tikzcd}
Y_g \arrow[r, "\psi"] \arrow[d] & \operatorname{Spec}(R) \arrow[d, "g"] \\
\mathscr{J} \arrow[r, "\phi"] & \mathscr{C}
\end{tikzcd}$$

We want to prove that ψ has a section. Since ϕ is smooth, ψ is also smooth. By Grothendieck [Gro67, 18.5.17], to show that ψ has a section, we just need to show that the closed fiber of ψ has a section (i.e., a rational point). But this closed fiber is smooth and nonempty (since ϕ is surjective); also its base field is separably closed since R is strictly Henselian. Hence by Bosma *et al.* [BLR90, Corollary 2.2.13], the closed fiber has an R-rational point. ∎

3.3. *Visible Elements of $H^1(K, A)$.* In this section, we produce a map $B(K)/nB(K) \to \operatorname{Vis}_J(H^1(K, A))$ with bounded kernel.

Lemma 3.7. *Let A and B be abelian subvarieties of an abelian variety J over a number field K such that $A \cap B$ is finite. Suppose n is a natural number such that*

$$\gcd(n, \#(J/B)(K)_{\mathrm{tor}} \cdot \#B(K)_{\mathrm{tor}}) = 1$$

and $B[n] \subset A$ as subgroup schemes of J. Then there is a natural map

$$\varphi : B(K)/nB(K) \to \mathrm{Vis}_J(H^1(K, A))$$

such that $\ker(\varphi) \subset J(K)/(B(K) + A(K))$. *If $A(K)$ has rank* 0, *then* $\ker(\varphi) = 0$ (*more generally*, $\ker(\varphi)$ *has order at most n^r where r is the rank of $A(K)$*).

Proof. First we produce a map $\varphi : B(K)/nB(K) \to \mathrm{Vis}(H^1(K, A))$ by using that $B[n] \subset A$ hence a certain map factors through multiplication by n. Then we use the snake lemma and our hypothesis that n does not divide the orders of certain torsion groups to bound the dimension of the kernel of φ.

The quotient J/A is an abelian variety C over K. The long exact sequence of Galois cohomology associated to the short exact sequence

$$0 \to A \to J \to C \to 0$$

begins

$$0 \to A(K) \to J(K) \to C(K) \xrightarrow{\delta} H^1(K, A) \to \cdots. \tag{3.1}$$

Let ψ be map $B \to C$ obtained by composing the inclusion $B \hookrightarrow J$ with the quotient map $J \to C$. Since $B[n] \subset A$, we see that ψ factors through multiplication by n, so the following diagram commutes:

$$\begin{tikzcd}
 & B \arrow[r, "n"] \arrow[d] \arrow[dr, "\psi"] & B \arrow[d] \\
A \arrow[r] & J \arrow[r] & C.
\end{tikzcd}$$

Using that $B[n](K) = \{0\}$, we obtain the following commutative diagram, all of whose rows and columns are exact:

$$\begin{tikzcd}
 & K_0 \arrow[d] & K_1 \arrow[d] & K_2 \arrow[d] & \\
0 \arrow[r] & B(K) \arrow[r, "n"] \arrow[d] & B(K) \arrow[r] \arrow[d] \arrow[dr, "\pi"] & B(K)/nB(K) \arrow[r] \arrow[d, "\varphi"] & 0 \\
0 \arrow[r] & J(K)/A(K) \arrow[r] \arrow[d] & C(K) \arrow[r] & \delta(C(K)) \arrow[r] & 0 \\
 & K_3, & & &
\end{tikzcd} \tag{3.2}$$

where K_0, K_1 and K_2 are the indicated kernels and K_3 is the indicated cokernel. Exactness of the top row expresses the fact that $B[n](K) = \{0\}$, and the bottom exact row arises from the exact sequence (3.1) above. The first vertical map $B(K) \to J(K)/A(K)$ is induced by the inclusion $B(K) \hookrightarrow J(K)$ composed with the quotient map $J(K) \to J(K)/A(K)$. The second vertical map $B(K) \to C(K)$ exists because the composition $B \hookrightarrow J \to C$ has kernel $B \cap A$, which contains $B[n]$, by assumption. The third vertical map exists because π contains $nB(K)$ in its kernel, so that π factors through $B(K)/nB(K)$.

Sequence (1.1) implies that the image of φ is contained in $\mathrm{Vis}_J(H^1(K, A))$. The snake lemma gives an exact sequence

$$K_0 \to K_1 \to K_2 \to K_3.$$

Because $B \to C$ has finite kernel, $K_1 \subset B(K)_{\mathrm{tor}}$. Since $B[n](K) = \{0\}$ and K_2 is an n-torsion group, the map $K_1 \to K_2$ is the 0 map. Thus, $K_2 = \ker(\varphi)$ is isomorphic to a subgroup of $K_3 = J(K)/(A(K) + B(K))$, as claimed.

Any torsion in the quotient $J(K)/B(K)$ is of order coprime to n because $J(K)/B(K)$ is a subgroup of $(J/B)(K)$, and $\gcd(n, \#(J/B)(K)_{\mathrm{tor}}) = 1$, by assumption. Thus if $A(K)$ is a torsion group, $K_3 = (J(K)/B(K))/A(K)$ has no nontrivial torsion of order dividing n, so when $A(K)$ has rank zero, $\ker(\varphi) = 0$.

Consider the map $\psi : A(K) \to J(K)/B(K)$. To show that $\ker(\phi)$ has order at most n^r, where r is the rank of $A(K)$, it suffices to show that $\mathrm{coker}(\psi)[n]$ has order at most n^r. To prove the latter statement, by the structure theorem for finite abelian groups, it suffices to prove it for the case when n is a power of a prime. Moreover, we may assume that $A(K)$ and $J(K)/B(K)$ have no prime-to-n torsion. Then $J(K)/B(K)$ is in fact torsion-free, and so we may also assume $A(K)$ is torsion-free. With these assumptions, the statement we want to prove follows easily by elementary group-theoretic arguments (in particular, by considering of the Smith normal form of the matrix representing ψ). ∎

3.4. *Proof of Theorem 3.1.*

Proof of Theorem 3.1. The proof proceeds in two steps. The first step is to use the hypothesis that $B[n] \subset A$ to produce a map $B(K)/nB(K) \to \mathrm{Vis}_J(H^1(K, A))[n]$. This was done in Section 3.3. The second step is to perform a local analysis at each place v of K in order to prove that the image of this map consists of locally trivial cohomology classes. We divide this local analysis into three cases:

1. When v is real archimedian, we use that $\gcd(2, n) = 1$. (We know that for any $p|n$ we have $p > 2$ because $1 \leqslant e_p < p - 1$, by assumption.)

2. When $\gcd(\mathrm{char}(v), n) = 1$, we use the result of Section 3.2 and a relationship between unramified cohomology and the cohomology of a component group.

3. When $\gcd(\mathrm{char}(v), n) \neq 1$, for each prime $p|n$, the reduction of J is abelian and by hypothesis $e_p < p - 1$, so we can apply an exactness theorem from [BLR90].

We now deduce that the image of $B(K)/nB(K)$ in $H^1(K, A)$ lies in $Ш(A)$. Fix an element $x \in B(K)$. To show that $\pi(x) \in Ш(A)$, it suffices to show that $\mathrm{res}_v(\pi(x)) = 0$ for all places v of K.

Case 1: *v real archimedian*. At a real archimedian place v, the restriction $\mathrm{res}_v(\pi(x))$ is killed by 2 and the odd n, hence $\mathrm{res}_v(\pi(x)) = 0$.

Case 2: $\gcd(\mathrm{char}(v), n) = 1$. Suppose that $\gcd(\mathrm{char}(v), n) = 1$. Let $m = c_{B,v} = \Phi_{B,v}(\mathbf{F}_v)$ be the Tamagawa number of B at v. The reduction of mx lies in the identity component of the closed fiber $\mathscr{B}_{\mathbf{F}_v}$ of the Néron model of B at v, so by Lemma 3.4, there exists $z \in B(K_v^{\mathrm{ur}})$ such that $nz = mx$. Thus the cohomology class $\mathrm{res}_v(\pi(mx))$ is defined by a cocycle that sends $\sigma \in \mathrm{Gal}(\overline{K_v}/K_v)$ to $\sigma(z) - z \in A(K_v^{\mathrm{ur}})$ (see diagram (3.2) for the definition of π). In particular, $\mathrm{res}_v(\pi(mx))$ is unramified at v. By Milne [Mil86, Proposition 3.8].

$$H^1(K_v^{\mathrm{ur}}/K_v, A(K_v^{\mathrm{ur}})) = H^1(K_v^{\mathrm{ur}}/K_v, \Phi_{A,v}(\bar{\mathbf{F}}_v)),$$

where $\Phi_{A,v}$ is the component group of A at v. The Herbrand quotient of a finite module is 1 (see, e.g., [Ser79, VIII.4.8]), so

$$\#\Phi_{A,v}(\mathbf{F}_v) = \#H^1(K_v^{\mathrm{ur}}/K_v, \Phi_{A,v}(\bar{\mathbf{F}}_v)).$$

Thus, the order of $\mathrm{res}_v(\pi(mx))$ divides both $\#\Phi_{A,v}(\mathbf{F}_v)$ and n. Since by assumption $\gcd(\#\Phi_{A,v}(\mathbf{F}_v), n) = 1$, it follows that $\mathrm{res}_v(\pi(mx)) = 0$, hence $m\,\mathrm{res}_v(\pi(x)) = 0$. Again, since the order of $\pi(x)$ divides n, and $\gcd(n, m) = 1$, we have $\mathrm{res}_v(\pi(x)) = 0$.

Case 3: $\gcd(\mathrm{char}(v), n) = p \neq 1$. Suppose that $\mathrm{char}(v) = p|n$. Let R be the ring of integers of K_v^{ur}, and let $\mathscr{A}$, $\mathscr{J}$, and $\mathscr{C}$ be the Néron models of A, J, and C, respectively. Since $e_p < p - 1$ and J has abelian reduction at v (since $p \nmid N$), by Bosch *et al.* [BLR90, Theorem 7.5.4(iii)], the induced sequence $0 \to \mathscr{A} \to \mathscr{J} \xrightarrow{\phi} \mathscr{C} \to 0$ is exact, which means that ϕ is faithfully flat and surjective with scheme-theoretic kernel $\mathscr{A}$. Since ϕ is faithfully flat with smooth kernel, ϕ is smooth (see, e.g., [BLR90, 2.4.8]). By Lemma 3.6, $\mathscr{J}(R) \to \mathscr{C}(R)$ is a surjection; i.e., $J(K_v^{\mathrm{ur}}) \to C(K_v^{\mathrm{ur}})$ is a surjection.

So $\mathrm{res}_v(\pi(x))$ is unramified, and again by Milne [Mil86, Proposition 3.8],

$$H^1(K_v^{\mathrm{ur}}/K_v, A) \cong H^1(K_v^{\mathrm{ur}}/K_v, \Phi_{A,v}(\bar{\mathbf{F}}_v)).$$

But $H^1(K_v^{\mathrm{ur}}/K_v, \Phi_{A,v}(\bar{\mathbf{F}}_v)) = \{0\}$, since $\Phi_{A,v}(\bar{\mathbf{F}}_v)$ is trivial, as A has good reduction at v (because $p \nmid N$). Thus $\mathrm{res}_v(\pi(x)) = 0$. ∎

4. SOME EXAMPLES

This section contains some examples of visible and invisible elements of Shafarevich–Tate groups. Section 4.1 uses Theorem 3.1 to produce nontrivial visible elements of $\text{Ш}(A)$, where A is a 20-dimensional modular abelian variety, thus giving evidence for the BSD conjecture. In Section 4.2 we show that an invisible Shafarevich–Tate group from [CM00] becomes visible at a higher level.

In [AS02], we describe the notation used (which is standard) and the algorithms that we used to carry out the computations described below. We also report on a large number of similar computations, which were performed using the second author's modular symbols package, which is part of Magma (see [BCP97]).

4.1. *Visibility in an Abelian Variety of Dimension 20.* Using the methods described in [AS02], we find that $S_2(\Gamma_0(389))$ contains exactly five Galois-conjugacy classes of newforms, and these are defined over extensions of $\mathbf{Q}$ of degrees 1, 2, 3, 6, and 20. Thus, $J = J_0(389)$ decomposes, up to isogeny, as a product $A_1 \times A_2 \times A_3 \times A_6 \times A_{20}$ of abelian varieties, where $d = \dim A_d$ and A_d is the quotient corresponding to the appropriate Galois-conjugacy class of newforms.

Next we consider the arithmetic of each A_d. Using [AS02], we find that

$$L(A_1, 1) = L(A_2, 1) = L(A_3, 1) = L(A_6, 1) = 0,$$

and

$$\frac{L(A_{20}, 1)}{\Omega_{A_{20}}} = \frac{5^2 \cdot 2^?}{97},$$

where $2^?$ is a power of 2. Using [AS02], we find that $\#A_{20}(Q) = 97$ and the Tamagawa number of A_{20} at 389 is also 97. The BSD Conjecture then predicts that $\#\text{Ш}(A_{20}) = 5^2 \cdot 2^?$. The following proposition provides support for this conjecture.

Proposition 4.1. *There is an inclusion*

$$(\mathbf{Z}/5\mathbf{Z})^2 \cong A_1(\mathbf{Q})/5A_1(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_J(\text{Ш}(A_{20}^\vee)).$$

Proof. Let $A = A_{20}^\vee, B = A_1^\vee = A_1$ and $J = A + B \subset J_0(389)$. Using algorithms in [AS02], we find that $A \cap B \cong (Z/4)^2 \times (Z/5Z)^2$, so $B[5] \subset$

A. Since 5 does not divide the numerator of $(389-1)/12$, it does not divide the Tamagawa numbers or the orders of the torsion subgroups of A, B, J, and J/B (we also verified this using a modular symbols computations), so Theorem 3.1 implies that there is an injective map

$$A_1(\mathbf{Q})/5A_1(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_J(\text{Ш}(A_{20}^{\vee})).$$

To finish, note that Cremona [Cre97] has verified that $A_1(\mathbf{Q}) \approx \mathbf{Z}\times\mathbf{Z}$. ∎

4.2. *Invisible Elements that Becomes Visible at Higher Level.* Consider the elliptic curve E of conductor $5389 = 17\cdot 317$ defined by the equation

$$y^2 + xy + y = x^3 - 35\,590x - 2\,587\,197.$$

In [CM00], Cremona and Mazur observe that the BSD conjecture implies that $\#\text{Ш}(E) = 9$, but they find that $\mathrm{Vis}_{J_0(5389)}(\text{Ш}(E)[3]) = \{0\}$. We will now verify, without assuming any conjectures, that $9|\#\text{Ш}(E)$ and that these 9 elements of $\text{Ш}(E)$ are visible in $J_0(5389\cdot 7)$.

First note that the mod 3 representation $\rho_{E,3}$ attached to E is irreducible because E is semistable and admits no 3-isogeny (according to [Cre]). The newform attached to E is

$$f_E = q + q^2 - 2q^3 - q^4 + 2q^5 - 2q^6 - 2q^7 + \cdots,$$

and $a_7^2 = (-2)^2 \equiv (7+1)^2 \pmod 3$, so Ribet's level-raising theorem [Rib90] implies that there is a newform g of level $7\cdot 5389$ that is congruent modulo 3 to f_E. This observation led us to the following proposition.

PROPOSITION 4.2. *Map E to $J_0(7\cdot 5389)$ by the sum of the two maps on Jacobians induced by the two degeneracy maps $X_0(7\cdot 5389) \to X_0(5389)$. The image E' of E in $J_0(7\cdot 5389)$ is 2-isogenous to E and*

$$(\mathbf{Z}/3\mathbf{Z})^2 \subset \mathrm{Vis}_{J_0(7\cdot 5389)}(\text{Ш}(E')).$$

Proof. It is easy to see from the discussion in [Rib90] that the kernel of the sum of the two degeneracy maps $J_0(5389) \to J_0(7\cdot 5389)$ is a group of 2-power order, so E' is isogenous to E via an isogeny of degree a power of 2.

Consider the elliptic curve F defined by $y^2 - y = x^3 + x^2 + 34x - 248$. Using Cremona's programs `tate` and `mwrank` we find that F has conductor $7\cdot 5389$, and that $F(\mathbf{Q}) \cong \mathbf{Z}\times\mathbf{Z}$. The Tamagawa numbers of F at 7, 17, and 317 are 1, 2, and 1, respectively. The newform attached to F is

$$f_F = q - 2q^2 + q^3 + 2q^4 - q^5 - 2q^6 - q^7 + \cdots$$

and, by Sturm [Stu87], we prove that $f_E(q) + f_E(q^7) \equiv f_F \pmod{3}$ by checking this congruence for the first $7632 = [\mathrm{SL}_2(\mathbf{Z}) : \Gamma_0(7 \cdot 5389)]/6$ terms. Since $2 \leqslant k < 3$ and $3 \nmid 7 \cdot 5389$, the first part of the multiplicity one theorem of [Edi92, Sect. 9] implies that $F[3] = E'[3]$.

Finally, we apply Theorem 3.1 with $A = E', B = F, J = A + B \subset J_0(7 \cdot 5389), N = 7 \cdot 5389$, and $n = 3$. It is routine to check the hypothesis. For example, the hypothesis that J/B has no $\mathbf{Q}$-rational 3-torsion can be checked as follows. Cremona's online tables imply that E admits no 3-isogeny, so $E[3]$ is irreducible. Since J/B is isogenous to E, the representation $(J/B)[3]$ is also irreducible, so $(J/B)(\mathbf{Q})[3] = \{0\}$. Thus, by Theorem 3.1, we have $(\mathbf{Z}/3\mathbf{Z})^2 \subset \mathrm{Vis}_J(\text{Ш}(E'))$. To finish the proof, note that $\mathrm{Vis}_J(\text{Ш}(E')) \subset \mathrm{Vis}_{J_0(7 \cdot 5389)}(\text{Ш}(E'))$. ∎

Since E' is 2-isogenous to E and $9 | \#\text{Ш}(E')$, it follows that $9 | \#\text{Ш}(E)$, as predicted by the BSD conjecture.

ACKNOWLEDGMENTS

We thank Barry Mazur for his generous guidance, Brian Conrad for his extensive assistance, Ralph Greenberg for suggesting the use of restriction of scalars in Section 1.3, Fabrizio Andreatta for suggesting that a semistability hypothesis was unnecessary in Theorem 3.1, and Loic Merel, Bjorn Poonen, and Ken Ribet for helpful conversations. The first author would like to thank the Mathematical Sciences Research Institute in Berkeley and the Institut des Hautes Études Scientifiques in France, and the second author the Max Planck Institute in Bonn, for their generous hospitality.

REFERENCES

[AS02] A. Agashe and W.A. Stein, Visible Evidence for the Birch and Swinnerton–Dyer Conjecture for Rank 0 Modular Abelian Varieties, preprint.

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, "Néron Models," Springer-Verlag, Berlin, 1990.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, The magma algebra system. I. The user language, *J. Symbolic Comput.* **24**, No. 3–4 (1997), 235–265; Computational Algebra and Number Theory, London, 1993.

[Cas63] J.W.S. Cassels, Arithmetic on curves of genus 1. V. Two counterexamples, *J. London Math. Soc.* **38** (1963), 244–248.

[Cre] J.E. Cremona, *Elliptic curves of conductor* $\leqslant$12000, `http://www.maths.nott.ac.uk/personal/jec/ftp/data/`.

[Cre97] J.E. Cremona, "Algorithms for Modular Elliptic Curves," 2nd ed., Cambridge Univ. Press, Cambridge, UK, 1997.

[CM00] J.E. Cremona and B. Mazur, Visualizing elements in the Shafarevich-Tate group, *Exp. Math.* **9**, No. 1 (2000), 13–28.

[Edi92] B. Edixhoven, The weight in Serre's conjectures on modular forms, *Invent. Math.* **109**, No. 3 (1992), 563–594.

[Gro65] A. Grothendieck, Éléments de géométric algébrique. IV. Étude locale des schémas et des morphismes de schémas. II, *Inst. Hautes Études Sci. Publ. Math.* No. 24 (1965), 231.

[Gro66] A. Grothendieck, Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. III, *Inst. Hautes Études Sci. Publ. Math.* No. 28 (1966), 255.

[Gro67] A. Grothendieck, Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas IV, *Inst. Hautes Études Sci. Publ. Math.* No. 32 (1967), 361.

[Gro70] A. Grothendieck, "Schémas en groupes. I: Propriétés générales des schémas en groupes," Springer-Verlag, Berlin, 1970.

[Kle01] T. Klenke, "Modular Varieties and Visibility," Ph.D. thesis, Harvard University, 2001.

[LT58] S. Lang and J. Tate, Principal homogeneous spaces over abelian varieties, *Amer. J. Math.* **80** (1958), 659–684.

[Maz99] B. Mazur, Visualizing elements of order three in the Shafarevich-Tate group, *Asian J. Math.* **3**, No. 1 (1999), 221–232.

[Mil86] J.S. Milne, "Arithmetic Quality Theorems," Academic Press Inc., Boston, MA, 1986.

[O'N01] C. O'Neil, The period-index obstruction for elliptic curves, *J. Number Theory*, to appear.

[Rib90] K.A. Ribet, Raising the levels of modular representations, *in* "Séminaire de Théorie des Nombres," Paris 1987–88, pp. 259–271, Birkhäuser, Boston, MA, 1990.

[Ser79] J-P. Serre, "Local Fields," Springer-Verlag, New York, 1979 (translated from the French by Marvin Jay Greenberg).

[Stu87] J. Sturm, On the congruence of modular forms, "Number Theory (New York, 1984–1985)," pp. 275–280, Springer, Berlin, 1987.

13 Visible Evidence For The Birch And Swinnerton-Dyer Conjecture For Modular Abelian Varieties Of Analytic Rank Zero, with A. Agashe

MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

VISIBLE EVIDENCE FOR THE BIRCH AND SWINNERTON-DYER CONJECTURE FOR MODULAR ABELIAN VARIETIES OF ANALYTIC RANK ZERO (WITH AN APPENDIX BY J. CREMONA AND B. MAZUR)

AMOD AGASHE AND WILLIAM STEIN

Abstract. This paper provides evidence for the Birch and Swinnerton-Dyer conjecture for analytic rank 0 abelian varieties A_f that are optimal quotients of $J_0(N)$ attached to newforms. We prove theorems about the ratio $L(A_f, 1)/\Omega_{A_f}$, develop tools for computing with A_f, and gather data about certain arithmetic invariants of the nearly 20000 abelian varieties A_f of level ≤ 2333. Over half of these A_f have analytic rank 0, and for these we compute upper and lower bounds on the conjectural order of $Ш(A_f)$. We find that there are at least 168 such that the Birch and Swinnerton-Dyer Conjecture implies that $Ш(A_f)$ is divisible by an odd prime, and we prove for 39 of these that the odd part of the conjectural order of $Ш(A_f)$ really divides $\#Ш(A_f)$ by constructing nontrivial elements of $Ш(A_f)$ using visibility theory. We also give other evidence for the conjecture. The appendix, by Cremona and Mazur, fills in some gaps in the theoretical discussion in their paper on visibility of Shafarevich-Tate groups of elliptic curves.

Contents

Received by the editor March 8, 2004.

1991 *Mathematics Subject Classification.* Primary 11G40; Secondary 11F11, 11G10, 14K15, 14H25, 14H40.

Key words and phrases. Birch and Swinnerton-Dyer conjecture, modular abelian variety, visibility, Shafarevich-Tate groups.

©1997 American Mathematical Society

1. Introduction

Let N be a positive integer, and f be a newform of weight 2 on $\Gamma_0(N)$. A construction due to Shimura associates to f an abelian variety quotient A_f of $J_0(N)$. We say that A_f has *analytic rank zero* if its L-function $L(A_f, s)$ is nonzero at $s = 1$. In this paper we give evidence for the Birch and Swinnerton-Dyer conjecture for analytic rank 0 abelian varieties A_f of arbitrary dimension. For such abelian varieties, the conjecture asserts that $A_f(\mathbf{Q})$ is finite, and gives a formula for the order of the Shafarevich-Tate group Ш(A_f).

Kolyvagin and Logachev proved in [KL89, KL92] that if $L(A_f, 1) \neq 0$, then $A_f(\mathbf{Q})$ and Ш(A_f) are both finite. To the best of our knowledge, Birch and Swinnerton-Dyer's formula for #Ш(A_f) has not been completely verified for a single abelian variety A_f of dimension greater than one. In [KL92, §1.6] Kolyvagin and Logachev remark that if one were able to compute the height of a certain Heegner point, their methods could be used to find an upper bound on #Ш(A_f), but we have not done this. Instead, in this paper we focus on computing *nonzero subgroups* of Ш(A_f) when the conjecture predicts that Ш(A_f) is nonzero.

Inspired by work of Cremona and Mazur (see [CM]), we had the idea to reverse their methods and prove, in some cases, that #Ш(A_f) is at least as big as predicted by the Birch and Swinnerton-Dyer conjecture. Instead of assuming that Ш(A_f) is as predicted by the conjecture and trying to understand whether or not it is visible in $J_0(N)$, we prove a theorem (see [AS02]) that allows us to sometimes construct the odd part of Ш(A_f) without assuming any conjectures. After developing algorithms that allow us to compute the conjectural order of Ш(A_f) in most cases, we analyzed the 19608 abelian varieties A_f of level ≤ 2333, and constructed the tables of Section 5. This resulted in the first systematic experimental evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of dimension greater than 2 (see [FpS$^+$01] for dimension 2).

This paper is organized as follows. In Section 2 we review background about modular abelian varieties and state the Birch and Swinnerton-Dyer conjecture. Section 3 explains the basic facts about quotients A_f of $J_0(N)$ that one needs to know in order to compute with them. In Section 4 we discuss a generalization of the Manin constant, derive a formula for the ratio $L(A_f, 1)/\Omega_{A_f}$, and bound the denominator of this ratio, thus giving some theoretical evidence towards the Birch and Swinnerton-Dyer conjecture. Section 5 reports on our construction of a table of 168 rank 0 abelian varieties A_f of level ≤ 2333 such that the Birch and

Swinnerton-Dyer conjecture predicts that $\#\text{Ш}(A_f)$ is divisible by an odd prime, and discusses what we computed to show that for 39 of the A_f there are *at least* as many elements of the odd part of $\#\text{Ш}(A_f)$ as predicted. The part of $\#\text{Ш}(A_f)$ that is coprime to the modular degree of A_f (which we define below) is a perfect square, and in the several cases where we could compute the odd part of the conjectured value of $\#\text{Ш}(A_f)$, we found the odd part to be a perfect square, which gives computational evidence for the conjecture. The appendix, written by Cremona and Mazur, fills in some gaps in the theoretical discussion in [CM].

Acknowledgment. It is a pleasure to thank Bryan Birch, Robert Coleman, Benedict Gross, Hednrik Lenstra, Dino Lorenzini, Loïc Merel, Bjorn Poonen, Ken Ribet, and John Tate for many helpful comments and discussions. Special thanks go to Barry Mazur for guiding our ideas on visibility and purchasing the second author a powerful computer, and to Allan Steel and David Kohel at Magma for their crucial computational support.

2. Background and Definitions

2.1. **Modular Forms.** Fix a positive integer N. The group

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : N \mid c \right\}$$

acts by linear fractional transformations on the extended complex upper halfplane $\mathfrak{h}^*$. As a Riemann surface, $X_0(N)(\mathbf{C})$ is the quotient $\Gamma_0(N)\backslash\mathfrak{h}^*$. There is a standard model for $X_0(N)$ over $\mathbf{Q}$ (see [Shi94, Ch. 6]), and the Jacobian $J_0(N)$ of $X_0(N)$ is an abelian variety over $\mathbf{Q}$ of dimension equal to the genus g of $X_0(N)$, which is equipped with an action of the Hecke algebra $\mathbf{T} = \mathbf{Z}[\ldots T_n \ldots]$. The space $S_2(\Gamma_0(N))$ of cuspforms of weight 2 on $\Gamma_0(N)$ is a module over $\mathbf{T}$ and $S_2(\Gamma_0(N)) \cong H^0(X_0(N), \Omega_{X_0(N)})$ as $\mathbf{T}$-modules.

2.2. **Abelian Varieties Attached to Newforms.** A *newform*

$$f = \sum_{n \geq 1} a_n q^n \in S_2(\Gamma_0(N))$$

is an eigenvector for $\mathbf{T}$ that is normalized so that $a_1 = 1$ and which lies in the orthogonal complement of the old subspace of $S_2(\Gamma_0(N))$. Let I_f denote the annihilator $\mathrm{Ann}_{\mathbf{T}}(f)$ of f in $\mathbf{T}$. Following Shimura [Shi73], attach to I_f the quotient

$$A_f = J_0(N)/I_f J_0(N),$$

which is an abelian variety over $\mathbf{Q}$ of dimension $[\mathbf{Q}(\ldots, a_n, \ldots) : \mathbf{Q}]$, which is equipped with a faithful action of $\mathbf{T}/I_f$. Moreover, A_f is an *optimal quotient* of $J_0(N)$, in the sense that $A_f^\vee \to J_0(N)$ is a closed immersion, or equivalently that the kernel of $J_0(N) \to A_f$ is connected (see [CS01, Prop. 3.3]).

Also, the complex torus $A_f(\mathbf{C})$ fits into the exact sequence

$$H_1(X_0(N), \mathbf{Z}) \to \mathrm{Hom}(S_2(\Gamma_0(N))[I_f], \mathbf{C}) \to A_f(\mathbf{C}) \to 0.$$

2.3. **The Birch and Swinnerton-Dyer Conjecture.** The conjecture of Birch and Swinnerton-Dyer makes sense for abelian varieties over fairly general global fields, but we only state a special case. This conjecture involves the L-function attached to $A = A_f$:

$$L(A, s) = \prod_{i=1}^{d} L(f^{(i)}, s) = \prod_{i=1}^{d} \left(\sum_{n \geq 1} \frac{a_n^{(i)}}{n^s} \right),$$

where $f^{(i)}$ is the ith Galois conjugate of f and $a_n^{(i)}$ is the ith Galois conjugate of a_n. It follows from work of Hecke that $L(A, s)$ has an analytic continuation to the whole complex plane and satisfies a functional equation. Birch and Swinnerton-Dyer made the following conjecture, which relates the rank of A to the order of vanishing of $L(A, s)$ at $s = 1$.

Conjecture 2.1 (Birch and Swinnerton-Dyer)**.** *The Mordell-Weil rank of A is equal to the order of vanishing of $L(A, s)$ at $s = 1$, i.e.,*

$$\dim(A(\mathbf{Q}) \otimes \mathbf{Q}) = \operatorname{ord}_{s=1} L(A, s).$$

Birch and Swinnerton-Dyer also furnished a conjectural formula for the order of the Shafarevich-Tate group

$$Ш(A) := \ker \left(H^1(\mathbf{Q}, A) \longrightarrow \prod_{\text{all places } v} H^1(\mathbf{Q}_v, A) \right).$$

(They only made their conjecture for elliptic curves, but Tate [Tat66] reformulated it a functorial way which makes sense for abelian varieties. See also [Lan91, §III.5] for another formulation.) We now state their conjecture in the special case when $L(A, 1) \neq 0$, where [KL89, KL92] implies that $Ш(A)$ is finite. The conjecture involves the Tamagawa numbers c_p of A (see Section 3.7), and the canonical volume Ω_A of $A(\mathbf{R})$ (see Section 4.2).

Conjecture 2.2 (Birch and Swinnerton-Dyer)**.** *Suppose $L(A, 1) \neq 0$. Then*

$$\frac{L(A, 1)}{\Omega_A} = \frac{\#Ш(A) \cdot \prod_{p \mid N} c_p}{\#A(\mathbf{Q})_{\text{tor}} \cdot \#A^\vee(\mathbf{Q})_{\text{tor}}},$$

where $A^\vee$ is the abelian variety dual of A.

Remark 2.3. Since $L(A, 1) \neq 0$, finiteness of $Ш(A)$ and the existence of the Cassels-Tate pairing implies that $\#Ш(A) = \#Ш(A^\vee)$, so Conjecture 2.2 can also be viewed as a formula for $\#Ш(A^\vee)$.

The algorithms outlined in this paper take advantage of the fact that A is attached to a newform in order to compute the conjectural order of $Ш(A)$ away from certain bad primes.

3. Explicit Approaches to Modular Abelian Varieties

We use the algorithms of this section to enumerate the A_f, compute information about the invariants of A_f that appear in Conjecture 2.2, and to verify the hypothesis of Theorem 3.13 in order to construct nontrivial subgroups of $Ш(A_f)$. The second author has implemented the algorithms discussed in this paper, and made many of them part of the Magma computer algebra system [BCP97].

In Section 3.1, we discuss modular symbols, which are the basic tool we use in many of the computations, and in Section 3.2 we discuss how we systematically enumerate modular abelian varieties. There is an analogue for A_f of the usual elliptic-curve modular degree, which we discuss in Section 3.3, and which we use to rule out the existence of visible elements of $\text{Ш}(A_f)$ of a certain order. In Section 3.4 we describe how to compute the intersection of two abelian varieties, which will be needed to verify the hypothesis of Theorem 3.13. In Sections 3.5 and 3.6, we describe standard methods for bounding the torsion subgroup of an abelian variety above and below. Section 3.7 reviews an algorithm for computing the odd part of the Tamagawa number c_p when $p \mid\mid N$, and discusses the Lenstra-Oort bound in the case when $p^2 \mid N$.

Unless otherwise stated, f is a newform, I_f its annihilator, and $A = A_f$ is the corresponding optimal quotient of $J_0(N)$.

3.1. **Modular Symbols.** Modular symbols are crucial to many algorithms for computing with modular abelian varieties, because they can be used to construct a *finite* presentation for $H_1(X_0(N), \mathbf{Z})$ in terms of paths between elements of $\mathbf{P}^1(\mathbf{Q}) = \mathbf{Q} \cup \{\infty\}$. They were introduced by Birch [Bir71] and studied by Manin, Mazur, Merel, Cremona, and others.

Let $\mathfrak{M}_2$ be the free abelian group with basis the set of all symbols $\{\alpha, \beta\}$, with $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, modulo the three-term relations

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0,$$

and modulo any torsion. The group $\mathrm{GL}_2(\mathbf{Q})$ acts on the left on $\mathfrak{M}_2$ by

$$g\{\alpha, \beta\} = \{g(\alpha), g(\beta)\},$$

where g acts on α and β by a linear fractional transformation. The space $\mathfrak{M}_2(\Gamma_0(N))$ of *modular symbols* for $\Gamma_0(N)$ is the quotient of $\mathfrak{M}_2$ by the subgroup generated by all elements of the form $x - g(x)$, for $x \in \mathfrak{M}_2$ and g in $\Gamma_0(N)$, modulo any torsion. A *modular symbol* for $\Gamma_0(N)$ is an element of this space, and we frequently denote the equivalence class that defines a modular symbol by giving a representative element.

Let $\mathfrak{B}_2(\Gamma_0(N))$ be the free abelian group with basis the finite set $\Gamma_0(N)\backslash\mathbf{P}^1(\mathbf{Q})$. The boundary map $\delta : \mathfrak{M}_2(\Gamma_0(N)) \to \mathfrak{B}_2(\Gamma_0(N))$ sends $\{\alpha, \beta\}$ to $[\beta] - [\alpha]$, where $[\beta]$ denotes the basis element of $\mathfrak{B}_2(\Gamma_0(N))$ corresponding to $\beta \in \mathbf{P}^1(\mathbf{Q})$. The *cuspidal modular symbols* are the kernel $\mathfrak{S}_2(\Gamma_0(N))$ of δ, and the integral homology $H_1(X_0(N), \mathbf{Z})$ is canonically isomorphic to $\mathfrak{S}_2(\Gamma_0(N))$.

Cremona's book [Cre97, §2.2] contains a concrete description of how to compute $\mathfrak{M}_2(\Gamma_0(N)) \otimes \mathbf{Q}$ using Manin symbols, which are a finite set of generators for $\mathfrak{M}_2(\Gamma_0(N))$. In general, the easiest way we have found to compute $\mathfrak{M}_2(\Gamma_0(N))$ is to compute $\mathfrak{M}_2(\Gamma_0(N)) \otimes \mathbf{Q}$, then compute the $\mathbf{Z}$-submodule of $\mathfrak{M}_2(\Gamma_0(N)) \otimes \mathbf{Q}$ generated by the Manin symbols.

3.2. **Enumerating Newforms.** Since $X_0(N)$ is defined over $\mathbf{Q}$ it is defined over $\mathbf{R}$, so complex conjugation acts on $X_0(N)(\mathbf{C})$ hence on the homology $H_1(X_0(N), \mathbf{Z})$. In terms of modular symbols, complex conjugation acts by sending $\{\alpha, \beta\}$ to $\{-\alpha, -\beta\}$. Let $H_1(X_0(N), \mathbf{Z})^+$ denote the $+1$-eigenspace for the action of the involution induced by complex conjugation, which we can compute using modular symbols. We list all newforms of a given level N by decomposing the new subspace of $H_1(X_0(N), \mathbf{Q})^+$ under the action of the the Hecke operators and listing the corresponding systems of Hecke eigenvalues (see [Ste02a]). First we compute the characteristic polynomial of T_2, and use it to break up the new space. We apply this process recursively with $T_3, T_5, \ldots$ until either we have exceeded the bound coming from [Stu87] (see [LS02]), or we have found a Hecke operator T_n whose characteristic polynomial is irreducible.

We *order the newforms* in a way that extends the ordering in [Cre97]: First sort by dimension, with smallest dimension first; within each dimension, sort in binary by the signs of the Atkin-Lehner involutions, e.g., $+++$, $++-$, $+-+$, $+--$, $-++$, etc. When two forms have the same Atkin-Lehner sign sequence, order by $|\operatorname{Tr}(a_p)|$ with ties broken by taking the positive trace first. We denote a Galois-conjugacy class of newforms by a bold symbol such as **389E**, which consists of a level and isogeny class, where **A** denotes the first class, **B** the second, **E** the fifth, **BB** the 28th, etc. As discussed in [Cre97, pg. 5], for certain small levels the above ordering, when restricted to elliptic curves, does not agree with the ordering used in the tables of [Cre97]. For example, our **446B** is Cremona's **446D**.

3.3. The Modular Degree.

Since A_f is an optimal quotient, the dual map $A_f^\vee \to J_0(N)$ is injective and the composite $\theta_f : A_f^\vee \to A_f$ has finite degree. The map θ_f is a polarization, so $\deg(\theta_f)$ is a perfect square (see Lemma 3.14). The *modular degree* of A_f is the square root of the degree of θ_f:

$$\operatorname{moddeg}(A_f) = \sqrt{\deg(\theta_f)}.$$

When $\dim A_f = 1$, $\operatorname{moddeg}(A_f)$ is the usual modular degree, i.e., the degree of $X_0(N) \to A_f$.

If M is an abelian group, let $M^* = \operatorname{Hom}_{\mathbf{Z}}(M, \mathbf{Z})$. The Hecke algebra acts in a natural way on $H_1(X_0(N), \mathbf{Z})$ and $H_1(X_0(N), \mathbf{Z})^*$, and we have a natural restriction map

$$r_f : H_1(X_0(N), \mathbf{Z})^*[I_f] \to (H_1(X_0(N), \mathbf{Z})[I_f])^*.$$

The following proposition leads to an algorithm for computing the modular degree.

Proposition 3.1. $\operatorname{coker}(r_f) \cong \ker(\theta_f)$, *so* $\operatorname{moddeg}(A_f) = \sqrt{\#\operatorname{coker}(r_f)}$.

The proposition is proved in [KS00]. The proof makes use of the *Abel-Jacobi theorem*, which realizes the Jacobian $J_0(N)(\mathbf{C})$ as a complex torus:

$$0 \to H_1(X_0(N), \mathbf{Z}) \to \operatorname{Hom}(S_2(\Gamma_0(N)), \mathbf{C}) \to J_0(N)(\mathbf{C}) \to 0,$$

where $H_1(X_0(N), \mathbf{Z})$ is embedded as a lattice of full rank in the complex vector space $\operatorname{Hom}(S_2(\Gamma_0(N)), \mathbf{C})$ using the integration pairing, and this description of $J_0(N)(\mathbf{C})$ is compatible with the action of Hecke operators.

3.4. Intersecting Complex Tori.

Let V be a finite dimensional complex vector space and let Λ be a lattice in V, so that $T = V/\Lambda$ is a complex torus. Suppose that V_A and V_B are subspaces of V such that $\Lambda_A = V_A \cap \Lambda$ and $\Lambda_B = V_B \cap \Lambda$ are lattices in V_A and V_B, respectively.

Proposition 3.2. *If $A \cap B$ is finite, then there is an isomorphism*

$$A \cap B \cong \left(\frac{\Lambda}{\Lambda_A + \Lambda_B}\right)_{\text{tor}}.$$

Proof. Extend the exact sequence

$$0 \to A \cap B \to A \oplus B \xrightarrow{(x,y)\mapsto x-y} T$$

to the following diagram:

$$\begin{tikzcd}
 & \Lambda_A \oplus \Lambda_B \arrow[r] \arrow[d] & \Lambda \arrow[r] \arrow[d] & \Lambda/(\Lambda_A + \Lambda_B) \arrow[d] \\
0 \arrow[r] \arrow[d] & V_A \oplus V_B \arrow[r] \arrow[d] & V \arrow[r] \arrow[d] & V/(V_A + V_B) \arrow[d] \\
A \cap B \arrow[r] & A \oplus B \arrow[r] & T \arrow[r] & T/(A+B).
\end{tikzcd}$$

The middle row is exact because $A \cap B$ is finite so $V_A \cap V_B = 0$.

Using the snake lemma, which connects the kernel $A \cap B$ of $A \oplus B \to T$ to the cokernel of $\Lambda_A \oplus \Lambda_B \to \Lambda$, we obtain an exact sequence

$$0 \to A \cap B \to \Lambda/(\Lambda_A + \Lambda_B) \to V/(V_A + V_B).$$

Since $V/(V_A + V_B)$ is a $\mathbf{C}$-vector space, the torsion part of $\Lambda/(\Lambda_A + \Lambda_B)$ must map to 0. No non-torsion in $\Lambda/(\Lambda_A + \Lambda_B)$ could map to 0, because if it did then $A \cap B$ would not be finite. The proposition follows. □

For abelian subvarieties of $J_0(N)$ attached to newforms, we use the proposition above as follows. The complex vector space $V = \mathrm{Hom}(S_2(\Gamma_0(N)), \mathbf{C})$ is the tangent space of $J_0(N)(\mathbf{C})$ at the identity. Setting $\Lambda = H_1(X_0(N), \mathbf{Z})$ and considering Λ as a lattice in V via the integration pairing, we have $J_0(N)(\mathbf{C}) \cong V/\Lambda$. Suppose f and g are non-conjugate newforms, and let I_f and I_g be their annihilators in the Hecke algebra $\mathbf{T}$, and let $A = A_f^\vee$ and $B = A_g^\vee$. Then $V_A = V[I_f]$ and $V_B = V[I_g]$ are the tangent spaces to A and B at the identity, respectively. The above proposition shows that the group $A \cap B$ is canonically isomorphic to $(\Lambda/(\Lambda_A + \Lambda_B))_{\mathrm{tor}}$. Here $\Lambda_A = \Lambda[I_f]$ and $\Lambda_B = \Lambda[I_g]$, because A_f and A_g are optimal quotients.

The following formula for the intersection of n subtori is obtained in a similar way to that of Proposition 3.2.

Proposition 3.3. *For $i = 1, \ldots, n$, with $n \geq 2$, let $A_i = V_i/\Lambda_i$ be a subtorus of $T = V/\Lambda$, and assume that each pairwise intersection $A_i \cap A_j$ is finite. Define a linear map*

$$f : V_1 \times \cdots \times V_n \longrightarrow V^{\oplus(n-1)}.$$

by $f(x_1, \ldots, x_n) = (x_1 - x_2,\, x_2 - x_3,\, x_3 - x_4,\, \ldots,\, x_{n-1} - x_n)$. Then

$$A_1 \cap \cdots \cap A_n \cong \left(\frac{\Lambda^{\oplus(n-1)}}{f(\Lambda_1 \oplus \cdots \oplus \Lambda_n)} \right)_{\mathrm{tor}}.$$

3.5. **Bounding the Torsion From Above.** In this section we recall the standard upper bound on the order of $\#A(\mathbf{Q})_{\mathrm{tor}}$, and illustrate its usefulness.

Let $f = \sum a_n q^n$ be a weight 2 newform on $\Gamma_1(N)$ with Nebentypus character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ (recall that f is a form on $\Gamma_0(N)$ if and only if $\varepsilon = 1$), and let $A = A_f$ be the corresponding optimal quotient of $J_1(N)$, as in [Shi73]. Shimura proved in [Shi94, Ch. 7] that the local Euler factor of A_f at p is

$$L_p(A_f, s) \;=\; \prod_{\sigma : K_f \hookrightarrow \overline{\mathbf{Q}}} \frac{1}{1 - \sigma(a_p) p^{-s} + \sigma(\varepsilon(p)) p^{1-2s}}$$

by showing that the characteristic polynomial F_p of Frobenius on any ℓ-adic Tate module of $A_{\mathbf{F}_p}$ (for $\ell \nmid pN$) is

$$F_p(X) = \prod_{\sigma : K_f \hookrightarrow \overline{\mathbf{Q}}} X^2 - \sigma(a_p)X + \sigma(\varepsilon(p))p,$$

where $K_f = \mathbf{Q}(\ldots, a_n, \ldots)$. Let $\mathbf{Q}(\varepsilon)$ be the field generated by the values of ε (note that $\mathbf{Q}(\varepsilon) \subset K_f$), and for any $p \nmid N$ let $G_p(X) \in \mathbf{Q}(\varepsilon)[X]$ be the characteristic polynomial of left multiplication by a_p on the $\mathbf{Q}(\varepsilon)$-vector space K_f, which is a polynomial of degree $d' = [K_f : \mathbf{Q}(\varepsilon)]$. Then

$$F_p(X) = \mathrm{Norm}_{\mathbf{Q}(\varepsilon)/\mathbf{Q}} \left(X^{d'} \cdot G_p \left(X + \frac{\varepsilon(p)p}{X} \right) \right),$$

so

$$\begin{aligned} \#A_{\mathbf{F}_p}(\mathbf{F}_p) &= \deg(1 - \mathrm{Frob}_p) = |\det(1 - \mathrm{Frob}_p)| \\ &= |F_p(1)| = |\,\mathrm{Norm}_{\mathbf{Q}(\varepsilon)/\mathbf{Q}}(G_p(1 + \varepsilon(p)p))|. \end{aligned}$$

If $p \nmid N$ is odd, standard facts about formal groups imply that the reduction map $A(\mathbf{Q})_{\mathrm{tor}} \to A_{\mathbf{F}_p}(\mathbf{F}_p)$ is injective, so

$$\#A(\mathbf{Q})_{\mathrm{tor}} \mid \gcd \left\{ \#A_{\mathbf{F}_p}(\mathbf{F}_p) \,:\, \text{primes } p \nmid 2N \right\}.$$

Likewise, since $A^\vee$ is isogenous to A, the same bound applies to $A^\vee(\mathbf{Q})_{\mathrm{tor}}$, since $A^\vee$ and A have the same L-series.

The upper bound is the same for every abelian variety isogenous to A, so it is not surprising that it is not sharp in general. For example, let E (resp., F) be the elliptic curve labeled **30A1** (resp. **30A2**) in Cremona's tables [Cre97]. Then E and F are isogenous, $E(\mathbf{Q}) \approx \mathbf{Z}/6\mathbf{Z}$, and $F(\mathbf{Q}) \approx \mathbf{Z}/12\mathbf{Z}$, so

$$12 \mid \gcd \left\{ \#E_{\mathbf{F}_p}(\mathbf{F}_p) \,:\, \text{primes } p \nmid 2N \right\}.$$

(Incidentally, since $\#E(\mathbf{F}_5) = 12$, the gcd is 12.) For answers to some related deep questions about this gcd, see [Kat81].

Example 3.4. Let

$$f = q + (-1 + \sqrt{2})q^2 + q^3 + (-2\sqrt{2} + 1)q^4 - 2\sqrt{2}q^5 + \cdots \in S_2(\Gamma_0(39))$$

be the form **39B**. Then $G_5(X) = X^2 - 8$, so

$$\#A_f(\mathbf{Q})_{\mathrm{tor}} \mid G_5(1 + 5) = 28.$$

We find in [FpS$^+$01] that A_f is isogenous to the Jacobian J of $y^2 + (x^3 + 1)y = -5x^4 - 2x^3 + 16x^2 - 12x + 2$ and that $\#J(\mathbf{Q}) = 28$. However A_f is not isomorphic to J since, as reported in Table 2 of [FpS$^+$01], the Tamagawa numbers of J are $c_3 = 28$, $c_{13} = 1$, whereas the methods of Section 3.7 below show that the Tamagawa numbers of A_f are $c_3 = 14$, $c_{13} = 2$. The authors do not know for sure whether $\#A_f(\mathbf{Q}) = 28$, but in Example 3.6 below we show that $14 \mid \#A_f(\mathbf{Q})$. (Also, using the computational techniques of this paper one sees that the Birch and Swinnerton-Dyer conjecture implies that $\#A_f(\mathbf{Q}) = 28$.)

Example 3.5. Let

$$f = q + (-\zeta_6 - 1)q^2 + (2\zeta_6 - 2)q^3 + \zeta_6 q^4 + (-2\zeta_6 + 1)q^5 + \cdots$$

be one of the two Galois-conjugate newforms in $S_2(\Gamma_1(13))$. This form has character $\varepsilon : (\mathbf{Z}/13\mathbf{Z})^* \to \mathbf{C}^*$ of order 6, and $A_f = J_1(13)$. We have $G_3(X) = X - 2\zeta_6 + 2$ and $\varepsilon(3) = -\zeta_6$, so

$$\begin{aligned}\#J_1(13)(\mathbf{Q})_{\mathrm{tor}} \mid \#J_1(13)(\mathbf{F}_3) &= |\operatorname{Norm}(G_3(1-3\zeta_6))| \\ &= |\operatorname{Norm}(-5\zeta_6+3)| = 19.\end{aligned}$$

In fact Ogg proved that $J_1(13)(\mathbf{Q})_{\mathrm{tor}} \approx \mathbf{Z}/19\mathbf{Z}$ (see [Ogg73] and [MT74]).

3.6. **Bounding the Torsion From Below.** A cusp $\alpha \in \Gamma_0(N)\backslash\mathbf{P}^1(\mathbf{Q}) \subset X_0(N)$ defines a point $(\alpha) - (\infty) \in J_0(N)(\overline{\mathbf{Q}})_{\mathrm{tor}}$. The rational cuspidal subgroup C of $J_0(N)(\mathbf{Q})_{\mathrm{tor}}$ generated by $\mathbf{Q}$-rational cusps is of interest because the order of the image of C in $A_f(\mathbf{Q})_{\mathrm{tor}}$ provides a lower bound on $\#A_f(\mathbf{Q})_{\mathrm{tor}}$. Stevens [Ste82, §1.3] computed the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the subgroup of $J_0(N)(\overline{\mathbf{Q}})$ generated by all cusps (and for other congruence subgroups besides $\Gamma_0(N)$). He found that $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on the cusps through $\mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q}) \cong (\mathbf{Z}/N\mathbf{Z})^*$, and that $d \in (\mathbf{Z}/N\mathbf{Z})^*$ acts by $x/y \mapsto x/(d'y)$, where $dd' \equiv 1 \pmod{N}$. Thus, e.g., $(0) - (\infty) \in J_0(N)(\mathbf{Q})_{\mathrm{tor}}$, and if N is square-free then all cusps are rational.

To compute the image of C in $A_f(\mathbf{Q})_{\mathrm{tor}}$, first make a list of inequivalent cusps using, e.g., the method described in [Cre97, §2.2, pg. 17]. Keep only the $\mathbf{Q}$-rational cusps, which can be determined using the result of Stevens above and [Cre97, Prop. 2.2.3] (when N is squarefree all cusps are rational). Next compute the subgroup $\mathcal{C}$ of $\mathfrak{M}_2(\Gamma_0(N))$ generated by modular symbols $\{\alpha, \infty\}$, where α is a $\mathbf{Q}$-rational cusp. The image of C in $A_f(\mathbf{Q})_{\mathrm{tor}}$ is isomorphic to the image of $\mathcal{C}$ in

$$P = \Phi_f(\mathfrak{M}_2(\Gamma_0(N)))/\Phi_f(\mathfrak{S}_2(\Gamma_0(N))),$$

where $\Phi_f : \mathfrak{M}_2(\Gamma_0(N)) \to \mathrm{Hom}(S_2(\Gamma_0(N))[I_f], \mathbf{C})$ is defined by the integration pairing. To keep everything rational, note that P can be computed using any map with the same kernel as Φ_f; for example, such a map can be constructed by finding a basis for $\mathrm{Hom}(\mathfrak{M}_2(\Gamma_0(N)), \mathbf{Q})[I_f]$ as described at the end of Section 4.2).

Example 3.6. Let the notation be as in Example 3.4. The cusps on $X_0(39)$ are represented by 0, ∞, $-1/9$, and $-4/13$, and since $N = 39$ is squarefree, these cusps are all rational. Using Magma we find that the image of C in $A_f(\mathbf{Q})_{\mathrm{tor}}$ is isomorphic to $\mathbf{Z}/14\mathbf{Z}$. Thus $A_f(\mathbf{Q})_{\mathrm{tor}}$ is isomorphic to one of $\mathbf{Z}/14\mathbf{Z}$, $\mathbf{Z}/28\mathbf{Z}$, or $\mathbf{Z}/14\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, but we do not know which.

Example 3.7. Let

$$f = q + \frac{1+\sqrt{5}}{2}q^2 + \frac{1-\sqrt{5}}{2}q^3 + \frac{5+\sqrt{5}}{2}q^4 + \cdots \in S_2(\Gamma_0(175))$$

be the form **175D**. The cusps of $X_0(175)$ are represented by

$$0,\ \infty,\ \frac{1}{25},\ \frac{1}{28},\ \frac{1}{30},\ \frac{1}{35},\ \frac{1}{45},\ \frac{1}{60},\ \frac{1}{65},\ \frac{1}{70},\ \frac{1}{105},\ \frac{1}{140}.$$

The $\mathbf{Q}$-rational cusps in this list are $0, \infty, \frac{1}{25}, \frac{1}{28}$, and these generate a subgroup of $A_f(\mathbf{Q})_{\mathrm{tor}}$ of order 2. (Incidentally, the group generated by all cusps, both rational and not, is isomorphic to $\mathbf{Z}/32\mathbf{Z}$.) Using a_p for $p \leq 17$ and the method of the previous section, we see that $\#A_f(\mathbf{Q})_{\mathrm{tor}} \mid 4$. The authors do not know if the cardinality is 2 or 4.

Example 3.8. The form **209C** is

$$f = q + \alpha q^2 + (1/2\alpha^4 - \alpha^3 - 5/2\alpha^2 + 4\alpha + 1)q^3 + (\alpha^2 - 2)q^4 + \cdots,$$

where $\alpha^5 - 2\alpha^4 - 6\alpha^3 + 10\alpha^2 + 5\alpha - 4 = 0$. As above, we find that $\#A_f(\mathbf{Q})_{\mathrm{tor}}$ divides 5. The image of the (rational) cuspidal subgroup in $A_f(\overline{\mathbf{Q}})_{\mathrm{tor}}$ is isomorphic to $\mathbf{Z}/5\mathbf{Z}$, so $A_f(\mathbf{Q})_{\mathrm{tor}} \approx \mathbf{Z}/5\mathbf{Z}$.

3.7. **Tamagawa Numbers.** Suppose $p \mid N$ and let $\Phi_{A,p}$ denote the component group of A at p, which is defined by the following exact sequence:

$$0 \to \mathcal{A}^0_{\mathbf{F}_p} \to \mathcal{A}_{\mathbf{F}_p} \to \Phi_{A,p} \to 0,$$

where $\mathcal{A}_{\mathbf{F}_p}$ is the closed fiber of the Néron model of A over $\mathbf{Z}_p$ and $\mathcal{A}^0_{\mathbf{F}_p}$ is the component of $\mathcal{A}_{\mathbf{F}_p}$ that contains the identity.

Definition 3.9. The *Tamagawa number* of A at p is

$$c_p = c_{A,p} = \#\Phi_{A,p}(\mathbf{F}_p).$$

When $p \mid\mid N$, the second author found a computable formula for $\#\Phi_{A,p}(\overline{\mathbf{F}}_p)$ and (sometimes only up to a power of 2) for $\#\Phi_{A,p}(\mathbf{F}_p)$. There is a discussion about how to compute this number in [KS00] and [CS01] contains a proof of the formula. Note also that in this case the Tamagawa number of A at p is the same as the Tamagawa number of $A^\vee$ at p.

When $p^2 \mid N$ the authors do not know an algorithm to compute c_p. However, in this case Lenstra and Oort (see [LO85]) proved that

$$\sum_{\ell \neq p} (\ell - 1) \operatorname{ord}_\ell(\#\Phi_{A,p}(\overline{\mathbf{F}}_p)) \leq 2 \dim(A_f),$$

so if $\ell \mid \#\Phi_{A,p}(\overline{\mathbf{F}}_p)$ then $\ell \leq 2 \cdot \dim(A_f) + 1$ or $\ell = p$. (Here $\operatorname{ord}_\ell(x)$ denotes the exponent of the largest power of ℓ that divides x.)

Example 3.10. Let f be **39B** as in Example 3.4. Running the algorithm of [KS00], we find that $c_3 = 14$ and $c_{13} = 2$.

Example 3.11. Let f be **175D** as in Example 3.7. Running the algorithm of [KS00], we find that $c_7 = 1$, and the Lenstra-Oort bound implies that the only possible prime divisors of c_5 are 2, 3, and 5.

3.8. **Visibility Theory.** We briefly recall visibility theory, which we will use to construct elements of Shafarevich-Tate groups. Section 6 contains another approach to the results reported in this section, but in the special case of elliptic curves.

Definition 3.12. Let $\iota : A \hookrightarrow J$ be an embedding of abelian varities over $\mathbf{Q}$. The *visible subgroup of* Ш(A) *with respect to the embedding* ι is

$$\mathrm{Vis}_J(Ш(A)) = \mathrm{Ker}(Ш(A) \to Ш(J)).$$

The following is a special case of Theorem 3.1 of [AS02].

Theorem 3.13. *Let A and B be abelian subvarieties of an abelian variety J over $\mathbf{Q}$ such that $A(\overline{\mathbf{Q}}) \cap B(\overline{\mathbf{Q}})$ is finite. Let N be an integer divisible by the residue characteristics of primes of bad reduction for J (e.g., the conductor of J). Suppose p is a prime such that*

$$p \nmid 2 \cdot N \cdot \#(J/B)(\mathbf{Q})_{\mathrm{tor}} \cdot \#B(\mathbf{Q})_{\mathrm{tor}} \cdot \prod_\ell c_{A,\ell} \cdot c_{B,\ell},$$

where $c_{A,\ell} = \#\Phi_{A,\ell}(\mathbf{F}_\ell)$ (resp., $c_{B,\ell}$) is the Tamagawa number of A (resp., B) at ℓ. Suppose furthermore that $B[p](\overline{\mathbf{Q}}) \subset A(\overline{\mathbf{Q}})$ as subgroups of $J(\overline{\mathbf{Q}})$. Then there is a natural map

$$\varphi : B(\mathbf{Q})/pB(\mathbf{Q}) \to \mathrm{Vis}_J(\text{Ш}(A))$$

such that $\dim_{\mathbf{F}_p} \ker(\varphi) \leq \dim_{\mathbf{Q}} A(\mathbf{Q}) \otimes \mathbf{Q}$.

We return to the situation where $A = A_f$ is an optimal quotient of $J_0(N)$ attached to a newform. In Proposition 3.15 below we show that $\mathrm{Vis}_{J_0(N)}(\text{Ш}(A^\vee))$ is annihilated by multiplication by $\mathrm{moddeg}(A)$ (see also [CM, p.19]). We first state a lemma; the outline of the proof was indicated to us by B. Poonen.

Lemma 3.14. *Let A be an abelian variety over k, where k is a field, and let $\lambda : A \to A^\vee$ be a polarization. Suppose either that k has characteristic 0 or that its characteristic does not divide the degree of λ. Then there is a finite abelian group H such that $\ker(\lambda) \approx H \times H$ as groups.*

Proof. We work in the setting of Section 16 of [Mil86], using the notation used there. Consider the pairing

$$e^\lambda : \mathrm{Ker}(\lambda) \times \mathrm{Ker}(\lambda) \to \mu_m \subseteq \overline{k}^*,$$

as in [Mil86, p. 135], where m is an integer that kills $\mathrm{Ker}(\lambda)$. We will show that this pairing is nondegenerate.

Suppose $a \in \mathrm{Ker}(\lambda)$ is such that $e^\lambda(a, a') = 1$ for all $a' \in \mathrm{Ker}(\lambda)$. Let $a'' \in A^\vee[m]$. There exists an isogeny $\lambda' : A^\vee \to A$ such that $\lambda' \circ \lambda$ is multiplication by m on A and $\lambda \circ \lambda'$ is multiplication by m on $A^\vee$ (to construct λ', note that λ' is the quotient map $A^\vee \to A^\vee/\lambda(A[m])$). Pick an element $b \in A(\overline{k})$ such that $\lambda b = a''$. Then $mb = \lambda'(\lambda b) = \lambda'(a'')$. So $\overline{e}_m(a, a'') = \overline{e}_m(a, \lambda b) = e^\lambda(a, \lambda' a'') = 0$ (note that $\lambda(\lambda' a'') = ma'' = 0$, so that $\lambda' a'' \in \mathrm{Ker}(\lambda)$). This is true for all $a'' \in A^\vee[m]$, so the non-degeneracy of $\overline{e}_m$ ([Mil86, p. 131]) implies that $a = 0$.

Similarly, suppose $a' \in \mathrm{Ker}(\lambda)$ is such that $e^\lambda(a, a') = 1$ for all $a \in \mathrm{Ker}(\lambda)$. Since e^λ is skew-symmetric ([Mil86, p. 135]), this implies that $e^\lambda(a', a) = 1$ for all $a \in \mathrm{Ker}(\lambda)$. Then by the previous paragraph, $a' = 0$. This finishes the proof of non-degeneracy.

As mentioned before, the pairing e^λ is skew-symmetric. It is alternating because it extends to pairings on Tate modules (denoted by e^λ_ℓ in [Mil86, p. 132]), and the latter take values in a torsion-free group, so there is no distinction between skew-symmetric and alternating.

Now the lemma follows from the fact that if G is a finite abelian group with an alternating nondegenerate pairing, then there is a finite abelian group H such that $G \approx H \times H$ as groups (e.g., see [Del01, Prop. 2]). □

Proposition 3.15. *Let $m_A = \mathrm{moddeg}(A)$. We have*

$$\mathrm{Vis}_{J_0(N)}(\text{Ш}(A^\vee)) \subset \text{Ш}(A^\vee)[m_A].$$

Proof. The polarization θ_f (from Section 3.3) is the composite map $A^\vee \to J_0(N) \to A$. Let e_A be the exponent of the finite group $\ker(\theta_f)$. By Lemma 3.14, multiplication by m_A kills $\ker(\theta_f)$, so $e_A \mid m_A$. Also θ_f factors through multiplication by e_A, so there is a map $\theta'_f : A \to A^\vee$ such that $\theta'_f \circ \theta_f$ is multiplication by e_A. If ϕ is a map of abelian varieties (over $\mathbf{Q}$), let ϕ_* denote the corresponding map on

Shafarevich-Tate groups. Since $\mathrm{Vis}_{J_0(N)}(Ш(A^\vee))$ is contained in $\ker((\theta_f)_*)$, it is also contained in

$$\ker((\delta' \circ \delta)_*) = Ш(A^\vee)[e_A] \subset Ш(A^\vee)[m_A].$$

□

Since $Ш(A^\vee)[n]$ is finite for any n, we obtain the following corollary.

Corollary 3.16. $\mathrm{Vis}_{J_0(N)}(Ш(A^\vee))$ *is finite.*

4. The Quotient $L(A,1)/\Omega_A$

Fix a newform $f \in S_2(\Gamma_0(N))$, let I_f be the annihilator of f in $\mathbf{T}$, and $A = A_f = J_0(N)/I_f J_0(N)$ the corresponding optimal quotient. Suppose for the rest of this section that $L(A,1) \neq 0$.

4.1. The Manin Constant.

When trying to compute the conjectural order of $Ш(A)$, we try to compute the quotient $L(A,1)/\Omega_A$, but find that it is easier to compute $c_A \cdot L(A,1)/\Omega_A$ where c_A is the Manin constant of A, which is defined as follows:

Definition 4.1 (Manin constant)**.** The *Manin constant* of A is

$$c_A = \#\left(\frac{S_2(\Gamma_0(N),\mathbf{Z})[I_f]}{H^0(\mathcal{A},\Omega_{\mathcal{A}/\mathbf{Z}})}\right) \in \mathbf{Z},$$

where we consider $H^0(\mathcal{A},\Omega_{\mathcal{A}/\mathbf{Z}})$ as a submodule of $S_2(\Gamma_0(N),\mathbf{Q})$ using

$$H^0(\mathcal{A},\Omega_{\mathcal{A}/\mathbf{Z}}) \to H^0(\mathcal{J},\Omega_{\mathcal{J}/\mathbf{Z}})[I_f] \to H^0(J,\Omega_{J/\mathbf{Q}})[I_f] \to S_2(\Gamma_0(N),\mathbf{Q})[I_f],$$

where $\mathcal{A}$ and $\mathcal{J}$ are the Néron models of A and J, respectively. (See [AS04] for a discussion of why the image of $H^0(\mathcal{A},\Omega_{\mathcal{A}/\mathbf{Z}})$ is contained in $S_2(\Gamma_0(N),\mathbf{Z})$.)

Theorem 4.2. *If $\ell \mid c_A$ is a prime then $\ell^2 \mid 4N$.*

Proof. Mazur proved this when $\dim A = 1$ in [Maz78, §4], and we generalized his proof in [AS04]. □

When $\dim A = 1$, Edixhoven [Edi91] obtained strong results towards the folklore conjecture that $c_A = 1$, and when A has arbitrary dimension the authors have made the following conjecture (see [AS04] for evidence):

Conjecture 4.3. $c_A = 1$.

4.2. A Formula for $L(A,1)/\Omega_A$.

If L and M are lattices in a real vector space V, then the *lattice index* $[L : M]$ is the absolute value of the determinant of a linear transformation of V taking L onto M. The lattice index satisfies the usual properties suggested by the notation, e.g., $[L : M] \cdot [M : N] = [L : N]$.

The *real volume* Ω_A is defined as follows. If L^* is a lattice in the cotangent space

$$T^* = H^0(A_{\mathbf{R}},\Omega_{A_{\mathbf{R}}}) = S_2(\Gamma_0(N),\mathbf{R})[I_f]$$

of $A_{\mathbf{R}}$, then L^* determines a lattice $L = \mathrm{Hom}(L^*,\mathbf{Z})$ in the tangent space $T = \mathrm{Hom}(T^*,\mathbf{R})$, and hence a measure on T by declaring that the quotient T/L has measure 1. Let $A(\mathbf{R})^0$ denote the identity component of $A(\mathbf{R})$. Then $A(\mathbf{R})^0$ inherits a measure by virtue of being viewed as $T/H_1(A(\mathbf{R})^0,\mathbf{Z})$, and we have

$$\mu_L(A(\mathbf{R})^0) = [L : H_1(A(\mathbf{R})^0,\mathbf{Z})].$$

We also set

$$\mu_L(A(\mathbf{R})) = \mu_L(A(\mathbf{R})^0) \cdot c_\infty,$$

where $c_\infty = \#(A(\mathbf{R})/A(\mathbf{R})^0)$. Let $\mathcal{A}$ be the Néron model of A (see [BLR90]). The Néron differentials $H^0(\mathcal{A}, \Omega_{\mathcal{A}/\mathbf{Z}})$ define a lattice Λ^* in T^*, and we define $\Omega_A = \mu_\Lambda(A(\mathbf{R}))$.

Lemma 4.4. $H_1(A(\mathbf{R})^0, \mathbf{Z}) \cong H_1(A(\mathbf{C}), \mathbf{Z})^+$.

Proof. This lemma is well known, but we give a proof for the reader's convenience (which was suggested by H. Lenstra and B. Poonen). We have the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & H_1(A(\mathbf{R})^0, \mathbf{Z}) \arrow[r] \arrow[d, "\psi"] & H_1(A(\mathbf{R})^0, \mathbf{R}) \arrow[r] \arrow[d, "\cong"] & A(\mathbf{R})^0 \arrow[r] \arrow[d, hook, "i"] & 0 \\
0 \arrow[r] & H_1(A(\mathbf{C}), \mathbf{Z})^+ \arrow[r] & H_1(A(\mathbf{C}), \mathbf{R})^+ \arrow[r, "\pi"] & A(\mathbf{C})^+ &
\end{tikzcd}$$

where the upper horizontal sequences is exact (we view the real torus $A(\mathbf{R})^0$ as the quotient of the tangent space at the identity by the first integral homology), and the lower horizontal sequence is exact because it is the beginning of the long exact sequence of $\mathrm{Gal}(\mathbf{R}/\mathbf{C})$-cohomology that arises from

$$0 \to H_1(A(\mathbf{C}), \mathbf{Z}) \to H_1(A(\mathbf{C}), \mathbf{R}) \to A(\mathbf{C}) \to 0.$$

The middle vertical map is an isomorphism because if it were not then its kernel would be an uncountable set that maps to 0 in $A(\mathbf{R})^0$. The snake lemma then yields an exact sequence

$$0 \to \ker(\psi) \to 0 \to 0 \to \mathrm{coker}(\psi) \to 0,$$

which implies that ψ is an isomorphism. □

Let

$$\Phi : H_1(X_0(N), \mathbf{Q}) \to \mathrm{Hom}(S_2(\Gamma_0(N))[I_f], \mathbf{C})$$

be the map induced by integration, scaled so that

$$\Phi(\{0, \infty\})(f) = L(f, 1)$$

(that $\{0, \infty\} \in H_1(X_0(N), \mathbf{Q})$ is the Manin-Drinfeld theorem, and that $\int_0^\infty f$ is a multiple of $L(f, 1)$ follows from the definition of $L(f, s)$ as a Mellin transform).

Theorem 4.5. *Recall that A is an abelian variety attached to a newform $f \in S_2(\Gamma_0(N))$, that c_∞ is the number of connected components of $A(\mathbf{R})$, that c_A is the Manin constant of A, that Ω_A is the Néron canonical volume of $A(\mathbf{R})$, and that Φ is the period mapping on homology induced by integrating homology classes on $X_0(N)$ against the $\mathbf{C}$-vector space spanned by the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of f. Then we have the following equation:*

$$c_\infty \cdot c_A \cdot \frac{L(A, 1)}{\Omega_A} = [\Phi(H_1(X_0(N), \mathbf{Z}))^+ : \Phi(\mathbf{T}\{0, \infty\})] \in \mathbf{Q},$$

where the lattice index on the right hand side should be interpreted as 0 if $\Phi(\mathbf{T}\{0, \infty\})$ has rank less than the dimension of A.

Proof. It is easier to compute with $\tilde{\Lambda}^* = S_2(\Gamma_0(N), \mathbf{Z})[I_f]$ than with Λ^*, so let $\tilde{\Omega}_A = \mu_{\tilde{\Lambda}}(A(\mathbf{R}))$. Note that $\tilde{\Omega}_A \cdot c_A = \Omega_A$, where c_A is the Manin constant. By Lemma 4.4 and Section 2.2,

$$\begin{aligned}\tilde{\Omega}_A &= c_\infty \cdot [\tilde{\Lambda} : H_1(A(\mathbf{R})^0, \mathbf{Z})] \\ &= c_\infty \cdot [\mathrm{Hom}(S_2(\Gamma_0(N), \mathbf{Z})[I_f], \mathbf{Z}) : \Phi(H_1(X_0(N), \mathbf{Z}))^+].\end{aligned}$$

For any ring R the pairing

$$\mathbf{T}_R \times S_2(\Gamma_0(N), R) \to R$$

given by $\langle T_n, f\rangle = a_1(T_n f)$ is perfect, so $(\mathbf{T}/I_f) \otimes R \cong \mathrm{Hom}(S_2(\Gamma_0(N), R)[I_f], R)$. Using this pairing, we may view Φ as a map

$$\Phi : H_1(X_0(N), \mathbf{Q}) \to (\mathbf{T}/I_f) \otimes \mathbf{C},$$

so that

$$\tilde{\Omega}_A = c_\infty \cdot [\mathbf{T}/I_f : \Phi(H_1(X_0(N), \mathbf{Z}))^+].$$

Note that $(\mathbf{T}/I_f) \otimes \mathbf{C}$ is isomorphic as a ring to a product of copies of $\mathbf{C}$, with one copy corresponding to each Galois conjugate $f^{(i)}$ of f. Let $\pi_i \in (\mathbf{T}/I_f) \otimes \mathbf{C}$ be the projector onto the subspace of $(\mathbf{T}/I_f) \otimes \mathbf{C}$ corresponding to $f^{(i)}$. Then $\Phi(\{0, \infty\}) \cdot \pi_i = L(f^{(i)}, 1) \cdot \pi_i$. Since the π_i form a basis for the complex vector space $(\mathbf{T}/I_f) \otimes \mathbf{C}$, we see that

$$\det(\Phi(\{0, \infty\})) = \prod_i L(f^{(i)}, 1) = L(A, 1).$$

Letting $H = H_1(X_0(N), \mathbf{Z})$, we have

$$\begin{aligned}[\Phi(H)^+ : \Phi(\mathbf{T}\{0, \infty\})] &= [\Phi(H)^+ : (\mathbf{T}/I_f) \cdot \Phi(\{0, \infty\})] \\ &= [\Phi(H)^+ : \mathbf{T}/I_f] \cdot [\mathbf{T}/I_f : \mathbf{T}/I_f \cdot \Phi(\{0, \infty\})] \\ &= \frac{c_\infty}{\tilde{\Omega}_A} \cdot \det(\Phi(\{0, \infty\})) \\ &= \frac{c_\infty c_A}{\Omega_A} \cdot L(A, 1),\end{aligned}$$

which proves the theorem. □

Theorem 4.5 was inspired by the case when A is an elliptic curve (see [Cre97, §II.2.8]) or the winding quotient of $J_0(p)$ (see [Aga99]), and it generalizes to forms of weight > 2 (see [Ste00]).

Theorem 4.5 is true with Φ replaced by any linear map with the same kernel as Φ. One way to find such a linear map with image in a $\mathbf{Q}$-vector space is to compute a basis $\varphi_1, \ldots \varphi_d$ for $\mathrm{Hom}(H_1(X_0(N), \mathbf{Q}), \mathbf{Q})[I_f]$ and let $\Phi = \varphi_1 \times \cdots \times \varphi_d$. Also, since $H_1(X_0(N), \mathbf{Z})^+$ and $\mathbf{T}\{0, \infty\}$ are contained in $H_1(X_0(N), \mathbf{Q})^+$, Theorem 4.5 implies that $L(A, 1)/\Omega_A \in \mathbf{Q}$, a fact well known to the experts (see [Gro94, Prop. 2.7] for the statement, but without proof).

4.3. The Denominator of $L(A, 1)/\Omega_A$.

In this section, we prove a result about the denominator of the rational number $L(A, 1)/\Omega_A$ and compare it to what is predicted by the Birch and Swinnerton-Dyer conjecture.

Proposition 4.6. *Let z be the point in $J_0(N)(\mathbf{Q})$ defined by the degree 0 divisor $(0) - (\infty)$ on $X_0(N)$, and let $n = n_f$ be the order of the image of z in $A(\mathbf{Q})$. Then the denominator of $c_\infty \cdot c_A \cdot L(A, 1)/\Omega_A$ divides n.*

Proof. Let x be the image of z in $A(\mathbf{Q})$, and let $I = \mathrm{Ann}_{\mathbf{T}}(x)$ be the ideal of elements of $\mathbf{T}$ that annihilate x. Since f is a newform, the Hecke operators T_p, for $p \mid N$, act as 0 or ± 1 on $A(\mathbf{Q})$ (see, e.g., [DI95, §6]). If $p \nmid N$, then a standard calculation (see, e.g., [Cre97, §2.8]) shows that $T_p(x) = (p+1)x$.

Let C be the cyclic subgroup of $A(\mathbf{Q})$ of order n generated by x. Consider the map $\mathbf{T} \to C$ given by $T_p \mapsto T_p(x)$. The kernel of this map is I, and the map is surjective because its image is an additive group that contains x, and C is the smallest such group. Thus the map induces an isomorphism $\mathbf{T}/I \xrightarrow{\cong} C$. □

Conjecture 2.2 predicts that

$$\#A(\mathbf{Q}) \cdot \#A^\vee(\mathbf{Q}) \cdot \frac{L(A,1)}{\Omega_A} = \#\text{Ш}(A) \cdot \prod c_p \in \mathbf{Z},$$

and since $n \mid \#A(\mathbf{Q})$, Proposition 4.6 implies that

$$c_\infty \cdot c_A \cdot \#A(\mathbf{Q}) \cdot \frac{L(A,1)}{\Omega_A} \in \mathbf{Z}.$$

Since c_∞ is a power of 2, and c_A is conjecturally 1 (if N is prime, then by Theorem 4.2 it is a power of 2), Proposition 4.6 provides theoretical evidence for Conjecture 2.2, and also reflects a surprising amount of cancellation between $\prod c_p$ and $\#A^\vee(\mathbf{Q})$.

5. Results and Conclusions

We computed all 19608 abelian varieties $A = A_f$ attached to newforms of level $N \leq 2333$. Interesting data about some of these abelian varieties is summarized in Tables 1–4, which use the notation described in this section.

Suppose that A is one of the 10360 of these for which $L(A,1) \neq 0$, so Conjecture 2.2 asserts that Ш(A) has order

$$\#\text{Ш}_? = \frac{L(A,1)}{\tilde{\Omega}_A \cdot c_A} \cdot \frac{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^\vee(\mathbf{Q})_{\mathrm{tor}}}{\prod_{p \mid N} c_p}.$$

(See Section 4.2 for the definition of $\tilde{\Omega}_A$ and c_A.)

For any rational number x, let x^{odd} be the *odd part* of x. If a and b are rational numbers with $a \neq 0$, we say that $a \mid b$ if b/a is an integer.

Define integers S_l and S_u such that

$$S_l \mid \mathrm{numer}(\#\text{Ш}_?^{\mathrm{odd}}) \mid S_u$$

as follows:

$\mathbf{S}_u$ The *upper bound* S_u is the odd part of the numerator of

$$\frac{L(A,1)}{\tilde{\Omega}_A} \cdot \frac{T^2}{\prod_{p \| N} c_p},$$

where T is the upper bound on $\#A(\mathbf{Q})$ and $\#A^\vee(\mathbf{Q})$ computed using Section 3.5 using a_p for $p \leq 17$. Since the Manin constant and the Tamagawa numbers are integers, S_u is an upper bound on the odd part of $\#\text{Ш}_?$.

$\mathbf{S}_l$ The *lower bound* S_l is defined as follows: Let $S_{l,1}$ be the odd part of the rational number

$$\frac{L(A,1)}{\tilde{\Omega}_A} \cdot \frac{\#C \cdot \#D}{\prod_{p \| N} c_p},$$

where $C \subset A(\mathbf{Q})_{\mathrm{tor}}$ and D is the part of C coprime to the modular degree of A. Usually C is the group generated by the image of $(0)-(\infty)$, and in all cases it contains this subgroup. More precisely, when A is an elliptic curve, we instead let C and D be the full torsion subgroup $A(\mathbf{Q})_{\mathrm{tor}}$, because it is easy to calculate. When A is not an elliptic curve it would be better to let C be the subgroup generated by all rational cusps, but the authors only realized this after completing the calculations, so we did not do this.

If N is square free, we let $S_l = S_{l,1}$. Otherwise, let $S_{l,2}$ be the largest part of $S_{l,1}$ coprime to all primes whose square divides N. This takes care of the Manin constant, which only involves primes whose square divides N. To take care of Tamagawa numbers, remove all primes $p \leq 2\dim(A)+1$ from $S_{l,2}$ to obtain S_l.

Remark 5.1. When N is square free we have

$$S_l \mid \#\text{Ш}_?^{\mathrm{odd}} \mid S_u$$

since c_A is a power of 2 and no Tamagawa numbers have been omitted from the formulas for S_l and S_u. For every $N \leq 2333$ we found that S_l is an integer, so when $N \leq 2333$ is squarefree, $\#\text{Ш}_?^{\mathrm{odd}}$ is an integer. Since Conjecture 2.2 asserts that $\#\text{Ш}_?$ is the order of a group, hence an integer, our data gives evidence for Conjecture 2.2.

Tables 1–4 list every A of level $N \leq 2333$ such that $S_l > 1$. The A column contains the label of A (see Section 3.2), and the next column (labeled dim) contains $\dim A$. A star next to the label for A indicates that we have proved that the odd part of $\#\text{Ш}(A)$ is at least as large as conjectured by the Birch and Swinnerton-Dyer conjecture. *This is the case for* 39 *of the* 168 *examples.* The columns labeled S_l contains the number S_l defined above. If $S_l = S_u$ then the column labeled S_u contains an = sign, and otherwise, it contains S_u (there are only 13 cases in which $S_u \neq S_l$). The column labeled $\mathrm{moddeg}(A)^{\mathrm{odd}}$ contains the odd part m of the modular degree of A, written as a product $\gcd(m, S_u)\cdot {\scriptstyle m/\gcd(S_u,m)}$, where ${\scriptstyle m/\gcd(S_u,m)}$ is shrunk to save space. The only non-square-free levels of A_f for which $S_l > 1$ are 1058, 1664, 2224, and 2264.

The column labeled B contains all B such that $L(B,1)=0$ and

$$\gcd(S_l, \#(A^\vee \cap \tilde{B}^\vee)) > 1.$$

(In retrospect, it would probably have been more interesting to list those B such that $\gcd(S_u, \#(A^\vee \cap \tilde{B}^\vee)) > 1$.) Here if $B = A_g$ for some newform g of level dividing N, and $\tilde{B}^\vee$ is the abelian subvariety of $J_0(N)$ generated by all images of $B^\vee$ under the degeneracy maps. Thus, e.g., when $B^\vee$ is of level N, $\tilde{B}^\vee = B^\vee$. The next column, labeled dim, contains the dimension of B.

The final two columns contain information about the relationship between A and B. The one labeled $A^\vee \cap \tilde{B}^\vee$ contains the abelian group structure of the indicated abelian group, where e.g., $[a^b c^d]$ means the abelian group $(\mathbf{Z}/a\mathbf{Z})^b \times (\mathbf{Z}/c\mathbf{Z})^d$. The column labeled Vis contains a divisor of the order of $\mathrm{Vis}_C(\text{Ш}(A^\vee))$, where $C = A^\vee + \tilde{B}^\vee$ (note that $\mathrm{Vis}_C(\text{Ш}(A^\vee)) \subset \mathrm{Vis}_{J_0(N)}(\text{Ш}(A^\vee))$).

The table is divided into three vertical regions, where the columns in the first region are about A only, the columns of the second region are about B only, and the third column is about the relationship between A and B.

5.1. **Example: Level 389.** We illustrate what is involved in computing the first line of Table 1. Using the method sketched in Section 3.2, we find that $S_2(\Gamma_0(389))$ contains exactly five Galois-conjugacy classes of newforms, and these are defined over extensions of $\mathbf{Q}$ of degrees 1, 2, 3, 6, and 20. Thus $J = J_0(389)$ decomposes, up to isogeny, as a product $A_1 \times A_2 \times A_3 \times A_6 \times A_{20}$ of abelian varieties, where $\dim A_d = d$ and A_d is the optimal quotient corresponding to the appropriate Galois-conjugacy class of newforms.

Next we consider the arithmetic of the A_d. Using Theorem 4.5 we find that

$$L(A_1, 1) = L(A_2, 1) = L(A_3, 1) = L(A_6, 1) = 0,$$

and

$$\frac{L(A_{20}, 1)}{\Omega_{A_{20}}} = \frac{5^2 \cdot 2^{11}}{97 \cdot c_A},$$

where c_A is the Manin constant attached to A_{20}, which, by Theorem 4.2, is of the form 2^n with $n \geq 0$. Using the algorithms of Sections 3.5, 3.6, 3.7, we find that $\#A_{20}(\mathbf{Q}) = c_{389} = 97$. Thus Conjecture 2.2 predicts that $\#Ш(A_{20}) = 5^2 \cdot 2^{11}/c_A$. The following proposition provides support for this conjecture.

Proposition 5.2. *There is a natural inclusion*

$$(\mathbf{Z}/5\mathbf{Z})^2 \cong A_1(\mathbf{Q})/5A_1(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_{J_0(389)}(Ш(A_{20}^\vee)).$$

Proof. Let $A = A_{20}^\vee$, $B = A_1^\vee$ and $J = A + B \subset J_0(389)$. Using Proposition 3.3, we find that $A \cap B \cong (\mathbf{Z}/4)^2 \times (\mathbf{Z}/5\mathbf{Z})^2$, so $B[5] \subset A$. Since 5 does not divide the numerator of $(389-1)/12$, it does not divide the Tamagawa numbers or the orders of the torsion groups, so Theorem 3.13 yields the asserted injection. To see that $(\mathbf{Z}/5\mathbf{Z})^2 \cong A_1(\mathbf{Q})/5A_1(\mathbf{Q})$ use the standard elliptic curves algorithms [Cre97]. □

5.2. **Invisible Elements of $Ш(A)$.** Tables 1–4 suggest that much of $Ш(A^\vee)$ is invisible in $J_0(N)$. This is because Proposition 3.15 implies that if a prime divides $\#Ш(A^\vee)$ but not $\mathrm{moddeg}(A^\vee)$ then $Ш(A^\vee)$ contains an element of order p that is invisible. We find many examples in the table where p divides the conjectural order of $Ш(A^\vee)$, but $p \nmid \mathrm{moddeg}(A^\vee)$.

Invisible elements might become visible at higher level (see [AS02, §4.3] for a discussion and example).

5.3. **The Part of $Ш(A)$ That Must be a Perfect Square.** When $\dim A = 1$, properties of the Cassels-Tate pairing imply that if $Ш(A)$ is finite then $\#Ш(A)$ is a perfect square, and the fact that one finds in examples (see [Cre97]) that $\#Ш_?$ is a perfect square is computational evidence for Conjecture 2.2.

In contrast, when the dimension is greater than one, Poonen and Stoll [PS99] discovered Jacobians J such that $Ш(J)$ has order twice a square, and the second author found for each prime $p < 25000$ an abelian variety A of dimension $p-1$ such that $\#Ш(A) = pn^2$ for some integer n (see [Ste02b]).

Proposition 5.3. *Let $A = A_f$ be a quotient of $J_0(N)$ and ℓ be a prime that does not divide the modular degree of A. Suppose that $Ш(A)[\ell^\infty]$ is finite. Then $\#Ш(A)[\ell^\infty]$ is a perfect square.*

Proof. The Cassels-Tate pairing (see [Tat63, §3]) induces a pairing

$$\phi : Ш(A)[\ell^\infty] \times Ш(A^\vee)[\ell^\infty] \to \mathbf{Q}/\mathbf{Z}.$$

Since $Ш(A)[\ell^\infty]$ is finite, it follows from [Tat63, Thm. 3.2] that $Ш(A^\vee)[\ell^\infty]$ is also finite and ϕ is non-degenerate. In particular, $\#Ш(A^\vee)[\ell^\infty] = \#Ш(A)[\ell^\infty]$.

Since $J_0(N)$ is a Jacobian, it possesses a canonical polarization arising from the theta divisor; this divisor is rational over $\mathbf{Q}$, since $X_0(N)$ always has a point over $\mathbf{Q}$ (the cusp ∞ is rational). This polarization induces a polarization $\theta : A^\vee \to A$, which also comes from a divisor that is rational over $\mathbf{Q}$. Hence, by [Tat63, Thm. 3.3] (see also [PS99, Thm. 5]), the pairing

$$\phi' : Ш(A^\vee)[\ell^\infty] \times Ш(A^\vee)[\ell^\infty] \to \mathbf{Q}/\mathbf{Z}$$

obtained by composing θ with the pairing ϕ above is alternating.

Since ℓ does not divide the modular degree of A, it does not divide the degree of the isogeny θ. Hence θ induces an isomorphism $Ш(A^\vee)[\ell^\infty] \xrightarrow{\cong} Ш(A)[\ell^\infty]$. Thus by the non-degeneracy of the pairing ϕ, the pairing ϕ' is also non-degenerate. Since ϕ' is also alternating, it follows from arguments similar to those in [Cas62, p. 260] that $\#Ш(A^\vee)[\ell^\infty]$ is a perfect square. Since $\#Ш(A)[\ell^\infty] = \#Ш(A^\vee)[\ell^\infty]$, we see that $\#Ш(A)[\ell^\infty]$ is also a perfect square.

□

For the entries in Tables 1–4, $Ш(A)$ is finite, so if $\ell \nmid \mathrm{moddeg}(A)$ then the ℓ-power part of $\#Ш(A)$ must be a perfect square. When $S_l = S_u$ and the level is square free, then S_l is the odd part of the conjectural order of $Ш(A)$. We found that S_l is a perfect square whenever $S_l = S_u$, which provides evidence for Conjecture 2.2.

Table 1. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	$\text{moddeg}(A)^{\text{odd}}$	B	dim	$A^\vee \cap \tilde{B}^\vee$	Vis
389E$*$	20	5^2	$=$	5	**389A**	1	$[20^2]$	5^2
433D$*$	16	7^2	$=$	$7\cdot 111$	**433A**	1	$[14^2]$	7^2
446F$*$	8	11^2	$=$	$11\cdot 359353$	**446B**	1	$[11^2]$	11^2
551H	18	3^2	$=$	169	NONE			
563E$*$	31	13^2	$=$	13	**563A**	1	$[26^2]$	13^2
571D$*$	2	3^2	$=$	$3^2\cdot 127$	**571B**	1	$[3^2]$	3^2
655D$*$	13	3^4	$=$	$3^2\cdot 9799079$	**655A**	1	$[36^2]$	3^4
681B	1	3^2	$=$	$3\cdot 125$	**681C**	1	$[3^2]$	–
707G$*$	15	13^2	$=$	$13\cdot 800077$	**707A**	1	$[13^2]$	13^2
709C$*$	30	11^2	$=$	11	**709A**	1	$[22^2]$	11^2
718F$*$	7	7^2	$=$	$7\cdot 5371523$	**718B**	1	$[7^2]$	7^2
767F	23	3^2	$=$	1	NONE			
794G$*$	12	11^2	$=$	$11\cdot 34986189$	**794A**	1	$[11^2]$	–
817E$*$	15	7^2	$=$	$7\cdot 79$	**817A**	1	$[7^2]$	–
959D	24	3^2	$=$	583673	NONE			
997H$*$	42	3^4	$=$	3^2	**997B**	1	$[12^2]$	3^2
					997C	1	$[24^2]$	3^2
1001F	3	3^2	$=$	$3^2\cdot 1269$	**1001C**	1	$[3^2]$	–
					91A	1	$[3^2]$	–
1001L	7	7^2	$=$	$7\cdot 2029789$	**1001C**	1	$[7^2]$	–
1041E	4	5^2	$=$	$5^2\cdot 13589$	**1041B**	2	$[5^2]$	–
1041J	13	5^4	$=$	$5^3\cdot 21120929983$	**1041B**	2	$[5^4]$	–
1058D	1	5^2	$=$	$5\cdot 483$	**1058C**	1	$[5^2]$	–
1061D	46	151^2	$=$	$151\cdot 10919$	**1061B**	2	$[2^2 302^2]$	–
1070M	7	$3\cdot 5^2$	$3^2\cdot 5^2$	$3\cdot 5\cdot 1720261$	**1070A**	1	$[15^2]$	–
1077J	15	3^4	$=$	$3^2\cdot 1227767047943$	**1077A**	1	$[9^2]$	–
1091C	62	7^2	$=$	1	NONE			
1094F$*$	13	11^2	$=$	$11^2\cdot 172446773$	**1094A**	1	$[11^2]$	11^2
1102K	4	3^2	$=$	$3^2\cdot 31009$	**1102A**	1	$[3^2]$	–
1126F$*$	11	11^2	$=$	$11\cdot 13990352759$	**1126A**	1	$[11^2]$	11^2
1137C	14	3^4	$=$	$3^2\cdot 64082807$	**1137A**	1	$[9^2]$	–
1141I	22	7^2	$=$	$7\cdot 528921$	**1141A**	1	$[14^2]$	–
1147H	23	5^2	$=$	$5\cdot 729$	**1147A**	1	$[10^2]$	–
1171D$*$	53	11^2	$=$	$11\cdot 81$	**1171A**	1	$[44^2]$	11^2
1246B	1	5^2	$=$	$5\cdot 81$	**1246C**	1	$[5^2]$	–
1247D	32	3^2	$=$	$3^2\cdot 2399$	**43A**	1	$[36^2]$	–
1283C	62	5^2	$=$	$5\cdot 2419$	NONE			
1337E	33	3^2	$=$	71	NONE			
1339G	30	3^2	$=$	5776049	NONE			
1355E	28	3	3^2	$3^2\cdot 2224523985405$	NONE			
1363F	25	31^2	$=$	$31\cdot 34889$	**1363B**	2	$[2^2 62^2]$	–
1429B	64	5^2	$=$	1	NONE			
1443G	5	7^2	$=$	$7^2\cdot 18525$	**1443C**	1	$[7^1 14^1]$	–
1446N	7	3^2	$=$	$3\cdot 17459029$	**1446A**	1	$[12^2]$	–

Table 2. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	$\mathrm{moddeg}(A)^{\mathrm{odd}}$	B	dim	$A^\vee \cap \tilde{B}^\vee$	Vis
1466H$*$	23	13^2	$=$	$13\cdot 25631993723$	**1466B**	1	$[26^2]$	13^2
1477C$*$	24	13^2	$=$	$13\cdot 57037637$	**1477A**	1	$[13^2]$	13^2
1481C	71	13^2	$=$	70825	NONE			
1483D$*$	67	$3^2\cdot 5^2$	$=$	$3\cdot 5$	**1483A**	1	$[60^2]$	$3^2\cdot 5^2$
1513F	31	3	3^4	$3\cdot 759709$	NONE			
1529D	36	5^2	$=$	535641763	NONE			
1531D	73	3	3^2	3	**1531A**	1	$[48^2]$	–
1534J	6	3	3^2	$3^2\cdot 635931$	**1534B**	1	$[6^2]$	–
1551G	13	3^2	$=$	$3\cdot 110659885$	**141A**	1	$[15^2]$	–
1559B	90	11^2	$=$	1	NONE			
1567D	69	$7^2\cdot 41^2$	$=$	$7\cdot 41$	**1567B**	3	$[4^4 1148^2]$	–
1570J$*$	6	11^2	$=$	$11\cdot 228651397$	**1570B**	1	$[11^2]$	11^2
1577E	36	3	3^2	$3^2\cdot 15$	**83A**	1	$[6^2]$	–
1589D	35	3^2	$=$	6005292627343	NONE			
1591F$*$	35	31^2	$=$	$31\cdot 2401$	**1591A**	1	$[31^2]$	31^2
1594J	17	3^2	$=$	$3\cdot 259338050025131$	**1594A**	1	$[12^2]$	–
1613D$*$	75	5^2	$=$	$5\cdot 19$	**1613A**	1	$[20^2]$	5^2
1615J	13	3^4	$=$	$3^2\cdot 13317421$	**1615A**	1	$[9^1 18^1]$	–
1621C$*$	70	17^2	$=$	17	**1621A**	1	$[34^2]$	17^2
1627C$*$	73	3^4	$=$	3^2	**1627A**	1	$[36^2]$	3^4
1631C	37	5^2	$=$	6354841131	NONE			
1633D	27	$3^6\cdot 7^2$	$=$	$3^5\cdot 7\cdot 31375$	**1633A**	3	$[6^4 42^2]$	–
1634K	12	3^2	$=$	$3\cdot 3311565989$	**817A**	1	$[3^2]$	–
1639G$*$	34	17^2	$=$	$17\cdot 82355$	**1639B**	1	$[34^2]$	17^2
1641J$*$	24	23^2	$=$	$23\cdot 1491344147471$	**1641B**	1	$[23^2]$	23^2
1642D$*$	14	7^2	$=$	$7\cdot 123398360851$	**1642A**	1	$[7^2]$	7^2
1662K	7	11^2	$=$	$11\cdot 16610917393$	**1662A**	1	$[11^2]$	–
1664K	1	5^2	$=$	$5\cdot 7$	**1664N**	1	$[5^2]$	–
1679C	45	11^2	$=$	6489	NONE			
1689E	28	3^2	$=$	$3\cdot 172707180029157365$	**563A**	1	$[3^2]$	–
1693C	72	1301^2	$=$	1301	**1693A**	3	$[2^4 2602^2]$	–
1717H$*$	34	13^2	$=$	$13\cdot 345$	**1717B**	1	$[26^2]$	13^2
1727E	39	3^2	$=$	118242943	NONE			
1739F	43	659^2	$=$	$659\cdot 151291281$	**1739C**	2	$[2^2 1318^2]$	–
1745K	33	5^2	$=$	$5\cdot 1971380677489$	**1745D**	1	$[20^2]$	–
1751C	45	5^2	$=$	$5\cdot 707$	**103A**	2	$[505^2]$	–
1781D	44	3^2	$=$	61541	NONE			
1793G$*$	36	23^2	$=$	$23\cdot 8846589$	**1793B**	1	$[23^2]$	23^2
1799D	44	5^2	$=$	201449	NONE			
1811D	98	31^2	$=$	1	NONE			
1829E	44	13^2	$=$	3595	NONE			
1843F	40	3^2	$=$	8389	NONE			
1847B	98	3^6	$=$	1	NONE			
1871C	98	19^2	$=$	14699	NONE			

Table 3. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	$\mathrm{moddeg}(A)^{\mathrm{odd}}$	B	dim	$A^\vee \cap \tilde{B}^\vee$	Vis
1877B	86	7^2	=	1	NONE			
1887J	12	5^2	=	$5 \cdot 10825598693$	**1887A**	1	$[20^2]$	–
1891H	40	7^4	=	$7^2 \cdot 44082137$	**1891C**	2	$[4^2 196^2]$	–
1907D$\ast$	90	7^2	=	$7 \cdot 165$	**1907A**	1	$[56^2]$	7^2
1909D$\ast$	38	3^4	=	$3^2 \cdot 9317$	**1909A**	1	$[18^2]$	3^4
1913B$\ast$	1	3^2	=	$3 \cdot 103$	**1913A**	1	$[3^2]$	3^2
1913E	84	$5^4 \cdot 61^2$	=	$5^2 \cdot 61 \cdot 103$	**1913A**	1	$[10^2]$	–
					1913C	2	$[2^2 610^2]$	–
1919D	52	23^2	=	675	NONE			
1927E	45	3^2	3^4	52667	NONE			
1933C	83	$3^2 \cdot 7$	$3^2 \cdot 7^2$	$3 \cdot 7$	**1933A**	1	$[42^2]$	3^2
1943E	46	13^2	=	62931125	NONE			
1945E$\ast$	34	3^2	=	$3 \cdot 571255479184807$	**389A**	1	$[3^2]$	3^2
1957E$\ast$	37	$7^2 \cdot 11^2$	=	$7 \cdot 11 \cdot 3481$	**1957A**	1	$[22^2]$	11^2
					1957B	1	$[14^2]$	7^2
1979C	104	19^2	=	55	NONE			
1991C	49	7^2	=	1634403663	NONE			
1994D	26	3	3^2	$3^2 \cdot 46197281414642501$	**997B**	1	$[3^2]$	–
1997C	93	17^2	=	1	NONE			
2001L	11	3^2	=	$3^2 \cdot 44513447$	NONE			
2006E	1	3^2	=	$3 \cdot 805$	**2006D**	1	$[3^2]$	–
2014L	12	3^2	=	$3^2 \cdot 126381129003$	**106A**	1	$[9^2]$	–
2021E	50	5^6	=	$5^2 \cdot 729$	**2021A**	1	$[100^2]$	5^4
2027C$\ast$	94	29^2	=	29	**2027A**	1	$[58^2]$	29^2
2029C	90	$5^2 \cdot 269^2$	=	$5 \cdot 269$	**2029A**	2	$[2^2 2690^2]$	–
2031H$\ast$	36	11^2	=	$11 \cdot 1014875952355$	**2031C**	1	$[44^2]$	11^2
2035K	16	11^2	=	$11 \cdot 218702421$	**2035C**	1	$[11^1 22^1]$	–
2038F	25	5	5^2	$5^2 \cdot 92198576587$	**2038A**	1	$[20^2]$	–
					1019B	1	$[5^2]$	–
2039F	99	$3^4 \cdot 5^2$	=	13741381043009	NONE			
2041C	43	3^4	=	61889617	NONE			
2045I	39	3^4	=	$3^3 \cdot 3123399893$	**2045C**	1	$[18^2]$	–
					409A	13	$[9370199679^2]$	–
2049D	31	3^2	=	29174705448000469937	NONE			
2051D	45	7^2	=	$7 \cdot 674652424406369$	**2051A**	1	$[56^2]$	–
2059E	45	$5 \cdot 7^2$	$5^2 \cdot 7^2$	$5^2 \cdot 7 \cdot 167359757$	**2059A**	1	$[70^2]$	–
2063C	106	13^2	=	8479	NONE			
2071F	48	13^2	=	36348745	NONE			
2099B	106	3^2	=	1	NONE			
2101F	46	5^2	=	$5 \cdot 11521429$	**191A**	2	$[155^2]$	–
2103E	37	$3^2 \cdot 11^2$	=	$3^2 \cdot 11 \cdot 874412923071571792611$	**2103B**	1	$[33^2]$	11^2
2111B	112	211^2	=	1	NONE			
2113B	91	7^2	=	1	NONE			
2117E$\ast$	45	19^2	=	$19 \cdot 1078389$	**2117A**	1	$[38^2]$	19^2

Table 4. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	moddeg$(A)^{\text{odd}}$	B	dim	$A^\vee \cap \tilde{B}^\vee$	Vis
2119C	48	7^2	=	89746579	NONE			
2127D	34	3^2	=	$3\cdot 18740561792121901$	**709A**	1	$[3^2]$	–
2129B	102	3^2	=	1	NONE			
2130Y	4	7^2	=	$7\cdot 83927$	**2130B**	1	$[14^2]$	–
2131B	101	17^2	=	1	NONE			
2134J	11	3^2	=	1710248025389	NONE			
2146J	10	7^2	=	$7\cdot 1672443$	**2146A**	1	$[7^2]$	–
2159E	57	13^2	=	31154538351	NONE			
2159D	56	3^4	=	233801	NONE			
2161C	98	23^2	=	1	NONE			
2162H	14	3	3^2	$3\cdot 6578391763$	NONE			
2171E	54	13^2	=	271	NONE			
2173H	44	199^2	=	$199\cdot 3581$	**2173D**	2	$[398^2]$	–
2173F	43	19^2	$3^2\cdot 19^2$	$3^2\cdot 19\cdot 229341$	**2173A**	1	$[38^2]$	19^2
2174F	31	5^2	=	$5\cdot 21555702093188316107$	NONE			
2181E	27	7^2	=	$7\cdot 7217996450474835$	**2181A**	1	$[28^2]$	–
2193K	17	3^2	=	$3\cdot 15096035814223$	**129A**	1	$[21^2]$	–
2199C	36	7^2	=	$7^2\cdot 13033437060276603$	NONE			
2213C	101	3^4	=	19	NONE			
2215F	46	13^2	=	$13\cdot 1182141633$	**2215A**	1	$[52^2]$	–
2224R	11	79^2	=	79	**2224G**	2	$[79^2]$	–
2227E	51	11^2	=	259	NONE			
2231D	60	47^2	=	91109	NONE			
2239B	110	11^4	=	1	NONE			
2251E∗	99	37^2	=	37	**2251A**	1	$[74^2]$	37^2
2253C∗	27	13^2	=	$13\cdot 14987929400988647$	**2253A**	1	$[26^2]$	13^2
2255J	23	7^2	=	15666366543129	NONE			
2257H	46	$3^6\cdot 29^2$	=	$3^3\cdot 29\cdot 175$	**2257A**	1	$[9^2]$	–
					2257D	2	$[2^2 174^2]$	–
2264J	22	73^2	=	73	**2264B**	2	$[146^2]$	–
2265U	14	7^2	=	$7^2\cdot 73023816368925$	**2265B**	1	$[7^2]$	–
2271I∗	43	23^2	=	$23\cdot 392918345997771783$	**2271C**	1	$[46^2]$	23^2
2273C	105	7^2	=	7^2	NONE			
2279D	61	13^2	=	96991	NONE			
2279C	58	5^2	=	1777847	NONE			
2285E	45	151^2	=	$151\cdot 138908751161$	**2285A**	2	$[2^2 302^2]$	–
2287B	109	71^2	=	1	NONE			
2291C	52	3^2	=	427943	NONE			
2293C	96	479^2	=	479	**2293A**	2	$[2^2 958^2]$	–
2294F	15	3^2	=	$3\cdot 6289390462793$	**1147A**	1	$[3^2]$	–
2311B	110	5^2	=	1	NONE			
2315I	51	3^2	=	$3\cdot 4475437589723$	**463A**	16	$[13426312769169^2]$	–
2333C	101	83341^2	=	83341	**2333A**	4	$[2^6 166682^2]$	–

6. Appendix by J. Cremona and B. Mazur: "Explaining" Shafarevich-Tate via Mordell-Weil

Introduction. In our article [CM] we discussed the notion of visibility and offered some tables of examples of that phenomenon. We gave, however, very little theoretical discussion in that article. Here we wish to take the opportunity to correct some gaps in our commentary on our tables and to offer the details of the proof of a general criterion that is sometimes useful to test visibility. Regarding Table 1 of [CM] we said that for each pair (E, p) that occurs there and for which there is a corresponding "F" on the table of the same conductor of E, the Shafarevich-Tate group of E is *explained* by the Mordell-Weil group of F, in the technical sense that we gave to the word *explained* in that article. Now this is indeed the case for all entries of our table such that E has semistable reduction at p and it is also the case for those entries where the conductor of F properly divides the conductor of E. We will review why this is so, below. It is also true that for each of the remaining 7 entries ($E =$ **2601H**, **2718D**, **2900D**, **3555E**, **3879E**, **3933A**, **5499E**) a nontrivial subgroup of the Shafarevich-Tate group of E is *explained* by the Mordell-Weil group of the corresponding F, but we wish to notify our readers that we have not yet checked whether or not *all* of the "Ш" of these 7 elliptic curves is so explained. These 7 cases deserve to be looked at (the issue being local at the prime 3 for all but **2900D**, where it is local at the prime 5). Regarding Table 2 of [CM], although our commentary in [CM] does not say this clearly, for all the entries E of that table for which there is a corresponding F of the same conductor we only have checked that $E[2] = F[2]$ in $J_0(N)$ and nothing more, except, of course, for those entries we particularly signal to have shown something less; namely, in the language of our article, that they "seem to satisfy a 2-congruence." In these latter cases where we signal that we have shown something less, W. Stein has checked that in fact $E[2] \neq F[2]$ in $J_0(N)$.

Let p be an odd prime number. If E is an (optimal) elliptic curve over $\mathbf{Q}$ of conductor N then E may be unambiguously identified (up to sign) with a subabelian variety of the modular jacobian $J_0(N)$ (over $\mathbf{Q}$). If (E, F, p) is an entry of Table 1 of [CM] such that E and F are of the same conductor N we checked that we have equality of the finite group schemes $E[p]_{/\mathbf{Q}} = F[p]_{/\mathbf{Q}}$ in $J_0(N)_{/\mathbf{Q}}$. For the remaining three entries we checked that there is an isomorphism of finite group schemes $\iota : E[p]_{/\mathbf{Q}} \cong F[p]_{/\mathbf{Q}}$. In both cases, identifying the two finite group schemes let H denote the common cohomology group,

$$H := H^1(G_{\mathbf{Q}}, E[p]_{/\mathbf{Q}}) = H^1(G_{\mathbf{Q}}, F[p]_{/\mathbf{Q}}),$$

and $S_E \subset H$, and $S_F \subset H$ the p-Selmer groups of, respectively, E and F. What we will show is that

Proposition 6.1. *For each of the entries (E, p) in Table 1 of* [CM] *such that p is a prime of semistable reduction for E and for which there is a "corresponding" F, we have*

$$S_E = S_F \subset H.$$

To discuss this, we need some notation.

Let $X := \mathrm{Spec}(\mathbf{Z})$, $Y := \mathrm{Spec}(\mathbf{Z}[1/p]) = X - \mathrm{Spec}(\mathbf{F}_p)$, and $\eta := \mathrm{Spec}(\mathbf{Q})$. Let $E_\eta := E$ be our elliptic curve over $\mathbf{Q}$ of conductor N, $E_{/X}$ the Néron model over X of E_η and $E^o_{/X} \subset E_{/X}$ the "connected component" of Néron (meaning the open subgroup scheme every fiber of which is connected). We have, of course, similar notation for the corresponding elliptic curve F. Let $E[p]_{/X}$ denote the closed subgroup scheme given as the kernel of multiplication by p in the Néron model: $E[p]_{/X} \subset E_{/X}$. We have, in general, that the restriction $E[p]_{/Y}$ of $E[p]_{/X}$ to the base Y is an étale quasi-finite flat group scheme; and if p^2 doesn't divide N we have that the group scheme $E[p]_{/X}$ is a quasi-finite flat group scheme [Gro72, Prop. 3.1(d), pg. 343]. The étale quasi-finite flat group scheme $E[p]_{/Y}$ can be characterized by the following features:

(i) Its generic fiber is the group scheme $E[p]_{/\eta} \subset J_0(N)_{/\eta}$,

and (one has a choice here) either:

(ii) $E[p]_{/Y} \subset J_0(N)_{/Y}$ is a closed étale quasi-finite flat) subgroup scheme,

or:

(ii') $E[p]_{/Y}$ enjoys the Néronian property over the base Y.

Similar statements hold for $F[p]_{/X}$.

Let Φ be the (punctual) sheaf of abelian groups for the flat topology over X which fits into the exact sequence (of abelian sheaves over X)

$$0 \to E^o \to E \to \Phi \to 0. \tag{1}$$

We will use the same notation to indicate the corresponding exact sequence of sheaves for the étale topology over X. Since E^0 and E are smooth group schemes, the long exact sequences of cohomology derived from the short exact sequence (1), viewed either as sheaves of abelian groups for the flat or étale topology, coincide; cf Section 11 *Appendice: Un théorème de comparaison de la cohomologie étale et de la cohomologie fppf* in [Gro68]. Thinking now of Φ as a sheaf for the étale topology, denote by Φ_ℓ its stalk at the prime ℓ. So Φ_ℓ is representable as a finite étale group scheme over the field $\mathbf{F}_\ell$. We have that

$$\Phi = \bigoplus_{\ell \mid N} (i_\ell)_* \Phi_\ell,$$

where $i_\ell : \mathrm{Spec}\mathbf{F}_\ell \hookrightarrow X$ is the natural closed immersion. We have an exact sequence

$$0 \to E^o(X) \to E(\mathbf{Q}) \to H^0(X,\Phi) \to H^1(X,E^o) \to H^1(X,E) \to H^1(X,\Phi), \tag{2}$$

where cohomology is computed for the étale topology. We have, for either topology,

$$H^i(X,\Phi) = \bigoplus_{\ell \mid N} H^i(\mathrm{Spec}(\mathbf{F}_\ell), \Phi_\ell).$$

Viewing (1) as an exact sequence of sheaves for the flat topology, and passing to the associated cohomology sequence we see that (2) may be thought of, ambiguously as computed for either the étale or the flat topology.

If p is an odd prime number, the p-primary component of the Shafarevich-Tate group of E is the p-primary component of the image of $H^1(X,E^o) \to H^1(X,E)$ (see the appendix to [Maz72]), or equivalently the intersection of the kernels of

$$H^1(X,E) \to H^1(\mathrm{Spec}(\mathbf{F}_\ell), \Phi_\ell).$$

Let p be an odd prime number. Let $E' \subset E$ be the open subgroup scheme of E which is the inverse image of $p\Phi \subset \Phi$, so that we have an exact sequence of sheaves for the flat (or étale) topology:

$$0 \to E' \to E \to \Phi/p\Phi \to 0, \tag{3}$$

and if p is a prime of semistable reduction for E (equivalently: p^2 doesn't divide N) we have an exact sequence of flat group schemes

$$0 \to E[p] \to E \to E' \to 0. \tag{4}$$

Put

$$E[p]^o_{/X} := E[p]_{/X} \bigcap E^o_{/X}.$$

Then $E[p]^o_{/X}$ is an open (quasi-finite) subgroup scheme of $E[p]_{/X}$. Let $\tilde{E}[p]_{/X}$ be any "intermediate" open (quasi-finite) subgroup scheme

$$E[p]^o_{/X} \subset \tilde{E}[p]_{/X} \subset E[p]_{/X}$$

so that we have the exact sequence of sheaves for the finite flat topology

$$0 \to \tilde{E}[p]_{/X} \to E[p]_{/X} \to \Psi \to 0, \tag{5}$$

with Ψ a subquotient of Φ.

Consider the following hypothesis:

$\mathbf{A}(E,p,\ell)$: The Galois module $\Phi_\ell/p\Phi_\ell$ is either trivial, or else is a non-constant cyclic Galois module over $\mathbf{F}_\ell$.

Let $\mathbf{A}(E,p)$ denote the conjunction of Hypotheses $\mathbf{A}(E,p,\ell)$ for all prime numbers ℓ, or equivalently, for all ℓ dividing N.

Lemma 6.2. *These are equivalent formulations of Hypothesis* $\mathbf{A}(E,p)$.

(a) $\Phi/p\Phi$ *is cohomologically trivial; that is,* $H^0(X,\Phi/p\Phi) = H^1(X,\Phi/p\Phi) = 0$.

(b) *If* Ψ *is any subquotient of* Φ, Ψ *is "p-cohomologically trivial" in the sense that the p-primary components of* $H^i(X,\Psi)$ *vanish for all* i.

Moreover, if $p \geq 5$, *or if* $p = 3$ *and* E *has no Néron fibers of type IV or IV*, the above conditions are equivalent to:*

(c) *For every* ℓ *at which* E *has split multiplicative reduction,* p *does not divide the order of the group of connected components of the Néron fiber of* E *at* ℓ.

Proof. The equivalence of Hypothesis $\mathbf{A}(E,p)$ with (a) and with (b) is straightforward using standard exact sequences plus the fact that the p-primary components of the (underlying abelian group of) Φ_ℓ is cyclic since $p > 2$; and noting that a (finite) G-module of prime order with nontrivial G-action has trivial cohomology. For (c) we are using that if $p > 2$ the p-primary component of Φ_ℓ vanishes for all primes ℓ of additive reduction for E except when $p = 3$ and the Néron fiber type of E at ℓ is IV or IV*. □

A morphism $G_1 \to G_2$ of flat (commutative, finite type) groups schemes over X will be said to *induce an isomorphism on p-cohomology* if the induced mappings

$$H^i(X,G_1) \otimes \mathbf{Z}_p \to H^i(X,G_2) \otimes \mathbf{Z}_p$$

are isomorphisms for all $i \geq 0$, where cohomology is computed for the flat topology.

Lemma 6.3. *Let* p *be an odd prime number for which* $\mathbf{A}(E,p)$ *holds. We have that the natural morphisms*

$$E^o_{/X} \to E'_{/X} \qquad \textit{and} \qquad E'_{/X} \to E_{/X}$$

induce isomorphisms on p-cohomology. If p *is of semistable reduction for* E, *we also have that*

$$\tilde{E}[p]_{/X} \to E[p]_{/X}$$

induces isomorphisms on p-cohomology, for any of the open subgroup schemes $\tilde{E}[p]_{/X}$ in $E[p]_{/X}$ described above.

Proof. These all sit in short exact sequences of sheaves of abelian groups for the flat topology over X where the third sheaf is p-cohomologically trivial. □

Corollary 6.4. *If $p > 2$ and $\mathbf{A}(E, p)$ holds we have natural isomorphisms*

$$H^0(X, E^o) \otimes \mathbf{Z}_p \cong H^0(X, E') \otimes \mathbf{Z}_p \cong E(\mathbf{Q}) \otimes \mathbf{Z}_p$$

and

$$Ш(E) \otimes \mathbf{Z}_p \cong H^1(X, E^o) \otimes \mathbf{Z}_p \cong H^1(X, E') \otimes \mathbf{Z}_p \cong H^1(X, E) \otimes \mathbf{Z}_p.$$

Corollary 6.5. *Let p be an odd prime number, semistable for E, and suppose that $\mathbf{A}(E, p)$ holds.*

(i) *The image of the natural (injective) coboundary mapping*

$$0 \to E(\mathbf{Q})/pE(\mathbf{Q}) \hookrightarrow H^1(G_{\mathbf{Q}}, E[p])$$

attached to the Kummer sequence is contained in the image of the natural injection

$$H^1(X, E[p]^o) \hookrightarrow H^1(G_{\mathbf{Q}}, E[p]).$$

(ii) *We have an exact sequence*

$$0 \to E(\mathbf{Q})/pE(\mathbf{Q}) \to H^1(X, \tilde{E}[p]) \to Ш(E)[p] \to 0$$

for any of the open subgroup schemes $\tilde{E}[p]_{/X} \subset E[p]_{/X}$ defined above.

(iii) *The image of $H^1(X, \tilde{E}[p]) \hookrightarrow H^1(G_{\mathbf{Q}}, E[p])$ is equal to the p-Selmer subgroup,*

$$S_p(E) \subset H^1(G_{\mathbf{Q}}, E[p]).$$

Proof. All this follows from straightforward calculations using the cohomological exact sequences associated to the exact sequences (1)–(5) in the light of the previous discussion. □

To set things up for our application, let us record the following:

Corollary 6.6. *Let $E_{/\mathbf{Q}}$ and $F_{/\mathbf{Q}}$ be elliptic curves over $\mathbf{Q}$. Let p be an odd prime number of semistable reduction for E and F, and for which $\mathbf{A}(E, p)$ and $\mathbf{A}(F, p)$ both hold. Define $\tilde{E}[p]_{/X} \subset E[p]_{/X}$ to be the open quasi-finite subgroup scheme whose restriction to Y is equal to $E[p]_{/Y}$ and whose fiber at $\mathbf{F}_p$ is equal to $E[p]^o_{/\mathbf{F}_p} = (E[p] \bigcap E^o)_{/\mathbf{F}_p}$. Define $\tilde{F}[p]_{/X}$ similarly. Suppose, finally, that we have an isomorphism of $G_{\mathbf{Q}}$-modules $\iota : F[p]_{/\mathbf{Q}} \cong E[p]_{/\mathbf{Q}}$ which extends to an injection of quasi-finite flat group schemes*

$$\tilde{F}[p]_{/X} \hookrightarrow \tilde{E}[p]_{/X}.$$

Letting

$$H := H^1(G_{\mathbf{Q}}, E[p]) = H^1(G_{\mathbf{Q}}, F[p])$$

(making the identification via ι) we have that the p-Selmer groups $S_p(E) \subset H$ and $S_p(F) \subset H$ are the same.

Proposition 6.7. *Let (E, F, p) be a triple which is an entry of Table 1 of* [CM]. *Suppose further that p is of semistable reduction for E and for F. Then, with the notation of the previous corollary, the p-Selmer groups $S_p(E) \subset H$ and $S_p(F) \subset H$ are the same. In the terminology of* [CM] *the Shafarevich-Tate group of E is explained by the Mordell-Weil group of F.*

Proof. As mentioned above, we have checked that $E[p]_{/\mathbf{Q}} = F[p]_{/\mathbf{Q}} \subset J_0(N)$ whenever the pair E and F (appearing as entry of Table 1 of [CM]) have the same conductor. We have checked that $E[p]_{/\mathbf{Q}} \cong F[p]_{/\mathbf{Q}}$ for the three entries where E and F have different conductor ($E =$ **2932A**, **3306B**, and **5136B**). We have checked that Hypothesis $A(E, p, \ell)$ and $A(F, p, \ell)$ hold for all quadruples (E, F, p, ℓ) such that (E, F, p) occurs as an entry in Table 1 of [CM] (even when p is not semistable for E and F) with the exception of the entry $(E, F, p, \ell) = (\mathbf{2366D}, \mathbf{2366E}, 3, 13)$.

Sublemma 6.8. *Under the hypotheses of our proposition, the isomorphism of $G_{\mathbf{Q}}$-modules $\iota : E[p]_{/\mathbf{Q}} \cong F[p]_{/\mathbf{Q}}$ extends to an injection of quasi-finite flat group schemes*

$$\tilde{E}[p]_{/X} \hookrightarrow \tilde{F}[p]_{/X}$$

which is an isomorphism except in two instances ($E =$ **3306B**, *and* **5136B***).*

Proof. First, since $\tilde{E}[p]_{/Y} = E[p]_{/Y}$, $\tilde{F}[p]_{/Y} = F[p]_{/Y}$, and, as we mentioned at the beginning, both of these quasi-finite, flat (étale) group schemes $F[p]_{/Y}$ and $E[p]_{/Y}$ enjoy the Néronian property, the isomorphism ι extends to an isomorphism $\tilde{E}[p]_{/Y} \cong \tilde{F}[p]_{/Y}$. The remaining question is then local about p. If p is of good reduction for E, then $\tilde{E}[p]_{/X_p}$ and $\tilde{F}[p]_{/X_p}$ are both finite flat group schemes of odd order, so by Fontaine's Theorem [Fon75], the isomorphism between their generic fibers extends to an isomorphism over X_p. (Compare: Theorem I.1.4 in [Maz77].) A standard result allows us to *patch* the isomorphism extending ι over Y with the isomorphism ("extending ι") over X_p to get the extension of ι to an isomorphism of group schemes over X, $\tilde{E}[p]_{/X} \cong \tilde{F}[p]_{/X}$. Now consider the case where p is of bad reduction. By the assumptions of our proposition, p is then of multiplicative reduction for E, and hence the fiber of E over $\mathbf{F}_p$ is a finite multiplicative type group scheme of order p. We therefore have that $\tilde{E}[p]_{/X_p}$ sits in an exact sequence

$$0 \to \mathcal{C}_{/X_p} \to \tilde{E}[p]_{/X_p} \to \mathcal{E}_{/X_p} \to 0 \tag{6}$$

where $\mathcal{C}_{/X_p}$ is a finite flat group scheme of order p (and with fiber of multiplicative type in characteristic p) and where $\mathcal{E}_{/X_p}$ is an étale quasi-finite group scheme, with trivial fiber in characteristic p.

Let us take a moment to recall (see [Maz78, Lem. 1.1]) the construction of such an exact sequence (6): working in the category of formal schemes, let $\hat{X}_p := \mathrm{Spf}(\mathbf{Z}_p)$, and let $\hat{\mathcal{C}}_{/\hat{X}_p}$ be the formal completion of the zero-section in $\tilde{E}[p]_{/X_p}$. One checks that $\hat{\mathcal{C}}_{/\hat{X}_p}$ may be identified with a finite flat formal group scheme over $\hat{X}_p$ which admits a closed immersion into the formal group scheme over $\hat{X}_p$ associated to $\tilde{E}[p]_{/X_p}$. A standard algebrization argument establishes that there is a (unique) finite flat subgroup scheme $\mathcal{C}_{/X_p} \subset \tilde{E}[p]_{/X_p}$ whose associated formal group scheme over $\hat{X}_p$ is $\hat{\mathcal{C}}_{/\hat{X}_p}$. The exact sequence (6) is then obtained by letting $\mathcal{E}_{/X_p}$ be the evident quotient (quasi-finite flat) group scheme, and noting that, by construction, its special fiber is trivial.

Now let us return to the proof of the sublemma. Since the restriction of ι to $\mathcal{C}_{/X_p}$ (a finite flat multiplicative type group scheme of order p) is injective over the generic point, it follows (by elementary considerations, or by Fontaine's Theorem cited above) that ι restricted to $\mathcal{C}_{/X_p}$ is an injection over X_p. Since $\mathcal{E}_{/X_p}$ has trivial fiber in characteristic p, ι is an injection as was to be proved. In all cases under consideration, then,

$$\iota : \tilde{E}[p]_{/X} \hookrightarrow \tilde{F}[p]_{/X}$$

is an injection. If E is of good reduction at p, or if F is of bad reduction at p, ι is therefore an isomorphism. The cases remaining are when E is of bad reduction at p and F is of good reduction (i.e., $E =$ **3306B**, and **5136B**) in which case we can only assert that ι is an injection. □

Returning to our proposition, suppose that $\tilde{F}[p]_{/X_p}$ is finite flat (which happens in the two cases signalled above: $E =$ **3306B**, and **5136B**). Then the isomorphism induced by ι on generic fibers

$$\tilde{E}[p]_{/\mathbf{Q}_p} \cong \tilde{F}[p]_{/\mathbf{Q}_p}$$

restricted to the $G_{\mathbf{Q}_p}$-stable subgroup $\mathcal{C}_{/\mathbf{Q}_p} \subset \tilde{E}[p]_{/\mathbf{Q}_p}$ extends to a morphism of the finite flat group scheme $\mathcal{C}_{/X_p}$ into $\tilde{F}[p]_{/X_p}$. This extended morphism $j : \mathcal{C}_{/X_p} \to \tilde{F}[p]_{X_p}$ is necessarily a closed immersion since $\mathcal{C}_{/X_p}$ is a multiplicative type finite flat group scheme. Since $\mathcal{E}_{/X_p}$ has trivial fiber in characteristic p an application of the standard patching argument (as used in the previous case) allows us to put together the isomorphism of group schemes over Y extending ι with the closed immersion j over X_p to get a closed immersion

$$\tilde{E}[p]_{/X} \hookrightarrow \tilde{F}[p]_{/X}.$$

Finally suppose that both E and F have multiplicative reduction at p. We then have exact sequences (6) for each of our quasi-finite flat group schemes $\tilde{E}[p]_{/X_p}$ and $\tilde{F}[p]_{/X_p}$. Let V denote their common generic fiber (identified via ι) considered as two-dimensional $\mathbf{F}_p$-vector space with $G_{\mathbf{Q_p}}$-action. Let $\mathcal{C}(E) \subset V$ and $\mathcal{C}(F) \subset V$ denote the one-dimensional subspaces given by the generic fibers of the finite flat subgroup schemes $\mathcal{C}_{/X_p}$ corresponding to the exact sequence (6) for for E and for F respectively. Suppose, first, that these one-dimensional $\mathbf{F}_p$-subspaces $\mathcal{C}(E)$ and $\mathcal{C}(F)$ are different. It then follows that the $G_{\mathbf{Q}_p}$-representation V splits as the direct sum of $\mathcal{C}(E)$ and $\mathcal{C}(F)$, both $\mathbf{F}_p$-subspaces being isomorphic, as $I_{\mathbf{Q_p}}$-modules to μ_p, where $I_{\mathbf{Q_p}} \subset G_{\mathbf{Q_p}}$ is the inertia subgroup of $G_{\mathbf{Q}_p}$. But this contradicts the fact that V is self-Cartier dual (under the Weil pairing). Consequently, $\mathcal{C}(E) = \mathcal{C}(F) \subset V$. From the above discussion it follows that we can extend ι to an isomorphism $\tilde{E}[p]_{/X_p} \cong \tilde{F}[p]_{/X_p}$.

Our proposition then follows (from Corollary 6.6) for all entries in Table 1 of [CM] where p is of semistable reduction for E once we produce special arguments to cover the three special cases $E =$ **3306B**, **5136B** and **2366D**. The first two of these cases are "special" because we only have an injection $\tilde{E}[p]_{/X_p} \hookrightarrow \tilde{F}[p]_{/X_p}$ and not an isomorphism. However, the cokernel of this morphism restricted to the fiber in characteristic 3 is, in both of these cases, a cyclic group with nontrivial $G_{\mathbf{Q_3}}$-action and hence is 3-cohomologically trivial. In particular, the injection $\tilde{E}[p]_{/X_p} \hookrightarrow \tilde{F}[p]_{/X_p}$ induces an isomorphism on flat cohomology over X, and the argument for these two cases proceeds as before. This leaves $(E, F, p) = (\mathbf{2366D}, \mathbf{2366E}, 3)$

which is the only example of an entry (E, F, p) in our table, where E has a $\mathbf{Q}$-rational point of order p, and (this is no accident) where Hypothesis $\mathbf{A}(E, p)$ and Hypothesis $\mathbf{A}(F, p)$ fail. (Indeed there are no other failures of Hypothesis $\mathbf{A}(F, p)$ for any of the (E, F, p)'s occurring in Table 1 of [CM] and only one other failure of Hypothesis $\mathbf{A}(E, p)$, which is for $(E, p, \ell) = (\mathbf{2932A}, 3, 2)$.)

Let us now deal with the case $(E, F, p) = (\mathbf{2366D}, \mathbf{2366E}, 3)$. The subgroup C of $\mathbf{Q}$-rational points of order 3 on E specialize in characteristic 13 to yield an isomorphism

$$C \cong \Phi_{13}$$

and the same for the subgroup of $\mathbf{Q}$-rational points of order 3 on F. We make use of this information to cut down the group schemes $\tilde{E}[3]_{/X}$ and $\tilde{F}[3]_{/X}$ and define open subgroup schemes: $\tilde{\tilde{E}}[3]_{/X} \subset \tilde{E}[3]_{/X}$ and $\tilde{\tilde{E}}[3]_{/X} \subset \tilde{E}[3]_{/X}$ by requiring that these closed immersions of subgroup schemes be isomorphisms outside characteristic 13, and that the "double-tilded" group schemes each have trivial fiber in characteristic 13. We get via the above argument an isomorphism of group schemes $\tilde{\tilde{E}}[3]_{/X} \cong \tilde{\tilde{F}}[3]_{/X}$ extending ι, and an identification of the 3-Selmer groups of E and F with $H^1(X, \tilde{\tilde{E}}[3])$ and $H^1(X, \tilde{\tilde{F}}[3])$ respectively. Our proposition is proved. □

It remains to say a few words about why, in the 7 cases of entries (E, F, p) in our Table 1 of [CM] for which p is a prime of additive reduction for E *some* nontrivial elements of the Shafarevich-Tate group of E are explained by the Mordell-Weil group of F. Briefly, the reason is as follows. By the *inflated* p-Selmer group of E (and of F) let us mean the subgroup of H obtained by insisting upon all the local Selmer conditions at primes different from p, but putting no condition at p. The p-Selmer group of E (and of F) are, in all 7 instances, $\mathbf{F}_p$-vector spaces of dimension 2 and therefore, the inflated p-Selmer groups are of dimensions either 2 or 3. Working over Y rather than over X, the above argument applied to these 7 remaining cases gives us an identification of the *inflated* p-Selmer groups of E and of F in H. But the true p-Selmer groups (vector spaces of dimension 2) being subspaces in a vector space of dimension ≤ 3 must have a nontrivial intersection.

References

[Aga99] A. Agashe, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. **328** (1999), no. 5, 369–374. MR **2000e**:11083

[AS02] A. Agashe and W. A. Stein, *Visibility of Shafarevich-Tate Groups of Abelian Varieties*, to appear in J. of Number Theory (2002).

[AS04] A. Agashe and W. A. Stein, *The manin constant, congruence primes, and the modular degree*, Preprint,
`http://modular.fas.harvard.edu/papers/manin-agashe/` (2004).

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[Bir71] B. J. Birch, *Elliptic curves over* $\mathbf{Q}$*: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, pp. 396–400. MR **47**:3395

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR **91i**:14034

[Cas62] J. W. S. Cassels, *Arithmetic on curves of genus* 1*. III. The Tate-Šafarevič and Selmer groups*, Proc. London Math. Soc. (3) **12** (1962), 259–296. MR 29 #1212

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997. MR **99e**:11068

[CM] J. E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. **9** (2000), no. 1, 13–28. MR 1 758 797

[CS01] B. Conrad and W. A. Stein, *Component groups of purely toric quotients*, Math. Res. Lett. **8** (2001), no. 5-6, 745–766. MR **2003f:**11087

[Del01] C. Delaunay, *Heuristics on Tate-Shafarevitch groups of elliptic curves defined over* $\mathbf{Q}$, Experiment. Math. **10** (2001), no. 2, 191–196.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133. MR **97g**:11044

[Edi91] B. Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Birkhäuser Boston, Boston, MA, 1991, pp. 25–39. MR **92a**:11066

[Eme01] M. Emerton, *Optimal quotients of modular Jacobians.* Preprint.

[FpS$^+$01] E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. Comp. **70** (2001), no. 236, 1675–1697. MR 1 836 926

[Fon75] J-M. Fontaine, *Groupes finis commutatifs sur les vecteurs de Witt*, C. R. Acad. Sci. Paris Sér. A-B **280** (1975), Ai, A1423–A1425. MR **51**:10353

[Gro94] B. H. Gross, *L-functions at the central critical point*, Motives (Seattle, WA, 1991), Amer. Math. Soc., Providence, RI, 1994, pp. 527–535. MR **95a**:11060

[Gro68] A. Grothendieck, *Le groupe de Brauer. III. Exemples et compléments*, Dix Exposés sur la Cohomologie des Schémas, North-Holland, Amsterdam, 1968, pp. 88–188. MR **39**:5586c

[Gro72] A. Grothendieck, *Modèles de Néron et monodromie* in *Groupes de monodromie en géométrie algébrique. I*, Springer-Verlag, Berlin, 1972, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 I), Dirigé par A. Grothendieck. Vol. 288. MR **50**:7134

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR **87j**:11057

[Kat81] N. M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. **62** (1981), no. 3, 481–502. MR **82d**:14025

[KL89] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Algebra i Analiz **1** (1989), no. 5, 171–196. MR **91c**:11032

[KL92] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of* Ш *over totally real fields*, Math. USSR Izvestiya **39** (1992), no. 1, 829–853. MR **93d**:11063

[KS00] D. R. Kohel and W. A. Stein, *Component Groups of Quotients of $J_0(N)$*, Proceedings of the 4th International Symposium (ANTS-IV), Leiden, Netherlands, July 2–7, 2000 (Berlin), Springer, 2000. MR 1 850 621

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR **93a**:11048

[LO85] H. W. Lenstra, Jr. and F. Oort, *Abelian varieties having purely additive reduction*, J. Pure Appl. Algebra **36** (1985), no. 3, 281–298. MR **86e**:14020

[LS02] Joan-C. Lario and René Schoof, *Some computations with Hecke rings and deformation rings*, Experiment. Math. **11** (2002), no. 2, 303–311, With an appendix by Amod Agashe and William Stein. MR **2004b:**11072

[Maz72] B. Mazur, *Rational points of abelian varieties with values in towers of number fields*, Invent. Math. **18** (1972), 183–266. MR **56**:3020

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978). MR **80c**:14015

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162. MR **80h**:14022

[MT74] B. Mazur and J. Tate, *Points of order* 13 *on elliptic curves*, Invent. Math. **22** (1973/74), 41–49. MR **50**:327

[Mil86] J. S. Milne, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[Ogg73] A. P. Ogg, *Rational points on certain elliptic modular curves*, Analytic number theory (Proc. Sympos. Pure Math., Vol XXIV, St. Louis Univ., St. Louis, Mo., 1972), Amer. Math. Soc., Providence, R.I., 1973, pp. 221–231. MR **49**:2743

[PS99] B. Poonen and M. Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2) **150** (1999), no. 3, 1109–1149. MR **2000m**:11048

[Shi73] G. Shimura, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544. MR **47**:6709

[Shi94] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1. MR **95e**11048

[Ste82] G. Stevens, *Arithmetic on modular curves*, Birkhäuser Boston Inc., Boston, Mass., 1982. MR **87b**:11050

[Ste00] W. A. Stein, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley (2000).

[Ste02a] W. A. Stein, *An introduction to computing modular forms using modular symbols*, to appear in an MSRI Proceedings (2002).

[Ste02b] W. A. Stein, *Shafarevich-tate groups of nonsquare order*, Proceedings of MCAV 2002, Progress of Mathematics (to appear).

[Stu87] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin, 1987, pp. 275–280. MR **88h**:11031

[Tat63] J. Tate, *Duality theorems in Galois cohomology over number fields*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 288–295. MR **31**:168

[Tat66] J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1966 (reprinted in 1995), pp. Exp. No. 306, 415–440. MR 1 610 977

Department of Mathematics, U niversity of Texas, Austin, Texas 78712
E-mail address: agashe@math.utexas.edu

Department of Mathematics, Harvard University, Cambridge, Massachussetts 02138
E-mail address: was@math.harvard.edu

14 $J_1(p)$ Has Connected Fibers, with B. Conrad and B. Edixhoven

$J_1(p)$ Has Connected Fibers

Brian Conrad, Bas Edixhoven, William Stein

Received: November 23, 2002

Communicated by Stephen Lichtenbaum

Abstract. We study resolution of tame cyclic quotient singularities on arithmetic surfaces, and use it to prove that for any subgroup $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ the map $X_H(p) = X_1(p)/H \to X_0(p)$ induces an injection $\Phi(J_H(p)) \to \Phi(J_0(p))$ on mod p component groups, with image equal to that of H in $\Phi(J_0(p))$ when the latter is viewed as a quotient of the cyclic group $(\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$. In particular, $\Phi(J_H(p))$ is always Eisenstein in the sense of Mazur and Ribet, and $\Phi(J_1(p))$ is trivial: that is, $J_1(p)$ has connected fibers. We also compute tables of arithmetic invariants of optimal quotients of $J_1(p)$.

2000 Mathematics Subject Classification: 11F11, 11Y40, 14H40
Keywords and Phrases: Jacobians of modular curves, Component groups, Resolution of singularities

Contents

1 Introduction

Let p be a prime and let $X_1(p)_{/\mathbf{Q}}$ be the projective smooth algebraic curve over $\mathbf{Q}$ that classifies elliptic curves equipped with a point of exact order p. Let $J_1(p)_{/\mathbf{Q}}$ be its Jacobian. One of the goals of this paper is to prove:

Theorem 1.1.1. *For every prime p, the Néron model of $J_1(p)_{/\mathbf{Q}}$ over $\mathbf{Z}_{(p)}$ has closed fiber with trivial geometric component group.*

This theorem is obvious when $X_1(p)$ has genus 0 (*i.e.*, for $p \leq 7$), and for $p = 11$ it is equivalent to the well-known fact that the elliptic curve $X_1(11)$ has j-invariant with a simple pole at 11 (the j-invariant is $-2^{12}/11$). The strategy of the proof in the general case is to show that $X_1(p)_{/\mathbf{Q}}$ has a regular proper model $\mathcal{X}_1(p)_{/\mathbf{Z}_{(p)}}$ whose closed fiber is geometrically integral. Once we have such a model, by using the well-known dictionary relating the Néron model of a generic-fiber Jacobian with the relative Picard scheme of a regular proper model (see [9, Ch. 9], esp. [9, 9.5/4, 9.6/1], and the references therein), it follows that the Néron model of $J_1(p)$ over $\mathbf{Z}_{(p)}$ has (geometrically) connected closed fiber, as desired. The main work is therefore to prove the following theorem:

Theorem 1.1.2. *Let p be a prime. There is a regular proper model $\mathcal{X}_1(p)$ of $X_1(p)_{/\mathbf{Q}}$ over $\mathbf{Z}_{(p)}$ with geometrically integral closed fiber.*

What we really prove is that if $X_1(p)^{\mathrm{reg}}$ denotes the minimal regular resolution of the normal (typically non-regular) coarse moduli scheme $X_1(p)_{/\mathbf{Z}_{(p)}}$, then a minimal regular contraction $\mathcal{X}_1(p)$ of $X_1(p)^{\mathrm{reg}}$ has geometrically integral closed fiber; after all the contractions of -1-curves are done, the component that remains corresponds to the component of $X_1(p)_{/\mathbf{F}_p}$ classifying étale order-p subgroups. When $p > 7$, so the generic fiber has positive genus, such a minimal regular contraction is the unique minimal regular proper model of $X_1(p)_{/\mathbf{Q}}$.

Theorem 1.1.2 provides natural examples of a finite map π between curves of arbitrarily large genus such that π does not extend to a morphism of the minimal regular proper models. Indeed, consider the natural map

$$\pi : X_1(p)_{/\mathbf{Q}} \to X_0(p)_{/\mathbf{Q}}.$$

When $p = 11$ or $p > 13$, the target has minimal regular proper model over $\mathbf{Z}_{(p)}$ with reducible geometric closed fiber [45, Appendix], while the source has minimal regular proper model with (geometrically) integral closed fiber, by Theorem 1.1.2. If the map extended, it would be proper and dominant (as source and target have unique generic points), and hence surjective. On the level of closed fibers, there cannot be a surjection from an irreducible scheme onto a reducible scheme. By the valuative criterion for properness, π is defined in codimension 1 on minimal regular proper models, so there are finitely many points of $\mathcal{X}_1(p)$ in codimension 2 where π cannot be defined.

Note that the fiber of $J_1(p)$ at infinity need not be connected. More specifically, a modular-symbols computation shows that the component group of $J_1(p)(\mathbf{R})$ has order 2 for $p = 17$ and $p = 41$. In contrast, A. Agashe has observed that [47, §1.3] implies that $J_0(p)(\mathbf{R})$ is always connected.

Rather than prove Theorem 1.1.2 directly, we work out the minimal regular model for $X_H(p)$ over $\mathbf{Z}_{(p)}$ for any subgroup $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ and use this to study the mod p component group of the Jacobian $J_H(p)$; note that $J_H(p)$ usually does not have semistable reduction. Our basic method is to use a variant on the classical Jung–Hirzebruch method for complex surfaces, adapted to the case of a proper curve over an arbitrary discrete valuation ring. We refer the reader to Theorem 2.4.1 for the main result in this direction; this is the main new theoretical contribution of the paper. This technique will be applied to prove:

Theorem 1.1.3. *For any prime p and any subgroup H of $(\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$, the natural surjective map $J_H(p) \to J_0(p)$ of Albanese functoriality induces an injection on geometric component groups of mod-p fibers, with the component* *group $\Phi(\mathcal{J}_H(p)_{/\overline{\mathbf{F}}_p})$ being cyclic of order $|H|/\gcd(|H|,6)$. In particular, the finite étale component-group scheme $\Phi(\mathcal{J}_H(p)_{/\mathbf{F}_p})$ is constant over $\mathbf{F}_p$.*

If we view the constant cyclic component group $\Phi(\mathcal{J}_0(p)_{/\mathbf{F}_p})$ as a quotient of the cyclic $(\mathbf{Z}/p)^\times/\{\pm 1\}$, then the image of the subgroup $\Phi(\mathcal{J}_H(p)_{/\mathbf{F}_p})$ in this quotient is the image of $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ in this quotient.

Remark 1.1.4. The non-canonical nature of presenting one finite cyclic group as a quotient of another is harmless when following images of subgroups under maps, so the final part of Theorem 1.1.3 is well-posed.

The constancy in Theorem 1.1.3 follows from the injectivity claim and the fact that $\Phi(\mathcal{J}_0(p)_{/\mathbf{F}_p})$ is constant. Such constancy was proved by Mazur-Rapoport [45, Appendix], where it is also shown that this component group for $J_0(p)$ is cyclic of the order indicated in Theorem 1.1.3 for $H = (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$.

Since the Albanese map is compatible with the natural map $\mathbf{T}_H(p) \to \mathbf{T}_0(p)$ on Hecke rings and Mazur proved [45, §11] that $\Phi(\mathcal{J}_0(p)_{/\overline{\mathbf{F}}_p})$ is Eisenstein as a $\mathbf{T}_0(p)$-module, we obtain:

Corollary 1.1.5. *The Hecke module $\Phi(\mathcal{J}_H(p)_{/\overline{\mathbf{F}}_p})$ is Eisenstein as a $\mathbf{T}_H(p)$-module* (i.e., *T_ℓ acts as $1+\ell$ for all $\ell \neq p$ and $\langle d\rangle$ acts trivially for all $d \in (\mathbf{Z}/p\mathbf{Z})^\times$*).

In view of Eisenstein results for component groups due to Edixhoven [18] and Ribet [54], [55] (where Ribet gives examples of non-Eisenstein component groups), it would be of interest to explore the range of validity of Corollary 1.1.5 when auxiliary prime-to-p level structure of $\Gamma_0(N)$-type is allowed. A modification of the methods we use should be able to settle this more general problem. In fact, a natural approach would be to aim to essentially reduce to the Eisenstein results in [54] by establishing a variant of the above injectivity result on component groups when additional $\Gamma_0(N)$ level structure is allowed away from p. This would require a new idea in order to avoid the crutch of cyclicity (the case of $\Gamma_1(N)$ seems much easier to treat using our methods because the relevant groups tend to be cyclic, though we have not worked out the details for $N > 1$), and preliminary calculations of divisibility among orders of component groups are consistent with such injectivity.

In order to prove Theorem 1.1.3, we actually first prove a surjectivity result:

Theorem 1.1.6. *The map of Picard functoriality $J_0(p) \to J_H(p)$ induces a surjection on mod p component groups, with the mod p component group for $J_H(p)$ having order $|H|/\gcd(|H|,6)$.*

In particular, each connected component of $\mathcal{J}_H(p)_{/\mathbf{F}_p}$ contains a multiple of the image of $(0)-(\infty) \in \mathcal{J}_0(p)(\mathbf{Z}_{(p)})$ in $\mathcal{J}_H(p)(\mathbf{F}_p)$.

Let us explain how to deduce Theorem 1.1.3 from Theorem 1.1.6. Recall [28, Exposé IX] that for a discrete valuation ring R with fraction field K and an abelian variety A over K over R, Grothendieck's biextension pairing sets up a bilinear pairing between the component groups of the closed fibers of the Néron models of A and its dual A'. Moreover, under this pairing the component-group map induced by a morphism $f : A \to B$ (to another abelian variety) has as an adjoint the component-group map induced by the dual morphism $f' : B' \to A'$. Since Albanese and Picard functoriality maps on Jacobians are dual to each other, the surjectivity of the Picard map therefore implies the injectivity of the Albanese map provided that the biextension pairings in question are perfect pairings (and then the description of the image of the resulting Albanese injection in terms of H as in Theorem 1.1.3 follows immediately from the order calculation in Theorem 1.1.6).

In general the biextension pairing for an abelian variety and its dual need not be perfect [8], but once it is known to be perfect for the $J_H(p)$'s then surjectivity of the Picard map in Theorem 1.1.6 implies the injectivity of the Albanese map as required in Theorem 1.1.3. To establish the desired perfectness, one can use either that the biextension pairing is always perfect in case of generic characteristic 0 with a perfect residue field [6, Thm. 8.3.3], or that surjectivity of the Picard map ensures that $J_H(p)$ has mod p component group of order prime to p, and the biextension pairing is always perfect on primary components prime to the residue characteristic [7, §3, Thm. 7].

It is probable that the results concerning the component groups $\Phi(\mathcal{J}_H(p)_{/\overline{\mathbf{F}}_p})$ and the maps between them that are proved in this article via models of $X_H(p)$ over $\mathbf{Z}_{(p)}$ can also be proved using [20, 5.4, Rem. 1], and the well-known stable model of $X_1(p)$ over $\mathbf{Z}_{(p)}[\zeta_p]$ that one can find for example in [30]. (This observation was prompted by questions of Robert Coleman.) However, such an approach does not give information on regular models of $X_H(p)$ over $\mathbf{Z}_{(p)}$. Hence we prefer the method of this paper.

Acknowledgements. This paper was inspired by the following question of Shuzo Takehashi: "What can be said about component groups of optimal quotients of $J_1(p)$?". Edixhoven was partially supported by the European Research Training Network Contract HPRN-CT-2000-00120 "arithmetic algebraic geometry", while Conrad and Stein are grateful to the NSF for support during work on this paper. Conrad is also grateful to the Sloan Foundation for support and to Henri Darmon, Paul Hacking, and Victor Kac. Dino Lorenzini made a large number of helpful remarks. Stein thanks Ken Ribet and Sheldon Kamienny for discussions about the torsion subgroup of $J_1(p)$, Allan Steel of Magma for computational support, and William R. Hearst III and Sun Microsystems for donating computers that were used for computations in this paper. The authors thank the referee for a thoughtful report.

1.2 Outline

Section 1.3 contains a few background notational remarks. In Section 2 we develop the basic Jung–Hirzebruch resolution technique in the context of tame cyclic quotient surface singularities. This includes mod-p singularities on many (coarse) modular curves when $p > 3$ and the p-power level structure is only on p-torsion. In Section 3, we recall some general results on moduli problems for elliptic curves and coarse moduli schemes for such problems. In Section 4, we use the results of Sections 2 and 3 to locate all the non-regular points on the coarse moduli scheme $X_H(p)_{/\mathbf{Z}_{(p)}}$ (*e.g.*, when H is trivial this is the set of $\mathbf{F}_p$-rational points $(E, 0)$ with $j = 0, 1728$). In Section 5, we use the Jung–Hirzebruch formulas to compute the minimal regular resolution $X_H(p)^{\mathrm{reg}}$ of $X_H(p)_{/\mathbf{Z}_{(p)}}$, and we use use a series of intersection number computations to obtain a regular proper model for $X_H(p)_{/\mathbf{Q}}$; from this, the desired results on component groups follow. We conclude in Section 6 with some computer computations concerning the arithmetic of $J_1(p)$ for small p, where (among other things) we propose a formula for the order of the torsion subgroup of $J_1(p)(\mathbf{Q})$.

To avoid using Weierstrass equations in proofs, we have sometimes argued more abstractly than is strictly necessary, but this has the merit of enabling us to treat cusps by essentially the same methods as the other points. We would prefer to avoid mentioning j-invariants, but it is more succinct to say "cases with $j = 0$" than it is to say "cases such that $\mathrm{Aut}(E_{/k})$ has order 6."

Because we generally use methods of abstract deformation theory, the same approach should apply to Drinfeld modular curves, as well as to cases with auxiliary level structure away from p (including mod p component groups of suitable Shimura curves associated to indefinite quaternion algebras over $\mathbf{Q}$, with p not dividing the discriminant). However, since a few additional technicalities arise, we leave these examples to be treated at a future time.

1.3 Notation and terminology

Throughout this paper, p denotes an arbitrary prime unless otherwise indicated. Although the cases $p \leq 3$ are not very interesting from the point of view of our main results, keeping these cases in mind has often led us to more conceptual proofs. We write $\Phi_p(T) = (T^p - 1)/(T - 1) \in \mathbf{Z}[T]$ to denote the pth cyclotomic polynomial (so $\Phi_p(T + 1)$ is p-Eisenstein).

We write $V^\vee$ to denote the dual of a vector space V, and we write $\mathcal{F}^\vee$ to denote the dual of a locally free sheaf $\mathcal{F}$.

If X and S' are schemes over a scheme S then $X_{/S'}$ and $X_{S'}$ denote $X \times_S S'$. If S is an integral scheme with function field K and X is a K-scheme, by a *model* of X (over S) we mean a flat S-scheme with generic fiber X.

By an *S-curve* over a scheme S we mean a flat separated finitely presented map $X \to S$ with fibers of pure dimension 1 (the fibral dimension condition need only be checked on generic fibers, thanks to [27, IV$_3$, 13.2.3] and a reduction to the noetherian case). Of course, when a map of schemes $X \to S$ is proper flat and finitely presented with geometrically connected generic fibers, then the other fibers are automatically geometrically connected (via reduction to the noetherian case and a Stein factorization argument). For purely technical reasons, we do *not* require S-curves to be proper or to have geometrically connected fibers. The main reason for this is that we want to use étale localization arguments on X without having to violate running hypotheses. The use of Corollary 2.2.4 in the proof of Theorem 2.4.1 illustrates this point.

2 RESOLUTION OF SINGULARITIES

Our eventual aim is to determine the component groups of Jacobians of intermediate curves between $X_1(p)$ and $X_0(p)$. Such curves are exactly the quotient curves $X_H(p) = X_1(p)/H$ for subgroups $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$, where we identify the group $\mathrm{Aut}_{\mathbf{Q}}(X_1(p)/X_0(p)) = \mathrm{Aut}_{\overline{\mathbf{Q}}}(X_1(p)/X_0(p))$ with $(\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ via the diamond operators (in terms of moduli, $n \in (\mathbf{Z}/p\mathbf{Z})^\times$ sends a pair (E, P) to the pair $(E, n \cdot P)$). The quotient $X_H(p)_{/\mathbf{Z}_{(p)}}$ is an arithmetic surface with tame cyclic quotient singularities (at least when $p > 3$).

After some background review in Section 2.1 and some discussion of generalities in Section 2.2, in Section 2.3 we will describe a class of curves that give rise to (what we call) *tame cyclic quotient singularities*. Rather than work with global quotient situations X/H, it is more convenient to require such quotient descriptions only on the level of complete local rings. For example, this is what one encounters when computing complete local rings on coarse modular curves: the complete local ring is a subring of invariants of the universal deformation ring under the action of a finite group, but this group-action might not be induced by an action on the global modular curve. In Section 2.4 we establish the Jung–Hirzebruch continued-fraction algorithm that minimally resolves tame cyclic quotient singularities on curves over an arbitrary discrete valuation ring. The proof requires the Artin approximation theorem, and for this reason we need to define the concept of a *curve* as in Section 1.3 without requiring properness or geometric connectivity of fibers.

We should briefly indicate here why we need to use Artin approximation to compute minimal resolutions. Although the end result of our resolution process is intrinsic and of étale local nature on the curve, the mechanism by which the proof gets there depends on coordinatization and is not intrinsic (*e.g.*, we do not blow-up at points, but rather along certain codimension-1 subschemes). The only way we can relate the general case to a coordinate-dependent calculation in a special case is to use Artin approximation to find a common étale neighborhood over the general case and a special case (coupled with the étale local nature of the intrinsic minimal resolution that we are seeking to describe).

These resolution results are applied in subsequent sections to compute a regular proper model of $X_H(p)_{/\mathbf{Q}}$ over $\mathbf{Z}_{(p)}$ in such a way that we can compute both the mod-p geometric component group of the Jacobian $J_H(p)$ and the map induced by $J_0(p) \rightarrow J_H(p)$ on mod-p geometric component-groups. In this way, we will prove Theorem 1.1.6 (as well as Theorem 1.1.2 in the case of trivial H).

2.1 Background review

Some basic references for intersection theory and resolution of singularities for connected proper flat regular curves over Dedekind schemes are [29, Exposé X], [13], and [41, Ch. 9].

If S is a connected Dedekind scheme with function field K and X is a normal S-curve, when S is excellent we can construct a resolution of singularities as follows: blow-up the finitely many non-regular points of X (all in codimension 2), normalize, and then repeat until the process stops. That this process always stops is due to a general theorem of Lipman [40]. For more general (*i.e.*, possibly non-excellent) S, and $X_{/S}$ with *smooth* generic fiber, the same algorithm works (including the fact that the non-regular locus consists of only finitely many closed points in closed fibers). Indeed, when $X_{/K}$ is smooth then the non-smooth locus of $X \to S$ is supported on finitely many closed fibers, so we may assume $S = \mathrm{Spec}(R)$ is local. We can then use Lemma 2.1.1 below to bring results down from $X_{/\widehat{R}}$ since $\widehat{R}$ is excellent.

See Theorem 2.2.2 for the existence and uniqueness of a canonical minimal regular resolution $X^{\mathrm{reg}} \to X$ for any connected Dedekind S when $X_{/K}$ smooth. A general result of Lichtenbaum [39] and Shafarevich [61] ensures that when $X_{/S}$ is also proper (with smooth generic fiber if S isn't excellent), by beginning with X^{reg} (or any regular proper model of $X_{/K}$) we can successively blow down -1-curves (see Definition 2.2.1) in closed fibers over S until there are no more such -1-curves, at which point we have reached a relatively minimal model among the regular proper models of $X_{/K}$. Moreover, when $X_{/K}$ is in addition geometrically integral with positive arithmetic genus (*i.e.*, $\mathrm{H}^1(X_{/K}, \mathscr{O}) \neq 0$), this is the unique relatively minimal regular proper model, up to unique isomorphism.

In various calculations below with proper curves, it will be convenient to work over a base that is complete with algebraically closed residue field. Since passage from $\mathbf{Z}_{(p)}$ to $W(\overline{\mathbf{F}}_p)$ involves base change to a strict henselization followed by base change to a completion, in order to not lose touch with the situation over $\mathbf{Z}_{(p)}$ it is useful to keep in mind that formation of the minimal regular proper model (when the generic fiber is smooth with positive genus) is compatible with base change to a completion, henselization, and strict henselization on the base. We will not really require these results, but we do need to use the key fact in their proof: certain base changes do not destroy regularity or normality (and so in particular commute with formation of normalizations). This is given by:

Lemma 2.1.1. *Let R be a discrete valuation ring with fraction field K and let X be a locally finite type flat R-scheme that has* regular *generic fiber. Let $R \to R'$ be an extension of discrete valuation rings for which $\mathfrak{m}_R R' = \mathfrak{m}_{R'}$ and the residue field extension $k \to k'$ is separable. Assume either that the fraction field extension $K \to K'$ is separable or that $X_{/K}$ is smooth* (*so either way, $X_{/K'}$ is automatically regular*).

For any $x' \in X' = X \times_R R'$ lying over $x \in X$, the local ring $\mathcal{O}_{X',x'}$ is regular (*resp. normal*) *if and only if the local ring $\mathcal{O}_{X,x}$ is regular* (*resp. normal*).

Proof. Since $\mathfrak{m}_R R' = \mathfrak{m}_{R'}$, the map $\pi : X' \to X$ induces $\pi_k : X_{/k} \times_k k' \to X_{/k}$ upon reduction modulo $\mathfrak{m}_R$. The separability of k' over k implies that π_k is a regular morphism. Thus, if x and x' lie in the closed fibers then $\mathcal{O}_{X,x} \to \mathcal{O}_{X',x'}$ is faithfully flat with regular fiber ring $\mathcal{O}_{X',x'}/\mathfrak{m}_x$. Consequently, X is regular at x if and only if X' is regular at x' [44, 23.7]. Meanwhile, if x and x' lie in the generic fibers then they are both regular points since the generic fibers are regular. This settles the regular case.

For the normal case, when X' is normal then the normality of X follows from the faithful flatness of π [44, Cor. to 23.9]. Conversely, when X is normal then to deduce normality of X' we use Serre's "$R_1 + S_2$" criterion. The regularity of X' in codimensions ≤ 1 is clear at points on the regular generic fiber. The only other points of codimension ≤ 1 on X' are the generic points of the closed fiber, and these lie over the (codimension 1) generic points of the closed fiber of X. Such points on X are regular since X is now being assumed to be normal, so the desired regularity on X' follows from the preceding argument. This takes care of the R_1 condition. It remains to check that points $x' \in X'$ in codimensions ≥ 2 contain a regular sequence of length 2 in their local rings. This is clear if x' lies on the regular generic fiber, and otherwise x' is a point of codimension ≥ 1 on the closed fiber. Thus, $x = \pi(x')$ is either a generic point of $X_{/k}$ or is a point of codimension ≥ 1 on $X_{/k}$. In the latter case the normal local ring $\mathcal{O}_{X,x}$ has dimension at least 2 and hence contains a regular sequence of length 2; this gives a regular sequence in the faithfully flat extension ring $\mathcal{O}_{X',x'}$. If instead x is a generic point of $X_{/k}$ then $\mathcal{O}_{X,x}$ is a regular ring. It follows that $\mathcal{O}_{X',x'}$ is regular, so we again get the desired regular sequence (since $\dim \mathcal{O}_{X',x'} \geq 2$). □

We wish to record an elementary result in intersection theory that we will use several times later on. First, some notation needs to be clarified: if X is a connected regular proper curve over a discrete valuation ring R with residue field k, and D and D' are two *distinct* irreducible and reduced divisors in the closed fiber, then

$$D.D' := \dim_k \mathrm{H}^0(D \cap D', \mathcal{O}) = \sum_{d \in D \cap D'} \dim_k \mathcal{O}_{D \cap D',d}.$$

This is generally larger than the length of the artin ring $\mathrm{H}^0(D \cap D', \mathcal{O})$, and is called the *$k$-length* of $D \cap D'$. If $F = \mathrm{H}^0(D, \mathcal{O}_D)$, then $D \cap D'$ is also an F-scheme, and so it makes sense to define

$$D._F D' = \dim_F \mathrm{H}^0(D \cap D', \mathcal{O}) = D.D'/[F : k].$$

We call this the *F-length* of $D \cap D'$. We can likewise define $D._{F'} D'$ for the field $F' = \mathrm{H}^0(D', \mathcal{O})$. If $D' = D$, we define the relative self-intersection $D._F D$ to be $(D.D)/[F : k]$ where $D.D$ is the usual self-intersection number on the k-fiber.

Theorem 2.1.2. *Let X be a connected regular proper curve over a discrete valuation ring, and let $P \in X$ be a closed point in the closed fiber. Let C_1, C_2 be two (possibly equal) effective divisors supported in the closed fiber of X, with each C_j passing through P, and let C_j' be the strict transform of C_j under the blow-up $\pi : X' = \mathrm{Bl}_P(X) \to X$. We write $E \simeq \mathbf{P}^1_{k(P)}$ to denote the exceptional divsor.*

We have $\pi^{-1}(C_j) = C_j' + m_j E$ where $m_j = \mathrm{mult}_P(C_j)$ is the multiplicity of the curve C_j at P. Also, $m_j = (C_j')._{k(P)}E$ and

$$C_1.C_2 = C_1'.C_2' + m_1 m_2 [k(P) : k].$$

Proof. Recall that for a regular local ring R of dimension 2 and any non-zero non-unit $g \in R$, the 1-dimensional local ring R/g has multiplicity (*i.e.*, leading coefficient of its Hilbert-Samuel polynomial) equal to the unique integer $\mu \geq 1$ such that $g \in \mathfrak{m}_R^{\mu}$, $g \notin \mathfrak{m}_R^{\mu+1}$.

We have $\pi^{-1}(C_j) = C_j' + m_j E$ for some positive integer m_j that we must prove is equal to the multiplicity $\mu_j = \mathrm{mult}_P(C_j)$ of C_j at P. We have $E._{k(P)}E = -1$, so $E.E = -[k(P) : k]$, and we also have $\pi^{-1}(C_j).E = 0$, so $m_j = (C_j'.E)/[k(P) : k] = (C_j')._{k(P)}E$. The strict transform C_j' is the blow-up of C_j at P, equipped with its natural (closed immersion) map into X'. The number m_j is the $k(P)$-length of the scheme-theoretic intersection $C_j' \cap E$; this is the fiber of $\mathrm{Bl}_P(C_j) \to C_j$ over P. Intuitively, this latter fiber is the scheme of tangent directions to C_j at P, but more precisely it is $\mathrm{Proj}(S_j)$, where

$$S_j = \bigoplus_{n \geq 0} \mathfrak{m}_j^n / \mathfrak{m}_j^{n+1},$$

and $\mathfrak{m}_j$ is the maximal ideal of $\mathcal{O}_{C_j,P} = \mathcal{O}_{X,P}/(f_j)$, with f_j a local equation for C_j at P. We have $\mathfrak{m}_j = \mathfrak{m}/(f_j)$ with $\mathfrak{m}$ the maximal ideal of $\mathcal{O}_{X,P}$. Since $f_j \in \mathfrak{m}^{\mu_j}$ and $f_j \notin \mathfrak{m}^{\mu_j+1}$,

$$S_j \simeq \mathrm{Sym}_{k(P)}(\mathfrak{m}/\mathfrak{m}^2)/\overline{f}_j = k(P)[u,v]/(\overline{f}_j)$$

with $\overline{f}_j$ denoting the nonzero image of f_j in degree μ_j. We conclude that $\mathrm{Proj}(S_j)$ has $k(P)$-length μ_j, so $m_j = \mu_j$. Thus, we may compute

$$\begin{aligned} C_1.C_2 = \pi^{-1}(C_1).\pi^{-1}(C_2) &= C_1'.C_2' + 2m_1 m_2 [k(P) : k] + m_1 m_2 E.E \\ &= C_1'.C_2' + m_1 m_2 [k(P) : k]. \end{aligned}$$

□

2.2 Minimal resolutions

It is no doubt well-known to experts that the classical technique of resolution for cyclic quotient singularities on complex surfaces [25, §2.6] can be adapted to the case of tame cyclic quotient singularities on curves over a complete equicharacteristic discrete valuation ring. We want the case of an arbitrary discrete valuation ring, and this seems to be less widely known (it is not addressed in the literature, and was not known to an expert in log-geometry with whom we consulted). Since there seems to be no adequate reference for this more general result, we will give the proof after some preliminary work (*e.g.*, we have to define what we mean by a *tame cyclic quotient singularity*, and we must show that this definition is applicable in many situations. Our first step is to establish the existence and uniqueness of a minimal regular resolution in the case of relative curves over a Dedekind base (the case of interest to us); this will eventually serve to make sense of the *canonical resolution* at a point.

Since we avoid properness assumptions, to avoid any confusion we should explicitly recall a definition.

Definition 2.2.1. Let $X \to S$ be a regular S-curve, with S a connected Dedekind scheme. We say that an integral divisor $D \hookrightarrow X$ in a closed fiber X_s is a -1-*curve* if D is proper over $k(s)$, $\mathrm{H}^1(D, \mathcal{O}_D) = 0$, and $\deg_k \mathcal{O}_D(D) = -1$, where $k = \mathrm{H}^0(D, \mathcal{O}_D)$ is a finite extension of $k(s)$.

By Castelnuovo's theorem, a -1-curve $D \hookrightarrow X$ as in Definition 2.2.1 is k-isomorphic to a projective line over k, where $k = \mathrm{H}^0(D, \mathcal{O}_D)$.

The existence and uniqueness of minimal regular resolutions is given by:

Theorem 2.2.2. *Let $X \to S$ be a normal S-curve over a connected Dedekind scheme S. Assume either that S is excellent or that $X_{/S}$ has smooth generic fiber.*

There exists a birational proper morphism $\pi : X^{\mathrm{reg}} \to X$ such that X^{reg} is a regular S-curve and there are no -1-curves in the fibers of π. Such an X-scheme is unique up to unique isomorphism, and every birational proper morphism $X' \to X$ with a regular S-curve X' admits a unique factorization through π. Formation of X^{reg} is compatible with base change to $\operatorname{Spec} \mathcal{O}_{S,s}$ *and* $\operatorname{Spec} \widehat{\mathcal{O}}_{S,s}$ *for closed points $s \in S$. For local S, there is also compatibility with ind-étale base change $S' \to S$ with local S' whose closed point is residually trivial over that of S.*

We remind that reader that, for technical reasons in the proof of Theorem 2.4.1, we avoid requiring curves to be proper and we do not assume the generic fiber to be geometrically connected. The reader is referred to [41, 9/3.32] for an alternative discussion in the proper case.

Proof. We first assume S to be excellent, and then we shall use Lemma 2.1.1 and some descent considerations to reduce the general case to the excellent case by passage to completions.

As a preliminary step, we wish to reduce to the proper case (to make the proof of uniqueness easier). By Nagata's compactification theorem [43] and the finiteness of normalization for excellent schemes, we can find a schematically dense open immersion $X \hookrightarrow \overline{X}$ with $\overline{X}_{/S}$ normal, proper, and flat over S (hence a normal S-curve). By resolving singularities along $\overline{X} - X$, we may assume the non-regular locus on $\overline{X}$ coincides with that on X. Thus, the existence and uniqueness result for X will follow from that for $\overline{X}$. The assertion on regular resolutions (uniquely) factorizing through π goes the same way. Hence, we now assume (for excellent S) that $X_{/S}$ is proper. We can also assume X to be connected.

By Lemma 2.1.1 and resolution for excellent surfaces, there exists a birational *proper* morphism $X' \to X$ with X' a regular proper S-curve. If there is a -1-curve in the fiber of X' over some (necessarily closed) point of X, then by Castelnuovo we can blow down the -1-curve and $X' \to X$ will factor through the blow-down. This blow-down process cannot continue forever, so we get the existence of $\pi : X^{\text{reg}} \to X$ with no -1-curves in its fibers.

Recall the Factorization Theorem for birational *proper* morphisms between regular connected S-curves: such maps factor as a composite of blow-ups at closed points in closed fibers. Using the Factorization Theorem, to prove uniqueness of π and the (unique) factorization through π for any regular resolution of X we just have to show that if $X'' \to X' \to X$ is a tower of birational proper morphisms with regular S-curves X' and X'' such that X' has no -1-curves in its fibers over X, then any -1-curve C in a fiber of $X'' \to X$ is necessarily contracted by $X'' \to X'$. Also, via Stein factorization we can assume that the proper normal connected S-curves X, X', and X'' with common generic fiber over S have geometrically connected fibers over S. We may assume that S is local. Since the map $q : X'' \to X'$ is a composite of blow-ups, we may assume that C meets the exceptional fiber E of the first blow-down $q_1 : X'' \to X''_1$ of a factorization of q. If $C = E$ we are done, so we may assume $C \neq E$. In this case we will show that X is regular, so again uniqueness holds (by the Factorization Theorem mentioned above).

The image $q_1(C)$ is an irreducible divisor on X''_1 with strict transform C, so by Theorem 2.1.2 we conclude that $q_1(C)$ has non-negative self-intersection number, so this self-intersection must be zero. Since $X''_1 \to S$ is its own Stein factorization, and hence has geometrically connected closed fiber, $q_1(C)$ must be the entire closed fiber of X''_1. Thus, X''_1 has irreducible closed fiber, and so the (surjective) proper birational map $X''_1 \to X$ is quasi-finite and hence finite. Since X and X''_1 are normal and connected (hence integral), it follows that $X''_1 \to X$ must be an isomorphism. Thus, X is regular, as desired.

With X^{reg} unique up to (obviously) unique isomorphism, for the base change compatibility we note that the various base changes $S' \to S$ being considered (to completions on S, or to local S' ind-étale surjective over local S and residually trivial at closed points), the base change $X^{\text{reg}}_{/S'}$ is regular and proper birational over the normal curve $X_{/S'}$ (see Lemma 2.1.1). Thus, we just have to check that the fibers of $X^{\text{reg}}_{/S'} \to X_{/S'}$ do not contain -1-curves. The closed-fiber situation is identical to that before base change, due to the residually trivial condition at closed points, so we are done.

Now suppose we do not assume S to be excellent, but instead assume $X_{/S}$ has smooth generic fiber. In this case all but finitely many fibers of $X_{/S}$ are smooth. Thus, we may reduce to the local case $S = \mathrm{Spec}(R)$ with a discrete valuation ring R. Consider $X_{/\widehat{R}}$, a normal $\widehat{R}$-curve by Lemma 2.1.1. Since $\widehat{R}$ is excellent, there is a minimal regular resolution

$$\pi : (X_{/\widehat{R}})^{\mathrm{reg}} \to X_{/\widehat{R}}.$$

By [40, Remark C, p. 155], the map π is a blow-up along a 0-dimensional closed subscheme $\widehat{Z}$ physically supported in the non-regular locus of $X_{/\widehat{R}}$. This $\widehat{Z}$ is therefore physically supported in the closed fiber of $X_{/\widehat{R}}$, yet $\widehat{Z}$ is artinian and hence lies in some infinitesimal closed fiber of $X_{/\widehat{R}}$. Since $X \times_R \widehat{R} \to X$ induces isomorphisms on the level of nth infinitesimal closed-fibers for all n, there is a unique 0-dimensional closed subscheme Z in X with $Z_{/\widehat{R}} = \widehat{Z}$ inside of $X_{/\widehat{R}}$.

Since the blow-up $\mathrm{Bl}_Z(X)$ satisfies

$$\mathrm{Bl}_Z(X)_{/\widehat{R}} \simeq \mathrm{Bl}_{\widehat{Z}}(X_{/\widehat{R}}) = (X_{/\widehat{R}})^{\mathrm{reg}},$$

by Lemma 2.1.1 we see that $\mathrm{Bl}_Z(X)$ is a regular S-curve. There are no -1-curves in its fibers over X since $\mathrm{Spec}\,\widehat{R} \to \mathrm{Spec}\,R$ is an isomorphism over $\mathrm{Spec}\,R/\mathfrak{m}$. This establishes the existence of $\pi : X^{\mathrm{reg}} \to X$, as well as its compatibility with base change to completions on S. To establish uniqueness of π, or more generally its universal factorization property, we must prove that certain birational maps from regular S-curves to X^{reg} are morphisms. This is handled by a standard graph argument that can be checked after faithfully flat base change to $\widehat{R}$ (such base change preserves regularity, by Lemma 2.1.1). Thus, the uniqueness results over the excellent base $\widehat{R}$ carry over to our original R. The same technique of base change to $\widehat{R}$ shows compatibility with ind-étale base change that is residually trivial over closed points.

□

One mild enhancement of the preceding theorem rests on a pointwise definition:

Definition 2.2.3. Let $X_{/S}$ be as in Theorem 2.2.2, and let $\Sigma \subseteq X$ be a finite set of closed points in closed fibers over S. Let U be an open in X containing Σ such that U does not contain the finitely many non-regular points of X outside of Σ. We define the *minimal regular resolution along* Σ to be the morphism $\pi_\Sigma : X_\Sigma \to X$ obtained by gluing $X - \Sigma$ with the part of X^{reg} lying over U (note: the choice of U does not matter, and X_Σ is not regular if there are non-regular points of X outside of Σ).

It is clear that the minimal regular resolution along Σ is compatible with local residually-trivial ind-étale base change on a local S, as well as with base change to a (non-generic) complete local ring on S. It is also uniquely characterized among normal S-curves C equipped with a proper birational morphism $\varphi : C \to X$ via the following conditions:

- π_Σ is an isomorphism over $X - \Sigma$,
- X_Σ is regular at points over Σ,
- X_Σ has no -1-curves in its fibers over Σ.

This yields the crucial consequence that (under some mild restrictions on residue field extensions) formation of X_Σ is étale-local on X. This fact is ultimately the reason we did not require properness or geometrically connected fibers in our definition of S-*curve*:

Corollary 2.2.4. *Let $X_{/S}$ be a normal S-curve over a connected Dedekind scheme S, and let $\Sigma \subseteq X$ be a finite set of closed points in closed fibers over S. Let $X' \to X$ be étale (so X' is an S-curve), and let Σ' denote the preimage of Σ. Assume that S is excellent or $X_{/S}$ has smooth generic fiber.*

If $X_\Sigma \to X$ denotes the minimal regular resolution along Σ, and $X' \to X$ is residually trivial over Σ, then the base change $X_\Sigma \times_X X' \to X'$ is the minimal regular resolution along Σ'.

Remark 2.2.5. The residual triviality condition over Σ is satisfied when S is local with separably closed residue field, as then all points of Σ have separably closed residue field (and so the étale $X' \to X$ must induce trivial residue field extensions over such points).

Proof. Since $X_\Sigma \times_X X'$ is étale over X_Σ, we conclude that $X_\Sigma \times_X X'$ is an S-curve that is regular along the locus over $\Sigma' \subseteq X'$, and its projection to X' is proper, birational, and an isomorphism over $X' - \Sigma'$. It remains to check that

$$X_\Sigma \times_X X' \to X' \tag{2.2.1}$$

has no -1-curves in the proper fibers over Σ'. Since $X' \to X$ is residually trivial over Σ (by hypothesis), so this is clear.

□

2.3 Nil-semistable curves

In order to compute minimal regular resolutions of the sort that arise on $X_H(p)$'s, it is convenient to study the following concept before we discuss resolution of singularities. Let S be a connected Dedekind scheme and let X be an S-curve.

Definition 2.3.1. For a closed point $s \in S$, a closed point $x \in X_s$ is *nil-semistable* if the reduced fiber-curve $X_s^{\rm red}$ is semistable over $k(s)$ at x and all of the analytic branch multiplicities through x are not divisible by $\mathrm{char}(k(s))$. If $X_s^{\rm red}$ is semistable for all closed points $s \in S$ and all irreducible components of X_s have multiplicity not divisible by $\mathrm{char}(k(s))$, X is a *nil-semistable curve* over S.

Considerations with excellence of the fiber X_s show that the number of analytic branches in Definition 2.3.1 may be computed on the formal completion at a point over x in $X_{s/k'}$ for any separably closed extension k' of $k(s)$. We will use the phrase "analytic branch" to refer to such (formal) branches through a point over x in such a geometric fiber over s.

As is well-known from [34], many fine moduli schemes for elliptic curves are nil-semistable.

Fix a closed point $s \in S$. From the theory of semistable curves over fields [24, III, §2], it follows that when $x \in X_s^{\mathrm{red}}$ is a semistable non-smooth point then the finite extension $k(x)/k(s)$ is separable. We have the following analogue of the classification of semistable curve singularities:

Lemma 2.3.2. *Let $x \in X_s$ be a closed point and let $\pi_s \in \mathcal{O}_{S,s}$ be a uniformizer.*

If x is a nil-semistable point at which X is regular, then the underlying reduced scheme of the geometric closed fiber over s has either one or two analytic branches at a geometric point over x, with these branches smooth at x. When moreover $k(x)/k(s)$ is separable and there is exactly one analytic branch at $x \in X_s$, with multiplicity m_1 in $\mathcal{O}^{\mathrm{sh}}_{X_s,x}$, then

$$\widehat{\mathcal{O}^{\mathrm{sh}}_{X,x}} \simeq \widehat{\mathcal{O}^{\mathrm{sh}}_{S,s}}[\![t_1,t_2]\!]/(t_1^{m_1}-\pi_s). \tag{2.3.1}$$

If there are two analytic branches (so $k(x)/k(s)$ is automatically separable), say with multiplicities m_1 and m_2 in $\mathcal{O}^{\mathrm{sh}}_{X_s,x}$, then

$$\widehat{\mathcal{O}^{\mathrm{sh}}_{X,x}} \simeq \widehat{\mathcal{O}^{\mathrm{sh}}_{S,s}}[\![t_1,t_2]\!]/(t_1^{m_1}t_2^{m_2}-\pi_s). \tag{2.3.2}$$

Conversely, if $\widehat{\mathcal{O}^{\mathrm{sh}}_{X,x}}$ admits one of these two explicit descriptions with the exponents not divisible by $\mathrm{char}(k(s))$, then x is a nil-semistable regular point on X with $k(x)/k(s)$ separable.

In view of this lemma, we call the exponents in the formal isomorphisms (2.3.1) and (2.3.2) the *analytic geometric multiplicities* of X_s at x (this requires $k(x)/k(s)$ to be separable). We emphasize that these exponents can be computed after base change to any separably closed extension of $k(s)$ when x is nil-semistable with $k(x)/k(s)$ separable.

Proof. First assume $x \in X_s^{\mathrm{red}}$ is a non-smooth semistable point and X is regular at x. Since $k(x)$ is therefore finite separable over $k(s)$, we can make a base change to the completion of a strict henselization of $\mathcal{O}_{S,s}$ to reduce to the case $S = \mathrm{Spec}(W)$ with a complete discrete valuation ring W having separably closed residue field k such that x a k-rational point. Since $\widehat{\mathcal{O}}_{X,x}$ is a 2-dimensional complete regular local W-algebra with residue field k, it is a quotient of $W[\![t_1,t_2]\!]$ and hence has the form $W[\![t_1,t_2]\!]/(f)$ where f is a regular parameter. The semistability condition and non-smoothness of $X^{\mathrm{red}}_{/k}$ at x imply

$$k[\![t_1,t_2]\!]/\mathrm{rad}(\overline{f}) = (k[\![t_1,t_2]\!]/(\overline{f}))_{\mathrm{red}} \simeq \widehat{\mathcal{O}}_{X^{\mathrm{red}}_{/k},x} \simeq k[\![u_1,u_2]\!]/(u_1u_2)$$

where $\overline{f} = f \bmod \mathfrak{m}_W$, so $\overline{f}$ has exactly two distinct irreducible factors and these have distinct (non-zero) tangent directions in $X^{\rm red}_{/k}$ through x. We can choose t_1 and t_2 to lift these tangent directions, so upon replacing f with a unit multiple we may assume $\overline{f} = t_1^{m_1} t_2^{m_2} \bmod \mathfrak{m}_W$ for some $m_1, m_2 \geq 1$ not divisible by $p = \mathrm{char}(k) \geq 0$. Let π be a uniformizer of W, so $f = t_1^{m_1} t_2^{m_2} - \pi g$ for some g, and g must be a unit since f is a regular parameter. Since some m_j is not divisible by p, and hence the unit g admits an m_jth root, by unit-rescaling of the corresponding t_j we get to the case $g = 1$.

In the case when $X^{\rm red}_s$ is smooth at x and $k(x)/k(s)$ is separable, we may again reduce to the case in which $S = \mathrm{Spec}\, W$ with complete discrete valuation ring W having separably closed residue field k and $k(x) = k$. In this case, there is just one analytic branch and we see by a variant of the preceding argument that the completion of $\mathcal{O}^{\rm sh}_{X,x}$ has the desired form.

The converse part of the lemma is clear.

□

In Definition 2.3.6, we shall give a local definition of the class of curve-singularities that we wish to resolve, but we will first work through some global considerations that motivate the relevance of the local Definition 2.3.6.

Assume X is *regular*, and let H be a finite group and assume we are given an action of H on $X_{/S}$ that is free on the scheme of generic points (*i.e.*, no non-identity element of H acts trivially on a connected component of X). A good example to keep in mind is the (affine) fine moduli scheme over $S = \mathrm{Spec}(\mathbf{Z}_{(p)})$ of $\Gamma_1(p)$-structures on elliptic curves equipped with auxiliary full level ℓ-structure for an odd prime $\ell \neq p$, and $H = \mathrm{GL}_2(\mathbf{F}_\ell)$ acting in the usual manner (see Section 3 for a review of these basic level structures).

We wish to work with a quotient S-curve $X' = X/H$, so we now also assume that X is quasi-projective Zariski-locally on S. Clearly $X \to X'$ is a finite H-equivariant map with the expected universal property; in the above modular-curve example, this quotient X' is the coarse moduli scheme $Y_1(p)$ over $\mathbf{Z}_{(p)}$. We also now assume that S is excellent or $X_{/K}$ is smooth, so that there are only finitely many non-regular points (all in codimension 2) and various results centering on resolution of singularities may be applied.

The S-curve X' has regular generic fiber (and even smooth generic fiber when $X_{/S}$ has smooth generic fiber), and X' is regular away from finitely many closed points in the closed fibers. Our aim is to understand the *minimal regular resolution* ${X'}^{\rm reg}$ of X', or rather to describe the geometry of the fibers of ${X'}^{\rm reg} \to X'$ over non-regular points x' satisfying a mild hypothesis on the structure of $X \to X'$ over x'.

We want to compute the minimal regular resolution for $X' = X/H$ at non-regular points x' that satisfy several conditions. Let $s \in S$ be the image of x', and let $p \geq 0$ denote the common characteristic of $k(x')$ and $k(s)$. Pick $x \in X$ over x'.

- We assume that X is nil-semistable at x (by the above hypotheses, X is also regular at x).

- We assume that the inertia group $H_{x|x'}$ in H at x (*i.e.*, the stablizer in H of a geometric point over x) has order not divisible by p (so this group acts semi-simply on the tangent space at a geometric point over x).

- When there are two analytic branches through x, we assume $H_{x|x'}$ does not interchange them.

These conditions are independent of the choice of x over x' and can be checked at a geometric point over x, and when they hold then the number of analytic branches through x coincides with the number of analytic branches through x' (again, we are really speaking about analytic branches on a geometric fiber over s).

Since p does not divide $|H_{x|x'}|$, it follows that $k(x')$ is the subring of invariants under the action of $H_{x|x'}$ on $k(x)$, so a classical theorem of Artin ensures that $k(x)/k(x')$ is separable (and even Galois). Thus, $k(x)/k(s)$ is separable if and only if $k(x')/k(s)$ is separable, and such separability holds when the point $x \in X_s^{\rm red}$ is semistable but not smooth. Happily for us, this separability condition over $k(s)$ is always satisfied (we are grateful to Lorenzini for pointing this out):

Lemma 2.3.3. *With notation and hypotheses as above, particularly with $x' \in X' = X/H$ a non-regular point, the extension $k(x')/k(s)$ is separable.*

Proof. Recall that, by hypothesis, $x \in X_s^{\rm red}$ is either a smooth point or an ordinary double point. If x is a non-smooth point on the curve $X_s^{\rm red}$, then the desired separability follows from the theory of ordinary double point singularities. Thus, we may (and do) assume that x is a smooth point on $X_s^{\rm red}$.

We may also assume S is local and strictly henselian, so $k(s)$ is separably closed and hence $k(x)$ and $k(x')$ are separably closed. Thus, $k(x) = k(x')$ and $H_{x|x'}$ is the physical stabilizer of the point $x \in X$. We need to show that the common residue field $k(x) = k(x')$ is separable over $k(s)$. If we let $X'' = X/H_{x|x'}$, then the image x'' of x in X'' has complete local ring isomorphic to that of $x' \in X'$, so we may replace X' with X'' to reduce to the case when H has order not divisible by p and x is in the fixed-point locus of H. By [20, Prop. 3.4], the fixed-point locus of H in X admits a closed-subscheme structure in X that is smooth over S. On the closed fiber this smooth scheme is finite and hence étale over $k(s)$, so its residue fields are separable over $k(s)$.

□

The following refinement of Lemma 2.3.2 is adapted to the $H_{x|x'}$-action, and simultaneously handles the cases of one and two (geometric) analytic branches through x'.

Lemma 2.3.4. *With hypotheses as above, there is an $\widehat{\mathcal{O}_{S,s}^{\rm sh}}$-isomorphism*

$$\widehat{\mathcal{O}_{X,x}^{\rm sh}} \simeq \widehat{\mathcal{O}_{S,s}^{\rm sh}}[\![t_1, t_2]\!]/(t_1^{m_1} t_2^{m_2} - \pi_s)$$

(with $m_1 > 0$, $m_2 \geq 0$) such that the $H_{x|x'}$-action looks like $h(t_j) = \chi_j(h)t_j$ for characters $\chi_1, \chi_2 : H_{x|x'} \to \widehat{\mathcal{O}^{\mathrm{sh}}_{S,s}}^{\times}$ that are the Teichmüller lifts of characters giving a decomposition of the semisimple $H_{x|x'}$-action on the 2-dimensional cotangent space at a geometric point over x. Moreover, $\chi_1^{m_1}\chi_2^{m_2} = 1$.

The characters χ_j also describe the action of $H_{x|x'}$ on the tangent space at (a geometric point over) x. There are two closed-fiber analytic branches through x when m_1 and m_2 are positive, and then the branch with formal parameter t_2 has multiplicity m_1 since

$$(k[\![t_1, t_2]\!]/(t_1^{m_1}t_2^{m_2}))[1/t_2] = k(\!(t_2)\!)[t_1]/(t_1^{m_1})$$

has length m_1. Likewise, when $m_2 > 0$ it is the branch with formal parameter t_1 that has multiplicity m_2.

Proof. We may assume $S = \operatorname{Spec} W$ with W a complete discrete valuation ring having separably closed residue field k and uniformizer π, so x is k-rational. Let $R = \widehat{\mathcal{O}^{\mathrm{sh}}_{X,x}} = \widehat{\mathcal{O}}_{X,x}$. We have seen in Lemma 2.3.2 that there is an isomorphism of the desired type as W-algebras, but we need to find better such t_j's to linearize the $H_{x|x'}$-action.

We first handle the easier case $m_2 = 0$. In this case there is only one minimal prime (t_1) over (π), so $h(t_1) = u_h t_1$ for a unique unit $u_h \in R^{\times}$. Since $t_1^{m_1} = \pi$ is $H_{x|x'}$ invariant, we see that $u_h \in \mu_{m_1}(R)$ is a Teichmüller lift from k (since $p \nmid m_1$). Thus, $h(t_1) = \chi_1(h)t_1$ for a character $\chi_1 : H_{x|x'} \to R^{\times}$ that is a lift of a character for $H_{x|x'}$ on $\mathrm{Cot}_x(X)$. Since $H_{x|x'}$ acts semisimply on the 2-dimensional cotangent space $\mathrm{Cot}_x(X)$ and there is a stable line spanned by $t_1 \bmod \mathfrak{m}_x^2$, we can choose t_2 to lift an $H_{x|x'}$-stable line complementary to the one spanned by $t_1 \bmod \mathfrak{m}_x^2$. If χ_2 denotes the Teichmüller lift of the character for $H_{x|x'}$ on this complementary line, then

$$h(t_2) = \chi_2(h)(t_2 + \delta_h)$$

with $\delta_h \in \mathfrak{m}_x^i$ for some $i \geq 2$. It is straightfoward to compute that

$$h \mapsto \delta_h \bmod \mathfrak{m}_x^{i+1}$$

is a 1-cocycle with values in the twisted $H_{x|x'}$-module $\chi_2^{-1} \otimes (\mathfrak{m}_x^i/\mathfrak{m}_x^{i+1})$. Changing this 1-cocycle by a 1-coboundary corresponds to adding an element of $\mathfrak{m}_x^i/\mathfrak{m}_x^{i+1}$ to $t_2 \bmod \mathfrak{m}_x^{i+1}$. Since

$$\mathrm{H}^1(H_{x|x'}, \chi_2^{-1} \otimes (\mathfrak{m}_x^i/\mathfrak{m}_x^{i+1})) = 0,$$

we can successively increase $i \geq 2$ and pass to the limit to find a choice of t_2 such that $H_{x|x'}$ acts on t_2 through the character χ_2. That is, $h(t_1) = \chi_1(h)t_1$ and $h(t_2) = \chi_2(h)t_2$ for all $h \in H_{x|x'}$. This settles the case $m_2 = 0$.

Now we turn to the more interesting case when also $m_2 > 0$, so there are two analytic branches through x. By hypothesis, the $H_{x|x'}$-action preserves the two minimal primes (t_1) and (t_2) over (π) in R. We must have $h(t_1) = u_h t_1$, $h(t_2) = v_h t_2$ for unique units $u_h, v_h \in R^\times$. Since $t_1^{m_1} t_2^{m_2} = \pi$, by applying h we get $u_h^{m_1} v_h^{m_2} = 1$.

Consider what happens if we replace t_2 with a unit multiple $t_2' = vt_2$, and then replace t_1 with the unit multiple $t_1' = v^{-m_2/m_1} t_1$ so as to ensure $t_1'^{m_1} t_2'^{m_2} = \pi$. Note that an m_1th root v^{-m_2/m_1} of the unit v^{-m_2} makes sense since k is separably closed and $p \nmid m_1$. The resulting map $W[\![t_1', t_2']\!]/(t_1'^{m_1} t_2'^{m_2} - \pi) \to R$ is visibly surjective, and hence is an isomorphism for dimension reasons. Switching to these new coordinates on R has the effect of changing the 1-cocycle $\{v_h\}$ by a 1-coboundary, and *every* 1-cocycle cohomologous to $\{v_h\}$ is reached by making such a unit multiple change on t_2.

By separately treating residue characteristic 0 and positive residue characteristic, an inverse limit argument shows that $\mathrm{H}^1(H_{x|x'}, U)$ vanishes, where $U = \ker(R^\times \twoheadrightarrow k^\times)$. Thus, the natural map $\mathrm{H}^1(H_{x|x'}, R^\times) \to \mathrm{H}^1(H_{x|x'}, k^\times)$ is injective. The $H_{x|x'}$-action on $k^\times$ is trivial since $H_{x|x'}$ acts trivially on W, so

$$\mathrm{H}^1(H_{x|x'}, k^\times) = \mathrm{Hom}(H_{x|x'}, k^\times) = \mathrm{Hom}(H_{x|x'}, k^\times_{\mathrm{tors}}),$$

with all elements in the torsion subgroup $k^\times_{\mathrm{tors}}$ of order not divisible by p and hence uniquely multiplicatively lifting into R. Thus,

$$\mathrm{H}^1(H_{x|x'}, R^\times) \to \mathrm{H}^1(H_{x|x'}, k^\times)$$

is bijective, and so replacing t_1 and t_2 with suitable unit multiples allows us to assume $h(t_2) = \chi_2(h) t_2$, with $\chi_2 : H_{x|x'} \to W^\times_{\mathrm{tors}}$ some homomorphism of order not divisible by p (since $H_{x|x'}$ acts trivially on $k^\times$ and $p \nmid |H_{x|x'}|$).

Since

$$1 = u_h^{m_1} v_h^{m_2} = u_h^{m_1} \chi_2(h)^{m_2}$$

and $p \nmid m_1$, we see that u_h is a root of unity of order not divisible by p. Viewing $k^\times_{\mathrm{tors}} \subseteq R^\times$ via the Teichmüller lifting, we conclude that $u_h \in k^\times_{\mathrm{tors}} \subseteq R^\times$. Thus, we can write $h(t_1) = \chi_1(h) t_1$ for a homomorphism $\chi_1 : H_{x|x'} \to W^\times_{\mathrm{tors}}$ also necessarily of order not divisible by p. The preceding calculation also shows that $\chi_1^{m_1} \chi_2^{m_2} = 1$ since $u_h^{m_1} v_h^{m_2} = 1$.

□

Although Lemma 2.3.4 provides good (geometric) coordinate systems for describing the inertia action, one additional way to simplify matters is to reduce to the case in which the tangent-space characters χ_1 and χ_2 are powers of each other. We wish to explain how this special situation is essentially the general case (in the presence of our running assumption that H acts freely on the scheme of generic points of X).

First, observe that $H_{x|x'}$ acts faithfully on the tangent space $T_x(X)$ at x. Indeed, if an element in $H_{x|x'}$ acts trivially on the tangent space $T_x(X)$, then by Lemma 2.3.4 it acts trivially on the completion of $\mathcal{O}^{\mathrm{sh}}_{X,x}$ and hence acts trivially on the corresponding connected component of the normal X. By hypothesis, H acts freely on the scheme of generic points of X, so we conclude that the product homomorphism

$$\chi_1 \times \chi_2 : H_{x|x'} \hookrightarrow k(x)_{\mathrm{sep}}^{\times} \times k(x)_{\mathrm{sep}}^{\times}, \tag{2.3.3}$$

is *injective* (where $k(x)_{\mathrm{sep}}$ is the separable closure of $k(x)$ used when constructing $\mathscr{O}_{X,x}^{\mathrm{sh}}$). In particular, $H_{x|x'}$ is a product of two cyclic groups (one of which might be trivial).

Lemma 2.3.5. *Let $\kappa_j = |\ker(\chi_j)|$. The characters $\chi_1^{\kappa_2}$ and $\chi_2^{\kappa_1}$ factor through a common quotient of $H_{x|x'}$ as faithful characters. When $H_{x|x'}$ is cyclic, this quotient is $H_{x|x'}$.*

In addition, $\kappa_2 | m_1$ and $\kappa_1 | m_2$.

The cyclicity condition on $H_{x|x'}$ will hold in our application to modular curves, as then even H is cyclic.

Proof. The injectivity of (2.3.3) implies that χ_1 is faithful on $\ker(\chi_2)$ and χ_2 is faithful on $\ker(\chi_1)$. Since $\chi_1^{m_1}\chi_2^{m_2} = 1$, we get $\kappa_2 | m_1$ and $\kappa_1 | m_2$ (even if $m_2 = 0$).

For the proof that the indicated powers of the χ_j's factor as faithful characters of a common quotient of $H_{x|x'}$, it is enough to focus attention on ℓ-primary parts for a prime ℓ dividing $|H_{x|x'}|$ (so $\ell \neq p$). More specifically, if G is an finite ℓ-group that is either cyclic or a product of two cyclic groups, and $\psi_0, \psi_1 : G \to \mathbf{Z}/\ell^n\mathbf{Z}$ are homomorphisms such that $\psi_0 \times \psi_1$ is injective (*i.e.*, $\ker(\psi_0) \cap \ker(\psi_1) = \{1\}$), then we claim that the $\psi_j^{\kappa_{1-j}}$'s factor as faithful characters on a common quotient of G, where $\kappa_j = |\ker(\psi_j)|$. If one of the ψ_j's is faithful (or equivalently, if the ℓ-group G is cyclic), this is clear. This settles the case in which G is cyclic, so we may assume G is a product of two non-trivial cyclic ℓ-groups and that both ψ_j's have non-trivial kernel. Since the ℓ-torsion subgroups $\ker(\psi_j)[\ell]$ must be non-trivial with trivial intersection, these must be distinct lines spanning $G[\ell]$. Passing to group $G/G[\ell]$ and the characters ψ_j^{ℓ} therefore permits us to induct on $|G|$.

□

By the lemma, we conclude that the characters $\chi_1' = \chi_1^{\kappa_2}$ and $\chi_2' = \chi_1^{\kappa_1}$ both factor faithfully through a common (cyclic) quotient $H_{x|x'}'$ of $H_{x|x'}$. Define $t_1' = t_1^{\kappa_2}$ and $t_2' = t_2^{\kappa_1}$. Since formation of $H_{x|x'}$-invariants commutes with passage to quotients on $\widehat{\mathscr{O}_{S,s}^{\mathrm{sh}}}$-modules, Lemma 2.3.4 shows that in order to compute the $H_{x|x'}$-invariants of $\widehat{\mathscr{O}}_{X',x'}^{\mathrm{sh}}$ it suffices to compute invariants on the level of $\widehat{\mathscr{O}_{S,s}^{\mathrm{sh}}}[\![t_1, t_2]\!]$ and then pass to a quotient. The subalgebra of invariants in $\widehat{\mathscr{O}_{S,s}^{\mathrm{sh}}}[\![t_1, t_2]\!]$ under the subgroup generated by $\ker(\chi_1)$ and $\ker(\chi_2)$ is $\widehat{\mathscr{O}_{S,s}^{\mathrm{sh}}}[\![t_1', t_2']\!]$, and $H_{x|x'}$ acts on this subalgebra through the quotient $H_{x|x'}'$ via the characters χ_1' and χ_2'. Letting $m_1' = m_1/\kappa_2$ and $m_2' = m_2/\kappa_1$ (so $m_2' = 0$ in the case of one analytic branch), we obtain the description

$$\widehat{\mathcal{O}^{\rm sh}_{X',x'}} = (\widehat{\mathcal{O}^{\rm sh}_{S,s}}[\![t'_1, t'_2]\!]/({t'_1}^{m'_1}{t'_2}^{m'_2} - \pi_s))^{H'_{x|x'}} \tag{2.3.4}$$

Obviously $\chi'_2 = {\chi'_1}^{r_{x|x'}}$ for a unique $r_{x|x'} \in (\mathbf{Z}/|H'_{x|x'}|\mathbf{Z})^\times$, as the characters χ'_j are both faithful on $H'_{x|x'}$.

Since $|H'_{x|x'}|$ and $r_{x|x'} \in (\mathbf{Z}/|H'_{x|x'}|\mathbf{Z})^\times$ are intrinsic to $x' \in X' = X/H$ and do not depend on x (or on a choice of $k(x)_{\rm sep}$), we may denote these two integers $n_{x'}$ and $r_{x'}$ respectively. We have $m'_1 + m'_2 r'_{x'} \equiv 0 \bmod n_{x'}$ since $1 = {\chi'_1}^{m'_1}{\chi'_2}^{m'_2} = {\chi'_1}^{m'_1+m'_2 r_{x'}}$ with χ'_1 faithful. Theorem 2.3.9 below shows that $n_{x'} > 1$, since x' is the non-regular.

If S were a smooth curve over $\mathbf{C}$, then the setup in (2.3.4) would be the classical cyclic surface quotient-singularity situation whose minimal regular resolution is most readily computed via toric varieties. That case motivates what to expect for minimal regular resolutions with more general S in §2.4, but rather than delve into a relative theory of toric varieties we can just use the classical case as a guide.

To define the class of singularities we shall resolve, let $X'_{/S}$ now be a *normal* (not necessarily connected) curve over a connected Dedekind scheme S. Assume moreover that either S is excellent or that $X'_{/S}$ has smooth generic fiber, so there are only finitely many non-regular points (all closed in closed fibers). Consider a closed point $s \in S$ with residue characteristic $p \geq 0$, and pick a closed point $x' \in X'_s$ such that X'_s has one or two (geometric) analytic branches at x'.

Definition 2.3.6. We say that a closed point x' in a closed fiber X'_s is a *tame cyclic quotient singularity* if there exists a positive integer $n > 1$ not divisible by $p = \operatorname{char}(k(s))$, a unit $r \in (\mathbf{Z}/n\mathbf{Z})^\times$, and integers $m'_1 > 0$ and $m'_2 \geq 0$ satisfying $m'_1 \equiv -rm'_2 \bmod n$ such that $\widehat{\mathcal{O}^{\rm sh}_{X',x'}}$ is isomorphic to the subalgebra of $\mu_n(k(s)_{\rm sep})$-invariants in $\widehat{\mathcal{O}^{\rm sh}_{S,s}}[\![t'_1, t'_2]\!]/({t'_1}^{m'_1}{t'_2}^{m'_2} - \pi_s)$ under the action $t'_1 \mapsto \zeta t'_1$, $t'_2 \mapsto \zeta^r t'_2$.

Remark 2.3.7. Note that when $X'_{/S}$ has a tame cyclic quotient singularity at $x' \in X'_s$, then $k(x')/k(s)$ is separable and x' is non-regular (by Theorem 2.3.9 below). Also, it is easy to check that the exponents m'_1 and m'_2 are necessarily the analytic branch multiplicities at x'. Note that the data of n and r is merely part of a presentation of $\widehat{\mathcal{O}}_{X',x'}$ as a ring of invariants, so it is not clear *a priori* that n and r are intrinsic to $x' \in X'$. The fact that n and r are uniquely determined by x' follows from Theorem 2.4.1 below, where we show that n and r arise from the structure of the minimal regular resolution of X' at x'.

Using notation as in the preceding global considerations, there is a very simple criterion for a nil-semistable $x' \in X/H$ to be a non-regular point: there should not be a line in $T_x(X)$ on which the inertia group $H_{x|x'}$ acts trivially. To prove this, we recall Serre's pseudo-reflection theorem [57, Thm. 1′]. This requires a definition:

Definition 2.3.8. Let V be a finite-dimensional vector space over a field k. An element σ of $\mathrm{Aut}_k(V)$ is called a *pseudo-reflection* if $\mathrm{rank}(1-\sigma) \leq 1$.

Theorem 2.3.9 (Serre). *Let A be a noetherian regular local ring with maximal ideal $\mathfrak{m}$ and residue field k. Let G be a finite subgroup of $\mathrm{Aut}(A)$, and let A^G denote the local ring of G-invariants of A. Suppose that:*

1. *The characteristic of k does not divide the order of G,*
2. *G acts trivially on k, and*
3. *A is a finitely generated A^G-module.*

Then A^G is regular if and only if the image of G in $\mathrm{Aut}_k(\mathfrak{m}/\mathfrak{m}^2)$ is generated by pseudo-reflections.

In fact, the "only if" implication is true without hypotheses on the order of G, provided A^G has residue field k (which is automatic when k is algebraically closed).

Remark 2.3.10. By Theorem 3.7(i) of [44] with $B = A$ and $A = A^G$, hypothesis 3 of Serre's theorem forces A^G to be noetherian. Serre's theorem ensures that x' as in Definition 2.3.6 is necessarily non-regular.

Proof. Since this result is not included in Serre's Collected Works, we note that a proof of the "if and only if" assertion can be found in [68, Cor. 2.13, Prop. 2.15]. The proof of the "only if" implication in [68] works without any conditions on the order of G as long as one knows that A^G has the same residue field as A. Such equality is automatic when k is algebraically closed. Indeed, the case of characteristic 0 is clear, and for positive characteristic we note that k is a priori finite over the residue field of A^G, so if equality were to fail then the residue field of A^G would be of positive characteristic with algebraic closure a finite extension of degree > 1, an impossibility by Artin-Schreier.

To see why everything still works without restriction on the order of G when we assume A^G is regular, note first that regularity of A^G ensures that $A^G \to A$ must be finite free, so even without a Reynolds operator we still have $(A \otimes_{A^G} A)^G = A$, where G acts on the left tensor factor. Hence, the proof of [68, Lemma 2.5] still works. Meanwhile, equality of residue fields for A^G and A makes the proof of [68, Prop. 2.6] still work, and then one easily checks that the proofs of [68, Thm. 2.8, Prop. 2.15(i)⇒(ii)] go through unchanged. □

The point of the preceding study is that in a *global* quotient situation $X' = X/H$ as considered above, one always has a tame cyclic quotient singularity at the image x' of a nil-semistable point $x \in X_s$ when x' is not regular (by Lemma 2.3.3, both $k(x)$ and $k(x')$ are automatically separable over $k(s)$ when such non-regularity holds). Thus, when computing complete local rings at geometric closed points on a coarse modular curve (in residue characteristic

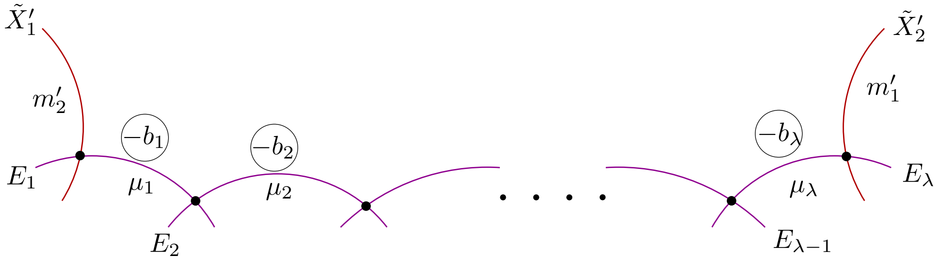


Figure 1: Minimal regular resolution of x'

> 3), we will naturally encounter a situation such as in Definition 2.3.6. The ability to explicitly (minimally) resolve tame cyclic quotient singularities in general will therefore have immediate applications to modular curves.

2.4 Jung–Hirzebruch resolution

As we noted in Remark 2.3.7, it is natural to ask whether the numerical data of n and $r \in (\mathbf{Z}/n\mathbf{Z})^\times$ in Definition 2.3.6 are intrinsic to $x' \in X'$. We shall see in the next theorem that this data is intrinsic, as it can be read off from the minimal regular resolution over x'.

Theorem 2.4.1. *Let $X'_{/S}$ be a normal curve over a local Dedekind base S with closed point s. Assume either that S is excellent or that $X'_{/S}$ has smooth generic fiber. Assume X' has a tame cyclic quotient singularity at a closed point $x' \in X'_s$ with parameters n and r (in the sense of Definition* 2.3.6*), where we represent $r \in (\mathbf{Z}/n\mathbf{Z})^\times$ by the unique integer r satisfying $1 \leq r < n$ and* $\gcd(r,n) = 1$. *Finally, assume either that $k(s)$ is separably closed or that all connected components of the regular compactification $\overline{X}'_K$ of the regular generic-fiber curve X'_K have positive arithmetic genus.*

Consider the Jung–Hirzebruch continued fraction expansion

$$\frac{n}{r} = b_1 - \cfrac{1}{b_2 - \cfrac{1}{\cdots - \cfrac{1}{b_\lambda}}} \tag{2.4.1}$$

with integers $b_j \geq 2$ for all j.

The minimal regular resolution of X' along x' has fiber over $k(x')_{\rm sep}$ whose underlying reduced scheme looks like the chain of E_j's as shown in Figure 1*, where:*

- *all intersections are transverse, with $E_j \simeq \mathbf{P}^1_{k(x')_{\rm sep}}$;*
- *$E_j.E_j = -b_j < -1$ for all j;*

- *E_1 is transverse to the strict transform $\widetilde{X}'_1$ of the global algebraic irreducible component X'_1 through x' with multiplicity m'_2 (along which t'_1 is a cotangent direction), and similarly for E_λ and the component $\widetilde{X}'_2$ with multiplicity m'_1 in the case of two analytic branches.*

Remark 2.4.2. The case $X'_2 = X'_1$ can happen, and there is no $\widetilde{X}'_1$ in case of one analytic branch (*i.e.*, in case $m'_2 = 0$).

We will also need to know the multiplicities μ_j of the components E_j in Figure 1, but this will be easier to give after we have proved Theorem 2.4.1; see Corollary 2.4.3.

The labelling of the E_j's indicates the order in which they arise in the resolution process, with each "new" E_j linking the preceding ones to the rest of the closed fiber in the case of one initial analytic branch. Keeping this picture in mind, we see that it is always the strict transform $\widetilde{X}'_2$ of the initial component with formal parameter t'_2 that occurs at the end of the chain, and this is the component whose multiplicity is m'_1.

Proof. We may assume S is local, and if S is not already excellent then (by hypothesis) X'_K is smooth and all connected components of its regular compactification have positive arithmetic genus. We claim that this positivity assumption is preserved by extension of the fraction field K. That is, if $\overline{C}$ is a connected regular proper curve over a field k with $\mathrm{H}^1(\overline{C}, \mathcal{O}_{\overline{C}}) \neq 0$ and C is a dense open in $\overline{C}$ that is k-smooth, then for any extension k'/k we claim that all connected components C'_i of the regular k'-curve $C' = C_{/k'}$ have compactification $\overline{C}'_i$ with $\mathrm{H}^1(\overline{C}'_i, \mathcal{O}_{\overline{C}'_i}) \neq 0$. Since the field $\mathrm{H}^0(\overline{C}, \mathcal{O}_{\overline{C}})$ is clearly finite separable over k, by using Stein factorization for $\overline{C}$ we may assume $\overline{C}$ is geometrically connected over k. Thus, $\overline{C}' = \overline{C}_{/k'}$ is a connected proper k'-curve with $\mathrm{H}^1(\overline{C}', \mathcal{O}_{\overline{C}'}) \neq 0$ and there is a dense open C' that is k'-smooth, and we want to show that the normalization of $\overline{C}'_{\mathrm{red}}$ has positive arithmetic genus. Since $\overline{C}'$ is generically reduced, the map from $\mathcal{O}_{\overline{C}'}$ to the normalization sheaf of $\mathcal{O}_{\overline{C}'_{\mathrm{red}}}$ has kernel and cokernel supported in dimension 0, and so the map on H^1's is an isomorphism. Thus, the normalization of $\overline{C}'_{\mathrm{red}}$ indeed has positive arithmetic genus.

We conclude that Lemma 2.1.1 and the base-change compatibility of Definition 2.2.3 (via Theorem 2.2.2) permit us to base-change to $\widehat{\mathcal{O}}_{S,s}$ without losing any hypotheses. Thus, we may assume $S = \operatorname{Spec} W$ with W a complete (hence excellent) discrete valuation ring. This brings us to the excellent case with all connected components of the regular compactification of X'_K having positive arithmetic genus when the residue field is not separably closed. If in addition $k(s)$ is not separably closed, then we claim that base-change to $\operatorname{Spec} W^{\mathrm{sh}}$ preserves all hypotheses, and so we can always get to the case of a separably closed residue field (in particular, we get to the case with $k(x')$ separably closed); see [24, p. 17] for a proof that strict henselization preserves excellence. We need to show that base change to W^{sh} commutes with the formation of the minimal regular resolution. This is a refinement on Theorem 2.2.2 because such base change is generally not residually trivial.

From the proof of Theorem 2.2.2 in the excellent case, we see that if $X' \hookrightarrow \overline{X}'$ is a Nagata compactification then the minimal resolution $X \to X'$ of X' is the part of the minimal regular resolution of $\overline{X}'$ that lies over X'. Hence, the base-change problem for $W \to W^{\rm sh}$ is reduced to the proper case. We may assume that X' is connected, so $\widetilde{W} = \mathrm{H}^0(X', \mathcal{O}_{X'})$ is a complete discrete valuation ring finite over W. Hence, $\widetilde{W}^{\rm sh} \simeq \widetilde{W} \otimes_W W^{\rm sh}$, so we may reduce to the case when $X' \to \operatorname{Spec} W$ is its own Stein factorization. In this proper case, the positivity condition on the arithmetic genus of the generic fiber allows us to use [41, 9/3.28] (which rests on a dualizing-sheaf criterion for minimality) to conclude that formation of the minimal regular resolution of X' is compatible with étale localization on W. A standard direct limit argument that chases the property of having a -1-curve in a fiber over X' thereby shows that the formation of the minimal regular resolution is compatible with ind-étale base change (such as $W \to W^{\rm sh}$). Thus, we may finally assume that W is excellent and has a separably closed residue field, and so we no longer need to impose a positivity condition on arithmetic genera of the connected components of the generic-fiber regular compactification.

The intrinsic numerical data for the *unique* minimal resolution (that is, the self-intersection numbers and multiplicities of components in the exceptional divisor for this resolution) may be computed in an étale neighborhood of x', by Corollary 2.2.4 and Remark 2.2.5, and the Artin approximation theorem is the ideal tool for finding a convenient étale neighborhood in which to do such a calculation. We will use the Artin approximation theorem to construct a special case that admits an étale neighborhood that is also an étale neighborhood of our given x', and so it will be enough to carry out the resolution in the special case. The absence of a good theory of minimal regular resolutions for complete 2-dimensional local noetherian rings prevents us from carrying out a proof entirely on $\widehat{\mathcal{O}}_{X',x'}$, and so forces us to use the Artin approximation theorem. It is perhaps worth noting at the outset that the reason we have to use Artin approximation is that the resolution process to be used in the special case will not be intrinsic (we blow up certain codimension-1 subschemes that depend on coordinates).

Here is the special case that we wish to analyze. Let $n > 1$ be a positive integer that is a unit in W, and choose $1 \le r < n$ with $\gcd(r, n) = 1$. Pick integers $m_1 \ge 1$ and $m_2 \ge 0$ satisfying $m_1 \equiv -rm_2 \bmod n$. For technical reasons, we do not require either of the m_j's to be units in W. To motivate things, let us temporarily assume that the residue field k of W contains a full set of nth roots of unity. Let $\mu_n(k)$ act on the *regular domain* $A = W[t_1, t_2]/(t_1^{m_1} t_2^{m_2} - \pi)$ via

$$[\zeta](t_1) = \zeta t_1, \ \ [\zeta](t_2) = \zeta^r t_2. \tag{2.4.2}$$

Since the $\mu_n(k)$-action in (2.4.2) is clearly free away from $t_1 = t_2 = \pi = 0$, the quotient

$$Z = (\mathrm{Spec}(A))/\mu_n(k) = \mathrm{Spec}(B)$$

(with $B = A^{\mu_n(k)}$) is normal and also is regular away from the image point $z \in Z$ of $t_1 = t_2 = \pi = 0$.

To connect up the special situation (Z, z) and the tame cyclic quotient singularity $x' \in X'_{/S}$, note that Lemma 2.3.4 shows that our situation is formally isomorphic to the algebraic $Z = \mathrm{Spec}(B)$ for a suitable such B and $n \in W^\times$. By the Artin approximation theorem, there is a common (residually trivial) connected étale neighborhood (U, u) of (Z, z) and (X', x'). That is, there is a pointed connected affine W-scheme $U = \mathrm{Spec}(A)$ that is a residually-trivial étale neighborhood of x' and of z. In particular, U is a connected normal W-curve. We can assume that u is the only point of U over z, and also the only point of U over x'. Keep in mind (*e.g.*, if $\gcd(m_1, m_2) > 1$) that the field K might not be separably closed in the function fields of U or Z, so the generic fibers of U and $Z = \mathrm{Spec}(B)$ over W might not be geometrically connected and U is certainly not proper over W in general.

The étale-local nature of the minimal regular resolution, as provided by Corollary 2.2.4 and Remark 2.2.5, implies that the minimal regular resolutions of (X', x') and (Z, z) have pullbacks to (U, u) that coincide with the minimal regular resolution of U along $\{u\}$. The fibers over u, x', z are all the same due to residual-triviality, so the geometry of the resolution fiber at x' is the same as that over z. Hence, we shall compute the minimal regular resolution $Z' \to Z$ at z, and will see that the fiber of Z' over z is as in Figure 1.

Let us now study (Z, z). Since n is a unit in W, the normal domain $B = A^{\mu_n(k)}$ is a quotient of $W[t_1, t_2]^{\mu_n(k)}$ via the natural map. Since the action of $\mu_n(k)$ as in (2.4.2) sends each monomial $t_1^{e_1} t_2^{e_2}$ to a constant multiple of itself, the ring of invariants $W[t_1, t_2]^{\mu_n(k)}$ is spanned over W by the invariant monomials. Clearly $t_1^{e_1} t_2^{e_2}$ is $\mu_n(k)$-invariant if and only if $e_1 + re_2 = nf$ for some integer f (so $e_2 \leq (n/r)f$), in which case $t_1^{e_1} t_2^{e_2} = u^f v^{e_2}$, where $u = t_1^n$ and $v = t_2/t_1^r$ are $\mu_n(k)$-invariant elements in the fraction field of $W[t_1, t_2]$. Note that even though v does not lie in $W[t_1, t_2]$, for any pair of integers i, j satisfying $0 \leq j \leq (n/r)i$ we have $u^i v^j \in W[t_1, t_2]$ and

$$W[t_1, t_2]^{\mu_n(k)} = \bigoplus_{0 \leq j \leq (n/r)i} W u^i v^j.$$

We have $t_1^{m_1} t_2^{m_2} = u^\mu v^{m_2}$ with $m_1 + rm_2 = n\mu$ (so $m_2 \leq (n/r)\mu$). Thus,

$$B = \frac{\bigoplus_{0 \leq j \leq (n/r)i} W u^i v^j}{(u^\mu v^{m_2} - \pi)}. \tag{2.4.3}$$

Observe that (2.4.3) makes sense as a definition of finite-type W-algebra, without requiring n to be a unit and without requiring that k contain any non-trivial roots of unity. It is clear that (2.4.3) is W-flat, as it has a W-module basis given by monomials $u^i v^j$ with $0 \leq j \leq (n/r)i$ and either $i < \mu$ or $j < m_2$. It is less evident if (2.4.3) is normal for any n, but we do not need this fact. We will inductively compute certain blow-ups on (2.4.3) *without restriction on* n or on the residue field, and the process will end at a resolution of singularities for $\operatorname{Spec} B$.

Before we get to the blowing-up, we shall show that $\operatorname{Spec} B$ is a W-curve and we will infer some properties of its closed fiber. Note that the map $K(u,v) \to K(t_1,t_2)$ defined by $u \mapsto t_1^n$, $v \mapsto t_2/t_1^r$ induces a W-algebra injection

$$\bigoplus_{0\leq j\leq (n/r)i} Wu^iv^j \to W[t_1,t_2] \tag{2.4.4}$$

that is *finite* because $t_1^n = u$ and $t_2^n = u^rv^n$. Thus, the left side of (2.4.4) is a 3-dimensional noetherian domain and passing to the quotient by $u^\mu v^{m_2} - \pi = t_1^{m_1}t_2^{m_2} - \pi$ yields a finite surjection

$$\operatorname{Spec}(W[t_1,t_2]/(t_1^{m_1}t_2^{m_2} - \pi)) \to \operatorname{Spec}(B). \tag{2.4.5}$$

Passing to the generic fiber and recalling that B is W-flat, we infer that $\operatorname{Spec}(B)$ is a W-curve with irreducible generic fiber, so $\operatorname{Spec}(B)$ is 2-dimensional and connected. We also have a finite surjection modulo π,

$$\operatorname{Spec}(k[t_1,t_2]/(t_1^{m_1}t_2^{m_2})) \to \operatorname{Spec}(B/\pi), \tag{2.4.6}$$

so the closed fiber of $\operatorname{Spec}(B)$ consists of at most two irreducible components (or just one when $m_2 = 0$), to be called the images of the t_1-axis and t_2-axis (where we omit mention of the t_1-axis when $m_2 = 0$). Since the t_2-axis is the preimage of the zero-scheme of $u = t_1^n$ under (2.4.6), we conclude that when $m_2 > 0$ the closed fiber $\operatorname{Spec}(B/\pi)$ does have two distinct irreducible components.

Inspired by the case of toric varieties, we will now compute the blow-up Z' of the W-flat $Z = \operatorname{Spec}(B)$ along the ideal (u, uv). Since

$$\operatorname{Spec}(W[t_1,t_2]/(t_1^{m_1}t_2^{m_2} - \pi, t_1^n, t_1^{n-r}t_2)) \to \operatorname{Spec}(B/(u,uv))$$

is a finite surjection and the source is supported in the t_2-axis of the closed fiber over $\operatorname{Spec}(W)$, it follows that $\operatorname{Spec}(B/(u,uv))$ is supported in the image of the t_2-axis of the closed fiber of $\operatorname{Spec}(B)$ over $\operatorname{Spec}(W)$. In particular, blowing up Z along (u,uv) does not affect the generic fiber of Z over W. Since Z is W-flat, it follows that the proper blow-up map $Z' \to Z$ is surjective.

There are two charts covering Z', $D_+(u)$ and $D_+(uv)$, where we adjoin the ratios $uv/u = v$ and $u/uv = 1/v$ respectively. Thus,

$$D_+(u) = \operatorname{Spec}(B[v]) = \operatorname{Spec}(W[u,v]/(u^\mu v^{m_2} - \pi))$$

is visibly regular and connected, and $D_+(uv) = \operatorname{Spec}(B[1/v])$ with

$$B[1/v] = \frac{\bigoplus_{j\leq(n/r)i,\, 0\leq i} Wu^iv^j}{(u^\mu v^{m_2} - \pi)}.$$

We need to rewrite this latter expression in terms of a more useful set of variables. We begin by writing (as one does when computing the Jung–Hirzebruch continued fraction for n/r)

$$n = b_1 r - r'$$

with $b_1 \geq 2$ and either $r = 1$ with $r' = 0$ or else $r' > 0$ with $\gcd(r, r') = 1$ (since $\gcd(n, r) = 1$). We will first treat the case $r' = 0$ (proving that $B[1/v]$ is also regular) and then we will treat the case $r' > 0$. Note that there is no reason to expect that p cannot divide r or r', even if $p \nmid n$, and it is for this reason that we had to recast the definition of B in a form that avoids the assumption that n is a unit in W. For similar reasons, we must avoid assuming m_1 or m_2 is a unit in W.

Assume $r' = 0$, so $r = 1$, $b_1 = n$, and $b_1\mu - m_2 = m_1$. Let $i' = b_1 i - j$ and $j' = i$, so i' and j' vary precisely over non-negative integers and $u^i v^j = (1/v)^{i'}(uv^{b_1})^{j'}$. Thus, letting $u' = 1/v$ and $v' = uv^{b_1}$ yields

$$B[1/v] = W[u', v']/({u'}^{b_1\mu - m_2}{v'}^{\mu} - \pi) = W[u', v']/({u'}^{m_1}{v'}^{\mu} - \pi),$$

which is regular. In the closed fiber of $Z' = \mathrm{Bl}_{(u,uv)}(Z)$ over $\mathrm{Spec}(W)$, let D_1 denote the v'-axis in $D_+(uv) = \mathrm{Spec}\, B[1/v]$ and when $m_2 > 0$ let D_2 denote the u-axis in $D_+(u)$. The multiplicities of D_1 and D_2 in Z'_k are respectively $m_1 = b_1\mu - m_2$ and m_2 (with multiplicity $m_2 = 0$ being a device for recording that there is no D_2). The exceptional divisor E is a projective line over k (with multiplicity μ and gluing data $u' = 1/v$) and hence the uniformizer π has divisor on $Z' = \mathrm{Bl}_{(u,uv)}(Z)$ given by

$$\mathrm{div}_{Z'}(\pi) = (b_1\mu - m_2)D_1 + \mu E + m_2 D_2 = m_1 D_1 + \mu E + m_2 D_2$$

(when $m_2 = 0$, the final term really is omitted).

It is readily checked that the D_j's each meet E transversally at a single k-rational point (suppressing D_2 when $m_2 = 0$). The intersection product $\mathrm{div}_{Z'}(\pi).E$ makes sense since E is proper over k, even though Z is not proper over W, and it must vanish because $\mathrm{div}_{Z'}(\pi)$ is principal, so by additivity of intersection products in the first variable (restricted to effective Cartier divisors for a fixed proper second variable such as E) we have

$$0 = \mathrm{div}_{Z'}(\pi).E = b_1\mu - m_2 + \mu(E.E) + m_2.$$

Thus, $E.E = -b_1$.

Now assume $r' > 0$. Since $n = b_1 r - r'$, the condition $0 \leq j \leq (n/r)i$ can be rewritten as $0 \leq i \leq (r/r')(b_1 i - j)$. Letting $j' = i$ and $i' = b_1 i - j$, we have $u^i v^j = {u'}^{i'}{v'}^{j'}$ with $u' = 1/v$ and $v' = uv^{b_1}$. In particular, $u^\mu v^{m_2} = {u'}^{b_1\mu - m_2}{v'}^{\mu}$. Thus,

$$B[1/v] = \frac{\bigoplus_{0 \leq j' \leq (r/r')i'} W{u'}^{i'}{v'}^{j'}}{({u'}^{b_1\mu - m_2}{v'}^{\mu} - \pi)}. \tag{2.4.7}$$

Note the similarity between (2.4.3) and (2.4.7) up to modification of parameters: replace (n, r, m_1, m_2, μ) with $(r, r', m_1, \mu, b_1\mu - m_2)$. The blow-up along $(u', u'v')$ therefore has closed fiber over $\mathrm{Spec}(W)$ with the following irreducible components: the v'-axis D_1 in $D_+(uv)$ with multiplicity $b_1\mu - m_2$, the u-axis D_2 in $D_+(u)$ with multiplicity m_2 (so this only shows up when $m_2 > 0$), and the exceptional divisor E that is a projective line (via gluing $u' = 1/v$) having multiplicity μ and meeting D_1 (as well as D_2 when $m_2 > 0$) transversally at a single k-rational point. We will focus our attention on $D_+(uv)$ (as we have already seen that the other chart $D_+(u)$ is regular), and in particular we are interested in the "origin" in the closed fiber of $D_+(uv)$ over $\mathrm{Spec}(W)$ where the projective line E meets D_1; near this origin, $D_+(uv)$ is an affine open that is given by the spectrum of (2.4.7).

If r were also a unit in W then $D_+(uv)$ would be the spectrum of the ring of $\mu_r(k)$-invariants in $W[t'_1, t'_2]/({t'_1}^{m_1}{t'_2}^{\mu} - \pi)$ with the action $[\zeta](t'_1) = \zeta t'_1$ and $[\zeta](t'_1) = \zeta^{r'} t'_2$ (this identification uses the identity $m_1 + r'\mu = r(b_1\mu - m_2)$), and without any restriction on r we at least see that (2.4.7) is an instance of the general (2.4.3) and that there is a natural finite surjection

$$\mathrm{Spec}(k[t'_1, t'_2]/({t'_1}^{m_1}{t'_2}^{\mu})) \to D_+(uv)_k.$$

On $D_+(uv)_k$, the component E of multiplicity μ is the image of the t'_1-axis and the component D_1 with multiplicity m_1 is the image of the t'_2-axis. As a motivation for what follows, note also that if $r \in W^\times$ then since $r > 1$ we see that the "origin" in $D_+(uv)_k$ is necessarily a non-regular point in the total space over $\mathrm{Spec}(W)$ (by Serre's Theorem 2.3.9).

We conclude (without requiring any of our integer parameters to be units in W) that if we make the change of parameters

$$(n, r, m_1, m_2, \mu) \rightsquigarrow (r, r', m_1, \mu, b_1\mu - m_2) \tag{2.4.8}$$

then $D_+(uv)$ is like the original situation (2.4.3) with a revised set of initial parameters. In particular, n is replaced by the strictly smaller $r > 1$, so the process will eventually end. Moreover, since $\mu > 0$ we see that the case $m_2 = 0$ is now "promoted" to the case $m_2 > 0$. When we make the blow-up at the origin in $D_+(uv)_k$, the strict transform E_1 of E plays the same role that D_2 played above, so E_1 is entirely in the regular locus and the new exceptional divisor E' has multiplicity $b_1\mu - m_2$ (this parameter plays the role for the second blow-up that μ played for the first blow-up, as one sees by inspecting our change of parameters in (2.4.8)).

As the process continues, nothing more will change around E_1, so inductively we conclude from the descriptions of the regular charts that the process ends at a regular connected W-curve with closed-fiber Weil divisor

$$\cdots + (b_1\mu - m_2)E' + \mu E_1 + m_2 D_2 + \ldots \tag{2.4.9}$$

(where we have abused notation by writing E' to denote the strict transform of E' in the final resolution, and this strict transform clearly has generic multiplicity $b_1\mu - m_2$). The omitted terms in (2.4.9) do not meet E_1, so we may form the intersection against E_1 to solve

$$0 = (b_1\mu - m_2) + \mu(E_1.E_1) + m_2$$

just as in the case $r' = 0$ (*i.e.*, $r = 1$), so $E_1.E_1 = -b_1$. Since

$$\frac{n}{r} = b_1 - \frac{1}{r/r'},$$

by induction on the length of the continued fraction we reach a regular resolution in the expected manner, with $E_j.E_j = -b_j$ for all j and the final resolution having fiber over $z \in Z$ looking exactly like in Figure 1. Note also that each new blow-up separates all of the previous exceptional lines from the (strict transform of the initial) component through z with multiplicity m_1. Since $-b_j \leq -2 < -1$ for all j, we conclude that at no stage of the blow-up process before the end did we have a regular scheme (otherwise there would be a -1-curve in a fiber over the original base Z). Thus, we have computed the minimal regular resolution at z.

□

We now compute the multiplicity μ_j in the closed fiber of ${X'}^{\mathrm{reg}}$ for each fibral component E_j over $x' \in X'$ in Figure 1. In order to compute the μ_j's, we introduce some notation. Let $n/r > 1$ be a reduced-form fraction with positive integers n and r, so we can write

$$n/r = [b_1, b_2, \ldots, b_\lambda]_{\mathrm{JH}} := b_1 - \cfrac{1}{b_2 - \cfrac{1}{\cdots - \cfrac{1}{b_\lambda}}}$$

as a Jung–Hirzebruch continued fraction, where $b_j \geq 2$ for all j. Define $P_j = P_j(b_1, \ldots, b_\lambda)$ and $Q_j = Q_j(b_1, \ldots, b_\lambda)$ by

$$P_{-1} = 0,\ \ Q_{-1} = -1,\ \ P_0 = 1,\ \ Q_0 = 0,$$

$$P_j = b_j P_{j-1} - P_{j-2},\ \ Q_j = b_j Q_{j-1} - Q_{j-2}$$

for all $j \geq 1$. Clearly P_j and Q_j are universal polynomials in $b_1, \ldots, b_j$, and by induction $P_j Q_{j-1} - Q_j P_{j-1} = -1$ and $Q_j > Q_{j-1}$ for all $j \geq 0$, so in particular $Q_j > 0$ for all $j > 0$. Thus,

$$[b_1, \ldots, b_\lambda]_{\mathrm{JH}} = \frac{P_\lambda(b_1, \ldots, b_\lambda)}{Q_\lambda(b_1, \ldots, b_\lambda)}$$

makes sense and P_λ/Q_λ is in reduced form. Thus, $P_\lambda = n$ and $Q_\lambda = r$ since the Q_j's are necessarily positive.

Corollary 2.4.3. *With hypotheses and notation as in Theorem* 2.4.1, *let* μ_j *denote the multiplicity of* E_j *in the fiber of* ${X'}^{\mathrm{reg}}$ *over* $k(x')_{\mathrm{sep}}$. *The condition* $r = 1$ *happens if and only if* $\lambda = 1$, *in which case* $\mu_1 = (m'_1 + m'_2)/n$.

If $r > 1$ (*so* $\lambda > 1$), *then the* μ_j*'s are the unique solution to the equation*

$$\begin{pmatrix} b_1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & b_2 & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & b_3 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & b_{\lambda-1} & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & b_\lambda \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \vdots \\ \mu_\lambda \end{pmatrix} = \begin{pmatrix} m'_2 \\ 0 \\ \vdots \\ 0 \\ m'_1 \end{pmatrix}. \tag{2.4.10}$$

Keeping the condition $r > 1$, *define* $P'_j = P_j(b_{\lambda-j+1}, \dots, b_\lambda)$, *so* $P'_\lambda = n$ *and* $P'_{\lambda-1} = Q_\lambda(b_1, \dots, b_\lambda) = r$. *If we let* $\widetilde{m}_2 = P'_{\lambda-1} m'_2 + m'_1 = r m'_2 + m'_1$, *then the* μ_j*'s are also the unique solution to*

$$\begin{pmatrix} P'_\lambda & 0 & 0 & \dots & 0 & 0 & 0 \\ -P'_{\lambda-2} & P'_{\lambda-1} & 0 & \dots & 0 & 0 & 0 \\ 0 & -P'_{\lambda-3} & P'_{\lambda-2} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -P'_1 & P'_2 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & P'_1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \vdots \\ \mu_\lambda \end{pmatrix} = \begin{pmatrix} \widetilde{m}_2 \\ m'_1 \\ \vdots \\ m'_1 \\ m'_1 \end{pmatrix}. \tag{2.4.11}$$

In particular, $\mu_1 = (r m'_2 + m'_1)/n$.

Note that in the applications with $X' = X/H$ as at the beginning of §2.3, the condition $\chi'_1 \neq \chi'_2$ (*i.e.*, $H'_{x|x'}$ does not act through scalars) is equivalent to the condition $r > 1$ in Corollary 2.4.3.

Proof. The value of μ_1 when $r = 1$ was established in the proof of Theorem 2.4.1, so now assume $r > 1$. On ${X'}^{\mathrm{reg}}$ (or rather, its base change to $\mathcal{O}^{\mathrm{sh}}_{S,s}$) we have

$$\mathrm{div}(\pi_s) = m'_1 \widetilde{X}'_2 + \sum_{j=1}^{\lambda} \mu_j E_j + m'_2 \widetilde{X}'_1 + \dots \tag{2.4.12}$$

where

- the $\widetilde{X}'_1$-term does not appear if there is only one analytic branch through x' (recall we also set $m'_2 = 0$ in this case),
- the $\widetilde{X}'_j$-terms are a single term when there are two analytic branches but only one global irreducible (geometric) component (in which case $m'_1 = m'_2$),
- the omitted terms "$\dots$" on the right side of (2.4.12) are not in the fiber over x' (and in particular do not intersect the E_j's).

Thus, the equations $E_j.\mathrm{div}(\pi_s) = 0$ and the intersection calculations in the proof of Theorem 2.4.1 (as summarized by Figure 1, including transversalities) immediately yield (2.4.10). By solving this system of equations by working up from the bottom row, an easy induction argument yields the reformulation (2.4.11).

□

To prove Theorems 1.1.2 and 1.1.6, the preceding general considerations will provide the necessary intersection-theoretic information on a minimal resolution. To apply Theorem 2.4.1 and Corollary 2.4.3 to the study of singularities at points x' on modular curves, we need to find the value of the parameter $r_{x'}$ in each case. This will be determined by studying universal deformation rings for moduli problems of elliptic curves.

3 The Coarse moduli scheme $X_1(p)$

Let p be a prime number. In this section we review the construction of the coarse moduli scheme $X_1(p)$ attached to $\Gamma_1(p)$ in terms of an auxiliary finite étale level structure which exhibits $X_1(p)$ as the compactification of a quotient of a fine moduli scheme. It is the fine moduli schemes whose completed local rings are well understood through deformation theory (as in [34]), and this will provide the starting point for our subsequent calculations of regular models and component groups.

3.1 Some general nonsense

As in [34, Ch. 4], for a scheme T we let (Ell/T) be the category whose objects are elliptic curves over T-schemes and whose morphisms are cartesian diagrams. The moduli problem $[\Gamma_1(p)]$ is the contravariant functor $(\mathrm{Ell}) \to (\mathrm{Sets})$ that to an elliptic curve $E_{/S}$ attaches the set of $P \in E(S)$ such that the relative effective Cartier divisor

$$[0] + [P] + [2P] + \cdots + [(p-1)P],$$

viewed as a closed subscheme of E, is a closed subgroup scheme. For any moduli problem $\mathcal{P}$ on (Ell/T) and any object $E_{/S}$ over a T-scheme, we define the functor $\mathcal{P}_{E/S}(S') = \mathcal{P}(E_{/S'})$ to classify "$\mathcal{P}$-structures" on base changes of $E_{/S}$. If $\mathcal{P}_{E/S}$ is representable (with some property $\mathbf{P}$ relative to S) for every $E_{/S}$, we say that $\mathcal{P}$ is *relatively representable* (with property $\mathbf{P}$). For example, $[\Gamma_1(p)]$ is relatively representable and finite locally free of degree p^2-1 on (Ell) for every prime p.

For $p \geq 5$, the moduli problem $[\Gamma_1(p)]_{/\mathbf{Z}[1/p]}$ is representable by a smooth affine curve over $\mathbf{Z}[1/p]$ [34, Cor. 2.7.3, Thm. 3.7.1, and Cor. 4.7.1]. For any elliptic curve $E_{/S}$ over an $\mathbf{F}_p$-scheme S, the point $P = 0$ is fixed by the automorphism -1 of $E_{/S}$, and is in $[\Gamma_1(p)](E/S)$ because $[0]+[P]+\cdots+[(p-1)P]$ is the kernel of the relative Frobenius morphism $F\colon E \to E^{(p)}$. Thus, $[\Gamma_1(p)]_{/\mathbf{Z}_{(p)}}$ is not rigid, so it is not representable.

As there is no fine moduli scheme associated to $[\Gamma_1(p)]_{/\mathbf{Z}_{(p)}}$ for any prime p, we let $X_1(p)$ be the compactified coarse moduli scheme $\overline{M}([\Gamma_1(p)]_{/\mathbf{Z}_{(p)}})$, as constructed in [34, Ch. 8]. This is a proper normal $\mathbf{Z}_{(p)}$-model of a smooth and geometrically connected curve $X_1(p)_{/\mathbf{Q}}$, but $X_1(p)$ is usually not regular. Nevertheless, the complete local rings on $X_1(p)$ are computable in terms of abstract deformation theory. Since $(\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ acts on isomorphism classes of $\Gamma_1(p)$-structures via

$$(E, P) \mapsto (E, a \cdot P) \simeq (E, -a \cdot P),$$

we get a natural action of this group on $X_1(p)$ which is readily checked to be a faithful action (*i.e.*, non-identity elements act non-trivially). Thus, for any subgroup $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ we get the modular curve $X_H(p) = X_1(p)/H$ which is a normal proper connected $\mathbf{Z}_{(p)}$-curve with smooth generic fiber $X_H(p)_{/\mathbf{Q}}$. When $p > 3$, the curve $X_H(p)$ has tame cyclic quotient singularities at its non-regular points.

In order to compute a minimal regular model for these normal curves, we need more information than is provided by abstract deformation theory: we need to keep track of *global* irreducible components on the geometric fiber mod p, whereas deformation theory will only tell us about the analytic branches through a point. Fortunately, in the case of modular curves $X_H(p)$, distinct analytic branches through a closed-fiber geometric point always arise from distinct global (geometric) irreducible components through the point. In order to review this fact, as well as to explain the connection between complete local rings on $X_H(p)$ and rings of invariants in universal deformation rings, we need to recall how $X_1(p)$ can be constructed from fine moduli schemes. Let us briefly review the construction process.

Pick a representable moduli problem $\mathcal{P}$ that is finite, étale, and Galois over $(\mathrm{Ell}/\mathbf{Z}_{(p)})$ with Galois group $G_{\mathcal{P}}$, and for which $M(\mathcal{P})$ is affine. For example (cf. [34, §4.5–4.6]) if $\ell \neq p$ is a prime with $\ell \geq 3$, we can take $\mathcal{P}$ to be the moduli problem $[\Gamma(\ell)]_{/\mathbf{Z}_{(p)}}$ that attaches to $E_{/S}$ the set of isomorphisms of S-group schemes

$$\phi\colon (\mathbf{Z}/\ell\mathbf{Z})^2_S \simeq E[\ell];$$

the Galois group $G_{\mathcal{P}}$ is $\mathrm{GL}_2(\mathbf{F}_\ell)$. Let $Y_1(p;\mathcal{P})$ be the fine moduli scheme $M([\Gamma_1(p)]_{/\mathbf{Z}_{(p)}}, \mathcal{P})$ that classifies pairs consisting of a $\Gamma_1(p)$-structure and a $\mathcal{P}$-structure on elliptic curves over variable $\mathbf{Z}_{(p)}$-schemes. The scheme $Y_1(p;\mathcal{P})$ is a flat affine $\mathbf{Z}_{(p)}$-curve. Let $Y_1(p)$ be the quotient of $Y_1(p;\mathcal{P})$ by the $G_{\mathcal{P}}$-action.

We introduce the global $\mathcal{P}$ rather than just use formal deformation theory throughout because on characteristic-p fibers we need to retain a connection between closed fiber irreducible components of global modular curves and closed fiber "analytic" irreducible components of formal deformation rings. The precise connection between global $\mathcal{P}$'s and infinitesimal deformation theory is given by the well-known:

Theorem 3.1.1. *Let k be an algebraically closed field of characteristic p and let $W = W(k)$ be its ring of Witt vectors. Let $z \in Y_1(p)_{/k}$ be a rational point. Let* $\mathrm{Aut}(z)$ *denote the finite group of automorphisms of the (non-canonically unique) $\Gamma_1(p)$-structure over k underlying z. Choose a $\mathcal{P}$-structure on the elliptic curve underlying z, with $\mathcal{P}$ as above, and let $z' \in Y_1(p;\mathcal{P})(k)$ be the corresponding point over z.*

The ring $\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$ is naturally identified with the formal deformation ring of z. Under the resulting natural action of $\mathrm{Aut}(z)$ *on $\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$, the subring of* $\mathrm{Aut}(z)$*-invariants is $\widehat{\mathcal{O}}_{Y_1(p)_W,z}$.*

For any subgroup $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$ equipped with its natural action on $Y_1(p)$, the stabilizer $H_{z'|z}$ of z' in H acts faithfully on the universal deformation ring $\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$ of z in the natural way, with subring of invariants $\widehat{\mathcal{O}}_{Y_H(p)_W,z}$.

Proof. Since $\mathcal{P}$ is étale and $Y_1(p;\mathcal{P})_W$ is a fine moduli scheme, the interpretation of $\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$ as a universal deformation ring is immediate. Since $Y_1(p)_W$ is the quotient of $Y_1(p;\mathcal{P})_W$ by the action of $G_{\mathcal{P}}$, it follows that $\widehat{\mathcal{O}}_{Y_1(p)_W,z}$ is identified with the subring of invariants in $\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$ for the action of the stabilizer of z' for the $G_{\mathcal{P}}$-action on $Y_1(p;\mathcal{P})_W$. We need to compute this stabilizer subgroup.

If $z' = (E_z, P_z, \iota)$ with supplementary $\mathcal{P}$-structure ι, then $g \in G_{\mathcal{P}}$ fixes z' if and only if (E_z, P_z, ι) is isomorphic to $(E_z, P_z, g(\iota))$. This says exactly that there exists an automorphism α_g of (E_z, P_z) carrying ι to $g(\iota)$, and such α_g is clearly unique if it exists. Moreover, any two $\mathcal{P}$-structures on E_z are related by the action of a unique $g \in G_{\mathcal{P}}$ because of the definition of $G_{\mathcal{P}}$ as the Galois group of $\mathcal{P}$ (and the fact that z is a geometric point). Thus, the stabilizer of z in $G_{\mathcal{P}}$ is naturally identified with $\mathrm{Aut}(E_z, P_z) = \mathrm{Aut}(z)$ (*compatibly* with actions on the universal deformation ring of z). The assertion concerning the H-action is clear.

□

Since $Y_1(p;\mathcal{P})$ is a regular $\mathbf{Z}_{(p)}$-curve [34, Thm. 5.5.1], it follows that its quotient $Y_1(p)$ is a normal $\mathbf{Z}_{(p)}$-curve. Moreover, by [34, Prop. 8.2.2] the natural map $j : Y_1(p) \to \mathbf{A}^1_{\mathbf{Z}_{(p)}}$ is finite, and hence it is also *flat* [44, 23.1]. In [34], $X_1(p)$ is *defined* to be the normalization of $Y_1(p)$ over the compactified j-line $\mathbf{P}^1_{\mathbf{Z}_{(p)}}$. Both $X_1(p)$ and $Y_1(p)$ are independent of the auxiliary choice of $\mathcal{P}$. The complex analytic theory shows that $X_1(p)$ has geometrically connected fibers over $\mathbf{Z}_{(p)}$, so the same is true for $Y_1(p)$ since the complete local rings at the cusps are analytically irreducible mod p (by the discussion in §4.2, especially the self-contained Lemma 4.2.4 and Lemma 4.2.5).

3.2 Formal parameters

To do deformation theory computations, we need to recall some canonical formal parameters in deformation rings. Fix an algebraically closed field k of characteristic p and let $W = W(k)$ denote its ring of Witt vectors. Let

$z \in Y_1(p)_{/k}$ be a k-rational point corresponding to an elliptic curve $E_{z/k}$ with $\Gamma_1(p)$-structure P_z.

For later purposes, it is useful to give a conceptual description of the 1-dimensional "reduced" cotangent space $\mathfrak{m}/(p, \mathfrak{m}^2)$ of $\mathcal{R}_z^0$, or equivalently the cotangent space to the equicharacteristic formal deformation functor of E_z:

Theorem 3.2.1. *The cotangent space to the equicharacteristic formal deformation functor of an elliptic curve E over a field k is canonically isomorphic to* $\mathrm{Cot}_0(E)^{\otimes 2}$.

Proof. This is just the dual of the Kodaira-Spencer isomorphism. More specifically, the cotangent space is isomorphic to $\mathrm{H}^1(E, (\Omega^1_{E/k})^\vee)^\vee$, and Serre duality identifies this latter space with

$$\begin{tikzcd} \mathrm{H}^0(E, (\Omega^1_{E/k})^{\otimes 2}) & \mathrm{H}^0(E, \Omega^1_{E/k})^{\otimes 2} \arrow[l, "\simeq"'] \arrow[r, equal] & \mathrm{Cot}_0(E)^{\otimes 2}, \end{tikzcd}$$

the first map being an isomorphism since $\Omega^1_{E/k}$ is (non-canonically) trivial. □

Let

$$\mathbf{E}_z \to \mathrm{Spec}(\mathcal{R}_z^0)$$

denote an algebraization of the universal deformation of E_z, so non-canonically $\mathcal{R}_z^0 \simeq W[\![t]\!]$ and (by Theorem 3.1.1) there is a unique local W-algebra map $\mathcal{R}_z^0 \to \mathcal{R}_z$ to the universal deformation ring $\mathcal{R}_z$ of (E_z, P_z) such that there is a (necessarily unique) isomorphism of deformations between the base change of $\mathbf{E}_z$ over $\mathcal{R}_z$ and the universal elliptic curve underlying the algebraized universal $\Gamma_1(p)$-structure deformation at z.

Now make the additional hypothesis $P_z = 0$, so upon choosing a formal coordinate $\underline{x}$ for the formal group of $\mathbf{E}_z$ it makes sense to consider the coordinate

$$x = \underline{x}(\mathbf{P}_z) \in \mathcal{R}_z$$

of the "point" $\mathbf{P}_z$ in the universal $\Gamma_1(p)$-structure over $\mathcal{R}_z$. We thereby get a natural local W-algebra map

$$W[\![x, t]\!] \to \mathcal{R}_z. \tag{3.2.1}$$

Theorem 3.2.2. *The natural map* (3.2.1) *is a surjection with kernel generated by an element f_z that is part of a regular system of parameters of the regular local ring $W[\![x,t]\!]$. Moreover, x and t span the* 2*-dimensional cotangent space of the target ring.*

Proof. The surjectivity and cotangent-space claims amount to the assertion that an artinian deformation whose $\Gamma_1(p)$-structure vanishes and whose t-parameter vanishes necessarily has $p = 0$ in the base ring (so we then have a constant deformation). The vanishing of p in the base ring is [34, 5.3.2.2]. Since the deformation ring $\mathcal{R}_z$ is a 2-dimensional regular local ring, the kernel of the surjection (3.2.1) is a height-1 prime that must therefore be principal with a generator that is part of a regular system of parameters. □

3.3 Closed-fiber description

For considerations in Section 5, we will need some more refined information, particularly a description of $f_z \bmod p$ in Theorem 3.2.2. To this end, we first need to recall some specialized moduli problems in characteristic p.

Definition 3.3.1. If $E_{/S}$ is an elliptic curve over an $\mathbf{F}_p$-scheme S, and $G \hookrightarrow E$ is a finite locally free closed subgroup scheme of order p, we shall say that G is a $(1,0)$-*subgroup* if G is the kernel of the relative Frobenius map $F_{E/S} : E \to E^{(p)}$ and G is a $(0,1)$-*subgroup* if the order p group scheme $E[p]/G \hookrightarrow E/G$ is the kernel of the relative Frobenius for the quotient elliptic curve E/G over S.

Remark 3.3.2. This is a special case of the more general concept of (a,b)-*cyclic subgroup* which is developed in [34, §13.4] for describing the mod p fibers of modular curves. On an ordinary elliptic curve over a field of characteristic p, an (a,b)-cyclic subgroup has connected-étale sequence with connected part of order p^a and étale part of order p^b.

Let $\mathcal{P}$ be a representable moduli problem over $(\mathrm{Ell}/\mathbf{Z}_{(p)})$ that is finite, étale, and Galois with $M(\mathcal{P})$ affine (as in §3.1). For $(a,b) = (1,0), (0,1)$, it makes sense to consider the subfunctor

$$[[\Gamma_1(p)]\text{-}(a,b)\text{-cyclic}, \mathcal{P}] \tag{3.3.1}$$

of points of $[\Gamma_1(p)_{/\mathbf{F}_p}, \mathcal{P}]$ whose $\Gamma_1(p)$-structure generates an (a,b)-cyclic subgroup. By [34, 13.5.3, 13.5.4], these subfunctors (3.3.1) are represented by closed subschemes of $Y_1(p;\mathcal{P})_{/\mathbf{F}_p}$ that intersect at exactly the supersingular points and have ordinary loci that give a covering of $Y_1(p;\mathcal{P})^{\mathrm{ord}}_{/\mathbf{F}_p}$ by open subschemes. Explicitly, we have an $\mathbf{F}_p$-scheme isomorphism

$$M([\Gamma_1(p)]\text{-}(0,1)\text{-cyclic}, \mathcal{P}) \simeq M([\mathrm{Ig}(p)], \mathcal{P}) \tag{3.3.2}$$

with a smooth (possibly disconnected) Igusa curve, where $[\mathrm{Ig}(p)]$ is the moduli problem that classifies $\mathbf{Z}/p\mathbf{Z}$-generators of the kernel of the relative Verschiebung $V_{E/S} : E^{(p)} \to E$, and the line bundle ω of relative 1-forms on the universal elliptic curve over $M(\mathcal{P})_{/\mathbf{F}_p}$ provides the description

$$M([\Gamma_1(p)]\text{-}(1,0)\text{-cyclic}, \mathcal{P}) \simeq \mathrm{Spec}((\mathrm{Sym}_{M(\mathcal{P})_{/\mathbf{F}_p}}\omega)/\omega^{\otimes(p-1)}) \tag{3.3.3}$$

as the cover obtained by locally requiring a formal coordinate of the level-p structure to have $(p-1)$th power equal to zero. The scheme (3.3.3) has generic multiplicity $p-1$ and has smooth underlying reduced curve $M(\mathcal{P})_{/\mathbf{F}_p}$.

We conclude that $Y_1(p;\mathcal{P})$ is $\mathbf{Z}_{(p)}$-smooth at points in

$$M([\Gamma_1(p)]\text{-}(0,1)\text{-cyclic}, \mathcal{P})^{\mathrm{ord}},$$

and near points in $M([\Gamma_1(p)]\text{-}(1,0)\text{-cyclic}, \mathcal{P})$ we can use a local trivialization of ω to find a nilpotent function X with a moduli-theoretic interpretation as the formal coordinate of the point in the $\Gamma_1(p)$-structure (with X^{p-1} arising as $\Phi_p(X+1) \bmod p$ along the ordinary locus). Thus, we get the "ordinary" part of:

Theorem 3.3.3. *Let k be an algebraically closed field of characteristic p, and $z \in Y_1(p)_{/k}$ a rational point corresponding to a $(1,0)$-subgroup of an elliptic curve E over k. Choose $z' \in Y_1(p;\mathcal{P})_{/k}$ over z. Let f_z be a generator of the kernel of the surjection $W[\![x,t]\!] \twoheadrightarrow \widehat{\mathcal{O}}_{Y_1(p;\mathcal{P}),z'}$ in* (3.2.1).

We can choose f_z so that

$$f_z \bmod p = \begin{cases} x^{p-1} & \text{if } E \text{ is ordinary,} \\ x^{p-1}t' & \text{if } E \text{ is supersingular,} \end{cases}$$

with p, x, t' a regular system of parameters in the supersingular case. In particular, $Y_1(p;\mathcal{P})^{\mathrm{red}}_{/k}$ has smooth irreducible components, ordinary double point singularities at supersingular points, and no other non-smooth points.

The significance of Theorem 3.3.3 for our purposes is that it ensures the regular $\mathbf{Z}_{(p)}$-curve $Y_1(p;\mathcal{P})_{\mathbf{Z}_{(p)}}$ is nil-semistable in the sense of Definition 2.3.1. In particular, for $p > 3$ and any subgroup $H \subseteq (\mathbf{Z}/p\mathbf{Z})^\times/\{\pm 1\}$, the modular curve $X_H(p)$ has tame cyclic quotient singularities away from the cusps.

Proof. The geometric irreducible components of $Y_1(p,\mathcal{P})^{\mathrm{red}}_{/k}$ are smooth curves (3.3.2) and (3.3.3) that intersect at exactly the supersingular points, and (3.3.3) settles the description of $f_z \bmod p$ in the ordinary case. It remains to verify the description of $f_z \bmod p$ at supersingular points z, for once this is checked then the two minimal primes (x) and (t') in the deformation ring at z must correspond to the k-fiber irreducible components of the smooth curves (3.3.2) and $(3.3.3)_{\mathrm{red}}$ through z', and these two primes visibly generate the maximal ideal at z' in the k-fiber so (3.3.2) and $(3.3.3)_{\mathrm{red}}$ intersect transversally at z' as desired.

Consider the supersingular case. The proof of [34, 13.5.4] ensures that we can choose f_z so that

$$f_z \bmod p = g_{(1,0)}g_{(0,1)}, \tag{3.3.4}$$

with $k[\![x,t]\!]/g_{(0,1)}$ the complete local ring at z' on the closed subscheme (3.3.2) and likewise for $k[\![x,t]\!]/g_{(1,0)}$ and (3.3.3). By (3.3.3), we can take $g_{(1,0)} = x^{p-1}$, so by (3.3.4) it suffices to check that the formally smooth ring $k[\![x,t]\!]/g_{(0,1)}$ does not have t as a formal parameter. In the proof of [34, 12.8.2], it is shown that there is a natural isomorphism between the moduli stack of Igusa structures and the moduli stack of $(p-1)$th roots of the Hasse invariant of elliptic curves over $\mathbf{F}_p$-schemes. Since the Hasse invariant commutes with base change and the Hasse invariant on the the universal deformation of a supersingular elliptic curve over $k[\![t]\!]$ has a simple zero [34, 12.4.4], by extracting a $(p-1)$th root we lose the property of t being a formal parameter if $p > 2$. We do not need the theorem for the supersingular case when $p = 2$, so we leave this case as an exercise for the interested reader.

□

4 Determination of non-regular points

Since the quotient $X_H(p)$ of the normal proper $\mathbf{Z}_{(p)}$-curve $X_1(p;\mathcal{P})$ is normal, there is a finite set of non-regular points in codimension-2 on $X_H(p)$ that we have to resolve to get a regular model. We will prove that the non-regular points on the nil-semistable $X_H(p)$ are certain *non-cuspidal* $\mathbf{F}_p$-rational points with j-invariants 0 and 1728, and that these singularities are tame cyclic quotient singularities when $p>3$, so Jung–Hirzebruch resolution in Theorem 2.4.1 will tell us everything we need to know about the minimal regular resolution of $X_H(p)$.

4.1 Analysis away from cusps

The only possible non-regular points on $X_H(p)$ are closed points in the closed fiber. We will first consider those points that lie in $Y_H(p)$, and then we will study the situation at the cusps. The reason for treating these cases separately is that the deformation theory of generalized elliptic curves is a little more subtle than that of elliptic curves. One can also treat the situation at the cusps by using Tate curves instead of formal deformation theory; this is the approach used in [34].

In order to determine the non-regular points on $Y_H(p)$, by Lemma 2.1.1 we only need to consider geometric points. By Theorem 3.1.1, we need a criterion for detecting when a finite group acting on a regular local ring has regular subring of invariants. The criterion is provided by Serre's Theorem 2.3.9 and leads to:

Theorem 4.1.1. *A geometric point $z=(E_z,P_z)\in Y_1(p)$ has non-regular image in $Y_H(p)$ if and only if it is a point in the closed fiber such that $|\mathrm{Aut}(E_z)|>2$, $P_z=0$, and $2|H|\nmid|\mathrm{Aut}(E_z)|$.*

In particular, when $p>3$ there are at most two non-regular points on $Y_H(p)$ and such points are $\mathbf{F}_p$-rational, while for $p\le 3$ (so H is trivial) the unique ($\mathbf{F}_p$-rational) supersingular point is the unique non-regular point.

Proof. Let k be an algebraically closed field of characteristic p and define $W=W(k)$; we may assume that z is a k-rational point. By Lemma 2.1.1, we may consider the situation after base change by $\mathbf{Z}_{(p)}\to W$. A non-regular point z must be a closed point on the closed fiber. Let z' be a point over z in $Y_1(p;\mathcal{P})(k)$. Let (E_z,P_z) be the structure arising from z.

First suppose $p>3$ and H is trivial. The group $\mathrm{Aut}_k(E_z)$ is cyclic of order prime to p, so the automorphism group $\mathrm{Aut}(z)$ of the $\Gamma_1(p)$-structure underlying z is also cyclic of order prime to p. By Theorems 3.1.1 and 2.3.9, the regularity of $\widehat{\mathcal{O}}_{Y_1(p)_W,z}$ is therefore equivalent to the existence of a stable line under the action of $\mathrm{Aut}(z)$ on the 2-dimensional cotangent space to the regular universal deformation ring $\mathcal{R}_z=\widehat{\mathcal{O}}_{Y_1(p;\mathcal{P})_W,z'}$ of the $\Gamma_1(p)$-structure z.

When the $\Gamma_1(p)$-structure z is étale (*i.e.*, $P_z\neq 0$), then the formal deformation theory for z is the same as for the underlying elliptic curve $E_z/\langle P_z\rangle$, whence the universal deformation ring is isomorphic to $W[\![t]\!]$. In such cases, p spans an $\mathrm{Aut}(z)$-invariant line in the cotangent space of the deformation ring. Even when H is not assumed to be trivial, this line is stable under the action of the stabilizer of z' the preimage of H in $(\mathbf{Z}/p\mathbf{Z})^\times$). Hence, we get regularity at z for any H when $p > 3$ and $P_z \neq 0$.

Still assuming $p > 3$, now drop the assumption of triviality on H but suppose that the $\Gamma_1(p)$-structure is not étale, so $z = (E_z, 0)$ and $\mathrm{Aut}(z) = \mathrm{Aut}_k(E_z)$. The preimage $H' \subseteq (\mathbf{Z}/p\mathbf{Z})^\times$ of H acts on the deformation ring $\mathcal{R}_z$ since $P_z = 0$. By Theorem 3.1.1 and Theorem 3.2.2, the cotangent space to $\mathcal{R}_z$ is canonically isomorphic to

$$\mathrm{Cot}_0(E_z) \oplus \mathrm{Cot}_0(E_z)^{\otimes 2}, \tag{4.1.1}$$

where this decomposition corresponds to the lines spanned by the images of x and t respectively. Conceptually, the first line in (4.1.1) arises from equicharacterisitc deformations of the point of order p on constant deformations of the elliptic curve E_z, and the second line arises from deformations of the elliptic curve without deforming the vanishing level structure P_z. These identifications are compatible with the natural actions of $\mathrm{Aut}(z) = \mathrm{Aut}(E_z)$.

Since $p > 3$, the action of $\mathrm{Aut}(E_z) = \mathrm{Aut}(z)$ on the line $\mathrm{Cot}_0(E_z)$ is given by a faithful (non-trivial) character $\overline{\chi}_{\mathrm{id}}$, and the other line in (4.1.1) is acted upon by $\mathrm{Aut}(E_z)$ via the character χ_{id}^2. The resulting representation of $\mathrm{Aut}(z)$ on $\mathrm{Cot}_0(E_z)^{\otimes 2}$ is trivial if and only if $\overline{\chi}_{\mathrm{id}}^2 = 1$, which is to say (by faithfulness) that $\mathrm{Aut}(E_z)$ has order 2 (*i.e.*, $j(E_z) \neq 0, 1728$). Since the H'-action is trivial on the line $\mathrm{Cot}_0(E_z)^{\otimes 2}$ (due to H' only acting on the level structure) and we are passing to invariants by the action of the group $H' \times \mathrm{Aut}(E_{z/k})$, by Serre's theorem we get regularity without restriction on H when $j(E_z) \neq 0, 1728$.

If $j(E_z) \in \{0, 1728\}$ then $|\mathrm{Aut}(E_z)| > 2$ and the cyclic H' acts on (4.1.1) through a representation $\psi \oplus 1$ with ψ a faithful character. The cyclic $\mathrm{Aut}(z)$ acts through a representation $\chi \oplus \chi^2$ with χ a faithful character, so $\chi^2 \neq 1$. The commutative group of actions on (4.1.1) generated by H' and $\mathrm{Aut}(z)$ is generated by pseudo-reflections if and only if the action of the cyclic $\mathrm{Aut}(z)$ on the first line is induced by the action of a subgroup of H'. That is, the order of χ must divide the order of ψ, or equivalently $|\mathrm{Aut}(z)|$ must divide $|H'| = 2|H|$. This yields exactly the desired conditions for non-regularity when $p > 3$.

Now suppose $p \leq 3$, so H is trivial. If $\mathrm{Aut}(E_{z/k}) = \{\pm 1\}$, so z is an ordinary point, then for $p = 3$ we can use the preceding argument to deduce regularity at z. Meanwhile, for $p = 2$ we see that $\mathcal{R}_z$ is formally smooth by Theorem 3.3.3, so the subring of invariants at z is formally smooth (by [34, p. 508]). It remains to check non-regularity at the unique (supersingular) point $z \in Y_1(p)_{/k}$ with $j = 0 = 1728$ in k.

By Serre's theorem, it suffices to check that the action of $\mathrm{Aut}(z) = \mathrm{Aut}(E_z)$ on (4.1.1) is not generated by pseudo-reflections, where E_z is the unique supersingular elliptic curve over k (up to isomorphism). The action of $\mathrm{Aut}(E_z)$ is through 1-dimensional characters, so the p-Sylow subgroup must act trivially. In both cases ($p = 2$ or 3) the group $\mathrm{Aut}(E_z)$ has order divisible by only two primes p and p', with the p'-Sylow of order > 2. This p'-Sylow must act through a faithful character on $\mathrm{Cot}_0(E_z)$ (use [20, Lemma 3.3] or [68, Lemma 2.16]), and hence this group also acts non-trivially on $\mathrm{Cot}_0(E_z)^{\otimes 2}$. It follows that this action is not generated by pseudo-reflections.

□

4.2 Regularity along the cusps

Now we check that $X_H(p)$ is regular along the cusps, so we can focus our attention on $Y_H(p)$ when computing the minimal regular resolution of $X_H(p)$. We will again use deformation theory, but now in the case of generalized elliptic curves. Throughout this section, p is an arbitrary prime.

Recall that a *generalized elliptic curve* over a scheme S is a proper flat map $\pi : E \to S$ of finite presentation equipped with a section $e : S \to E^{\mathrm{sm}}$ into the relative smooth locus and a map

$$+ : E^{\mathrm{sm}} \times_S E \to E$$

such that

- the geometric fibers of π are smooth genus 1 curves or Néron polygons;
- $+$ restricts to a commutative group scheme structure on E^{sm} with identity section e;
- $+$ is an action of E^{sm} on E such that on singular geometric fibers with at least two "sides", the translation action by each rational point in the smooth locus induces a rotation on the graph of irreducible components.

Since the much of the basic theory of Drinfeld structures was developed in [34, Ch. 1] for arbitrary smooth separated commutative group schemes of relative dimension 1, it can be applied (with minor changes in proofs) to the smooth locus of a generalized elliptic curve. In this way, one can merge the "affine" moduli-theoretic $\mathbf{Z}$-theory in [34] with the "proper" moduli-theoretic $\mathbf{Z}[1/N]$-theory in [15]. We refer the reader to [21] for further details on this synthesis.

The main deformation-theoretic fact we need is an analogue of Theorem 3.2.1:

Theorem 4.2.1. *An irreducible generalized elliptic curve C_1 over a perfect field k of characteristic $p > 0$ admits a universal deformation ring that is abstractly isomorphic to $W[\![t]\!]$, and the equicharacteristic cotangent space of this deformation ring is canonically isomorphic to* $\mathrm{Cot}_0(C_1^{\mathrm{sm}})^{\otimes 2}$.

Proof. The existence and abstract structure of the deformation ring are special cases of [15, III, 1.2]. To describe the cotangent space intrinsically, we wish to put ourselves in the context of deformation theory of proper flat curves. Infinitesimal deformations of C_1 admit a unique generalized elliptic curve structure once we fix the identity section [15, II, 2.7], and any two choices of identity section are uniquely related by a translation action. Thus, the deformation theory for C_1 as a generalized elliptic (*i.e.*, marked) curve coincides with its deformation theory as a flat (unmarked) curve. In particular, the tangent space to this deformation functor is canonically identified with $\mathrm{Ext}^1_{C_1}(\Omega^1_{C_1/k}, \mathcal{O}_{C_1})$ [56, §4.1.1].

Since the natural map $\Omega^1_{C_1/k} \to \omega_{C_1/k}$ to the invertible relative dualizing sheaf is injective with finite-length cokernel (supported at the singularity),

$$\mathrm{Ext}^1_{C_1}(\omega_{C_1/k}, \mathcal{O}_{C_1}) \simeq \mathrm{Ext}^1_{C_1}(\omega^{\otimes 2}_{C_1/k}, \omega_{C_1/k}) \simeq \mathrm{H}^0(C_1, \omega^{\otimes 2}_{C_1/k})^\vee,$$

with the final isomorphism provided by Grothendieck duality. Thus, the cotangent space to the deformation functor is identified with $\mathrm{H}^0(C_1, \omega^{\otimes 2}_{C_1/k})$. Since $\omega_{C_1/k}$ is (non-canonically) trivial, just as for elliptic curves, we get a canonical isomorphism

$$\mathrm{H}^0(C_1, \omega^{\otimes 2}_{C_1/k}) \simeq \mathrm{H}^0(C_1, \omega_{C_1/k})^{\otimes 2} \simeq \mathrm{Cot}_0(C_1^{\mathrm{sm}})^{\otimes 2}$$

(the final isomorphism defined via pullback along the identity section).

□

Definition 4.2.2. A $\Gamma_1(N)$*-structure* on a generalized elliptic curve $E \to S$ is an "S-ample" Drinfeld $\mathbf{Z}/N\mathbf{Z}$-structure on E^{sm}; *i.e.*, a section $P \in E^{\mathrm{sm}}[N](S)$ such that the relative effective Cartier divisor

$$D = \sum_{j \in \mathbf{Z}/N\mathbf{Z}} [jP]$$

in E^{sm} is a subgroup scheme which meets all irreducible components of all geometric fibers.

If $E_{/S}$ admits a $\Gamma_1(N)$-structure, then the non-smooth geometric fibers must be d-gons for various $d|N$. In case $N = p$ is prime, this leaves p-gons and 1-gons as the only options. The importance of Definition 4.2.2 is the following analogue of Theorem 3.1.1:

Theorem 4.2.3. *Let k be an algebraically closed field of characteristic $p > 0$, and $W = W(k)$. The points of $X_1(p)_{/k} - Y_1(p)_{/k}$ correspond to isomorphism classes of $\Gamma_1(p)$-structures on degenerate generalized elliptic curves over k with 1 or p sides.*

For $z \in X_1(p)_{/k} - Y_1(p)_{/k}$, there exists a universal deformation ring $\mathcal{S}_z$ for the $\Gamma_1(p)$-structure z, and $\widehat{\mathcal{O}}_{X_1(p)_W, z}$ is the subring of $\mathrm{Aut}(z)$*-invariants in $\mathcal{S}_z$.*

Proof. In general, $\Gamma_1(p)$-structures on generalized elliptic curves form a proper flat Deligne-Mumford stack $\overline{M}_{\Gamma_1(p)}$ over $\mathbf{Z}_{(p)}$ of relative dimension 1, and this stack is smooth over $\mathbf{Q}$ and is normal (as one checks via abstract deformation theory). For our purposes, the important point is that if we choose an odd prime $\ell \neq p$ then we can define an evident $[\Gamma_1(p), \Gamma(\ell)]$-variant on Definition 4.2.2 (imposing an ampleness condition on the combined level structure), and the open locus of points with trivial geometric automorphism group is a scheme (as it is an algebraic space quasi-finite over the j-line). This locus fills up the entire stack $\overline{M}_{[\Gamma_1(p),\Gamma(\ell)]}$ over $\mathbf{Z}_{(p)}$, so this stack is a scheme.

The resulting normal $\mathbf{Z}_{(p)}$-flat proper scheme $\overline{M}_{[\Gamma_1(p),\Gamma(\ell)]}$ is finite over the j-line, whence it must *coincide* with the scheme $X_1(p;[\Gamma(\ell)])$ as constructed in [34] by the *ad hoc* method of normalization of the fine moduli scheme $Y_1(p;[\Gamma(\ell)])$ over the j-line. We therefore get a map

$$\overline{M}_{[\Gamma_1(p),\Gamma(\ell)]} = X_1(p;[\Gamma(\ell)]) \to X_1(p)$$

that *must* be the quotient by the natural $\mathrm{GL}_2(\mathbf{F}_\ell)$-action on the source. Since complete local rings at geometric points on a Deligne-Mumford stack coincide with universal formal deformation rings, we may conclude as in the proof of Theorem 3.1.1.

□

We are now in position to argue just as in the elliptic curve case: we shall work out the deformation rings in the various possible cases and for $p \neq 2$ we will use Serre's pseudo-reflection theorem to deduce regularity of $X_1(p)$ along the cusps on the closed fiber. A variant on the argument will also take care of $p = 2$.

As in the elliptic curve case, it will suffice to consider geometric points. Thus, there will be two types of $\Gamma_1(p)$-structures (E, P) to deform: E is either a p-gon or a 1-gon.

Lemma 4.2.4. *Let E_0 be a p-gon over an algebraically closed field k of characteristic p, and $P_0 \in E_0^{\mathrm{sm}}(k)$ a $\Gamma_1(p)$-structure. The deformation theory of (E_0, P_0) coincides with the deformation theory of the 1-gon generalized elliptic curve $E_0/\langle P_0\rangle$.*

Note that in the p-gon case, the point $P_0 \in E_0^{\mathrm{sm}}(k)$ generates the order-p constant component group of E_0^{sm}, so the group scheme $\langle P_0\rangle$ generated by P_0 is visibly étale and the quotient $E_0/\langle P_0\rangle$ makes sense (as a generalized elliptic curve) and is a 1-gon.

Proof. For any infinitesimal deformation (E, P) of (E_0, P_0), the subgroup scheme H generated by P is finite étale, and it makes sense to form the quotient E/H as a generalized elliptic curve deformation of the 1-gon E_0/H_0 (with $H_0 = \langle P_0\rangle$). Since any finite étale cover of a generalized elliptic curve admits a unique compatible generalized elliptic curve structure once we fix a lift of the identity section and demand geometric connectedness of fibers over the base [15, II, 1.17], we see that the deformation theory of (E_0, H_0) (ignoring P) is equivalent to the deformation theory of the 1-gon E_0/H_0. The deformation theory of a 1-gon is formally smooth of relative dimension 1 [15, III, 1.2], and upon specifying (E, H) deforming (E_0, H_0) the étaleness of H ensures the existence and uniqueness of the choice of $\Gamma_1(p)$-structure P generating H such that

P lifts P_0 on E_0. That is, the universal deformation ring for (E_0, P_0) coincides with that of E_0/H_0.

□

In the 1-gon case, there is only one (geometric) possibility up to isomorphism: the pair $(C_1, 0)$ where C_1 is the standard 1-gon (over an algebraically closed field k of characteristic p). For this, we have an analogue of (4.1.1):

LEMMA 4.2.5. *The universal deformation ring of the $\Gamma_1(p)$-structure $(C_1, 0)$ is isomorphic to the regular local ring $W[\![t]\!][\![X]\!]/\Phi_p(X+1)$, with cotangent space canonically isomorphic to*

$$\mathrm{Cot}_0(C_1^{\mathrm{sm}}) \oplus \mathrm{Cot}_0(C_1^{\mathrm{sm}})^{\otimes 2}.$$

Proof. Since the p-torsion on C_1^{sm} is isomorphic to μ_p, upon fixing an isomorphism $C_1^{\mathrm{sm}}[p] \simeq \mu_p$ there is a unique compatible isomorphism $C^{\mathrm{sm}}[p] \simeq \mu_p$ for any infinitesimal deformation C of C_1. Thus, the deformation problem is that of endowing a $\mathbf{Z}/p\mathbf{Z}$-generator to the μ_p inside of deformations of C_1 (as a generalized elliptic curve). By Theorem 4.2.3, this is the scheme of generators of μ_p over the universal deformation ring $W[\![t]\!]$ of C_1.

The scheme of generators of μ_p over $\mathbf{Z}$ is $\mathbf{Z}[Y]/\Phi_p(Y)$, so we obtain $W[\![t]\!][Y]/\Phi_p(Y)$ as the desired (regular) deformation ring. Now just set $X = Y - 1$. The description of the cotangent space follows from Theorem 4.2.1.

□

Since C_1 has automorphism group (as a generalized elliptic curve) generated by the unique extension $[-1]$ of inversion from C_1^{sm} to all of C_1, we conclude that $\mathrm{Aut}(C_1, 0)$ is generated by $[-1]$. This puts us in position to carry over our earlier elliptic-curve arguments to prove:

THEOREM 4.2.6. *The scheme $X_H(p)$ is regular along its cusps.*

Proof. As usual, we may work after making a base change by $W = W(k)$ for an algebraically closed field k of characteristic $p > 0$. Let $z \in X_1(p)_{/k}$ be a cusp whose image z_H in $X_H(p)_{/k}$ we wish to study. Let H' be the preimage of H in $(\mathbf{Z}/p\mathbf{Z})^\times$, and let H'_z be the maximal subgroup of H' that acts on the deformation space for z (*e.g.*, $H'_z = H'$ if the level structure P_z vanishes). By Theorem 4.2.3, the ring $\widehat{\mathcal{O}}_{X_H(p), z_H}$ is the subring of invariants under the action of $\mathrm{Aut}(z) \times H'_z$ on the formal deformation ring for z. By Theorem 4.2.1 and Lemma 4.2.4 (as well as [34, p. 508]), this deformation ring is regular (even formally smooth) in the p-gon case. In the 1-gon case, Lemma 4.2.5 ensures that the deformation ring is regular (and even formally smooth when $p = 2$). Thus, for $p \neq 2$ we may use Theorem 2.3.9 to reduce the problem for $p \neq 2$ to checking that the action of $\mathrm{Aut}(z) \times H'_z$ on the 2-dimensional cotangent space to the deformation functor has an invariant line.

In the p-gon case, the deformation ring is $W[\![t]\!]$ and the cotangent line spanned by p is invariant. In the 1-gon case, Lemma 4.2.5 provides a functorial description of the cotangent space to the deformation functor and from this it is clear that the involution $[-1]$ acts with an invariant line $\mathrm{Cot}_0(z)^{\otimes 2}$ when $p \neq 2$ and that H'_z also acts trivially on this line.

To take care of $p = 2$ (for which H is trivial), we just have to check that any non-trivial W-algebra involution ι of $W[\![T]\!]$ has regular subring of invariants. In fact, for $T' = T\iota(T)$ the subring of invariants is $W[\![T']\!]$ by [34, p. 508].

□

5 The Minimal resolution

We now are ready to compute the minimal regular resolution $X_H(p)^{\mathrm{reg}}$ of $X_H(p)$. Since $X_H(p)_{/\mathbf{Q}}$ is a projective line when $p \leq 3$, both Theorem 1.1.2 and Theorem 1.1.6 are trivial for $p \leq 3$. Thus, from now on we assume $p > 3$. We have found all of the non-regular points (Theorem 4.1.1): the $\mathbf{F}_p$-rational points of $(1,0)$-type such that $j \in \{0, 1728\}$, provided that $|H|$ is not divisible by 3 (resp. 2) when $j = 0$ (resp. $j = 1728$). Theorem 3.3.3 provides the necessary local description to carry out Jung–Hirzebruch resolution at these points. These are tame cyclic quotient singularities (since $p > 3$). Moreover, the closed fiber of $X_H(p)$ is a nil-semistable curve that consists of two irreducible components that are geometrically irreducible, as one sees by considering the (1,0)-cyclic and (0,1)-cyclic components.

5.1 General considerations

There are four cases, depending on $p \equiv \pm 1, \pm 5 \bmod 12$ as this determines the behavior of the j-invariants 0 and 1728 in characteristic p (*i.e.*, supersingular or ordinary). This dichotomy between ordinary and supersingular cases corresponds to Jung–Hirzebruch resolution with either one or two analytic branches.

Pick a point $z = (E, 0) \in X_1(p)(\mathbf{F}_p)$ with $j = 0$ or 1728 corresponding an elliptic curve E over $\overline{\mathbf{F}}_p$ with automorphism group of order > 2. Let $z_H \in X_H(p)(\mathbf{F}_p)$ be the image of z. By Theorem 4.1.1, we know that z_H is non-regular if and only if $|H|$ is odd for $j(E) = 1728$, and if and only if $|H|$ is not divisible by 3 for $j(E) = 0$.

There is a single irreducible component through z_H in the ordinary case (arising from either (3.3.2) or (3.3.3)), while there are two such (transverse) components in the supersingular case, and to compute the generic multiplicities of these components in $X_H(p)_{/\overline{\mathbf{F}}_p}$ we may work with completions because the irreducible components through z_H are analytically irreducible (even smooth) at z_H.

Let C' and C denote the irreducible components of $X_H(p)_{/\overline{\mathbf{F}}_p}$, with C' corresponding to étale level p-structures. Since the preimage of H in $(\mathbf{Z}/p\mathbf{Z})^\times$ (of order $2|H|$) acts generically freely (resp. trivially) on the preimage of C' (resp. of C) in a fine moduli scheme over $X_H(p)_{/\overline{\mathbf{F}}_p}$ obtained by adjoining some prime-to-p level structure, ramification theory considerations and Theorem 3.3.3 show that the components C' and C in $X_H(p)_{/\overline{\mathbf{F}}_p}$ have respective multiplicities of 1 and $(p-1)/2|H| = [(\mathbf{Z}/p\mathbf{Z})^\times/\{\pm1\} : H]$. Moreover, by Theorem 3.3.3 we see that z_H lies on C when it is an ordinary point.

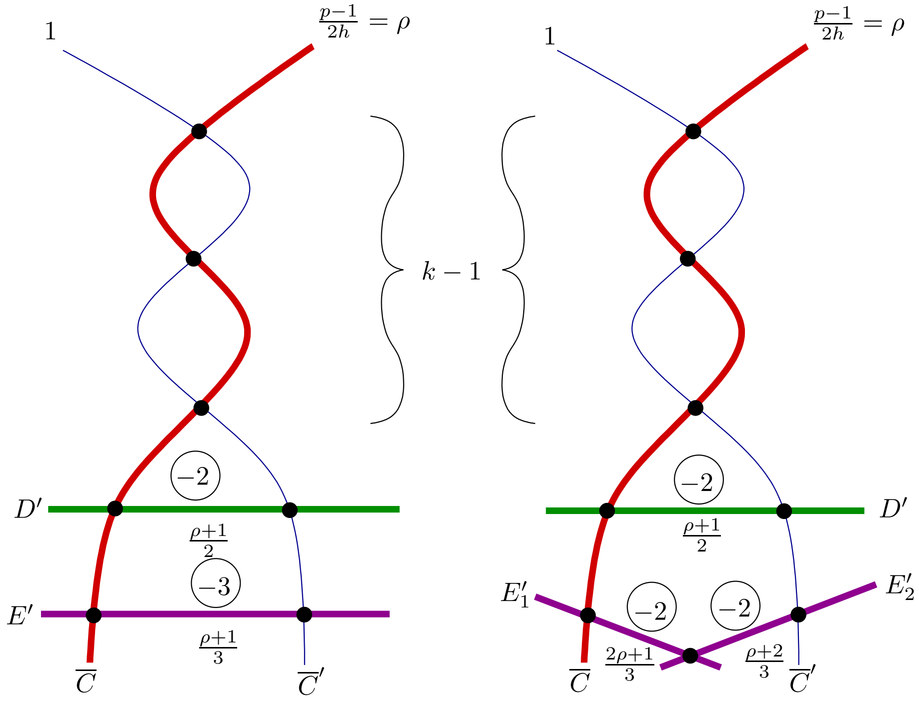
5.2 The case $p \equiv -1 \bmod 12$

We are now ready to resolve the singularities on $X_H(p)_{/W}$ with $W = W(\overline{\mathbf{F}}_p)$. We will first carry out the calculation in the case $p \equiv -1 \pmod{12}$, so 0 and 1728 are supersingular j-values. In this case $(p-1)/2$ is not divisible by 2 or 3, so $|H|$ is automatically not divisible by 2 or 3 (so we have two non-regular points).

Write $p = 12k-1$ with $k \geq 1$. By the Deuring Mass Formula [34, Cor. 12.4.6] the components C and C' meet in $(p-11)/12 = k-1$ geometric points away from the two supersingular points with $j = 0, 1728$. Consider one of the two non-regular supersingular points z_H. The complete local ring at z_H on $X_H(p)_W$ is the subring of invariants for the commuting actions of $\mathrm{Aut}(z)$ and the preimage $H' \subseteq (\mathbf{Z}/p\mathbf{Z})^\times$ of H on the universal deformation ring $\mathcal{R}_z$ of the $\Gamma_1(p)$-structure z. Note that the actions of H' and $\mathrm{Aut}(z)$ on $\mathcal{R}_z$ have a common involution. The action of H' on the tangent space fixes one line and acting through a faithful character on the other line (see the proof of Theorem 4.1.1), so by Serre's Theorem 2.3.9 the subring of H'-invariants in $\mathcal{R}_z$ is regular. By Lemma 2.3.5 and the subsequent discussion there, the subring of H'-invariant has the form $W[\![x', t']\!]/(x'^{(p-1)/|H'|}t' - p)$ with $\mathrm{Aut}(z)/\{\pm1\}$ acting on the tangent space via $\chi^{|H|} \oplus \chi$ for a faithful character χ of $\mathrm{Aut}(z)/\{\pm1\}$. Let $h = |H|$, so $\rho := (p-1)/2h$ is the multiplicity of C in $X_H(p)_{/\overline{\mathbf{F}_p}}$.

When $j(z_H) = 1728$ the character χ is quadratic, so we apply Theorem 2.4.1 and Corollary 2.4.3 with $n = 2$, $r = 1$, $m'_1 = 1$, $m'_2 = \rho$. The resolution has a single exceptional fiber D' that is transverse to the strict transforms $\overline{C}$ and $\overline{C}'$, and D' has self-intersection -2 and multiplicity $(m'_1 + m'_2)/2 = (\rho+1)/2$. When $j(z_H) = 0$ the character χ is cubic, so we apply Theorem 2.4.1 with $n = 3$, $m'_1 = 1$, $m'_2 = \rho$, and $r = h \bmod 3$. That is, $r = 1$ when $h \equiv 1 \bmod 6$ and $r = 2$ when $h \equiv -1 \bmod 6$. In the case $r = 1$ we get a single exceptional fiber E' in the resolution, transverse to $\overline{C}$ and $\overline{C}'$ with self-intersection -3 and multiplicity $(\rho+1)/3$ (by Corollary 2.4.3). This is illustrated in Figure 2(a). In the case $r = 2$ we use the continued fraction $3/2 = 2 - 1/2$ to see that the resolution of z_H has exceptional fiber with two components E'_1 and E'_2, and these have self-intersection -2 and transverse intersections as shown in Figure 2(b) with respective multiplicities $(2\rho+1)/3$ and $(\rho+2)/3$ by Corollary 2.4.3. This completes the computation of the minimal regular resolution $X_H(p)'$ of $X_H(p)$ when $p \equiv -1 \bmod 12$.

To compute the intersection matrix for the closed fiber of $X_H(p)'$, we need to compute some more intersection numbers. For $h \equiv 1 \bmod 6$ we let μ and ν denote the multiplicities of D' and E' in $X_H(p)'$, and for $h \equiv -1 \bmod 6$ we



(a) $h \equiv 1 \bmod 6$

(b) $h \equiv -1 \bmod 6$

Figure 2: Minimal regular resolution $X_H(p)'$ of $X_H(p)$, $p = 12k-1$, $k \geq 1$, $h = |H|$

define μ in the same way and let ν_j denote the multiplicity of E'_j in $X_H(p)'$. In other words,

$$\mu = (\rho+1)/2,\ \nu = (\rho+1)/3,\ \nu_1 = (2\rho+1)/3,\ \nu_2 = (\rho+2)/3.$$

Thus,

$$\overline{C}' + \rho\overline{C} + \mu D' + \nu E' \equiv 0, \tag{5.2.1}$$

so if we intersect (5.2.1) with $\overline{C}$ and use the identities

$$\rho = (6k-1)/h,\ \ \overline{C}'.\overline{C} = k-1 = (h\rho-5)/6,$$

we get

$$\overline{C}.\overline{C} = -1-(h-\varepsilon)/6$$

where $\varepsilon = \pm 1 \equiv h \bmod 6$. In particular, $\overline{C}.\overline{C} < -1$ unless $h=1$ (*i.e.*, unless H is trivial). We can also compute the self-intersection for $\overline{C}'$, but we do not need it.

When H is trivial, so $\overline{C}$ is a -1-curve, we can contract $\overline{C}$ and then by Theorem 2.1.2 and Figure 2 the self-intersection numbers for the components D' and E' drop to -1 and -2 respectively. Then we may contract D', so E' becomes a -1-curve, and finally we end with a single irreducible component (coming from $\overline{C}'$). This proves Theorem 1.1.2 when $p \equiv -1 \bmod 12$.

Returning to the case of general H, let us prove Theorem 1.1.6 for $p \equiv -1 \bmod 12$. Since $\overline{C}'$ has multiplicity 1 in the closed fiber of $X_H(p)'$, we can use the following special case of a result of Lorenzini [9, 9.6/4]:

Lemma 5.2.1 (Lorenzini). *Let X be a regular proper flat curve over a complete discrete valuation ring R with algebraically closed residue field and fraction field K. Assume that $X_{/K}$ is smooth and geometrically connected. Let $X_1, \dots, X_m$ be the irreducible components of the closed fiber $\overline{X}$ and assume that some component X_{i_0} occurs with multiplicity 1 in the closed fiber divisor.*

The component group of the Néron model of the Jacobian $\mathrm{Pic}^0_{X_K/K}$ has order equal to the absolute value of the $(m-1)\times(m-1)$ minor of the intersection matrix $(X_i.X_j)$ obtained by deleting the i_0th row and column.

The intersection submatrices formed by the ordered set $\{\overline{C}, D', E'\}$ for $h \equiv 1 \bmod 6$ and by $\{\overline{C}, D', E'_1, E'_2\}$ for $h \equiv -1 \bmod 6$ are given in Figure 3. The absolute value of the determinant is h in each case, so by Lemma 5.2.1 the order of the component group $\Phi(\mathcal{J}_H(p)_{/\mathbf{F}_p})$ is $h = |H| = |H|/\gcd(|H|, 6)$.

To establish Theorem 1.1.6 for $p \equiv -1 \bmod 12$, it remains to show that the natural Picard map $J_0(p) \to J_H(p)$ induces a surjection on mod-p geometric component groups. We outline a method that works for general p but that we will (for now) carry out only for $p \equiv -1 \bmod 12$, as we have only computed the intersection matrix in this case.

$$\begin{array}{cccc} & \overline{C} & D' & E' \\ \overline{C} & -1-\frac{(h-1)}{6} & 1 & 1 \\ D' & 1 & -2 & 0 \\ E' & 1 & 0 & -3 \end{array}$$

(a) $h \equiv 1 \bmod 6$

$$\begin{array}{ccccc} & \overline{C} & D' & E'_1 & E'_2 \\ \overline{C} & -1-\frac{(h+1)}{6} & 1 & 1 & 0 \\ D' & 1 & -2 & 0 & 0 \\ E'_1 & 1 & 0 & -2 & 1 \\ E'_2 & 0 & 0 & 1 & -2 \end{array}$$

(b) $h \equiv -1 \bmod 6$

Figure 3: Submatrices of intersection matrix for $X_H(p)'$, $p \equiv -1 \bmod 12$

The component group for $J_0(p)$ is generated by $(0)-(\infty)$, where (0) classifies the 1-gon with standard subgroup $\mu_p \hookrightarrow \mathbf{G}_m$ in the smooth locus, and (∞) classifies the p-gon with subgroup $\mathbf{Z}/p\mathbf{Z} \hookrightarrow (\mathbf{Z}/p\mathbf{Z}) \times \mathbf{G}_m$ in the smooth locus. The generic-fiber Picard map induced by the coarse moduli scheme map

$$X_H(p)_{/\mathbf{Z}_{(p)}} \to X_0(p)_{/\mathbf{Z}_{(p)}}$$

pulls $(0)-(\infty)$ back to a divisor

$$P \ - \sum_{j=1}^{(p-1)/2|H|} P'_i \tag{5.2.2}$$

where the P'_i's are $\mathbf{Q}$-rational points whose (cuspidal) reduction lies in the component $\overline{C}'$ classifying étale level-structures and P is a point with residue field $(\mathbf{Q}(\zeta_p)^+)^H$ whose (cuspidal) reduction lies in the component $\overline{C}$ classifying multiplicative level-structures. This description is seen by using the moduli interpretation of cusps (*i.e.*, Néron polygons) and keeping track of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-actions, and it is valid for any prime p (*e.g.*, the $\Gamma_1(p)$-structures on the standard 1-gon consistute a principal homogenous space for the action of $\mathrm{Gal}(\mathbf{Q}(\mu_p)/\mathbf{Q})$, so they give a single closed point P on $X_H(p)_{/\mathbf{Q}}$ with residue field $(\mathbf{Q}(\zeta_p)^+)^H$).

To apply (5.2.2), we need to recall some general facts (see [9, 9.5/9, 9.6/1]) concerning the relationship between the closed fiber of a regular proper model X of a smooth geometrically connected curve X_η and the component group Φ of (the Néron model of) the Jacobian of X_η, with the base equal to the spectrum of a discrete valuation ring R with algebraically closed residue field. If $\{X_i\}_{i\in I}$ is the set of irreducible components in the closed fiber of X, then we can form a complex

$$\mathbf{Z}^I \xrightarrow{\alpha} \mathbf{Z}^I \xrightarrow{\beta} \mathbf{Z}$$

where $\mathbf{Z}^I$ is the free group on the X_i's, the map α is defined by the intersection matrix $(X_i.X_j)$, and β sends each standard basis vector to the multiplicity of the corresponding component in the closed fiber. The cokernel $\ker(\beta)/\mathrm{im}(\alpha)$

is naturally identified with the component group Φ via the map $\mathrm{Pic}(X) \to \mathbf{Z}^I$ that assigns to each invertible sheaf $\mathcal{L}$ its tuple of partial degrees $\deg_{X_i}(\mathcal{L})$.

By using [9, 9.1/5] to compute such line-bundle degrees, one finds that the Néron-model integral point associated to the pullback divisor in (5.2.2) has reduction whose image in $\Phi(\mathcal{J}_H(p)_{/\overline{\mathbf{F}}_p})$ is represented by

$$\frac{[\mathbf{Q}(P):\mathbf{Q}]}{\mathrm{mult}(\overline{C})}\cdot\overline{C} - \sum_{i=1}^{(p-1)/2|H|}\overline{C}' = \overline{C} - \frac{p-1}{2|H|}\cdot\overline{C}' \tag{5.2.3}$$

when this component group is computed by using the regular model $X_H(p)'$ that we have found for $p \equiv -1 \bmod 12$ (the same calculation will work for all other p's, as we shall see).

The important property emerging from this calculation is that one of the coefficients in (5.2.3) is ±1, so an element in $\ker(\beta)$ that is a $\mathbf{Z}$-linear combination of $\overline{C}$ and $\overline{C}'$ *must* be a multiple of (5.2.3) and hence is in the image of $\Phi(\mathcal{J}_0(p))$ under the Picard map. Thus, to prove that the component group for $J_0(p)$ surjects onto the component group for $J_H(p)$, it suffices to check that any element in $\ker(\beta)$ can be modified modulo $\mathrm{im}(\alpha)$ to lie in the $\mathbf{Z}$-span of $\overline{C}$ and $\overline{C}'$.

Since the matrix for α is the intersection matrix, it suffices (and is even necessary) to check that the submatrix $M_{\overline{C},\overline{C}'}$ of the intersection matrix given by the rows labelled by the irreducible components other than $\overline{C}$ and $\overline{C}'$ is a *surjective* matrix over $\mathbf{Z}$. Indeed, such surjectivity ensures that we can always subtract a suitable element of $\mathrm{im}(\alpha)$ from any element of $\ker\beta$ to kill coefficients away from $\overline{C}$ and $\overline{C}'$ in a representative for an element in $\Phi \simeq \ker(\beta)/\mathrm{im}(\alpha)$. The surjectivity assertion over $\mathbf{Z}$ amounts to requiring that the matrix $M_{\overline{C},\overline{C}'}$ have top-degree minors with gcd equal to 1. It is enough to check that those minors that avoid the column coming from $\overline{C}'$ have gcd equal to 1. Thus, it is enough to check that in Figure 3 the matrix of rows beneath the top row has top-degree minors with gcd equal to 1. This is clear in both cases. In particular, this calculation (especially the analysis of (5.2.3)) yields the following result when $p \equiv -1 \bmod 12$:

Corollary 5.2.2. *Let $\rho = (p-1)/2|H|$. The degree-0 divisor $\overline{C} - \rho\overline{C}'$ represents a generator of the mod-p component group of $J_H(p)$.*

The other cases $p \equiv 1, \pm 5 \bmod 12$ will behave similarly, with Corollary 5.2.2 being true for all such p. The only differences in the arguments are that cases with $|H|$ divisible by 2 or 3 can arise and we will sometimes have to use the "one branch" version of Jung–Hirzebruch resolution to resolve non-regular ordinary points.

5.3 The case $p \equiv 1 \bmod 12$.

We have $p = 12k + 1$ with $k \geq 1$, so $(p-1)/2 = 6k$. In this case 0 and 1728 are both ordinary j-invariants, so the number of supersingular points is

$(p-1)/12 = k$ by the Deuring Mass Formula. The minimal regular resolution $X_H(p)'$ of $X_H(p)$ is illustrated in Figure 4, depending on the congruence class of $h = |H|$ modulo 6. When h is divisible by 6 there are no non-regular points, so $X_H(p)' = X_H(p)_{/W}$ is as in Figure 4(a). When h is even but not divisible by 3 there is only the non-regularity at $j = 0$ to be resolved, as shown in Figures 4(b),(c). The case of odd h is given in Figures 4(d)–(f), and these are all easy applications of Theorem 2.4.1 and Corollary 2.4.3. We illustrate by working out the case $h \equiv 5 \bmod 6$, for which there are two ordinary singularities to resolve.

Arguing much as in the case $p \equiv -1 \bmod 12$, but now with a "one branch" situation at ordinary points, the ring to be resolved is formally isomorphic to the ring of invariants in $W[\![x', t']\!]/(x'^{(p-1)/2|H|} - p)$ under an action of the cyclic $\mathrm{Aut}(z)/\{\pm 1\}$ with a tangent-space action of $\chi^{|H|} \oplus \chi$ for a faithful character χ. At a point with $j = 1728$ we have quadratic χ, $n = 2$, $r = 1$. Using the "one branch" version of Theorem 2.4.1 yields the exceptional divisor D' as illustrated in Figure 4(f), transverse to $\overline{C}$ with self-intersection -2 and multiplicity $\rho/2$. At a point with $j = 0$ we have a cubic χ, so $n = 3$. Since $h \equiv 2 \bmod 3$ when $h \equiv 5 \bmod 6$, we have $r = 2$. Since $3/2 = 2 - 1/2$, we get exceptional divisors E'_1 and E'_2 with transverse intersections as shown and self-intersections of -2. The "outer" component E'_1 has multiplicity $\rho/3$ and the "inner" component E'_2 has multiplicity $2\rho/3$. Once again we will suppress the calculation of $\overline{C}'.\overline{C}'$ since it is not needed.

We now proceed to analyze the component group for each value of $h \bmod 6$. Since $\overline{C}'$ has multiplicity 1 in the closed fiber, we can carry out the same strategy that was used for $p \equiv -1 \bmod 12$, resting on Lemma 5.2.1. When $h \equiv 0 \bmod 6$, there are only the components $\overline{C}$ and $\overline{C}'$ in the closed fiber of $X_H(p)' = X_H(p)$, with $\overline{C}.\overline{C} = -h/6$. Thus, the component group has the expected order $|H|/6$ and since there are no additional components we are done in this case.

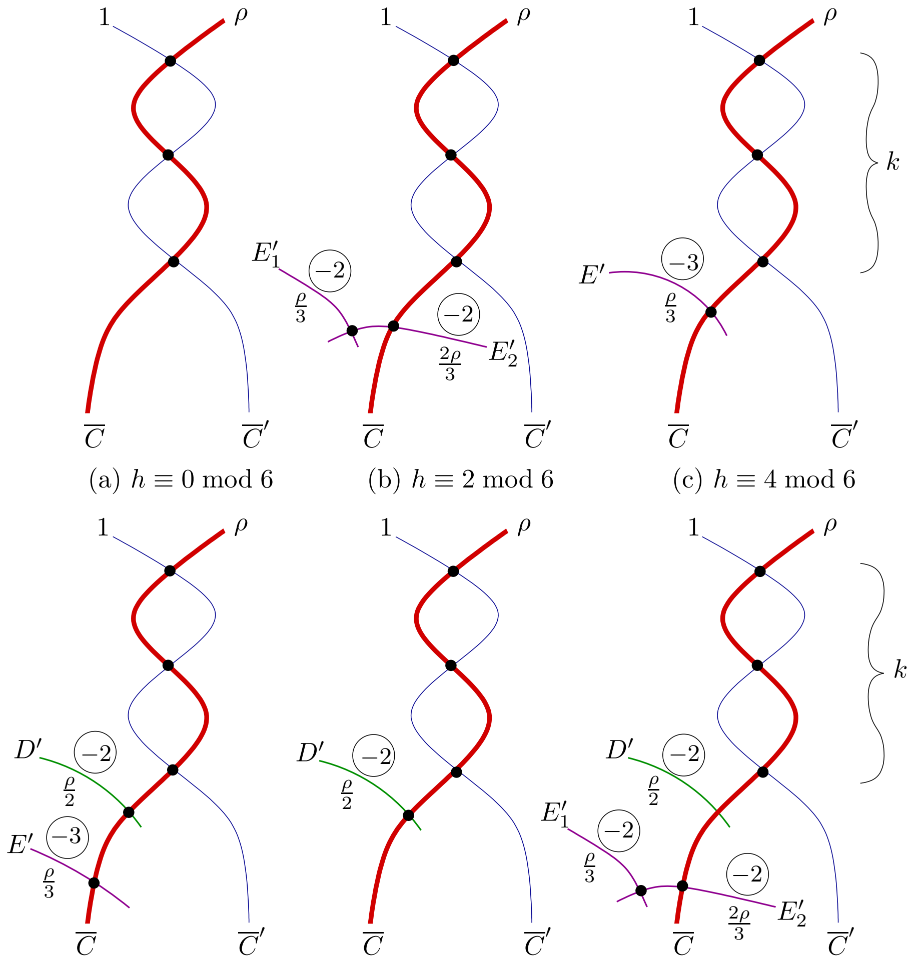
If $h \equiv 1 \bmod 6$, one finds that the submatrix of the intersection matrix corresponding to the ordered set $\{\overline{C}, D', E'\}$ is

$$\begin{pmatrix} -(h+5)/6 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 0 & -3 \end{pmatrix}$$

with absolute determinant $h = |H|/\gcd(|H|, 6)$ as desired, and the bottom two rows have 2×2 minors with gcd equal to 1. Moreover, in the special case $h = 1$ we see that $\overline{C}$ is a -1-curve, and after contracting this we contract D' and E' in turn, leaving us with only the component $\overline{C}'$. This proves Theorem 1.1.2 for $p \equiv 1 \bmod 12$.

For $h \equiv 2 \bmod 6$, the submatrix indexed by $\{\overline{C}, E'_1, E'_2\}$ is

$$\begin{pmatrix} -(h+4)/6 & 0 & 1 \\ 0 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$$



(a) $h \equiv 0 \bmod 6$

(b) $h \equiv 2 \bmod 6$

(c) $h \equiv 4 \bmod 6$

(d) $h \equiv 1 \bmod 6$

(e) $h \equiv 3 \bmod 6$

(f) $h \equiv 5 \bmod 6$

Figure 4: Minimal regular resolution $X_H(p)'$, $p = 12k + 1$, $k \geq 1$, $h = |H|$, $\rho = (p-1)/2h$

with absolute determinant $h/2 = |H|/\gcd(|H|, 6)$, and the bottom two rows have 2×2 minors with gcd equal to 1. The cases $h \equiv 3, 4 \bmod 6$ are even easier, since there are just two components to deal with, $\{\overline{C}, D'\}$ and $\{\overline{C}, E'\}$ with corresponding matrices

$$\begin{pmatrix} -(h+3)/6 & 1 \\ 1 & -2 \end{pmatrix}, \qquad \begin{pmatrix} -(h+2)/6 & 1 \\ 1 & -3 \end{pmatrix}$$

that yield the expected results.

For the final case $h \equiv -1 \bmod 6$, the submatrix indexed by the ordered set of components $\{\overline{C}, D', E'_1, E'_2\}$ is

$$\begin{pmatrix} -(h+7)/6 & 1 & 0 & 1 \\ 1 & -2 & 0 & 0 \\ 0 & 0 & -2 & 1 \\ 1 & 0 & 1 & -2 \end{pmatrix}$$

with absolute determinant $h = |H|/\gcd(|H|, 6)$ and gcd 1 for the 3×3 minors along the bottom three rows. The case $p \equiv 1 \bmod 12$ is now settled.

5.4 The cases $p \equiv \pm 5 \bmod 12$

With $p = 12k + 5$ for $k \geq 0$, we have $(p-1)/2 = 6k + 2$, so $h = |H|$ is not divisible by 3. Thus, the supersingular $j = 0$ is always non-regular and the ordinary $j = 1728$ is non-regular for even h.

Using Theorem 2.4.1 and Corollary 2.4.3, we obtain a minimal regular resolution depending on the possibilities for $h \bmod 6$ not divisible by 3, as given in Figure 5.

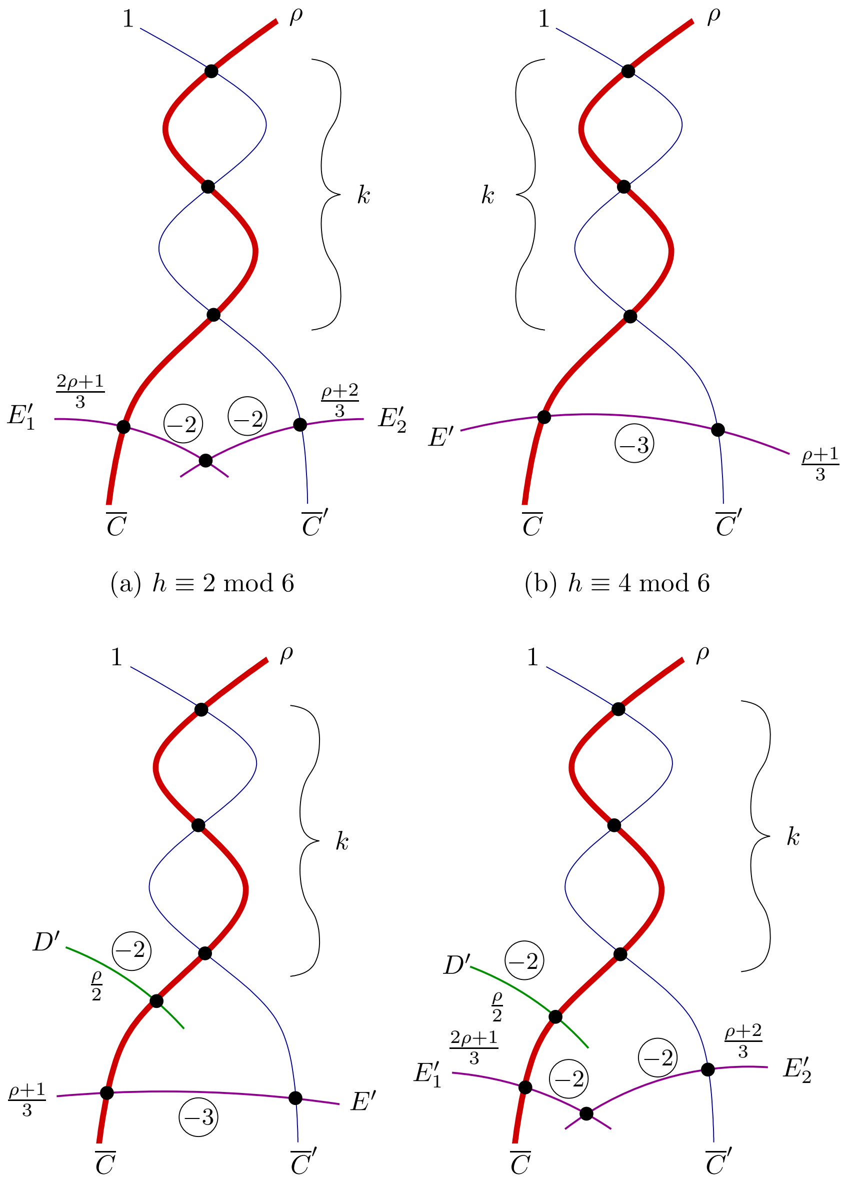
From Figure 5 one easily carries out the computations of the absolute determinant and the gcd of minors from the intersection matrix, just as we have done in earlier cases, and in all cases one gets $|H|/\gcd(|H|, 6)$ for the absolute determinant and the gcd of the relevant minors is 1. Also, the case $h = 1$ has $\overline{C}$ as a -1-curve, and successive contractions end at an integral closed fiber, so we have established Theorems 1.1.2 and 1.1.6 for the case $p \equiv 5 \bmod 12$.

When $p = 12k - 5$ with $k \geq 1$, so $(p-1)/2 = 6k - 3$ is odd, we have that $h = |H|$ is odd. Thus, $j = 1728$ does give rise to a non-regular point, but the behavior at $j = 0$ depends on $h \bmod 6$. The usual applications of Jung–Hirzebruch resolution go through, and the minimal resolution has closed-fiber diagram as in Figure 6, depending on odd $h \bmod 6$, and both Theorem 1.1.2 and Theorem 1.1.6 drop out just as in the preceding cases.

6 The Arithmetic of $J_1(p)$

Our theoretical results concerning component groups inspired us to carry out some arithmetic computations in $J_1(p)$, and this section summarizes this work.

In Section 6.1 we recall the Birch and Swinnerton-Dyer conjecture, as this motivates many of our computations, and then we describe some of the theory



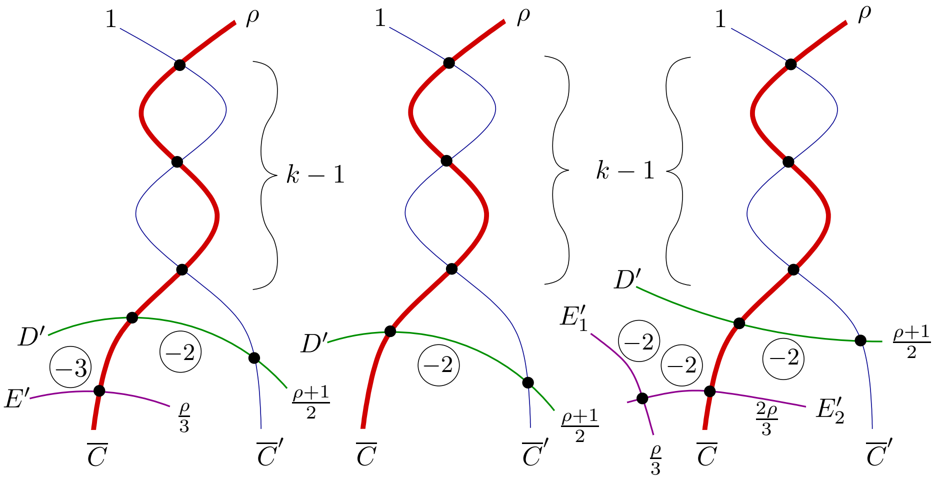
(a) $h \equiv 2 \bmod 6$

(b) $h \equiv 4 \bmod 6$

(c) $h \equiv 1 \bmod 6$

(d) $h \equiv -1 \bmod 6$

Figure 5: Minimal regular resolution $X_H(p)'$, $p = 12k + 5$, $k \geq 0$, $h = |H|$, $\rho = (p-1)/2h$



(a) $h \equiv 1 \bmod 6$ (b) $h \equiv 3 \bmod 6$ (c) $h \equiv 5 \bmod 6$

Figure 6: Minimal regular resolution $X_H(p)'$, $p = 12k-5$, $k \geq 1$, $h = |H|$, $\rho = (p-1)/2h$

behind the computations that went into computing the tables of Section 6.6. In Section 6.2 we find all p such that $J_1(p)$ has rank 0. We next discuss tables of certain arithmetic invariants of $J_1(p)$ and we give a conjectural formula for $|J_1(p)(\mathbf{Q})_{\mathrm{tor}}|$, along with some evidence. In Section 6.3 we investigate Jacobians of intermediate curves $J_H(p)$ associated to subgroups of $(\mathbf{Z}/p\mathbf{Z})^\times$, and in Section 6.4 we consider optimal quotients A_f of $J_1(p)$ attached to newforms. In Section 6.4.1 we describe the lowest-level modular abelian variety that (assuming the Birch and Swinnerton-Dyer conjecture) should have infinite Mordell-Weil group but to which the general theorems of Kato, Kolyvagin, *et al.*, do not apply.

6.1 Computational methodology

We used the third author's modular symbols package for our computations; this package is part of [10] V2.10-6. See Section 6.5 for a description of how to use Magma to compute the tables. For the general theory of computing with modular symbols, see [14] and [63].

Remark 6.1.1. Many of the results of this section assume that a Magma program running on a computer executed correctly. Magma is complicated software that runs on physical hardware that is subject to errors from both programming mistakes and physical processes, such as cosmic radiation. We thus make the running *assumption* for the rest of this section that the computations below were performed correctly. To decrease the chance of hardware errors such as the famous Pentium bug (see [17]), we computed the tables in

Section 6.6 on three separate computers with different CPU architectures (an AMD Athlon 2000MP, a Sun Fire V480 which was donated to the third author by Sun Microsystems, and an Intel Pentium 4-M laptop).

Let A be a modular abelian variety over $\mathbf{Q}$, *i.e.*, a quotient of $J_1(N)$ for some N. We will make frequent reference to the following special case of the general conjectures of Birch and Swinnerton-Dyer:

CONJECTURE 6.1.2 (BSD CONJECTURE). *Let* $Ш(A)$ *be the Shafarevich-Tate group of* A*, let* $c_p = |\Phi_{A,p}(\mathbf{F}_p)|$ *be the Tamagawa number at* p *for* A*, and let* Ω_A *be the volume of* $A(\mathbf{R})$ *with respect to a generator of the invertible sheaf of top-degree relative differentials on the Néron model* $A_{/\mathbf{Z}}$ *of* A *over* $\mathbf{Z}$*. Let* $A^\vee$ *denote the abelian variety dual of* A*. The group* $Ш(A)$ *is finite and*

$$\frac{L(A,1)}{\Omega_A} = \frac{|Ш(A)| \cdot \prod_{p|N} c_p}{|A(\mathbf{Q})| \cdot |A^\vee(\mathbf{Q})|},$$

where we interpret the right side as 0 *in case* $A(\mathbf{Q})$ *is infinite.*

Remark 6.1.3. The hypothesis that A is modular implies that $L(A,s)$ has an analytic continuation to the whole complex plane and a functional equation of a standard type. In particular, $L(A,1)$ makes sense. Also, when $L(A,1) \neq 0$, [32, Cor. 14.3] implies that $Ш(A)$ is finite.

Let $\{f_1,\ldots,f_n\}$ be a set of newforms in $S_2(\Gamma_1(N))$ that is $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-stable. Let I be the Hecke-algebra annihilator of the subspace generated by $f_1,\ldots,f_n$. *For the rest of Section* 6.1*, we assume that* $A = A_I = J_1(N)/IJ_1(N)$ *for such an* I. Note that A is an *optimal quotient* in the sense that $IJ_1(N) = \ker(J_1(N) \to A)$ is an abelian subvariety of $J_1(N)$.

6.1.1 BOUNDING THE TORSION SUBGROUP

To obtain a multiple of the order of the torsion subgroup $A(\mathbf{Q})_{\mathrm{tor}}$, we proceed as follows. For any prime $\ell \nmid N$, the algorithm of [3, §3.5] computes the characteristic polynomial $f \in \mathbf{Z}[X]$ of Frob_ℓ acting on any p-adic Tate module of A with $p \neq \ell$. To compute $|A(\mathbf{F}_\ell)|$, we observe that

$$|A(\mathbf{F}_\ell)| = \deg(\mathrm{Frob}_\ell - 1) = \det(\mathrm{Frob}_\ell - 1),$$

and this is the value of the characteristic polynomial of Frob_ℓ at 1. For any prime $\ell \nmid 2N$, the reduction map $A(\mathbf{Q})_{\mathrm{tor}} \to A(\mathbf{F}_\ell)$ is injective, so $|A(\mathbf{Q})_{\mathrm{tor}}|$ divides

$$T = \gcd\{|A(\mathbf{F}_\ell)| : \ \ell < 60 \text{ and } \ell \nmid 2N\}.$$

(If N is divisible by all primes up to 60, let $T = 0$. In all of the examples in this paper, N is prime and so $T \neq 0$.) The injectivity of reduction mod ℓ on the finite group $A(\mathbf{Q})_{\mathrm{tor}}$ for any prime $\ell \neq 2$ is well known and follows from the determination of the torsion in a formal group (see, *e.g.*, the appendix to [33] and [59, §IV.6–9]).

The cardinality $|A(\mathbf{F}_\ell)|$ does not change if A is replaced by a $\mathbf{Q}$-isogenous abelian variety B, so we do not expect in general that $|A(\mathbf{Q})_{\mathrm{tor}}| = T$. (For much more on relationships between $|A(\mathbf{Q})_{\mathrm{tor}}|$ and T, see [33, p. 499].) When we refer to an upper bound on torsion, T is the (multiplicative) upper bound that we have in mind.

The number 60 has no special significance; we had to make some choice to do computations, and in practice the sequence of partial gcd's rapidly stabilizes. For example, if $A = J_1(37)$, then the sequence of partial gcd's is:

$$15249085236272475, 802583433488025, 160516686697605, \ldots$$

where the term 160516686697605 repeats for all $\ell < 1000$.

6.1.2 The Manin index

Let p be a prime, let $\Omega_{A/\mathbf{Z}}$ denote the sheaf of relative 1-forms on the Néron model of A over $\mathbf{Z}$, and let I be the annihilator of A in the Hecke algebra $\mathbf{T} \subset \mathrm{End}(J_1(N))$. For a subring $R \subset \mathbf{C}$, let $S_2(\Gamma_1(N), R)$ be the R-module of cusp forms whose Fourier expansion at ∞ lies in $R[\![q]\!]$. The natural surjective Hecke-equivariant morphism $J_1(N) \to J_1(N)/IJ_1(N) = A$ induces (by pullback) a Hecke-equivariant injection $\Psi_A : \mathrm{H}^0(A_{/\mathbf{Z}}, \Omega_{A/\mathbf{Z}}) \hookrightarrow S_2(\Gamma_1(N), \mathbf{Q})$ whose image lies in $S_2(\Gamma_1(N), \mathbf{Q})[I]$. (Here we identify $S_2(\Gamma_1(N), \mathbf{Q})$ with $\mathrm{H}^0(X_1(N), \Omega_{X_1(N)/\mathbf{Q}}) = \mathrm{H}^0(J_1(N), \Omega_{J_1(N)/\mathbf{Q}})$ in the usual manner.)

Definition 6.1.4 (Manin index). The *Manin index of A* is

$$c = [S_2(\Gamma_1(N), \mathbf{Z})[I] : \Psi_A(\mathrm{H}^0(A_{/\mathbf{Z}}, \Omega_{A/\mathbf{Z}}))] \in \mathbf{Q}.$$

Remark 6.1.5. We name c after Manin, since he first studied c, but only in the context of elliptic curves. When $X_0(N) \to A$ is an optimal elliptic-curve quotient attached to a newform f, the usual Manin constant of A is the rational number c such that $\pi^*(\omega_A) = \pm c \cdot f \mathrm{d}q/q$, where ω_A is a basis for the differentials on the Néron model of A. The usual Manin constant equals the Manin index, since $S_2(\Gamma_1(N), \mathbf{Z})[I]$ is generated as a $\mathbf{Z}$-module by f.

A priori, the index in Definition 6.1.4 is only a generalized lattice index in the sense of [12, Ch. 1, §3], which we interpret as follows. In [12], for any Dedekind domain R, the *lattice index* is defined for any two finite free R-modules V and W of the same rank ρ that are embedded in a ρ-dimensional $\mathrm{Frac}(R)$-vector space U. The lattice index is the fractional R-ideal generated by the determinant of any automorphism of U that sends V isomorphically onto W. In Definition 6.1.4, we take $R = \mathbf{Z}$, $U = S_2(\Gamma_1(N), \mathbf{Q})[I]$, $V = S_2(\Gamma_1(N), \mathbf{Z})[I]$, and $W = \Psi_A(\mathrm{H}^0(A_{/\mathbf{Z}}, \Omega_{A/\mathbf{Z}}))$. Thus, c is the absolute value of the determinant of any linear transformation of $S_2(\Gamma_1(N), \mathbf{Q})[I]$ that sends $S_2(\Gamma_1(N), \mathbf{Z})[I]$ onto $\Psi_A(\mathrm{H}^0(A_{/\mathbf{Z}}, \Omega_{A/\mathbf{Z}}))$. In fact, it is not necessary to consider lattice indexes, as the following lemma shows (note we will use lattices indices later in the statement of Proposition 6.1.10).

Lemma 6.1.6. *The Manin index c of A is an integer.*

Proof. Let $X_\mu(N)$ be the coarse moduli scheme over $\mathbf{Z}$ that classifies isomorphism classes of pairs $(E/S, \alpha)$, with $\alpha : \mu_N \hookrightarrow E^{\mathrm{sm}}$ a closed subgroup in the smooth locus of a generalized elliptic curve E with irreducible geometric fibers E_s. This is a smooth $\mathbf{Z}$-curve that is not proper, and it is readily constructed by combining the work of Katz-Mazur and Deligne-Rapoport (see §9.3 and §12.3 of [16]). There is a canonical $\mathbf{Z}$-point $\infty \in X_\mu(N)(\mathbf{Z})$ defined by the standard 1-gon equipped with the canonical embedding of μ_N into the smooth locus $\mathbf{G}_m$, and the theory of the Tate curve provides a canonical isomorphism between $\mathrm{Spf}(\mathbf{Z}[\![q]\!])$ and the formal completion of $X_\mu(N)$ along ∞.

There is an isomorphism between the smooth proper curves $X_1(N)$ and $X_\mu(N)$ over $\mathbf{Z}[1/N]$ because the open modular curves $Y_1(N)$ and $Y_\mu(N)$ coarsely represent moduli problems that may be identified over the category of $\mathbf{Z}[1/N]$-schemes via the map

$$(E, P) \mapsto (E/\langle P\rangle, E[N]/\langle P\rangle),$$

where $E[N]/\langle P\rangle$ is identified with μ_N via the Weil pairing on $E[N]$. For our purposes, the key point (which follows readily from Tate's theory) is that under the moduli-theoretic identification of the analytification of the $\mathbf{C}$-fiber of $X_\mu(N)$ with the analytic modular curve $X_1(N)$ via the trivialization of $\mu_N(\mathbf{C})$ by means of $\zeta_N = e^{\pm 2\pi\sqrt{-1}/N}$, the formal parameter q at the $\mathbf{C}$-point ∞ computes the standard analytic q-expansion for weight-2 cusp forms on $\Gamma_1(N)$. The reason we consider $X_\mu(N)$ rather than $X_1(N)$ is simply because we want a smooth $\mathbf{Z}$-model in which the analytic cusp ∞ descends to a $\mathbf{Z}$-point.

Let $\phi : J_1(N) \to A$ be the Albanese quotient map over $\mathbf{Q}$, and pass to Néron models over $\mathbf{Z}$ (without changing the notation). Since $X_\mu(N)$ is $\mathbf{Z}$-smooth, there is a morphism $X_\mu(N) \to J_1(N)$ over $\mathbf{Z}$ that extends the usual morphism sending ∞ to 0. We have a map $\Psi : \mathrm{H}^0(A, \Omega) \to \mathbf{Z}[\![q]\!]\mathrm{d}q/q$ of $\mathbf{Z}$-modules defined by composition

$$\mathrm{H}^0(A, \Omega) \to \mathrm{H}^0(J_1(N), \Omega) \to \mathrm{H}^0(X_\mu(N), \Omega) \xrightarrow{q-\exp} \mathbf{Z}[\![q]\!]\frac{\mathrm{d}q}{q}.$$

The map Ψ is injective, since it is injective after base extension to $\mathbf{Q}$ and each group above is torsion free. The image of Ψ in $\mathbf{Z}[\![q]\!]\mathrm{d}q/q$ is a finite free $\mathbf{Z}$-module, contained in the image of $S = S_2(\Gamma_1(N), \mathbf{Z})$, the sub-$\mathbf{Z}$-module of $S_2(\Gamma_1(N), \mathbf{C})$ of those elements whose analytic q-expansion at ∞ has coefficients in $\mathbf{Z}$. Since Ψ respects the action of Hecke operators, the image of Ψ is contained in $S[I]$, so the lattice index c is an integer. □

We make the following conjecture:

Conjecture 6.1.7. *If $A = A_f$ is a quotient of $J_1(N)$ attached to a single Galois-conjugacy class of newforms, then $c = 1$.*

Manin made this conjecture for one-dimensional optimal quotients of $J_0(N)$. Mazur bounded c in some cases in [46], Stevens considered c for one-dimensional quotients of $J_1(N)$ in [65], González and Lario considered c for $\mathbf{Q}$-curves in [26], Agashe and Stein considered c for quotients of $J_0(N)$ of dimension bigger than 1 in [4], and Edixhoven proved integrality results in [19, Prop. 2] and [22, §2].

Remark 6.1.8. We only make Conjecture 6.1.7 when A is attached to a *single* Galois-conjugacy class of newforms, since the more general conjecture is false. Adam Joyce [31] has recently used failure of multiplicity one for $J_0(p)$ to produce examples of optimal quotients A of $J_1(p)$, for $p = 431$, 503, and 2089, whose Manin indices are divisible by 2. Here, A is isogenous to a product of two elliptic curves, so A is not attached to a single Galois-orbit of newforms.

Remark 6.1.9. The question of whether or not c is an isogeny-invariant is not meaningful in the context of this paper because we only define the Manin index for optimal quotients.

6.1.3 Computing L-ratios

There is a formula for $L(A_f,1)/\Omega_{A_f}$ in [3, §4.2] when A_f is an optimal quotient of $J_0(N)$ attached to a single Galois conjugacy class of newforms. In this section we describe that formula; it applies to our quotient A of $J_1(N)$.

Recall our running hypothesis that $A = A_I$ is an optimal (new) quotient of $J_1(N)$ attached to a Galois conjugacy class of newforms $\{f_1, \ldots, f_n\}$. Let

$$\Psi : \mathrm{H}_1(X_1(N), \mathbf{Q}) \to \mathrm{Hom}(S_2(\Gamma_1(N))[I], \mathbf{C})$$

be the linear map that sends a rational homology class γ to the functional $\int_\gamma$ on the subspace $S_2(\Gamma_1(N))[I]$ in the space of holomorphic 1-forms on $X_1(N)$.

Let $\mathbf{T} \subset \mathrm{End}(\mathrm{H}_1(X_1(N), \mathbf{Q}))$ be the ring generated by all Hecke operators. Since the $\mathbf{T}$-module $H = \mathrm{Hom}(S_2(\Gamma_1(N))[I], \mathbf{C})$ has a natural $\mathbf{R}$-structure (and even a natural $\mathbf{Q}$-structure), it admits a natural $\mathbf{T}$-linear and $\mathbf{C}$-semilinear action by complex conjugation. If M is a $\mathbf{T}$-submodule of H, let M^+ denote the $\mathbf{T}$-submodule of M fixed by complex conjugation.

Let c be the Manin index of A as in Section 6.1.2, let c_∞ be the number of connected components of $A(\mathbf{R})$, let Ω_A be the volume of $A(\mathbf{R})$ as in Conjecture 6.1.2, and let $\{0, \infty\} \in \mathrm{H}_1(X_1(N), \mathbf{Q})$ be the rational homology class whose integration functional is integration from 0 to $i\infty$ along the i-axis (for the precise definition of $\{0, \infty\}$ and a proof that it lies in the rational homology see [38, Ch. IV §1–2]).

Proposition 6.1.10. *Let $A = A_I$ be an optimal quotient of $J_1(N)$ attached to a Galois-stable collection of newforms. With notation as above, we have*

$$c_\infty \cdot c \cdot \frac{L(A,1)}{\Omega_A} = [\Psi(\mathrm{H}_1(X_1(N), \mathbf{Z}))^+ : \Psi(\mathbf{T}\{0, \infty\})], \tag{6.1.1}$$

where the index is a lattice index as discussed in Section 6.1.7 (*in particular, $L(A,1) = 0$ if and only if $\Psi(\mathbf{T}\{0,\infty\})$ has smaller rank than $\mathrm{H}_1(X_1(N), \mathbf{Z})^+$*).

Proof. It is straightforward to adapt the argument of [3, §4.2] with $J_0(N)$ replaced by $J_1(N)$ (or even $J_H(N)$), but one must be careful when replacing A_f with A. The key observation is that if $f_1, \ldots, f_n$ is the unique basis of normalized newforms corresponding to A, then $L(A, s) = L(f_1, s) \cdots L(f_n, s)$. □

Remark 6.1.11. This equality (6.1.1) need not hold if oldforms are involved, even in the $\Gamma_0(N)$ case. For example, if $A = J_0(22)$, then $L(A, s) = L(J_0(11), s)^2$, but two copies of the newform corresponding to $J_0(11)$ do not form a basis for $S_2(\Gamma_0(22))$.

We finish this section with some brief remarks on how to compute the rational number $c \cdot L(A, 1)/\Omega_A$ using (6.1.1) and a computer. Using modular symbols, one can explicitly compute with $\mathrm{H}_1(X_1(N), \mathbf{Z})$. Though the above lattice index involves two lattices in a complex vector space, the index is unchanged if we replace Ψ with any linear map to a $\mathbf{Q}$-vector space such that the kernel is unchanged (see [3, §4.2]). Such a map may be computed via standard linear algebra by finding a basis for $\mathrm{Hom}(\mathrm{H}_1(X_1(N), \mathbf{Q}), \mathbf{Q})[I]$.

To compute c_∞, use the following well-known proposition; we include a proof for lack of an adequate published reference.

Proposition 6.1.12. *For an abelian variety A over $\mathbf{R}$,*

$$c_\infty = 2^{\dim_{\mathbf{F}_2} A[2](\mathbf{R}) - d},$$

where $d = \dim A$ and $c_\infty := |A(\mathbf{R})/A^0(\mathbf{R})|$.

Proof. Let $\Lambda = \mathrm{H}_1(A(\mathbf{C}), \mathbf{Z})$, so the exponential uniformization of $A(\mathbf{C})$ provides a short exact sequence

$$0 \to \Lambda \to \mathrm{Lie}(A(\mathbf{C})) \to A(\mathbf{C}) \to 0.$$

There is an evident action of $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$ on all terms via the action on $A(\mathbf{C})$, and this short exact sequence is Galois-equivariant because A is defined over $\mathbf{R}$. Let Λ^+ be the subgroup of Galois-invariants in Λ, so we get an exact cohomology sequence

$$0 \to \Lambda^+ \to \mathrm{Lie}(A(\mathbf{R})) \to A(\mathbf{R}) \to \mathrm{H}^1(\mathrm{Gal}(\mathbf{C}/\mathbf{R}), \Lambda) \to 0$$

because higher group cohomology for a finite group vanishes on a $\mathbf{Q}$-vector space (such as the Lie algebra of $A(\mathbf{C})$). The map $\mathrm{Lie}(A(\mathbf{R})) \to A(\mathbf{R})$ is the exponential map for $A(\mathbf{R})$, and so its image is $A(\mathbf{R})^0$. Thus, Λ^+ has $\mathbf{Z}$-rank equal to $\dim A$ and

$$A(\mathbf{R})/A(\mathbf{R})^0 \simeq \mathrm{H}^1(\mathrm{Gal}(\mathbf{C}/\mathbf{R}), \Lambda).$$

To compute the size of this H^1, consider the short exact sequence

$$0 \to \Lambda \overset{2}{\to} \Lambda \to \Lambda/2\Lambda \to 0$$

of Galois-modules. Since $\Lambda/n\Lambda \simeq A[n](\mathbf{C})$ as Galois-modules for any $n \neq 0$, the long-exact cohomology sequence gives an isomorphism

$$A[2](\mathbf{R})/(\Lambda^+/2\Lambda^+) \simeq \mathrm{H}^1(\mathrm{Gal}(\mathbf{C}/\mathbf{R}), \Lambda).$$

□

Remark 6.1.13. Since the canonical isomorphism

$$A[n](\mathbf{C}) \simeq \mathrm{H}_1(A(\mathbf{C}), \mathbf{Z})/n\mathrm{H}_1(A(\mathbf{C}), \mathbf{Z})$$

is $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$-equivariant, we can identify $A[2](\mathbf{R})$ with the kernel of $\overline{\tau} - 1$ where $\overline{\tau}$ is the mod-2 reduction of the involution on $\mathrm{H}_1(A(\mathbf{C}), \mathbf{Z})$ induced by the action τ of complex conjugation on $A(\mathbf{C})$. In the special case when A is a quotient of some $J_1(N)$, and we choose a connected component of $\mathbf{C} - \mathbf{R}$ to uniformize $Y_1(N)$ in the usual manner, then via the $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$-equivariant isomorphism $\mathrm{H}_1(J_1(N)(\mathbf{C}), \mathbf{Z}) \simeq \mathrm{H}_1(X_1(N)(\mathbf{C}), \mathbf{Z})$ we see that $\mathrm{H}_1(A(\mathbf{C}), \mathbf{Z})$ may be computed by modular symbols and that the action of τ on the modular symbol is $\{\alpha, \beta\} \mapsto \{-\alpha, -\beta\}$. This makes $A[2](\mathbf{R})$, and hence c_∞, readily computable via modular symbols.

6.2 Arithmetic of $J_1(p)$

6.2.1 The Tables

For $p \leq 71$, the first part of Table 1 (on page 393) lists the dimension of $J_1(p)$ and the rational number $L = c \cdot L(J_1(p), 1)/\Omega_{J_1(p)}$. Table 1 also gives an upper bound T (in the sense of divisibility) on $|J_1(p)(\mathbf{Q})_{\mathrm{tor}}|$ for $p \leq 71$, as discussed in §6.1.1.

When $L \neq 0$, Conjecture 6.1.2 and the assumption that $c = 1$ imply that the numerator of L divides $c_p \cdot |Ш(A)|$, that in turn divides T^2L. For every $p \neq 29$ with $p \leq 71$, we found that $T^2L = 1$. For $p = 29$, we have $T^2L = 2^{12}$; it would be interesting if the isogeny-invariant T overestimates the order of $J_1(29)(\mathbf{Q})_{\mathrm{tor}}$ or if $Ш(J_1(29))$ is nontrivial.

6.2.2 Determination of positive rank

Proposition 6.2.1. *The primes p such that $J_1(p)$ has positive rank are the same as the primes for which $J_0(p)$ has positive rank:*

$$p = 37, 43, 53, 61, 67, \textit{ and all } p \geq 73.$$

Proof. Proposition 2.8 of [45, §III.2.2, p. 147] says: "Suppose $g^+ > 0$ (which is the case for all $N > 73$, as well as $N = 37, 43, 53, 61, 67$). Then the Mordell-Weil group of J_+ is a torsion-free group of infinite order (*i.e.* of positive rank)." Here, N is a prime, g^+ is the genus of the Atkin-Lehner quotient $X_0(N)^+$ of $X_0(N)$, and J_+ is isogenous to the Jacobian of $X_0(N)^+$. This is essentially correct, except for the minor oversight that $g^+ > 0$ also when $N = 73$ (this is stated correctly on page 34 of [45]).

By Mazur's proposition $J_0(p)$ has positive algebraic rank for all $p \geq 73$ and for $p = 37, 43, 53, 61, 67$. The sign in the functional equation for $L(J_+, s)$ is -1, so

$$L(J, 1) = L(J_+, 1)L(J_-, 1) = 0 \cdot L(J_-, 1) = 0$$

for all p such that $g^+ > 0$. Using (6.1.1) we see that $L(J, 1) \neq 0$ for all p such that $g^+ = 0$, which by Kato (see [32, Cor. 14.3]) or Kolyvagin–Logachev (see [36]) implies that J has rank 0 whenever $g^+ = 0$. Thus $L(J_0(p), 1) = 0$ if and only if $J_0(p)$ has positive rank.

Work of Kato (see [32, Cor. 14.3]) implies that if $J_1(p)$ has analytic rank 0, then $J_1(p)$ has algebraic rank 0. It thus suffices to check that $L(J_1(p), 1) \neq 0$ for the primes p such that $J_0(p)$ has rank 0. We verify this by computing $c \cdot L(J_1(p), 1)/\Omega_{J_1(p)}$ using (6.1.1), as illustrated in Table 1.

□

If we instead consider composite level, it is not true that $J_0(N)$ has positive analytic rank if and only if $J_1(N)$ has positive analytic rank. For example, using (6.1.1) we find that $J_0(63)$ has analytic rank 0, but $J_1(63)$ has positive analytic rank. Closer inspection using Magma (see the program below) shows that there is a two-dimensional new quotient A_f with positive analytic rank, where $f = q + (\omega - 1)q^2 + (-\omega - 2)q^3 + \cdots$, and $\omega^3 = 1$. It would be interesting to prove that that the algebraic rank of A_f is positive.

```
> M := ModularSymbols(63,2);
> S := CuspidalSubspace(M);
> LRatio(S,1);        // So J_0(63) has rank 0
1/384

> G<a,b> := DirichletGroup(63,CyclotomicField(6));
> e := a^5*b;
> M := ModularSymbols([e],2,+1);
> S := CuspidalSubspace(M);
> LRatio(S,1);      // This step takes some time.
0
> D := NewformDecomposition(S);
> LRatio(D[1],1);
0
> qEigenform(D[1],5);
q + (-2*zeta_6 + 1)*q^2 + (-2*zeta_6 + 1)*q^3 - q^4 + O(q^5)
```

6.2.3 Conjectural order of $J_1(\mathbf{Q})_{\mathrm{tor}}$

For any Dirichlet character ε modulo N, define Bernoulli numbers $B_{2,\varepsilon}$ by

$$\sum_{a=1}^{N} \frac{\varepsilon(a) t e^{at}}{e^{Nt}-1} = \sum_{k=0}^{\infty} \frac{B_{k,\varepsilon}}{k!} t^k.$$

We make the following conjecture.

Conjecture 6.2.2. *Let $p \geq 5$ be prime. The rational torsion subgroup $J_1(p)(\mathbf{Q})_{\mathrm{tor}}$ is generated by the differences of $\mathbf{Q}$-rational cusps on $X_1(p)$. Equivalently* (*see below*), *for any prime $p \geq 5$,*

$$|J_1(p)(\mathbf{Q})_{\mathrm{tor}}| = \frac{p}{2^{p-3}} \cdot \prod_{\varepsilon \neq 1} B_{2,\varepsilon} \tag{6.2.1}$$

where the product is over the nontrivial even Dirichlet characters ε of conductor dividing p.

Due to how we defined $X_1(p)$, its $\mathbf{Q}$-rational cusps are exactly its cusps lying over the cusp $\infty \in X_0(p)(\mathbf{Q})$ (corresponding to the standard 1-gon equipped with the subgroup μ_p in its smooth locus $\mathbf{G}_m$) via the second standard degeneracy map

$$(E, P) \mapsto (E/\langle P\rangle, E[p]/\langle P\rangle).$$

In [49] Ogg showed that $|J_1(13)(\mathbf{Q})| = 19$, verifying Conjecture 6.2.2 for $p = 13$. The results of [37] are also relevant to Conjecture 6.2.2, and suggest that the rational torsion of $J_1(p)$ is cuspidal. Let $C(p)$ be the conjectural order of $J_1(p)(\mathbf{Q})_{\mathrm{tor}}$ on the right side of (6.2.1). In [37, p. 153], Kubert and Lang prove that $C(p)$ is equal to the order of the group generated by the differences of $\mathbf{Q}$-rational cusps on $X_1(p)$ (in their language, these are viewed as the cusps that lie over $0 \in X_0(p)(\mathbf{Q})$ via the first standard degeneracy map

$$(E, P) \mapsto (E, \langle P\rangle)),$$

and so $C(p)$ is *a priori* an integer that moreover divides $|J_1(p)(\mathbf{Q})_{\mathrm{tor}}|$.

Table 1 provides evidence for Conjecture 6.2.2. Let $T(p)$ be the upper bound on $J_1(p)(\mathbf{Q})_{\mathrm{tor}}$ (see Table 1). For all $p \leq 157$, we have $C(p) = T(p)$ except for $p = 29$, 97, 101, 109, and 113, where $T(p)/C(p)$ is 2^6, 17, 2^4, 3^7, and $2^{12} \cdot 3^2$, respectively. Thus Conjecture 6.2.2 is true for $p \leq 157$, except possibly in these five cases, where the deviation is consistent with the possibility that $T(p)$ is a nontrivial multiple of the true order of the torsion subgroup (recall that $T(p)$ is an isogeny-invariant, and so it is not surprising that it may be too large).

6.3 Arithmetic of $J_H(p)$

For each divisor d of $p-1$, let $H = H_d$ denote the unique subgroup of $(\mathbf{Z}/p\mathbf{Z})^\times$ of order $(p-1)/d$. The group of characters whose kernel contains H_d is exactly the group of characters of order dividing d. Since the linear fractional transformation associated to $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ acts trivially on the upper half plane, we lose nothing (for the computations that we will do in this section) if we assume that $-1 \in H$, and so $|H|$ is even.

For any subgroup H of $(\mathbf{Z}/p\mathbf{Z})^\times$ as above, let J_H be the Jacobian of $X_H(p)$, as in Section 1. For each $p \leq 71$, Table 2 lists the dimension of $J_H = J_H(p)$, the rational number $L = c \cdot L(J_H, 1)/\Omega_{J_H}$, an upper bound T on $|J_H(\mathbf{Q})_{\mathrm{tor}}|$, the conjectural multiple T^2L of $|\text{Ш}(J_H)| \cdot c_p$, and $c_p = |\Phi(J_H)|$. We compute $|\Phi(J_H)(\mathbf{F}_p)| = |\Phi(J_H)(\overline{\mathbf{F}}_p)|$ using Theorem 1.1.3. Note that Table 2 omits the data for $d = (p-1)/2$, since $J_H = J_1(p)$ for such d, so the corresponding data is therefore already contained in Table 1.

When $L \neq 0$, we have $T^2L = |\Phi(J_H)|$ in all but one case. The exceptional case is $p = 29$ and $d = 7$, where $T^2L = 2^6$, but $|\Phi(J_H)| = 1$; probably T overestimates the torsion in this case. In the following proposition we use this observation to deduce that $|\text{Ш}(J_H)| = c = 1$ in some cases.

Proposition 6.3.1. *Suppose that $p \leq 71$ is a prime and $d \mid (p-1)$ with $(p-1)/d$ even. Let J_H be the Jacobian of $X_H(p)$, where H is the subgroup of $(\mathbf{Z}/p\mathbf{Z})^\times$ of order $(p-1)/d$. Assume that Conjecture* 6.1.2 *is true, and if $p = 29$ then assume that $d \neq 7, 14$. If $L(J_H, 1) \neq 0$, then $|\text{Ш}(J_H)| = 1$ and $c = 1$.*

It is not interesting to remove the condition $p \leq 71$ in the statement of the proposition, since when $p > 71$ the quantity $L(J_H, 1)$ automatically vanishes (see Proposition 6.2.1). It is probably not always the case that $|\text{Ш}(J_H)| = 1$; for example, Conjecture 6.1.2 and the main result of [1] imply that 7^2 divides $|\text{Ш}(J_0(1091))|$.

Proof. We deduce the proposition from Tables 1–3 as follows. Using Conjecture 6.1.2 we have

$$c \cdot |\text{Ш}(J_H)| = c \cdot \frac{L(J_H, 1)}{\Omega_{J_H} \cdot |\Phi(J_H)|} \cdot |J_H(\mathbf{Q})_{\mathrm{tor}}|^2. \tag{6.3.1}$$

Let T denote the torsion bound on $J_H(\mathbf{Q})_{\mathrm{tor}}$ as in Section 6.1.1 and let $L = c \cdot L(J_H, 1)/\Omega_{J_H}$, so the right side of (6.3.1) divides $T^2L/|\Phi(J_H)|$. An inspection of the tables shows that $T^2L/|\Phi(J_H)| = 1$ for J_H satisfying the hypothesis of the proposition (in the excluded cases $p = 29$ and $d = 7, 14$, the quotient equals 2^6 and 2^{12}, respectively). Since $c \in \mathbf{Z}$, we conclude that $c = |\text{Ш}(J_H)| = 1$.

□

Remark 6.3.2. Theorem 1.1.3 is an essential ingredient in the proof of Proposition 6.3.1 because we used Theorem 1.1.3 to compute the Tamagawa factor c_p.

6.4 Arithmetic of newform quotients

Tables 4–5 at the end of this paper contain arithmetic information about each newform abelian variety quotient A_f of $J_1(p)$ with $p \leq 71$.

The first column gives a label determining a Galois-conjugacy class of newforms $\{f, \ldots\}$, where **A** corresponds to the first class, **B** to the second, *etc.*, and the classes are ordered first by dimension and then in lexicographical order by the sequence of nonegative integers $|\operatorname{tr}(a_2(f))|, |\operatorname{tr}(a_3(f))|, |\operatorname{tr}(a_5(f))|, \ldots$. (WARNING: This ordering does not agree with the one used by Cremona in [14]; for example, our **37A** is Cremona's **37B**.) The next two columns list the dimension of A_f and the order of the Nebentypus character of f, respectively. The fourth column lists the rational number $L = L(A_f, 1)/\Omega_{A_f}$, and the fifth lists the product T^2L, where T is an upper bound (as in Section 6.1.1) on the order of $A_f(\mathbf{Q})_{\mathrm{tor}}$. The sixth column, labeled "modular kernel", lists invariants of the group of $\overline{\mathbf{Q}}$-points of the kernel of the polarization $A_f^\vee \hookrightarrow J_1(p) \to A_f$; this kernel is computed by using an algorithm based on Proposition 6.4.1 below. The elementary divisors of the kernel are denoted with notation such as $[2^2 14^2]$ to denote

$$\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/14\mathbf{Z} \times \mathbf{Z}/14\mathbf{Z}.$$

Proposition 6.4.1. *Suppose $A = A_I$ is an optimal quotient of $J = J_1(N)$ attached to the annihilator I of a Galois-stable collection of newforms. The group of $\overline{\mathbf{Q}}$-points of the kernel of the natural map $A^\vee \hookrightarrow J \to A$ is isomorphic to the cokernel of the natural map*

$$\operatorname{Hom}(\mathrm{H}_1(X_1(N), \mathbf{Z}), \mathbf{Z})[I] \to \operatorname{Hom}(\mathrm{H}_1(X_1(N), \mathbf{Z})[I], \mathbf{Z}).$$

Proof. The proof is the same as [35, Prop. 1].

□

It is possible to compute the modular kernel by using the formula in this proposition, together with modular symbols and standard algorithms for computing with finitely generated abelian groups.

We do not give T in Tables 4–5, since in all but six cases $T^2L \neq 0$, hence T^2L and L determine T. The remaining six cases are **37B**, **43A**, **53A**, **61A**, **61B**, and **67C**, and in all these cases $T = 1$.

Remark 6.4.2. If $A = A_f$ is an optimal quotient of $J_1(p)$ attached to a newform, then the tables do not include the toric, additive, and abelian ranks of the closed fiber of the Néron model of A over $\mathbf{F}_p$, since they are easy to determine from other data about A as follows. If $\varepsilon(f) = 1$, then the toric rank is $\dim(A)$, since A is isogenous to an abelian subvariety of $J_0(p)$ and so A has purely toric reduction over $\mathbf{Z}_p$. Now suppose that $\varepsilon(f)$ is nontrivial, so A is isogenous to an abelian subvariety of the abelian variety $J_1(p)/J_0(p)$ that has potentially good reduction at p. Hence the toric rank of A is zero, and inertia $I_p \subset G_p = \operatorname{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ acts with finite image on the $\mathbf{Q}_\ell$-adic Tate module V_ℓ of A for any $\ell \neq p$. Hence V_ℓ splits as a nontrivial direct sum of simple representations of I_p. Let V' be a factor of V_ℓ corresponding to a simple summand K

of $\mathbf{T} \otimes \mathbf{Q}_\ell$, where $\mathbf{T}$ is the Hecke algebra. Since the Artin conductor of the 2-dimensional K-representation V'_ℓ is p, the $\overline{\mathbf{Q}}_\ell[I_p]$-module $\overline{\mathbf{Q}}_\ell \otimes_{\mathbf{Q}_\ell} V'$ is the direct sum of the trivial representation and the character $\varepsilon(f) : (\mathbf{Z}/p\mathbf{Z})^\times \to \overline{\mathbf{Q}}_\ell^\times$ viewed as a character of G_p via the identification $\mathrm{Gal}(\mathbf{Q}_p(\zeta_p)/\mathbf{Q}_p) = (\mathbf{Z}/p\mathbf{Z})^\times$. This implies that the abelian rank as well as the additive rank are both equal to half of the dimension of A.

6.4.1 The Simplest example not covered by general theory

The prime $p = 61$ is the only prime $p \leq 71$ such that the maximal quotient of $J_1(p)$ with positive analytic rank is not a quotient of $J_0(p)$. Let ε be a Dirichlet character of conductor 61 and order 6. Consider the abelian variety A_f attached to the newform

$$f = q + (e^{2\pi i/3} - 1)q^2 - 2q^3 + \cdots$$

that lies in the 6-dimensional $\mathbf{C}$-vector space $S_2(\Gamma_1(61), \varepsilon)$. Using Proposition 6.1.10, we see that $L(f, 1) = 0$.

It would be interesting to show that A_f has positive algebraic rank, since A_f is not covered by the general theorems of Kolyvagin, Logachev, and Kato concerning Conjecture 6.1.2. This example is the simplest example in the following sense: every elliptic curve over $\mathbf{Q}$ is a quotient of some $J_0(N)$, and an inspection of Tables 4–5 for any integer $N < 61$ shows that the maximal quotient of $J_1(N)$ with positive analytic rank is also a quotient of $J_0(N)$.

The following observation puts this question in the context of $\mathbf{Q}$-curves, and may be of some use in a direct computation to show that A_f has positive algebraic rank. Since $\overline{f} = f \otimes \varepsilon^{-1}$, Shimura's theory (see [62, Prop. 8]) supplies an isogeny $\varphi : A_f \to A_f$ defined over the degree-6 abelian extension of $\mathbf{Q}$ cut out by $\ker(\varepsilon)$. Using φ, one sees that A_f is isogenous to a product of two elliptic curves. According to Enrique Gonzalez-Jimenez (personal communication) and Jordi Quer, if $t^6 + t^5 - 25t^4 + 8t^3 + 123t^2 - 126t + 27 = 0$, so t generates the degree 6 subfield of $\mathbf{Q}(\zeta_{61})$ corresponding to ε, then one of the elliptic-curve factors of A_f has equation $y^2 = x^3 + c_4 x + c_6$, where

$$\begin{aligned} c_4 &= \frac{1}{3}(-321 + 738t - 305t^2 - 196t^3 + 47t^4 + 13t^5), \\ c_6 &= \frac{1}{3}(-4647 + 6300t + 996t^2 - 1783t^3 - 432t^4 - 14t^5). \end{aligned}$$

6.4.2 Can Optimal Quotients Have Nontrivial Component Group?

Let p be a prime. Component groups of optimal quotients of $J_0(p)$ are well-understood in the sense of the following theorem of Emerton [23]:

Theorem 6.4.3 (Emerton). *If $A_1, \ldots, A_n$ are the distinct optimal quotients of $J_0(p)$ attached the Galois-orbits of newforms, then the product of the orders of the component groups of the A_i's equals the order of the component* *group of* $J_0(p)$*,* i.e.*, the numerator of* $(p-1)/12$*. Moreover, the natural maps* $\Phi(J_0(p)) \to \Phi(A_i)$ *are surjective.*

Shuzo Takehashi asked a related question about $J_1(p)$:

Question 6.4.4 (Takehashi). Suppose $A = A_f$ is an optimal quotient of $J_1(p)$ attached to a newform. What can be said about the component group of A? In particular, is the component group of A necessarily trivial?

Since $J_1(p)$ has trivial component group (see Theorem 1.1.1), the triviality of the component group of A is equivalent to the surjectivity of the natural map from $\Phi(J_1(p))$ to $\Phi(A_f)$.

The data in Tables 4–5 sheds little light on Question 6.4.4. The following are the A_f's that have nonzero $L = c \cdot L(A_f, 1)/\Omega$ with numerator divisible by an odd prime: **37D**, **37F**, **43C**, **43F**, **53D**, **61E**, **61F**, **61G**, **61J**, **67D**, **67E**, and **67G**. For each of these, Conjecture 6.1.2 implies that $c \cdot Ш(A_f) \cdot c_p$ is divisible by an odd prime. However, it seems difficult to deduce which factors in the product are not equal to 1. We remark that for each A_f listed above such that the numerator of L is exactly divisible by p, there is a rank-1 elliptic curve E over $\mathbf{Q}$ such that $E[p] \subset A$, so methods as in [2] may shed light on this problem.

6.5 Using Magma to compute the tables

In this section, we describe how to use Magma V2.10-6 (or later) to compute the entries in Tables 1–5 at the end of this paper.

6.5.1 Computing Table 1: Arithmetic of $J_1(p)$

Let p be a prime. The following Magma code illustrates how to compute the two rows in Table 1 corresponding to p $(=19)$. Note that the space of cuspidal modular symbols has dimension $2\dim J_1(p)$.

```
> p := 19;
> M := ModularSymbols(Gamma1(p));
> S := CuspidalSubspace(M);
> S;
Modular symbols space of level 19, weight 2, and dimension
14 over Rational Field (multi-character)
> LRatio(S,1);
1/19210689
> Factorization(19210689);
[ <3, 4>, <487, 2> ]
> TorsionBound(S,60);
4383
```

Remark 6.5.1. It takes less time and memory to compute $c \cdot L(J_1(p), 1)/\Omega$ in $\mathbf{Q}^\times/2^{\mathbf{Z}}$, and this is done by replacing `M:=ModularSymbols(Gamma1(p))` with

`M:=ModularSymbols(Gamma1(p),2,+1)`. A similar remark applies to all computations of L-ratios in the sections below.

6.5.2 Computing Tables 2–3: Arithmetic of $J_H(p)$

Let p be a prime, d a divisor of $p-1$ such that $(p-1)/d$ is even, and H the subgroup of $(\mathbf{Z}/N\mathbf{Z})^{\times}$ of order $(p-1)/d$. We use Theorem 1.1.3 and commands similar to the ones in Section 6.5.1 to fill in the entries in Tables 2–3. The following code illustrates computation of the second row of Table 2 for $p = 19$.

```
> p := 19;
> [d : d in Divisors(p-1) | IsEven((p-1) div d)];
[ 1, 3, 9 ]
> d := 3;
> M := ModularSymbolsH(p,(p-1) div d, 2, 0);
> S := CuspidalSubspace(M);
> S;
Modular symbols space of level 19, weight 2, and dimension 2
over Rational Field (multi-character)
> L := LRatio(S,1); L;
1/9
> T := TorsionBound(S,60); T;
3
> T^2*L;
1
> Phi := d / GCD(d,6);  Phi;
1
```

It takes about ten minutes to compute all entries in Table 2–3 using an Athlon 2000MP-based computer.

6.5.3 Computing Tables 4–5

Let p be a prime number. To compute the modular symbols factors corresponding to the newform optimal quotients A_f of $J_1(p)$, we use the `NewformDecomposition` command. To compute the modular kernel, we use the command `ModularKernel`. The following code illustrates computation of the second row of Table 4 corresponding to $p = 19$.

```
> p := 19;
> M := ModularSymbols(Gamma1(19));
> S := CuspidalSubspace(M);
> D := NewformDecomposition(S);
> D;
[
    Modular symbols space for Gamma_0(19) of weight 2 and
    dimension 2 over Rational Field,
    Modular symbols space of level 19, weight 2, and
    dimension 12 over Rational Field (multi-character)
]
> A := D[2];
> Dimension(A) div 2;
6
> Order(DirichletCharacter(A));
9
> L := LRatio(A,1); L;
1/2134521
> T := TorsionBound(A,60);
> T^2*L;
1
> Invariants(ModularKernel(A));
[ 3, 3 ]
```

It takes about 2.5 hours to compute all entries in Tables 4–5, except that the entries corresponding to $p = 71$, using an Athlon 2000MP-based computer. The $p = 71$ entry takes about 3 hours.

6.6 Arithmetic tables

The notation in Tables 1–5 below is explained in Section 6.

Table 1: Arithmetic of $J_1(p)$

$J_1(p)$	dim	$c \cdot L(J_1(p),1)/\Omega$
11	1	$1/5^2$
13	2	$1/19^2$
17	5	$1/2^6 \cdot 73^2$
19	7	$1/3^4 \cdot 487^2$
23	12	$1/11^2 \cdot 37181^2$
29	22	$1/2^{12} \cdot 3^2 \cdot 7^2 \cdot 43^2 \cdot 17837^2$
31	26	$1/2^4 \cdot 5^4 \cdot 7^2 \cdot 11^2 \cdot 2302381^2$
37	40	0
41	51	$1/2^8 \cdot 5^2 \cdot 13^2 \cdot 31^4 \cdot 431^2 \cdot 250183721^2$
43	57	0
47	70	$1/23^2 \cdot 139^2 \cdot 82397087^2 \cdot 12451196833^2$
53	92	0
59	117	$1/29^2 \cdot 59^2 \cdot 9988553613691393812358794271^2$
61	126	0
67	155	0
71	176	$1/5^2 \cdot 7^2 \cdot 31^2 \cdot 113^2 \cdot 211^2 \cdot 281^2 \cdot 701^4 \cdot 12713^2 \cdot$ $1307084991922565572906 1^2$

$J_1(p)$	Torsion Bound
11	5
13	19
17	$2^3 \cdot 73$
19	$3^2 \cdot 487$
23	$11 \cdot 37181$
29	$2^{12} \cdot 3 \cdot 7 \cdot 43 \cdot 17837$
31	$2^2 \cdot 5^2 \cdot 7 \cdot 11 \cdot 2302381$
37	$3^2 \cdot 5 \cdot 7 \cdot 19 \cdot 37 \cdot 73 \cdot 577 \cdot 17209$
41	$2^4 \cdot 5 \cdot 13 \cdot 31^2 \cdot 431 \cdot 250183721$
43	$2^2 \cdot 7 \cdot 19 \cdot 29 \cdot 463 \cdot 1051 \cdot 416532733$
47	$23 \cdot 139 \cdot 82397087 \cdot 12451196833$
53	$7 \cdot 13 \cdot 85411 \cdot 96331 \cdot 379549 \cdot 641949283$
59	$29 \cdot 59 \cdot 9988553613691393812358794271$
61	$5 \cdot 7^2 \cdot 11^2 \cdot 19 \cdot 31 \cdot 2081 \cdot 2801 \cdot 40231 \cdot 411241 \cdot 514216621$
67	$11 \cdot 67 \cdot 193 \cdot 661^2 \cdot 2861 \cdot 8009 \cdot 11287 \cdot 9383200455691459$
71	$5 \cdot 7 \cdot 31 \cdot 113 \cdot 211 \cdot 281 \cdot 701^2 \cdot 12713 \cdot 13070849919225655729061$

Table 2: Arithmetic of $J_H(p)$

p	d	dim	$L = c \cdot L(J_H, 1)/\Omega$	$T =$ Torsion Bound	T^2L	$\|\Phi(J_H)\|$
11	1	1	$1/5$	5	5	5
13	1	0	1	1	1	1
	2	0	1	1	1	1
	3	0	1	1	1	1
17	1	1	$1/2^2$	2^2	2^2	2^2
	2	1	$1/2^3$	2^2	2	2
	4	1	$1/2^4$	2^2	1	1
19	1	1	$1/3$	3	3	3
	3	1	$1/3^2$	3	1	1
23	1	2	$1/11$	11	11	11
29	1	2	$1/7$	7	7	7
	2	4	$1/3^2{\cdot}7$	$3{\cdot}7$	7	7
	7	8	$1/2^6{\cdot}7^2{\cdot}43^2$	$2^6{\cdot}7{\cdot}43$	2^6	1
31	1	2	$1/5$	5	5	5
	3	6	$1/2^4{\cdot}5{\cdot}7^2$	$2^2{\cdot}5{\cdot}7$	5	5
	5	6	$1/5^4{\cdot}11^2$	$5^2{\cdot}11$	1	1
37	1	2	0	3	0	3
	2	4	0	$3{\cdot}5$	0	3
	3	4	0	$3{\cdot}7$	0	1
	6	10	0	$3{\cdot}5{\cdot}7{\cdot}37$	0	1
	9	16	0	$3^2{\cdot}7{\cdot}19{\cdot}577$	0	1
41	1	3	$1/2{\cdot}5$	$2{\cdot}5$	$2{\cdot}5$	$2{\cdot}5$
	2	5	$1/2^6{\cdot}5$	$2^3{\cdot}5$	5	5
	4	11	$1/2^8{\cdot}5{\cdot}13^2$	$2^4{\cdot}5{\cdot}13$	5	5
	5	11	$1/2{\cdot}5^2{\cdot}431^2$	$2{\cdot}5{\cdot}431$	2	2
	10	21	$1/2^6{\cdot}5^2{\cdot}31^4{\cdot}431^2$	$2^3{\cdot}5{\cdot}31^2{\cdot}431$	1	1

Table 3: Arithmetic of $J_H(p)$ (continued)

p	d	dim	$L = c \cdot L(J_H, 1)/\Omega$	$T =$ Torsion Bound	T^2L	$\lvert\Phi(J_H)\rvert$
43	1	3	0	7	0	7
	3	9	0	$2^2\cdot 7\cdot 19$	0	7
	7	15	0	$7\cdot 29\cdot 463$	0	1
47	1	4	$1/23$	23	23	23
53	1	4	0	13	0	13
	2	8	0	$7\cdot 13$	0	13
	13	40	0	$13\cdot 96331\cdot 379549$	0	1
59	1	5	$1/29$	29	29	29
61	1	4	0	5	0	5
	2	8	0	$5\cdot 11$	0	5
	3	12	0	$5\cdot 7\cdot 19$	0	5
	5	16	0	$5\cdot 2801$	0	1
	6	26	0	$5\cdot 7^2\cdot 11\cdot 19\cdot 31$	0	5
	10	36	0	$5\cdot 11^2\cdot 2081\cdot 2801$	0	1
	15	56	0	$5\cdot 7\cdot 19\cdot 2801\cdot 514216621$	0	1
67	1	5	0	11	0	11
	3	15	0	$11\cdot 193$	0	11
	11	45	0	$11\cdot 661\cdot 2861\cdot 8009$	0	1
71	1	6	$1/5\cdot 7$	$5\cdot 7$	$5\cdot 7$	$5\cdot 7$
	5	26	$1/5^2\cdot 7\cdot 31^2\cdot 211^2$	$5\cdot 7\cdot 31\cdot 211$	7	7
	7	36	$1/5\cdot 7^2\cdot 113^2\cdot 12713^2$	$5\cdot 7\cdot 113\cdot 12713$	5	5

Table 4: Arithmetic of Optimal Quotients A_f of $J_1(p)$

A_f	dim	ord(ε)	$L = c \cdot L(A_f, 1)/\Omega$	T^2L	modular kernel
11A	1	1	$1/5^2$	1	[]
13A	2	6	$1/19^2$	1	[]
17A	1	1	$1/2^4$	1	$[2^2]$
17B	4	8	$1/2^2 \cdot 73^2$	1	$[2^2]$
19A	1	1	$1/3^2$	1	$[3^2]$
19B	6	9	$1/3^2 \cdot 487^2$	1	$[3^2]$
23A	2	1	$1/11^2$	1	$[11^2]$
23B	10	11	$1/37181^2$	1	$[11^2]$
29A	2	2	$1/3^2$	1	$[14^4]$
29B	2	1	$1/7^2$	1	$[2^2 14^2]$
29C	6	7	$1/2^6 \cdot 43^2$	2^6	$[2^{10} 14^2]$
29D	12	14	$1/2^6 \cdot 17837^2$	2^6	$[2^8 14^4]$
31A	2	1	$1/5^2$	1	$[3^2 15^2]$
31B	4	5	$1/5^2 \cdot 11^2$	1	$[3^6 15^2]$
31C	4	3	$1/2^4 \cdot 7^2$	1	$[5^4 15^4]$
31D	16	15	$1/2302381^2$	1	$[15^8]$
37A	1	1	$1/3^2$	1	$[12^2]$
37B	1	1	0	0	$[36^2]$
37C	2	2	$2/5^2$	2	$[18^4]$
37D	2	3	$3/7^2$	3	$[6^2 18^2]$
37E	4	6	$1/37^2$	1	$[3^4 18^4]$
37F	6	9	$3/577^2$	3	$[2^6 6^2 102^4]$
37G	6	9	$1/3^2 \cdot 19^2$	1	$[2^8 34^2 102^2]$
37H	18	18	$1/73^2 \cdot 17209^2$	1	$[2^{12} 6^{12}]$
41A	2	2	$1/2^4$	1	$[20^4]$
41B	3	1	$1/2^2 \cdot 5^2$	1	$[2^2 20^4]$
41C	6	4	$1/2^2 \cdot 13^2$	1	$[5^2 10^{10}]$
41D	8	10	$1/31^4$	1	$[4^{12} 20^4]$
41E	8	5	$1/431^2$	1	$[4^{12} 20^4]$
41F	24	20	$1/250183721^2$	1	$[2^{20} 10^{12}]$
43A	1	1	0	0	$[42^2]$
43B	2	1	$2/7^2$	2	$[3^2 42^2]$
43C	2	3	$3/2^4$	3	$[35^2 105^2]$
43D	4	3	$1/19^2$	1	$[7^4 105^4]$
43E	6	7	$1/29^2$	1	$[3^8 39^2 273^2]$
43F	6	7	$7/463^2$	7	$[3^8 39^2 273^2]$
43G	36	21	$1/1051^2 \cdot 416532733^2$	1	$[3^{12} 21^{12}]$

Table 5: Arithmetic of Optimal Quotients A_f of $J_1(p)$ (continued)

A_f	dim	ord(ε)	$L = c \cdot L(A_f, 1)/\Omega$	T^2L	modular kernel
47A	4	1	$1/23^2$	1	$[23^6]$
47B	66	23	$1/139^2 \cdot 82397087^2 \cdot 12451196833^2$	1	$[23^6]$
53A	1	1	0	0	$[52^2]$
53B	3	1	$2/13^2$	2	$[2^2 26^2 52^2]$
53C	4	2	$2/7^2$	2	$[26^8]$
53D	36	13	$13/96331^2 \cdot 379549^2$	13	$[2^{66} 26^6]$
53E	48	26	$1/85411^2 \cdot 641949283^2$	1	$[2^{64} 26^8]$
59A	5	1	$1/29^2$	1	$[29^8]$
59B	112	29	$1/59^2 \cdot 998855361369139381235879427 1^2$	1	$[29^8]$
61A	1	1	0	0	$[60^2]$
61B	2	6	0	0	$[55^4]$
61C	3	1	$2/5^2$	2	$[6^2 30^2 60^2]$
61D	4	2	$2/11^2$	2	$[30^8]$
61E	8	3	$3/7^2 \cdot 19^2$	3	$[10^8 30^8]$
61F	8	6	$11^2/7^2 \cdot 31^2$	11^2	$[10^8 30^4 330^4]$
61G	12	5	$5/2801^2$	5	$[6^{18} 30^6]$
61H	16	10	$1/11^2 \cdot 2081^2$	1	$[3^8 6^{16} 30^8]$
61I	32	15	$1/514216621^2$	1	$[2^{40} 6^8 30^{16}]$
61J	40	30	$5^2/40231^2 \cdot 411241^2$	5^2	$[2^{32} 6^{12} 30^{20}]$
67A	1	1	1	1	$[165^2]$
67B	2	1	$2^2/11^2$	2^2	$[6^2 330^2]$
67C	2	1	0	0	$[66^4]$
67D	10	11	$11/2861^2$	11	$[3^{16} 7521^2 82731^2]$
67E	10	3	$3^2/193^2$	3^2	$[11^{10} 33^{10}]$
67F	10	11	$1/661^2$	1	$[3^{16} 4623^2 50853^2]$
67G	20	11	$11/8009^2$	11	$[3^{36} 240999^4]$
67H	100	33	$1/67^2 \cdot 661^2 \cdot 11287^2 \cdot 9383200455691459^2$	1	$[3^{60} 33^{20}]$
71A	3	1	$1/7^2$	1	$[5^2 35^2 315^2]$
71B	3	1	$1/5^2$	1	$[7^2 35^2 315^2]$
71C	20	5	$1/31^2 \cdot 211^2$	1	$[7^{30} 35^{10}]$
71D	30	7	$1/113^2 \cdot 12713^2$	1	$[5^{50} 35^{10}]$
71E	120	35	$1/281^2 \cdot 701^4 \cdot 1307084991922565572906 1^2$	1	$[5^{20} 35^{40}]$

References

[1] A. Agashé, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. 328 (1999), no. 5, 369–374.

[2] A. Agashe and W. Stein, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory 97 (2002), no. 1, 171–185.

[3] ______, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank* 0, to appear in Math. Comp.

[4] ______, *The Manin constant, congruence primes, and the modular degree*, in progress.

[5] M. Artin, *Algebraic approximation of structures over complete local rings*, Publ. Math. IHES, 36 (1969), 23–58.

[6] L. Bégueri, *Dualité sur un corps local à corps résiduel algébriquement clos*, Mém. Soc. Math. France (1980/81), no. 4.

[7] A. Bertapelle, *On perfectness of Grothendieck's pairing for the ℓ-parts of component groups*, J. Reine Angew. Math., 538 (2001), 223–236.

[8] A. Bertapelle and S. Bosch, *Weil restriction and Grothendieck's duality conjecture*, Journal of algebraic geometry, 9 (2000), 155–164.

[9] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.

[10] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. 24 (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[11] J. W. S. Cassels and E. V. Flynn, *Prolegomena to a middlebrow arithmetic of curves of genus* 2, Cambridge University Press, Cambridge, 1996.

[12] J. W. S. Cassels, A. Fröhlich, *Algebraic Number Theory*, Academic Press, London, 1967.

[13] T. Chinburg "Minimal models of curves over Dedekind rings" in *Arithmetic Geometry* (Cornell/Silverman ed.), Springer-Verlag, Berlin, 1986.

[14] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[15] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques" in *Modular Functions of One Variables II*, Lecture Notes in Mathematics 349, Springer-Verlag, Berlin, 1973.

[16] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[17] A. Edelman, *The mathematics of the Pentium division bug*, SIAM Rev. 39 (1997), no. 1, 54–67.

[18] S. J. Edixhoven, *L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein"*, Astérisque (1991), no. 196–197, 7–8, 159–170 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[19] ______, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Birkhäuser Boston, Boston, MA, 1991, pp. 25–39.

[20] ______, *Néron models and tame ramification*, Compositio Math., 81 (1992), 291–306.

[21] ______, *Modular parameterizations at primes of bad reduction*, in preparation.

[22] ______, *Comparison of integral structures on spaces of modular forms of weight two, and computation of spaces of forms mod* 2 *of weight one*, preprint.

[23] M. Emerton, *Optimal Quotients of Modular Jacobians*, to appear in Math. Ann.

[24] E. Freitag and R. Kiehl, *Étale cohomology and the Weil conjectures*, Springer-Verlag, Berlin, 1988.

[25] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies 131, Princeton University Press, 1993.

[26] J. González and J.-C. Lario, $\mathbf{Q}$-*curves and their Manin ideals*, Amer. J. Math. 123 (2001), no. 3, 475–503.

[27] A. Grothendieck, *Éléments de géométrie algébrique.* IV$_4$. Étude locale des schémas et des morphismes de schémas, Inst. Hautes Études Sci. Publ. Math. (1966), no. 28.

[28] ______, *Groupes de monodromie en géométrie algébrique. I*, Springer-Verlag, Berlin, 1973, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 I), Lecture Notes in Mathematics, Vol. 288.

[29] ______, *Groupes de monodromie en géométrie algébrique. II*, Springer-Verlag, Berlin, 1973, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 II), Dirigé par P. Deligne et N. Katz, Lecture Notes in Mathematics, Vol. 340.

[30] B. H. Gross, *A tameness criterion for Galois representations associated to modular forms* (*mod p*), Duke Math. J. 61 (1990), no. 2, 445–517.

[31] A. Joyce, *The Manin Constant of an Optimal Quotient of* $J_0(431)$, preprint, 2003.

[32] K. Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, 244 page preprint.

[33] N. M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. 62 (1981), no. 3, 481–502.

[34] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

[35] D. R. Kohel and W. A. Stein, *Component Groups of Quotients of* $J_0(N)$, Proceedings of the 4th International Symposium (ANTS-IV), Leiden, Netherlands, July 2–7, 2000 (Berlin), Springer, 2000.

[36] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Algebra i Analiz 1 (1989), no. 5, 171–196.

[37] D. S. Kubert and S. Lang, *Modular units*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], vol. 244, Springer-Verlag, New York, 1981.

[38] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[39] S. Lichtenbaum, *Curves over discrete valuation rings*, American Journal of Mathematics, 90 (1968), 380–405.

[40] J. Lipman, *Desingularization of two-dimensional schemes*, Annals of Mathematics, 107 (1978), 151–207.

[41] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford University Press, Oxford, 2002.

[42] Q. Liu and D. Lorenzini, *Models of curves and finite covers*, Compositio Math., 118 (1999), 61–102.

[43] W. Lütkebohmert, *On compactification of schemes*, Manuscripta Mathematica, 80 (1993), 95–111.

[44] H. Matsumura, *Commutative ring theory*, Cambridge University Press, Cambridge, 1986, Translated from the Japanese by M. Reid.

[45] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186.

[46] ———, *Rational isogenies of prime degree* (*with an appendix by D. Goldfeld*), Invent. Math. 44 (1978), no. 2, 129–162.

[47] L. Merel, *L'accouplement de Weil entre le sous-groupe de Shimura et le sous-groupe cuspidal de $J_0(p)$*, J. Reine Angew. Math. 477 (1996), 71–115.

[48] A. P. Ogg, *Hyperelliptic modular curves*, Bull. Soc. Math. France 102 (1974), 449–462.

[49] ______, *Rational points on certain elliptic modular curves*, Analytic number theory (Proc. Sympos. Pure Math., Vol XXIV, St. Louis Univ., St. Louis, Mo., 1972), Amer. Math. Soc., Providence, R.I., 1973, pp. 221–231.

[50] A. P. Ogg, *Rational points of finite order on elliptic curves*, Invent. Math. 12 (1971), 105–111.

[51] B. Poonen and M. Stoll, *The Cassels–Tate pairing on polarized abelian varieties*, Ann. of Math. (2) 150 (1999), no. 3, 1109–1149.

[52] D. Popescu, *General Néron desingularization and approximation*, Nagoya Math Journal, 104 (1986), pp. 85–115.

[53] Michel Raynaud, *Schémas en groupes de type $(p, \dots, p)$*, Bull. Soc. Math. France 102 (1974), 241–280.

[54] K. Ribet, "On the component groups and the Shimura subgroup of $J_0(N)$", exposé 6, Sém. Th. Nombres, Université Bordeaux, 1987–88.

[55] ______, "Irreducible Galois representations arising from component groups of Jacobians" in *Elliptic curves, modular forms, and Fermat's Last Theorem*, International Press, 1995.

[56] M. Schlessinger, *Infinitesimal deformations of singularities*, 1964 Harvard Ph.D. thesis, unpublished.

[57] J-P. Serre, *Groupes finis d'automorphismes d'anneaux locaux réguliers*, Colloque d'Algèbre (Paris, 1967), Exp. 8 (1968), 11.

[58] ______, *Cohomologie des groupes discrets*, Prospects in mathematics (Proc. Sympos., Princeton Univ., Princeton, N.J., 1970), Princeton Univ. Press, Princeton, N.J., 1971, pp. 77–169. Ann. of Math. Studies, No. 70.

[59] ______, *Lie algebras and Lie groups*, second ed., Lecture Notes in Mathematics, vol. 1500, Springer-Verlag, Berlin, 1992, 1964 lectures given at Harvard University.

[60] ______, *Œuvres. Collected papers. IV*, Springer-Verlag, Berlin, 2000, 1985–1998.

[61] I. Shafarevich, *Lectures on minimal models and birational transformations of two-dimensional schemes*, Tata Institute: Bombay, 1966.

[62] G. Shimura, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan 25 (1973), no. 3, 523–544.

[63] ______, *An introduction to computing modular forms using modular symbols*, to appear in an MSRI proceedings.

[64] ______, *Shafarevich-Tate groups of nonsquare order*, to appear in proceedings of MCAV 2002, Progress of Mathematics.

[65] G. Stevens, *Stickelberger elements and modular parametrizations of elliptic curves*, Invent. Math. 98 (1989), no. 1, 75–106.

[66] R. Swan, *Néron-Popescu desingularization*, Lectures in Algebra and Geometry 2, International Press, Cambridge (1998), pp. 135–192.

[67] J. Tate, *Duality theorems in Galois cohomology over number fields*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 288–295.

[68] J. Watanabe, *Some remarks on Cohen-Macaulay rings with many zero divisors and an application*, J. Algebra 39 (1976), no. 1, 1–14.

Brian Conrad
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109-1109
USA
bdconrad@umich.edu

Bas Edixhoven
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
The Netherlands
edix@math.leidenuniv.nl

William A. Stein
Department of Mathematics
Harvard University
Cambridge, MA 02138
USA
was@math.harvard.edu

15 Approximation of Eigenforms of Infinite Slope by Eigenforms of Finite Slope, with R. Coleman

Approximation of Eigenforms of Infinite Slope by Eigenforms of Finite Slope

Robert F. Coleman William A. Stein

April 21, 2003

Contents

1 Introduction

Fix a prime p. Consider a classical newform

$$F = \sum_{n\geq 1} a_n q^n \in S_k\left(\Gamma_1(Np^t), \overline{\mathbf{Q}}_p\right)$$

where k and N are positive integers and $p \nmid N$ is a prime (by a *newform* we mean a Hecke eigenform that lies in the new subspace and is normalized so that $a_1 = 1$). The *slope* of F is $\mathrm{ord}_p(a_p)$, where $\mathrm{ord}_p(p) = 1$. By [Shi94, Prop. 3.64], the twist

$$F^\chi = \sum \chi(n) a_n q^n$$

of F by any Dirichlet character χ of conductor dividing p is an eigenform on $\Gamma_1(Np^{\max\{t+1,2\}})$. This twist has infinite slope.

In Section 2, we prove that if F has finite slope then it is possible to approximate F^χ arbitrarily closely by (classical) finite slope *eigenforms*. Assuming refinements of standard conjectures, the best estimate we obtain for the smallest weight of an approximating eigenforms is exponential in the approximating modulus p^A. Section 4 contains computations that suggest that the best estimates should have weight that is linear in p^A.

One motivation for the question of approximation of infinite slope eigenforms by finite slope eigenforms is the desire to understand the versal deformation space of a residual modular representation [Maz89] (the deformation space of an irreducible representation is universal [Maz89] as is the deformation space of a residual pseudo-representation [CM98]). In [GM98] (see also [Maz97], and [Böc01] for a generalization), it was shown that the Zariski closure of the locus of finite slope modular deformations of an absolutely irreducible "totally unobstructed" residual modular representation is Zariski dense in the associated representation space but very little is known about the topological closure of this locus. For example, it is not known if it contains any nonempty open sets. Our result implies that it contains tamely ramified twists of modular deformations. We also show in Section 3.1 that a result of Hatada implies that in at least one (albeit not irreducible) case it does not contain all modular deformations.

Our investigation began with with our answer in Section 2 to a question of Jochnowitz. The idea of studying the p-adic variation of modular forms began with Serre [Ser73] and was since developed by Katz [Kat75] and Hida [Hid86] (see also [Gou88] for a sketch of the theory). It follows, in particular, from their work, that one can approximate all forms on $X_0(p^n)$ with forms on the j-line $X_0(1)$, but *not* necessarily with *eigenforms*.

We prove the above result about twists in Section 2, then state some questions about approximation by finite slope forms in Section 2.1. We explain how to reinterpret Hatada's result in Section 3.1, then present the results of our computations in Section 4.

Based on the results and computations discussed in this article, Mazur has suggested that it may be the case that an infinite slope eigenform can be approximated by finite slope eigenforms only if the corresponding representation is what he calls *tamely semistable* (i.e., semistable, in the sense of [CF00], after a tame extension).

Acknowledgments. The authors thank Naomi Jochnowitz for provoking this line of thought and for interesting conversations, Barry Mazur for helpful comments and questions, Frank Calegari for conversations, Loïc Merel for his comments on an early draft of this paper, and the referee for a brilliant report.

2 Approximating Teichmüller Twists of Finite Slope Eigenforms

This section is the theoretical heart of the paper. We prove that the infinite slope eigenforms obtained as twists of finite slope eigenforms by powers of the Teichmüller character can always be approximated by finite slope eigenforms. We first show that certain overconvergent eigenforms of sufficiently close weight are congruent and have the same slope. Then we use the θ operator on overconvergent forms to deduce the main result (Theorem 2.1) below.

Let p be a prime. All eigenforms in this section will be cusp forms with coefficients in $\overline{\mathbf{Q}}_p$ normalized so that $a_1 = 1$. Suppose $F = \sum_{n\geq 1} a_n q^n$ is an eigenform and $\chi : (\mathbf{Z}/M\mathbf{Z})^* \to \mathbf{C}_p^*$ is a Dirichlet character with modulus M, which we extend to $\mathbf{Z}/M\mathbf{Z}$ by setting $\chi(n) = 0$ if $(n, M) \neq 1$. Then the twist of F by χ is the eigenform

$$F^{\chi} = \sum_{n\geq 1} \chi(n) a_n q^n.$$

Let $\omega\colon (\mathbf{Z}/p\mathbf{Z})^* \to \mathbf{Z}_p^*$ be the Teichmüller character (so $\omega(n) \equiv n \pmod{p}$). The following theorem concerns finite slope approximations of twists of F by powers of ω. For example, it concerns the twist

$$F^{\omega^0} = \sum_{(n,p)=1} a_n(F) q^n$$

of F by the trivial character mod p, which we call the "p-deprivation" of F and which has infinite slope.

Theorem 2.1. *Suppose F is a classical eigenform on $X_1(Np^t)$, $t \geq 1$, over $\overline{\mathbf{Q}}_p$ of weight k, character ψ, and finite slope at p. Let $A \in \mathbf{Z}_{>0}$ and $r, s \in \mathbf{Z}_{\geq 0}$ with $r, s < p-1$. Then there exists a classical finite slope eigenform G on $X_1(Np^t)$ with $G(q) \equiv F^{\omega^r}(q) \pmod{p^A}$ such that G has weight congruent to $k + 2r - s$ modulo $p-1$ and character $\psi \cdot \omega^s$.*

(The slope of G will be at least A, since the pth Fourier coefficient of F^{ω^r} is 0.)

Let $\mathbf{q} = 4$ if $p = 2$ and p otherwise. Let $\tau : \mathbf{Z}_p^* \to \mathbf{C}_p^*$ be the character of finite order such that $a \equiv \tau(a) \pmod{\mathbf{q}}$. We only need to assume that $F = \sum_{n\geq 1} a_n q^n$ is an overconvergent eigenform of tame level N of finite slope with arithmetic weight-character $\kappa\colon a \to \chi(a)\langle\langle a\rangle\rangle^k$, where χ is a character of finite order whose conductor divides Np^t, k is a possibly negative integer, and $\langle\langle a\rangle\rangle = a/\tau(a)$. (For example, if F is a classical eigenform of weight k and character ψ, then $\chi = \psi\omega^k$.) Recall that the collection of continuous characters on $\mathbf{Z}_p^*$ is a metric space, with

$$d(\rho, \psi) = \max\{|\rho(a) - \psi(a)| : a \in \mathbf{Z}_p^*\},$$

where $|\,|$ is the absolute value on $\mathbf{C}_p$ normalized so that $|p| = 1/p$. We need,

Proposition 2.2. *Suppose $L \in \mathbf{Z}_{\geq 0}$ and H is an overconvergent eigenform of tame level N, finite slope and weight-character κ. Then if γ is a weight-character sufficiently close to κ there exists an overconvergent eigenform R of weight-character γ with the same slope as H such that*

$$H(q) \equiv R(q) \pmod{p^L}.$$

Proof. We will use the notation of the "*R-families*" section (in §B5) of [Col97b]. In particular, B is an affinoid disk in weight space containing κ and X is an affinoid finite over B such that $A(X)$ is generated by the images of the "Hecke operators" $T(n)$. Moreover, if $x \in X$ and $\eta_x\colon A(X) \to \mathbf{C}_p$ is the corresponding homomorphism, then

$$F_x(q) = \sum_{n \geq 1} \eta_x(T(n))q^n$$

is the q-expansion of an overconvergent finite slope eigenform and finally there is a point $y \in X$ such that $F_y(q) = H(q)$. Note that X is a subdomain of the eigencurve of tame level N (although the eigencurves of level $N > 1$ are not yet defined in the literature).

The ring $A^0(X)$ is finite over $A^0(B)$ by Corollary 6.4.1/5 of [BGR84]. Let $f_1, \ldots, f_n$ be generators. Let f_0 be a uniformizing parameter on B so that $A(B) = \mathbf{C}_p\langle f_0\rangle$, where $\mathbf{C}_p\langle f_0\rangle$ is the ring of power series in f_0 whose coefficients tend to 0 with their degree. Let $Z_L(y)$ be the following Weiersträss subdomain of X:

$$\{x \in X\colon |f_i(x) - f_i(y)| \leq p^{-L}, 0 \leq i \leq n\}.$$

Since the functions $x \to \eta_x(T(n))$ lie in $A^0(X)$, it follows that if $x \in Z_L(y)$, then

$$F_x(q) \equiv H(q) \pmod{p^L}.$$

Finally, since $Z_L(y)$ is a subdomain of X and X is finite over B, the map from $Z_L(y)$ to B is quasi-finite. It follows from Proposition A5.5 of [Col97b] that its image in B is a subdomain. Since κ is the image of y, its image contains a disk around y. □

Proof of Theorem 2.1. Let α be the slope of F. It follows from Proposition 2.2 that if $m \in \mathbf{Z}$ is sufficiently small p-adically there exists an overconvergent eigenform K of tame level N, weight-character $\chi \cdot \langle\langle\ \rangle\rangle^{k-m}$ and slope α such that $K(q) \equiv F(q)$ (mod p^A). Suppose $m \geq k$. Then, by Proposition 4.3 of [Col96] (see also [Col97a]) if $F_1 = \theta^{m-k+1}K$, then F_1 is an overconvergent eigenform of weight-character

$$\kappa_1 := \omega^{2(m-k+1)} \cdot \chi \cdot \langle\langle\ \rangle\rangle^{k_1},$$

where $k_1 = m - k + 2$, and F_1 has finite slope $\alpha_1 = \alpha + m - k + 1$. Applying this same process to F_1, for $\ell \in \mathbf{Z}$ sufficiently small p-adically such that $\ell \geq k_1$, we obtain an overconvergent finite slope eigenform F_2 of weight-character κ_2, where $\kappa_2 = \omega^{2\ell} \cdot \chi \cdot \langle\langle\ \rangle\rangle^{k_2}$ and where $k_2 = \ell - k_1 + 2 = k + \ell - m$, such that if $F_2(q) = \sum_{n\geq 1} b_n q^n$, then

$$\begin{aligned} b_n &\equiv n^{\ell-k_1+1} n^{m-k+1} a_n \\ &\equiv n^\ell a_n \pmod{p^A}. \end{aligned}$$

The latter is congruent to $\omega^r(n)a_n \pmod{p^A}$ if $\ell \equiv r \pmod{\varphi(p^A)}$ and $\ell + v(a_p) \geq A$. It follows from [Col96, §8], [Col97a], and [Col97b] that if c is an integer sufficiently small p-adically, such that $c + k_2 > v(b_p) + 1$ (note that $v(b_p)$ is the slope of F_2 so is finite) there exists a classical eigenform G on $X_1(Np^t)$ of weight $k_2 + c = k + \ell - m + c$, slope $v(b_p)$ and character $\omega^{m+r-c} \cdot \psi$ such that $G(q) \equiv F_2(q) \equiv F^{\omega^r}(q) \pmod{p^A}$. We can choose c so that $m + r - c \equiv s \pmod{p-1}$ and then $k_2 + c \equiv k + 2r - s \mod (p-1)$. □

The following corollary addresses a question of Jochnowitz, which motivated this entire investigation:

Corollary 2.3. *Suppose R is a classical eigenform of weight k on $X_1(N)$, let $A \in \mathbf{Z}_{>0}$, and let $r \in \mathbf{Z}_{\geq 0}$ with $r < p - 1$. Then there exists a classical eigenform S on $X_1(N)$ of weight congruent to $k + 2r$ modulo $p - 1$ such that $S(q) \equiv R^{\omega^r}(q)$ $\pmod{p^A}$.*

Proof. Suppose the F in Theorem 2.1 is one of the old eigenforms associated to R on $X_1(Np)$ and $s = 0$. Let G be a classical eigenform of weight $c + k_2$ as mentioned in the proof of the theorem, but suppose $c + k_2 > 2v(b_p) + 1$. Then G is old of weight congruent to $k \mod (p-1)$ and G is congruent to an eigenform S of the same weight on $X_1(N)$ modulo $p^{v(b_p)}$. Since $b_p \equiv 0 \pmod{p^A}$, we obtain the corollary. □

Remark 2.4. Assuming a natural refinement of the Gouvêa-Mazur conjectures, the best estimate we obtain for the weight of H in the above proof is exponential in p^A. Computational evidence suggests that the best estimates should have weights that are linear in p^A (see Section 4).

Remark 2.5. Jochnowitz and Mazur have independently observed that the above argument can be used to prove the following result: *Suppose F is an overconvergent eigenform of arithmetic weight-character κ, which is a limit of overconvergent eigenforms of finite slope. If $\iota \colon \mathbf{Z}_p^* \to \mathbf{Z}_p^*$ is the identity character, then the twist $F^{\iota/\kappa}(q)$ of F by ι/κ, which is the q-expansion of a convergent eigenform of weight-character ι^2/κ, is the limit of overconvergent eigenforms of finite slope.*

Remark 2.6. One can also approach the p-deprivation (the twist by the 0th power of Teichmüller) of a finite slope eigenform F by using the evil twins of eigenforms approaching F.

2.1 Questions

Some natural questions arise:

1. Is every p-adic convergent eigenform which is the limit of finite slope overconvergent eigenforms an overconvergent eigenform? (We can show the twist of an overconvergent eigenform by a Dirichlet character is overconvergent.)

2. Which infinite slope eigenforms are limits of finite slope eigenforms?

3. If $F(q)$ is the q-expansion of an overconvergent eigenform of weight-character κ, is $F^{\iota/\kappa}(q)$ the q-expansion of an overconvergent eigenform of weight-character

ι^2/κ (recall that ι is the identity character $\mathbf{Z}_p^* \xrightarrow{\sim} \mathbf{Z}_p^*$)? Another closely related question is as follows: Suppose ρ is the representation of the absolute Galois group of $\mathbf{Q}$ attached to an overconvergent eigenform and let χ denote the cyclotomic character. Then is the representation $\rho \otimes \chi \cdot \det(\rho)^{-1}$ attached to an overconvergent eigenform?

3 An Infinite Slope Eigenform that is Not Approximable

In Section 3.1, we prove an extension to higher level of a theorem of Hatada about the possibilities for systems of Hecke eigenvalues modulo 8. We use this result to deduce that the normalized weight 2 cusp form on $X_0(32)$ is not 2-adically approximable by normalized eigenforms of tame level 1 and finite slope. In Section 3.2 we give an example of an infinite slope eigenform of level 27 that computer computations suggest cannot be approximated by finite slope forms. For related investigations, see [CE03].

3.1 An Extension of a Theorem of Hatada

Theorem 3.1. *If $F = \sum a_n q^n$ is a normalized cuspidal newform over $\mathbf{C}_2$ of finite slope on $X_0(2^n)$, then $a_2 \equiv 0 \pmod 8$ and $a_p \equiv p+1 \pmod 8$ for all odd primes p.*

Proof. Suppose F has weight k and finite slope α. The assumption that F has finite slope implies $n \leq 1$. If $n = 0$ the assertion of Theorem 3.1 was proved by Hatada in [Hat79], so we may assume that $n = 1$ and $\alpha = (k-2)/2$ (in general, the slope of a newform on $\Gamma_0(p)$ of weight k is $(k-2)/2$). Note that $\alpha \geq 3$ since there are no newforms on $X_0(2)$ of weight < 8. It follows from Theorems A of [Col97b] (see §B2 of [Col97b] for the extension to $p = 2$) and Theorem B5.7 of [Col97b] that if j is an integer sufficiently close 2-adically to k, then there exists a classical normalized cuspidal eigenform G on $X_0(2)$ of weight j and slope α such that

$$G(q) \equiv F(q) \pmod 8.$$

If in addition we assume that $j > 2(\alpha + 1)$, then G must be old (since the slope of a newform of weight j is $(j-2)/2 \neq \alpha$). Thus there is a cuspidal eigenform $H = \sum b_n q^n$ of level 1 such that G is a linear combination of $H(q)$ and $H(q^2)$. More precisely,

$$G(q) = H(q) - \rho H(q^2)$$

where ρ is a root of $P(X) = X^2 - b_2 X + 2^{j-1}$. By Hatada's theorem $\mathrm{ord}_2(b_2) \geq 3$, and $j \geq 12$, so the slopes of the Newton polygon of $P(X)$ at 2 are both at least 3. Thus $G(q) \equiv H(q) \pmod 8$, which proves the theorem because H has level 1. □

Corollary 3.2. *Let G be the normalized weight 2 cusp form on $X_0(32)$. Then G is not 2-adically approximable by normalized eigenforms of tame level 1 and finite slope.*

Proof. If F_{32} were approximable there would have to be a normalized eigenform F on $X_0(2)$ such that $F_{32}(q) \equiv F(q) \pmod 8$. However, $F_{32}(q) = \sum_{n=1}^{\infty} a_n q^n$ where,

$$a_p = \begin{cases} 2x & \text{if } p = x^2 + y^2, \quad \text{written so } x + y \equiv x^2 \pmod 4 \\ 0 & \text{otherwise.} \end{cases}$$

As $a_3 = 0 \not\equiv 4 \pmod 8$, we see from Theorem 3.1 that F does not exist. □

Remark 3.3. If $p \equiv 1 \pmod 4$ then the coefficient of a_p in F_{32} agrees modulo 8 with $p + 1$. If p is 3 mod 4 it does not because for F_{32} the coefficient vanishes. What is happening is that there is a reducible mod 8 pseudo-representation (namely the trivial one-dimensional representation plus the cyclotomic character) such that any finite slope level 2^n form gives this pseudo-representation mod 8. Conversely the mod 8 representation associated to F_{32} is the direct sum of the quadratic character associated to $\mathbf{Q}(i)$ and the cyclotomic character. Hence the congruence works when $p = 1 \mod 4$ but not otherwise.

3.2 Another Eigenform that Conjecturally Cannot be Approximated

In this section we consider an infinite slope eigenform that is not a Teichmüller twist of a finite slope eigenform. We conjecture that this eigenform cannot be approximated arbitrarily closely by finite slope eigenforms.

Conjecture 3.4. *There are exactly five residue classes in* $(\mathbf{Z}/9\mathbf{Z})[[q]]$ *of normalized eigenforms in* $S_k(\Gamma_0(N))$ *where* $k \geq 1$ *and* $N = 1, 3, 9$. *They are given in the following table, where the indicated weight is the smallest weight where that system of eigenvalues occurs (the level is* 1 *in each case):*

Weight	[$a_2, a_3, \ldots, a_{43}$ mod 9]
12	[3, 0, 6, 5, 3, 8, 0, 2, 6, 3, 8, 2, 6, 5]
16	[0, 0, 0, 2, 0, 2, 0, 2, 0, 0, 2, 2, 0, 2]
20	[6, 0, 3, 8, 6, 5, 0, 2, 3, 6, 5, 2, 3, 8]
24	[6, 0, 3, 5, 6, 8, 0, 2, 3, 6, 8, 2, 3, 5]
32	[3, 0, 6, 8, 3, 5, 0, 2, 6, 3, 5, 2, 6, 8]

The system of eigenvalues mod 9 *associated to the weight* 2 *form* F *on* $X_0(27)$ *is*

```
[ 0, 0, 0, 8, 0, 5, 0, 2, 0, 0, 5, 2, 0, 8 ],
```

so we conjecture that there is no eigenform f *on* $\Gamma_0(N)$ *with* $N \mid 9$ *such that* $f \equiv F \pmod 9$.

As evidence, we verified that each of the mod 9 reductions of each newform of level 1 and weight $k \leq 74$ has one of the five systems of Hecke eigenvalues listed in the table. We also verified that all newforms of levels 3 and 9 and weight $k \leq 40$ have corresponding system of eigenvalues mod 9 in the above table. We checked using the method described in Section 4 that there is no newform of level 1 with weight $k \leq 300$ that approximates the weight 2 form on $X_0(27)$ modulo 9.

We now make some remarks about pseudo-representations when $p = 3$. Let

$$\chi : \mathbf{Z}/27\mathbf{Z} \to \mathbf{Z}/9\mathbf{Z}$$

be the mod 9 cyclotomic character, so χ has order 6 and if $\gcd(n, 3) = 1$ then $\chi(n) = n \in \mathbf{Z}/9\mathbf{Z}$. The pseudo-representation corresponding to a form of weight k giving the system of eigenvalues in the table in Conjecture 3.4 are

Weight	Pseudo-representation
12	$\chi^2 \oplus \chi^3$
16	$1 \oplus \chi^3$
20	$\chi^3 \oplus \chi^4$
24	$1 \oplus \chi^5$
32	$1 \oplus \chi$
$S_2(\Gamma_0(27))$	$\chi^2 \oplus \chi^5$

Note that the square of any pseudo-representation of level 1 in the above table has 1 as an eigenvalue, but the square of the pseudo-representation attached to $S_2(\Gamma_0(27))$ does not have 1 as an eigenvalue. Also,

$$F \equiv f_{16} \otimes \chi^2 \pmod{9},$$

where f_{16} is of weight 16. The order of χ^2 is 3, so χ^2 is not a power of the Teichmüller character (which has order 2) and Theorem 2.1 does not apply.

Further computations *suggest* that the pseudo-representations attached to forms of level 1 with coefficients in $\mathbf{Z}_9$ are

Weight	Pseudo-representations
$k \equiv 0 \pmod{6}$	$1 \oplus \chi^5, \quad \chi^2 \oplus \chi^3$
$k \equiv 2 \pmod{6}$	$1 \oplus \chi, \quad \chi^3 \oplus \chi^4$
$k \equiv 4 \pmod{6}$	$1 \oplus \chi^3$

The pseudo-representations attached to forms of level 27 with coefficients in $\mathbf{Z}_9$ seem to be

Weight	Pseudo-representations
$k \equiv 0 \pmod{6}$	$\chi \oplus \chi^4$
$k \equiv 2 \pmod{6}$	$\chi^2 \oplus \chi^5$
$k \equiv 4 \pmod{6}$	$\chi \oplus \chi^2, \quad \chi^4 \oplus \chi^5$

Also note that if $\chi^i \oplus \chi^j$ is one of the pseudo-representations of level 27 in the table, then the sum of the orders of χ^i and χ^j is 9, whereas at level 1 the sum of the orders is at most 7.

4 Computations About Approximating Infinite Slope Eigenforms

In this section, we investigate computationally how well certain infinite slope form can be approximated by finite slope eigenforms.

4.1 A Question About Families

The following question is an analogue of [GM92, §8] but for eigenforms of infinite slope. Fix a prime p and an integer N with $(N,p)=1$.

Question 4.1. Suppose $f \in S_{k_0}(\Gamma_0(Np^r))$ is an eigenform having infinite slope (note that f need not be a newform). Is there a "family" of eigenforms $\{f_k\}$, with $f_k \in S_k(\Gamma_0(Np))$, where the weights k run through an arithmetic progression

$$k \in \mathcal{K} = \{k_0 + mp^{\nu}(p-1) \text{ for } m = 1,2,\ldots\}$$

for some integer ν, such that

$$f_k \equiv f \pmod{p^n},$$

where $n = \mathrm{ord}_p(k-k_0)+1$? (When $p=2$ set $n=\mathrm{ord}_2(k-k_0)+2$.)

Our question differs from the one in [GM92, §8] because there the form being approximated has finite slope, whereas our form f does not. We know, as discussed in the previous section, that our question sometimes has a negative answer since it might not be possible to approximate f at all.

4.2 An Approximation Bound

Let

$$f = \sum_{n\geq 1} a_n q^n \in K[[q]]$$

be a q-expansion with coefficients that generate a number field K. Fix a prime p and an even integer $k \geq 2$. In order to gather some data about Question 4.1, we now define a reasonably easy to compute upper bound on how well f can be approximated by an eigenform in $S_k(\Gamma_0(p))$. Suppose $\ell \geq 1$, let F be the characteristic polynomial of T_ℓ acting on the space $S_k(\Gamma_0(p))$ of classical cusp forms of weight k and tame level 1, and let H be the characteristic polynomial of $a_\ell \in K$. Let G be the resultant of $F(Y)$ and $H(X+Y)$ with respect to the variable Y, normalized so that G is monic. Thus the roots of G are the differences $\alpha - \beta$ where α runs through the roots of F and β runs through the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of a_ℓ. We can easily compute the p-valuations of the roots of G without finding the roots, because the p-valuations of the roots are the slopes of the newton polygon of G. Let $m_\ell \in \mathbf{Q}\cup\{\infty\}$ be the *maximum* of the slopes of the Newton polygon of G. Let

$$c_k(f,r) = \min\{m_\ell : \ \ell \leq r \text{ is prime}\}.$$

We note that computing $c_k(f,r)$ requires knowing only the characteristic polynomials of Hecke operators T_ℓ on $S_k(\Gamma_0(p))$ and of a_ℓ for primes $\ell \leq r$.

Proposition 4.2. *If there is a normalized eigenform $g \in S_k(\Gamma_0(p))$ such that $f \equiv g \pmod{p^A}$, then $A \leq c_k(f,r)$ for any r.*

Proof. To see this observe that $c_k(f,r)$ is the minimum of the

$$\mathrm{ord}_p(a_n(f) - a_n(g))$$

where $1 \leq n \leq r$ and g runs through all normalized eigenforms in $S_k(\Gamma_0(p))$, and we run through all possible embeddings of f and g into $\overline{\mathbf{Z}}_p[[q]]$. □

The motivation for our definition of $c_k(f,r)$ is that it is straightforward to compute from characteristic polynomials of Hecke operators, even when the coefficients of f lie in a complicated number field. The number $c_k(f,r)$ could overestimate the true extent to which f is approximated by an eigenform in $S_k(\Gamma_0(p))$ in at least two ways:

1. There is an $r' > r$ such that $c_k(f,r') < c_k(f,r)$.

2. No *single* eigenform g is congruent to f, but each coefficient of f is congruent to some coefficient of some eigenform g.

4.3 Some Data About Approximations

Let p be a prime and $f \in S_{k_0}(\Gamma_0(p^r))$ be a newform of infinite slope. Suppose that the answer to Question 4.1 for f is yes. If k is a weight (in the arithmetic progression) then there should be an eigenform $f_k \in S_k(\Gamma_0(p))$ such that $f_k \equiv f \pmod{p^{n+1}}$ where $n = \mathrm{ord}_p(k-k_0)$. Thus we should have

$$\mathrm{ord}_p(k-k_0) + 1 \leq c_k(f,r)$$

for all $r > 1$ and all k in an arithmetic progression $\mathcal{K} = \{k_0 + mp^\nu(p-1) \text{ for } m = 0, 1, 2, \ldots\}$. (When $p = 2$ we should have $\mathrm{ord}_2(k-k_0) + 2 \leq c_k(f,r)$.)

The following or the results of some computations of $c_k(f,r)$.

$\mathbf{p = 2}$:

1. For $k_0 = 6, 10, 12, 14, 16, 20$ let $f \in S_{k_0}(\Gamma_0(4))$ be the unique newform. Then for all k with $k_0 < k \leq 100$ we have $c_k(f, 47) = \mathrm{ord}_2(k-k_0) + 2$.

2. For $k_0 = 18, 22$ let $f \in S_{k_0}(\Gamma_0(4))$ be the unique, up to Galois conjugacy, newform. Then for all k with $k_0 < k \leq 100$ we have $c_k(f,7) = \mathrm{ord}_2(k-k_0)+2$.

3. Let $f \in S_4(\Gamma_0(8))$ be the unique newform. For most $4 < k \leq 100$ we have $c_k(f, 47) = \mathrm{ord}_2(k-k_0) + 2$. However, in this range if $\mathrm{ord}_2(k-k_0) \geq 4$ then $c_k(f, 47) = 5$ Since $\mathrm{ord}_2(68-4) + 2 = 8$, this is a problem; perhaps this form is not approximated. Very similar behavior occurs for the newforms in $S_6(\Gamma_0(8))$, $S_8(\Gamma_0(8))$, and $S_4(\Gamma_0(16))$.

4. For the two newforms $f \in S_6(\Gamma_0(16))$, we have $c_k(f, 47) \leq 3$ for all $k < 100$, so these f probably can not be approximated by finite slope forms.

5. Let f be the 2-deprivation of the unique normalized eigenform in $S_{k_0}(\Gamma_0(1))$ for $k_0 = 12, 16, 18, 20, 22, 26$. Then $c_k(f, 47) = \mathrm{ord}_2(k - k_0) + 2$ for $12 < k \leq 100$. Same statement for $k_0 = 24, 28$ for the 2-deprivation of one of the Galois conjugates and $c_k(f, 47)$ replaced by $c_k(f, 7)$.

$\mathbf{p = 3}$:

1. Suppose f is a newform in $S_{k_0}(\Gamma_0(9))$ for $k_0 \leq 12$. Then for $k_0 < k \leq 100$ we have $c_k(f, 47) = \mathrm{ord}_3(k - k_0) + 1$, except possibly for the nonrational form of weight 8, where we have only checked that $c_k(f, 7) \geq \mathrm{ord}_3(k - k_0) + 1$.

2. Let f be the twist of a newform in $S_{k_0}(\Gamma_0(1))$ by ω_3 for $k_0 \leq 32$. Then $c_k(f, 7) \geq \mathrm{ord}_3(k - k_0) + 1$ for $k_0 < k \leq 100$, with equality usually.

3. Let f be the newform in $S_2(\Gamma_0(45))$ of tame level 5. Then $c_{2+(3-1)3^n}(f, 7) = n+1$ for $n = 0, 1, 2, 3$ (here we are testing congruences with forms in $S_k(\Gamma_0(15))$).

$\mathbf{p = 5}$:

1. Let $f = q + q^2 + \cdots \in S_4(\Gamma_0(25))$ be a newform. Then $c_{4+4}(f, 7) = 1$, $c_{4+4\cdot 5}(f, 7) = 2$, and $c_{4+4\cdot 5^2}(f, 7) = 3$. Same result for the newform $f = q + 4q^2 + \cdots \in S_4(\Gamma_0(25))$.

2. Let $f = q - q^2 + \cdots \in S_2(\Gamma_0(2 \cdot 25))$. Then $c_{2+4}(f, 7) = 1$ and $c_{2+4\cdot 5}(f, 7) = 2$, where we are testing congruences with forms in $S_k(\Gamma_0(10))$.

3. Let f be one of the newforms in $S_2(\Gamma_0(5^3))$ defined over a quadratic extension of $\mathbf{Q}$. Then $c_{2+4}(f, 7) = c_{2+4\cdot 5}(f, 7) = c_{2+4\cdot 5^2}(f, 2) = 1/2$. Thus it seems unlikely that f can be approximated by forms of finite slope.

$\mathbf{p = 7}$:

1. Let $f \in S_2(\Gamma_0(49))$ be the newform. Then $c_{2+6}(f, 7) = 1$ and $c_{2+6\cdot 7}(f, 7) = 2$. Same statement for the form $f = q - q^2 \in S_4(\Gamma_0(49))$ at weights $4 + 6$ and $4 + 6 \cdot 7$.

The data and results of this paper suggests the following:

Guess 4.3. Let p be a prime and N an integer coprime to p. Then the eigenforms on $X_0(Np^t)$ that can be approximated by finite-slope eigenforms are exactly the eigenforms on $X_0(Np^2)$. Suppose f is an infinite slope eigenform that can be approximated by finite slope eigenforms and f has weight k_0. Then for any $k > k_0$ with $k \equiv k_0 \pmod{p-1}$, there is an eigenform f_k on $X_0(Np)$ of weight k such that $f \equiv f_k \pmod{p^n}$ where $n = \mathrm{ord}_p(k - k_0) + 1$ (or $+2$ if $p = 2$). (In general one might have to restrict to n sufficiently large.)

References

[Böc01] G. Böckle, *On the density of modular points in universal deformation spaces*, Amer. J. Math. **123** (2001), no. 5, 985–1007.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[BGR84] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis: A systematic approach to rigid analytic geometry*, Springer-Verlag, Berlin, 1984.

[CE03] F. Calegari and M. Emerton, *The Hecke Algebra* $\mathbf{T}_k$ *has Large Index*, Preprint, 2003.

[CM98] R. Coleman and B. Mazur, *The Eigencurve*, Galois representations in arithmetic algebraic geometry (Durham, 1996), Cambridge Univ. Press, Cambridge, 1998, pp. 1–113.

[Col96] R. F. Coleman, *Classical and overconvergent modular forms*, Invent. Math. **124** (1996), no. 1-3, 215–241.

[Col97a] R. F. Coleman, *Classical and overconvergent modular forms of higher level*, J. Théor. Nombres Bordeaux **9** (1997), no. 2, 395–403.

[Col97b] R. F. Coleman, *p-adic Banach spaces and families of modular forms*, Invent. Math. **127** (1997), no. 3, 417–479.

[CF00] P. Colmez and J.-M. Fontaine, *Construction des représentations p-adiques semi-stables*, Invent. Math. **140** (2000), no. 1, 1–43.

[GM92] F. Gouvêa and B. Mazur, *Families of modular eigenforms*, Math. Comp. **58** (1992), no. 198, 793–805.

[GM98] F. Q. Gouvêa and B. Mazur, *On the density of modular representations*, Computational perspectives on number theory (Chicago, IL, 1995), Amer. Math. Soc., Providence, RI, 1998, pp. 127–142.

[Gou88] F. Q. Gouvêa, *Arithmetic of p-adic modular forms*, Springer-Verlag, Berlin, 1988.

[Hat79] K. Hatada, *Eigenvalues of Hecke operators on* $\mathrm{SL}(2,\mathbf{Z})$, Math. Ann. **239** (1979), no. 1, 75–96.

[Hat01] K. Hatada, *On classical and l-adic modular forms of levels* Nl^m *and* N, J. Number Theory **87** (2001), no. 1, 1–14.

[Hid86] H. Hida, *Iwasawa modules attached to congruences of cusp forms*, Ann. Sci. École Norm. Sup. (4) **19** (1986), no. 2, 231–273.

[Kat73] N. M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 69–190. Lecture Notes in Mathematics, Vol. 350.

[Kat75] N. M. Katz, *Higher congruences between modular forms*, Ann. of Math. (2) **101** (1975), 332–367.

[Maz89] B. Mazur, *Deforming Galois representations*, Galois groups over $\mathbf{Q}$ (Berkeley, CA, 1987), Springer, New York, 1989, pp. 385–437.

[Maz97] B. Mazur, *An "infinite fern" in the universal deformation space of Galois representations*, Collect. Math. **48** (1997), no. 1-2, 155–193, Journées Arithmétiques (Barcelona, 1995).

[Ser73] J-P. Serre, *Formes modulaires et fonctions zêta p-adiques*, Proceedings of the International Summer School, University of Antwerp, RUCA, July 17–August 3, 1972 (Berlin), Springer, 1973, pp. 191–268. Lecture Notes in Math., Vol. 350.

[Shi94] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

[Stu87] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin, 1987, pp. 275–280.

16 ShafarevichTate Groups of Nonsquare Order

Shafarevich–Tate Groups of Nonsquare Order

William A. Stein

Abstract. Let A denote an abelian variety over $\mathbb{Q}$. We give the first known examples in which $\#Ш(A/\mathbb{Q})$ is neither a square nor twice a square. For example, let E be the elliptic curve $y^2 + y = x^3 - x$ of conductor 37. We prove that for every odd prime $p < 25000$ (with $p \neq 37$), there is a twist A of $E \times \cdots \times E$ ($p-1$ copies) such that $\#Ш(A/\mathbb{Q}) = pn^2$ for some integer n. We prove this by showing under certain hypothesis on E and p that there is an exact sequence

$$0 \to E(\mathbb{Q})/pE(\mathbb{Q}) \to Ш(A/\mathbb{Q})[p^\infty] \to Ш(E/K)[p^\infty] \to Ш(E/\mathbb{Q})[p^\infty] \to 0,$$

where K is a certain abelian extension of $\mathbb{Q}$ of degree p.

1. Introduction

The Shafarevich–Tate group of an abelian variety A over a number field F is

$$Ш(A/F) := \operatorname{Ker}\left(H^1(F, A) \to \bigoplus_{\text{all } v} H^1(F_v, A)\right).$$

What are the possibilities for the group structure of $Ш(A/F)$? It is conjectured that $Ш(A/F)$ is finite and this is known in some cases.

Theorem 1.1 (Kato, Kolyvagin, Wiles, et al.)**.** *Suppose A is an elliptic curve over $\mathbb{Q}$. (1) If $\operatorname{ord}_{s=1} L(A, s) \leq 1$, then $Ш(A/\mathbb{Q})$ is finite. (2) If χ is a character of the Galois group of an abelian extension K of $\mathbb{Q}$ and $L(A, \chi, 1) \neq 0$, then the χ-component of $Ш(A/K) \otimes_{\mathbb{Z}} \mathbb{Z}[\chi]$ is finite. (Here $\mathbb{Z}[\chi]$ is generated by the image of χ.)*

The Cassels–Tate pairing $Ш(A/F) \times Ш(A^\vee/F) \to \mathbb{Q}/\mathbb{Z}$ imposes strong constraints on the structure of $Ш(A/F)$.

Theorem 1.2 (Tate, Flach)**.** *Let p be a prime and suppose that there is a polarization $\lambda : A \to A^\vee$ of degree coprime to p. If $p = 2$ assume also that λ arises from an F-rational divisor on A (this hypothesis is automatic if A is an elliptic curve, but can fail in general). If $Ш(A/F)[p^\infty]$ is finite then $\#Ш(A/F)[p^\infty]$ is a perfect square.*

Proof. If λ is F-rational, the Cassels–Tate pairing on $Ш(A/F)[p^\infty]$ (induced by λ) is nondegenerate and alternating (see [Tat63]), so $\#Ш(A/F)[p^\infty]$ is a perfect square. Even when λ is not F-rational, the Cassels–Tate pairing is nondegenerate and antisymmetric (see [Fla90]), which when p is odd implies that $\#Ш(A/F)[p^\infty]$ is a perfect square. □

It is tempting to conjecture that $\#Ш(A/F)$ is always a perfect square. Perhaps squareness is a fundamental property of Shafarevich–Tate groups? While implementing algorithms based on [PS97] for computing with Jacobians of hyperelliptic curves, M. Stoll was shocked to discover an example of an abelian variety of dimension two such that $\#Ш(A/F)[2^\infty] = 2$. This was surprising because, for example, one finds in the literature [SD67, pg.149] the following statement: "[The group $Ш(A/F)$] is conjectured to be finite, and Tate [26] has shown that if it is finite its order is a perfect square." Stoll and B. Poonen discovered what hid behind this and other similar examples in which $\#Ш(A/F)$ is twice a perfect square.

An algebraic curve X of genus g over a local field k is *deficient* if X has no k-rational divisor of degree $g-1$.

Theorem 1.3 (Poonen-Stoll [PS99])**.** *Suppose A is the Jacobian of an algebraic curve over F that is deficient at an odd number of places. If $\#Ш(A/F)$ is finite, then $\#Ш(A/F)$ is twice a square.*

For example, they prove that the Jacobian J of the nonsingular projective curve defined by

$$y^2 = -3(x^2+1)(x^2-6x+1)(x^2+6x+1)$$

has Shafarevich–Tate group of order 2 (to see that $\#Ш(J) \mid 2$ they observe that J is isogenous to a product of CM elliptic curves and apply a theorem of Rubin; see [PS99, Prop. 27] for details). Also, Jordan and Livné [JL99] give an infinite family of Atkin–Lehner quotients of Shimura curves which are deficient at an odd number of places.

Though $\#Ш(A/F)$ need not be square, one might still be tempted to conjecture that $Ш(A/F)$ must have order either a square or twice a square. Let p be an odd prime. In this paper, we construct (under certain hypotheses that are satisfied for $p < 25000$) abelian varieties A such that $\#Ш(A/\mathbb{Q}) = pn^2$ for some integer n. For example (see Section 3):

Theorem 1.4. *Let E be the elliptic curve $y^2 + y = x^3 - x$ of conductor 37. For every odd prime $p < 25000$ (with $p \neq 37$), there is a twist A of $E^{\times(p-1)}$ such that $\#Ш(A/\mathbb{Q}) = pn^2$ for some integer n.*

This paper was originally motivated by the problem of relating the conjecture of Birch and Swinnerton-Dyer about the ranks of elliptic curves E to the Birch and Swinnerton-Dyer formula for the orders $\#Ш(A)$ for abelian varieties A of analytic rank 0.

Let p be a prime. Under suitable hypotheses, we construct an abelian variety A and a natural map $E(\mathbb{Q})/pE(\mathbb{Q}) \hookrightarrow Ш(A/\mathbb{Q})$. Thus if $E(\mathbb{Q}) \cong \mathbb{Z}$ then $Ш(A/\mathbb{Q})$ has a natural subgroup of order p, and no other natural subgroup of order p presents itself. Moreover, when E is defined by $y^2 + y = x^3 - x$, the Birch and Swinnerton-Dyer formula predicts that $Ш(A/\mathbb{Q})[3]$ is of order 3. Further investigation led to the results of this paper.

Acknowledgement: It is a pleasure to thank Kevin Buzzard, Frank Calegari, Sol Friedberg, Benedict Gross, Emmanuel Kowalski, Barry Mazur, Bjorn Poonen, and David Rohrlich for their helpful comments, and in particular Michael Stoll for Lemma 2.10 and Cristian González for carefully reading this paper, making many comments, and sending me a proof of Proposition 2.13.

1.1. Notation

If G is an abelian group and n is an integer, then $G[n]$ denotes the subgroup of elements of order n and $G[n^\infty]$ is the subgroup of elements of order any power of n. We refer to elliptic curves using the notation of [C97].

2. Construction of Nonsquare Shafarevich–Tate Groups

For the rest of this paper we will work with an elliptic curve E over $\mathbb{Q}$. Aside from the significant use of known cases of the Birch and Swinnerton-Dyer conjecture below, much of the construction should generalize to the situation when E is replaced by a principally polarized abelian variety over a global field.

For the rest of this section, fix an elliptic curve E over $\mathbb{Q}$. By [BCDT01], E is modular so there is a newform $f = \sum_{n=1}^{\infty} a_n q^n$ of level equal to the conductor $N = N_E$ of E such that $L(E,s) = L(f,s)$. For each prime $q \mid N$, the Tamagawa number c_q of E at q is the order of the group of rational components of the special fiber of the Néron model of E at q.

2.1. Twisting By Characters of Prime Order

Let p be a prime number. For any prime $\ell \equiv 1 \pmod{p}$, let

$$\chi_{p,\ell} : (\mathbb{Z}/\ell\mathbb{Z})^* \to \mu_p \subset \mathbb{C}^*$$

be one of the $p-1$ Galois-conjugate Dirichlet characters of order p and conductor ℓ.

Conjecture 2.1. *Suppose p is a prime such that $\rho_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[p])$ is surjective. Then there exists a prime $\ell \nmid N$ such that $L(E,\chi_{p,\ell},1) \neq 0$, $\ell \equiv 1 \pmod{p}$ and $a_\ell \not\equiv \ell + 1 \pmod{p}$.*

Remarks 2.2.

1. Formulas involving modular symbols imply that $L(E,\chi_{p,\ell},1) \neq 0$ if and only if $L(E,\chi_{p,\ell}^\sigma,1) \neq 0$ for any $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugate $\chi_{p,\ell}^\sigma$ of $\chi_{p,\ell}$.
2. J. Fearnley proved related nonvanishing results when $L(E,1) \neq 0$ in [Fea01].
3. If E is the elliptic curve $y^2 + y = x^3 - x$ of conductor 37 and rank 1, then $\ell = 41$ is the only $\ell \equiv 1 \pmod{5}$ with $\ell < 1000$ for which $L(E,\chi_{5,\ell},1) = 0$.

The following proposition gives evidence for Conjecture 2.1 for the lowest-conductor elliptic curves of ranks 1, 2, and 3.

Proposition 2.3. *Conjecture 2.1 is true for the rank* 1 *elliptic curve* **37A** *for every odd* $p < 25000$ *(with* $p \neq 37$*). The conjecture is true for the rank* 2 *curve* **389A** *for every odd* $p < 1000$ *(with* $p \neq 389$*). The conjecture is true for the rank* 3 *curve* **5077A** *for every odd* $p < 1000$.

Proof. Consider the modular symbol

$$e_{p,\ell} = \sum_{a\in(\mathbb{Z}/\ell\mathbb{Z})^*} \chi_{p,\ell}(a)\cdot\left\{0,\ \frac{a}{\ell}\right\} \in H_1(X_0(N),\mathbb{Q}(\zeta_p)).$$

Then $L(E,\chi_{p,\ell},1)\neq 0$ if and only if the image of $e_{p,\ell}$ under

$$H_1(X_0(N),\mathbb{Q}(\zeta_p)) \to H_1(E,\mathbb{Q}(\zeta_p))$$

is nonzero. In any particular case, we can use modular symbols to determine whether or not this image is nonzero.

When p is large, it is difficult to compute in the field $\mathbb{Q}(\zeta_p)$, so instead we compute in the residue class field $\mathbb{F}_\ell = \mathbb{Z}[\zeta_p]/\mathfrak{m} \cong Z/\ell\mathbb{Z}$, where $\mathfrak{m}$ is one of the maximal ideals of $\mathbb{Z}[\zeta_p]$ that lies over ℓ. (Note that ℓ splits completely in $\mathbb{Z}[\zeta_p]$ because $\ell \equiv 1 \pmod{p}$.) After reducing modulo $\mathfrak{m}$, we compute the image of

$$\overline{e}_{p,\ell} = \sum_{a\in(\mathbb{Z}/\ell\mathbb{Z})^*} a^{(\ell-1)/p}\cdot\left\{0,\ \frac{a}{\ell}\right\} \in H_1(X_0(N),\mathbb{F}_\ell)$$

in $H_1(E,\mathbb{F}_\ell)$. If it is nonzero, then the image of $e_{p,\ell}$ in $H_1(E,\mathbb{Q}(\zeta_p))$ is nonzero.

A big computation (that takes hundreds of hours using Magma [BCP97]) shows that the image of $\overline{e}_{p,\ell}$ is nonzero in the cases asserted by the proposition. So the reader can carry out similar computations, we include the following Magma V2.10-6 code, which illustrates verification of the proposition for **37A** for $p < 100$:

```
procedure VerifyConjecture(E, p)
   assert Type(E) eq CrvEll;
   assert Type(p) eq RngIntElt and IsPrime(p) and IsOdd(p);
   N := Conductor(E);
   assert N mod p ne 0;
   M := ModularSymbols(E,+1);  // takes a long time if N large!
   ell := 3; t := Cputime();
   printf "p=%o: ", p;
   while true do
      while (ell mod p ne 1)  or  (N mod ell eq 0) or
       TraceOfFrobenius(ChangeRing(E,GF(ell))) mod p eq (ell+1) do
         ell := NextPrime(ell);
      end while;
      k := FiniteField(ell);
      printf "trying ell=%o...",ell;
      psi := DirichletGroup(ell,k).1;
      eps := psi^(Order(psi) div p);  // order p character
      M_k := BaseExtend(M,k);
      phi := RationalMapping(M_k);
      e := TwistedWindingElement(M_k,1,eps);
      if phi(e) ne 0 then
         printf " success! (%o seconds)\n", Cputime(t);
         return;
      end if;
      printf "failed. ";
      ell := NextPrime(ell);
   end while;
end procedure;

E := EllipticCurve([0,0,1,-1,0]);  // 37A
for p in [q : q in [3..100] | IsPrime(q) and q ne 37] do
   VerifyConjecture(E,p);
end for;
```

The above input results in the following abbreviated output:

```
p=3: trying ell=7... success! (0.021 seconds)
p=5: trying ell=11... success! (0.039 seconds)
p=7: trying ell=29... success! (0.121 seconds)
...
p=89: trying ell=179... success! (0.739 seconds)
p=97: trying ell=389... success! (1.491 seconds)
```

□

2.2. A Restriction of Scalars Exact Sequence

As above, E is an elliptic curve over $\mathbb{Q}$. Let p be any prime (note that $p = 2$ is allowed). Suppose $\ell \equiv 1 \pmod{p}$ is another prime and that $\ell \nmid N_E$. Let $K \subset \mathbb{Q}(\mu_\ell)$ be the abelian extension of $\mathbb{Q}$ that corresponds to $\chi_{p,\ell}$ (thus K is the unique subfield of $\mathbb{Q}(\mu_\ell)$ of degree p).

Let $R = \mathrm{Res}_{K/\mathbb{Q}}(E_K)$ be the restriction of scalars down to $\mathbb{Q}$ of E viewed as an elliptic curve over K. Thus R is an abelian variety over $\mathbb{Q}$ of dimension $p = [K : \mathbb{Q}]$. It is characterized by the fact that it represents the following functor on $\mathbb{Q}$-schemes S:

$$S \mapsto E_K(S_K).$$

As a Galois module,

$$R(\overline{\mathbb{Q}}) = E(\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Z}[\mathrm{Gal}(K/\mathbb{Q})],$$

where $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on $\sum P_\sigma \otimes \sigma$ by

$$\tau\left(\sum P_\sigma \otimes \sigma\right) = \sum \tau(P_\sigma) \otimes \tau_{|K} \cdot \sigma,$$

where $\tau_{|K}$ is the image of τ in $\mathrm{Gal}(K/\mathbb{Q})$.

Proposition 2.4. *The identity map induces a closed immerion $\iota : E \hookrightarrow R$, and the trace* $\mathrm{Tr} : K \to \mathbb{Q}$ *induces a surjection* $\mathrm{Tr} : R \to E$ *whose kernel is geometrically connected. Thus we have an exact sequence of abelian varieties*

$$0 \to A \to R \xrightarrow{\mathrm{Tr}} E \to 0. \tag{1}$$

Proof. The existence of ι and Tr follows from Yoneda's lemma. The map ι is induced by the functorial inclusion $E(S) \hookrightarrow E_K(S_K) = R(S)$, so ι is injective. The Tr map is induced by the functorial trace map on points $R(S) = E_K(S_K) \xrightarrow{\mathrm{Tr}} E(S)$.

To verify that Ker(Tr) is geometrically connected, we base extend the exact sequence (1) to $\overline{\mathbb{Q}}$. First, note that there is an isomorphism

$$R_{\overline{\mathbb{Q}}} \cong E_{\overline{\mathbb{Q}}} \times \cdots \times E_{\overline{\mathbb{Q}}}.$$

After base extension, we identify the trace map with the summation map

$$+ : E_{\overline{\mathbb{Q}}} \times \cdots \times E_{\overline{\mathbb{Q}}} \longrightarrow E_{\overline{\mathbb{Q}}}.$$

Let $n = [K : \mathbb{Q}]$. The map defined by

$$(a_1, \ldots, a_{n-1}) \mapsto \left(a_1, a_2, \ldots, a_{n-1}, -\sum_{i=1}^{n-1} a_i \right),$$

is an isomorphism from $E_{\overline{\mathbb{Q}}}^{\times(n-1)}$ to $\mathrm{Ker}(+) = \mathrm{Ker}(\mathrm{Tr}_{\overline{\mathbb{Q}}})$. Thus $\mathrm{Ker}(\mathrm{Tr}_{\overline{\mathbb{Q}}})$ is isomorphic to a product of copies of $E_{\overline{\mathbb{Q}}}$, and hence is connected. □

Corollary 2.5. $\iota(E) \cap \mathrm{Ker}(\mathrm{Tr}) = \iota(E)[p]$.

Proof. The composition $\mathbb{Q} \hookrightarrow K \xrightarrow{\mathrm{Tr}} \mathbb{Q}$ is multiplication by p, so the composition $E \xrightarrow{\iota} R \xrightarrow{\mathrm{Tr}} E$ is also multiplication by p. Since $\iota(E) \cap \mathrm{Ker}(\mathrm{Tr})$ is the kernel of $\mathrm{Tr} \circ \iota = [p]$, it equals $E[p]$. □

Lemma 2.6. *The abelian varieties A_K, R_K, and $(R/\iota(E))_K$ are all isomorphic to a product of copies of E_K.*

Proposition 2.7. *The exact sequence* $0 \to A \to R \to E \to 0$ *of Proposition 2.4 extends to an exact sequence* $0 \to \mathcal{A} \to \mathcal{R} \to \mathcal{E} \to 0$ *of Néron models over* $\mathbb{Z}$.

Proof. We use results of [BLR90, Ch. 7] and the fact that formation of Néron models commutes with unramified base change (see [BLR90, §1.2, Prop. 2]) to prove that for every prime q, the complex

$$0 \to \mathcal{A}_{\mathbb{Z}_q} \to \mathcal{R}_{\mathbb{Z}_q} \to \mathcal{E}_{\mathbb{Z}_q} \to 0 \tag{2}$$

is exact.

First suppose that $q \neq \ell$, and let $\mathfrak{q}$ be a prime of K lying over q. We use the fact that formation of Néron models commutes with unramified base extension and check exactness of (2) after base extension to the unramified extension $\mathcal{O}_{K,\mathfrak{q}}$ of $\mathbb{Z}_q$. By Lemma 2.6, the generic fiber of the base extension of (2) to $\mathcal{O}_{K,\mathfrak{q}}$ is

$$0 \to E_{K,\mathfrak{q}}^{\oplus(n-1)} \to E_{K,\mathfrak{q}}^{\oplus n} \xrightarrow{\Sigma} E_{K,\mathfrak{q}} \to 0.$$

Thus the corresponding complex of Néron models over $\mathcal{O}_{K,\mathfrak{q}}$ is

$$0 \to \mathcal{E}_{\mathcal{O}_{K,\mathfrak{q}}}^{\oplus(n-1)} \to \mathcal{E}_{\mathcal{O}_{K,\mathfrak{q}}}^{\oplus n} \xrightarrow{\Sigma} \mathcal{E}_{\mathcal{O}_{K,\mathfrak{q}}} \to 0,$$

which is exact, since it is exact on S-points for *any* ring S.

Suppose that $q = \ell$. Since $p \neq \ell$, [BLR90, Prop. 7.5.3 (a)] asserts that the sequence $0 \to \mathcal{A}_{\mathbb{Z}_q} \to \mathcal{R}_{\mathbb{Z}_q} \to \mathcal{E}_{\mathbb{Z}_q}$ is exact. Since $p \neq q$, the map $[p] : \mathcal{E}_{\mathbb{Z}_q} \to \mathcal{E}_{\mathbb{Z}_q}$ is an étale morphism of smooth schemes. Since E has good reduction at q, we also know that the fibers of $\mathcal{E}_{\mathbb{Z}_q}$ are geometrically connected, so $[p]$ is surjective (for more details, see the proof of [AS02, Lem. 3.2]). It follows that $\mathcal{R}_{\mathbb{Z}_q} \to \mathcal{E}_{\mathbb{Z}_q}$ is surjective.

□

2.3. The Cokernel of Trace

Let ℓ be a prime as in Conjecture 2.1. This section is devoted to computing the cokernel of the trace map $R(\mathbb{Q}) \to E(\mathbb{Q})$. Note that $R(\mathbb{Q}) = E(K)$, so this cokernel is also $E(\mathbb{Q})/\operatorname{Tr}_{K/\mathbb{Q}}(E(K))$.

Lemma 2.8. *Let K_ℓ denote the completion of K at the totally ramified prime of K lying over ℓ. Then $E(K)[p] = E(K_\ell)[p] = 0$.*

Proof. The characteristic polynomial of $\mathrm{Frob}_\ell \in \mathrm{Gal}(\mathbb{Q}_\ell^{\mathrm{ur}}/\mathbb{Q}_\ell)$ on $E[p] = E(\mathbb{Q}_\ell^{\mathrm{ur}})[p]$ is $x^2 - a_\ell x + \ell \in \mathbb{F}_p[x]$. By hypothesis $a_\ell \not\equiv \ell + 1 \pmod{p}$, so $+1$ is not a root of $x^2 - a_\ell x + \ell$ hence

$$E(\mathbb{Q}_\ell)[p] = E(\mathbb{Q}_\ell^{\mathrm{ur}})[p]^{\mathrm{Frob}_\ell - 1} = 0.$$

Since K is totally ramified at ℓ and E has good reduction at ℓ, $E(K_\ell)[p] = 0$ as well, so $E(K)[p] = 0$, as required. □

Proposition 2.9. $\mathrm{Coker}(R(\mathbb{Q}) \to E(\mathbb{Q})) \cong E(\mathbb{Q})/pE(\mathbb{Q})$.

Proof. By Corollary 2.5 the the image of $\iota(E(\mathbb{Q})) \subset R(\mathbb{Q})$ in $E(\mathbb{Q})$ is $pE(\mathbb{Q})$, so the cokernel of $R(\mathbb{Q}) \to E(\mathbb{Q})$ is a quotient of $E(\mathbb{Q})/pE(\mathbb{Q})$. Thus it suffices to prove that $R(\mathbb{Q})/\iota(E(\mathbb{Q}))$ is *finite* of order coprime to p.

We have an exact sequence $0 \to E \to R \to A' \to 0$, with A' an abelian variety that is isogenous to A (in fact, A' is the abelian variety dual of A since R is self dual, but we will not use this fact.) The L-series of A' is $\prod_{i=1}^{p-1} L(E, \chi_{p,\ell}^i, s)$, so by hypothesis $L(A', 1) \neq 0$ and it follows from Kato's theorem (see [Rub98, §8.1]) that $A'(\mathbb{Q})$ is finite. Thus $R(\mathbb{Q})/\iota(E(\mathbb{Q}))$ is finite since $R(\mathbb{Q})/\iota(E(\mathbb{Q})) \subset A'(\mathbb{Q})$. By Lemma 2.6, $A'_K \approx E_K^{\times(p-1)}$ and by Lemma 2.8 $E(K)[p] = 0$, so $A'(\mathbb{Q})[p] = 0$, which proves the proposition. □

2.4. Étale Cohomology and Shafarevich–Tate Groups

Fix an elliptic curve E over $\mathbb{Q}$ and a prime $p \nmid \prod c_{E,q}$.

In this section, we use results mostly due to Mazur to relate the Shafarevich–Tate groups of A, R, and E to certain étale cohomology groups. We maintain the notation and assumptions of the previous sections, except that we abuse notation slightly and let $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{E}$ also denote the étale sheaves on $\mathrm{Spec}(\mathbb{Z})$ defined by the Néron models $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{E}$. Let $\mathcal{B}$ be either $\mathcal{A}$, $\mathcal{R}$, or $\mathcal{E}$ and let $B = \mathcal{B}_{\mathbb{Q}}$ be the corresponding abelian variety. Let $H^q(\mathbb{Z}, \mathcal{B})$ be the qth étale cohomology group of $\mathcal{B}$.

Lemma 2.10. *There is an isomorphism* $B(\mathbb{Q}_\ell)[p] \cong \mathcal{B}(\mathbb{F}_\ell)[p]$.

Proof. This follows from [ST68, Lem. 2, pg. 495], but we sketch a proof for the convenience of the reader. Let $B^1(\mathbb{Q}_\ell)$ denote the kernel of the natural reduction map $r : B(\mathbb{Q}_\ell) \to \mathcal{B}(\mathbb{F}_\ell)$. Using formal groups and that $p \neq \ell$, one sees that $[p] : B^1(\mathbb{Q}_\ell) \to B^1(\mathbb{Q}_\ell)$ is an isomorphism. Since $\mathcal{B}$ is smooth over $\mathbb{Q}_\ell$, Hensel's lemma (see [BLR90, §2.3 Prop. 5]) implies that the reduction map is surjective, so we obtain an exact sequence

$$0 \to B^1(\mathbb{Q}_\ell) \to B(\mathbb{Q}_\ell) \to \mathcal{B}(\mathbb{F}_\ell) \to 0.$$

The snake lemma applied to the multiplication-by-p diagram attached to this exact sequence yields the exact sequence

$$0 \to B(\mathbb{Q}_\ell)[p] \to \mathcal{B}(\mathbb{F}_\ell)[p] \to 0 \to B(\mathbb{Q}_\ell)/pB(\mathbb{Q}_\ell) \to \mathcal{B}(\mathbb{F}_\ell)/p\mathcal{B}(\mathbb{F}_\ell) \to 0,$$

which proves the lemma. □

The *Tamagawa number* of B at a prime q is $c_{B,q} = \#\Phi_{B,q}(\mathbb{F}_q)$, where $\Phi_{B,q}$ is the component group of the closed fiber of the Néron model of B at q.

Lemma 2.11. $p \nmid c_{B,q}$.

Proof. First suppose $q = \ell$. The cokernel of $\mathcal{B}(\mathbb{F}_\ell) \to \Phi_{B,\ell}(\mathbb{F}_\ell)$ is contained in $H^1(\mathbb{F}_\ell, \mathcal{B}^0)$, which is 0 by Lang's theorem (see [Lan56] or [Ser88, §VI.4]), so if $\Phi_{B,\ell}(\mathbb{F}_\ell)[p] \neq 0$ then $\mathcal{B}(\mathbb{F}_\ell)[p] \neq 0$. But by Lemmas 2.6, 2.8, and 2.10,

$$\mathcal{B}(\mathbb{F}_\ell)[p] \cong \mathcal{B}(\mathbb{Q}_\ell)[p] \subset \mathcal{B}(K_\ell)[p] \cong E(K_\ell)[p] \times \cdots \times E(K_\ell)[p] = 0.$$

Next suppose that $q \neq \ell$. Since formation of Néron models commutes with unramified base extension, we have

$$\Phi_{B,q}(\overline{\mathbb{F}}_q)[p] \cong \Phi_{E,q}(\overline{\mathbb{F}}_q)[p] \times \cdots \times \Phi_{E,q}(\overline{\mathbb{F}}_q)[p] = 0,$$

by our hypotheses on p. □

Following the appendix to [Maz72], let

$$\Sigma(B/\mathbb{Q}) = \ker\left(H^1(\mathbb{Q}, B) \to \bigoplus_{\text{all finite } q} H^1(\mathbb{Q}_q, B) \right),$$

where the sum is over all finite primes q of $\mathbb{Q}$. If p is an odd prime, then $\Sigma(B/\mathbb{Q})[p^\infty] = Ш(B/\mathbb{Q})[p^\infty]$; also one can see easily using Tate cohomology for the cyclic group $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ that

$$\Sigma(B/\mathbb{Q})[2]/Ш(B/\mathbb{Q})[2] \subset H^1(\mathbb{R}, B(\mathbb{C})) \cong B(\mathbb{R})/B(\mathbb{R})^0,$$

where $B(\mathbb{R})/B(\mathbb{R})^0$ has order 2^e for some $e \leq \dim B$.

Proposition 2.12 (Mazur)**.** *Suppose that* $a_\ell \not\equiv \ell + 1 \pmod{p}$*. If* p *is odd, then*

$$H^1(\mathbb{Z}, \mathcal{B})[p^\infty] \cong Ш(B/\mathbb{Q})[p^\infty].$$

Also, $\#H^1(\mathbb{Z}, \mathcal{B})[2^\infty]/Ш(B/\mathbb{Q})[2^\infty]$ *divides* $\#(B(\mathbb{R})/B(\mathbb{R})^0)$*.*

Proof. It follows from the appendix to [Maz72] that there is an exact sequence

$$0 \to \Sigma(B)[p^\infty] \to H^1(\mathbb{Z}, \mathcal{B})[p^\infty] \to \bigoplus_{\text{all finite } q} H^1\left(\mathbb{F}_q, \Phi_{B,q}(\overline{\mathbb{F}}_q)\right)[p^\infty], \tag{3}$$

where $\Phi_{B,q}$ is the component group of the fiber of $\mathcal{B}$ at q. By [Ser79, VIII.4.8],

$$\#H^1(\mathbb{F}_q, \Phi_{B,q}(\overline{\mathbb{F}}_q)) = \#\Phi_{B,q}(\mathbb{F}_q) = c_{B,q},$$

so the proposition follows from Lemma 2.11. □

Proposition 2.13. $H^2(\mathbb{Z}, \mathcal{A})[p] = 0$*.*

Proof. We apply the lemmas in [Sch83, §III.6]. Note that A has good reduction at p by [Mil72, Prop. 1], and $H^1(\mathbb{Z}, \mathcal{A})[p^\infty]$ is finite by Kato's theorem (see [Rub98, §8.1]) and Proposition 2.12. In the proof of Proposition 2.9, we showed that $A'(\mathbb{Q})$ is finite of order coprime to p, where A' is the abelian variety dual of A. We now use[1] Lemma 7 of [Sch83, §III.6], which because $A'(\mathbb{Q})[p] = 0$ implies that $H^2(\mathbb{Z}, \mathcal{A}[p^\infty]) = 0$ (Schneider uses H^q_{fpqf}, but this is not a problem since étale and fpqf cohomology agree on the smooth scheme $\mathcal{A}$.) It is easy to see (see, e.g., the proof of Lemma 6 of [Sch83, §III.6]) that the natural map $H^q(\mathbb{Z}, \mathcal{A}[p^\infty]) \to H^q(\mathbb{Z}, \mathcal{A})[p^\infty]$ is surjective for any $q > 0$, in particular, for $q = 2$, so $H^2(\mathbb{Z}, \mathcal{A})[p^\infty] = 0$ which proves the proposition. □

2.5. The Main Theorem

Fix an elliptic curve E over $\mathbb{Q}$ and a prime $p \nmid \prod c_{E,q}$ such that $\rho_{E,p} : G_\mathbb{Q} \to \mathrm{Aut}(E[p])$ is surjective. If $p = 2$ assume also that $E(\mathbb{R})$ is connected. Assume that ℓ is one of the primes whose existence is predicted by Conjecture 2.1. Let A and R be the corresponding abelian varieties, which fit into an exact sequence $0 \to A \to R \to E \to 0$, and recall that $L(A, 1) \neq 0$ so $A(\mathbb{Q})$ and $Ш(A/\mathbb{Q})$ are both finite (by [Rub98, §8.1] and [Kat, Cor. 14.3]).

[1] Note that the proof of Lemma 7 of [Sch83, §III.6] relies on a theorem of Artin and Mazur whose proof they never published; generalizations of this theorem have been published by McCallum [McC86, §5] and Milne [Mil86, §III.3.4], and Mazur assures the author that he and Milne both know the proof of Artin-Mazur duality well.

Theorem 2.14. *There is an exact sequence*

$$0 \to E(\mathbb{Q})/pE(\mathbb{Q}) \to Ш(A/\mathbb{Q})[p^\infty] \to Ш(E/K)[p^\infty] \to Ш(E/\mathbb{Q})[p^\infty] \to 0.$$

In particular, if E has odd rank and $Ш(E/\mathbb{Q})[p^\infty]$ is finite, then $\#Ш(A/\mathbb{Q})[p^\infty]$ is not a perfect square.

Proof. By Proposition 2.7 we have an exact sequence of étale sheaves

$$0 \to \mathcal{A} \to \mathcal{R} \to \mathcal{E} \to 0,$$

which gives rise to an exact sequence of étale cohomology groups

$$H^0(\mathbb{Z}, \mathcal{R}) \to H^0(\mathbb{Z}, \mathcal{E}) \to H^1(\mathbb{Z}, \mathcal{A}) \to H^1(\mathbb{Z}, \mathcal{R}) \to H^1(\mathbb{Z}, \mathcal{E}) \to H^2(\mathbb{Z}, \mathcal{A}).$$

We have

$$H^0(\mathbb{Z}, \mathcal{R}) = \mathcal{R}(\mathbb{Z}) = R(\mathbb{Q})$$

and likewise for $\mathcal{E}$, so by Propositions 2.9, 2.12, and 2.13 we obtain an exact sequence

$$0 \to E(\mathbb{Q})/pE(\mathbb{Q}) \to Ш(A/\mathbb{Q})[p^\infty] \to Ш(R/\mathbb{Q})[p^\infty] \to Ш(E/\mathbb{Q})[p^\infty] \to 0.$$

By Shapiro's lemma, there is an isomorphism $Ш(R/\mathbb{Q}) \cong Ш(E/K)$ (see [AS02, §1.3]), which yields the claimed exact sequence.

Kato's theorem ([Rub98, §8.1] and [Kat, Cor. 14.3]) implies that $Ш(E/K)[p^\infty]$ is finite (for the trivial character use our hypothesis that $Ш(E/\mathbb{Q})[p^\infty]$ is finite, and for the nontrivial characters use our hypothesis that $L(E, \chi_{p,\ell}, 1) \neq 0$). Theorem 1.2 then implies that $\#Ш(E/K)[p^\infty]$ is a perfect square. If $E(\mathbb{Q})$ has odd rank then $\#(E(\mathbb{Q})/pE(\mathbb{Q}))$ is an odd power of p (since $E[p]$ is irreducible), so $\#Ш(A/\mathbb{Q})[p^\infty]$ cannot be a perfect square. □

Remark 2.15. In the language of visibility of Shafarevich-Tate groups (see [CM00]), Theorem 2.14 asserts that the visible subgroup of $Ш(A)$ with respect to the embedding $A \hookrightarrow R$ is canonically isomorphic to the Mordell-Weil quotient $E(\mathbb{Q})/pE(\mathbb{Q})$.

Proposition 2.16. *If $q \neq p$ is a prime, then*

$$Ш(E/K)[q^\infty] \cong Ш(E/\mathbb{Q})[q^\infty] \oplus Ш(A/\mathbb{Q})[q^\infty]. \tag{4}$$

In particular, if $Ш(E/\mathbb{Q})[q^\infty]$ is finite, then $Ш(A/\mathbb{Q})[q^\infty]$ has order a perfect square.

Proof. The intersection of E and A in R is $E[p]$, so the summation map $E \times A \to R$ is an isogeny with kernel $E[p]$. Considering the long exact sequence associated to $0 \to E[p] \to E \times A \to R \to 0$, we see that

$$H^1(\mathbb{Q}, E \times A)[q^\infty] \cong H^1(\mathbb{Q}, R)[q^\infty], \tag{5}$$

and likewise for any completion $\mathbb{Q}_v$ of $\mathbb{Q}$. We then obtain (4) by combining (5) with the fact that cohomology commutes with products and that $H^1(\mathbb{Q}, R) \cong H^1(K, E)$.

If $Ш(E/\mathbb{Q})[q^\infty]$ is finite, then since $Ш(A/\mathbb{Q})[q^\infty]$ is finite (since $L(A, 1) \neq 0$, by construction), it follows from (4) that $Ш(E/K)[q^\infty]$ is finite. We have by Theorem 1.2 that both $Ш(E/K)[q^\infty]$ and $Ш(E/\mathbb{Q})[q^\infty]$ have order a perfect square, so (4) implies that $Ш(A/\mathbb{Q})[q^\infty]$ has order a perfect square. □

3. An Example

Combining Proposition 2.3, Theorem 2.14, and Proposition 2.16 yields the following theorem.

Theorem 3.1. *Let E be the elliptic curve $y^2 + y = x^3 - x$ of conductor 37. For every odd prime $p < 25000$ (with $p \neq 37$), there is a twist A of $E^{\times(p-1)}$ such that $\#\text{Ш}(A/\mathbb{Q}) = pn^2$ for some integer n.*

Remark 3.2. Using the elliptic curve of conductor 43 in place of E one can construct an abelian variety A with $\text{Ш}(A/\mathbb{Q}) = 37n^2$ for some integer n.

Though unnecessary for Theorem 3.1, we prove below that $\text{Ш}(E/\mathbb{Q}) = 0$, which removes our dependence on Proposition 2.13. We show that $\text{Ш}(E/\mathbb{Q})[p^\infty] = 0$ for all odd p using [Kol90, Thm. A], and we use a 2-descent (with [CrB]) to see that $\text{Ш}(E/\mathbb{Q})[2] = 0$.

Theorem 3.3 (Kolyvagin)**.** *Let E be an elliptic curve and let $L = \mathbb{Q}(\sqrt{-D})$ be an imaginary quadratic field of odd discriminant $-D$, where all primes dividing the conductor of E split, and assume that $D \neq 3, 4$. If the Heegner point $y_L \in E(L)$ has infinite order (equivalently, by* [GZ86]*, $L'(E/L, 1) \neq 0$), then $\#\text{Ш}(E/L) \mid t \cdot [E(L) : \mathbb{Z}y_L]^2$, where the only primes that divide t are 2 or primes where $\rho_{E,p}$ is not surjective.*

By [C97], E is isolated in its isogeny class, so $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \text{Aut}(E[p])$ is surjective for all primes p (see [RS01, §1.4]) hence t is a power of 2. Let $L = \mathbb{Q}(\sqrt{-7})$. To compute $[E(L) : \mathbb{Z}y_L]$ up to a power of 2 we use the Gross-Zagier formula and the fact that $[E(L) : E(\mathbb{Q}) + E^D(\mathbb{Q})]$ is a power of 2. By [GZ86, Thm. 6.3],

$$h(y_L) = \frac{u^2|D|^{\frac{1}{2}}}{\|\omega_f\|} L'(E,1)L(E^D,1),$$

where $D = -7$, $u = 1$, and $\|\omega_f\|$ is the Peterson norm of the newform f corresponding to E. Generators for the period lattice of E are $\omega_1 \sim 2.993459$ and $\omega_2 \sim 2.451389i$, so $\|\omega_f\| \sim 7.338133$. The quadratic twist E^D is the curve **1813B1** in [CrA], and $E^D(\mathbb{Q}) = 0$. From [CrA] we find that $L'(E,1) \sim 0.306000$ and $L(E^D,1) \sim 1.853076$, so $h(y_L) \sim 0.204446$. The height of a generator of $E(\mathbb{Q})$ is $\sim 0.051111 \sim h(y_L)/4$, so $[E(L) : \mathbb{Z}y_L]$ is a power of 2. (As a double check, and to avoid dependence on the Gross-Zagier formula, we wrote a program using [BCP97] to compute Heegner points and found that $y_L = (0,0)$, which is a generator for $E(\mathbb{Q})$.) Thus $\#\text{Ш}(E/L)$ is a power of 2.

To connect $\text{Ш}(E/L)$ with $\text{Ш}(E/\mathbb{Q})$, use the inflation-restriction exact sequence

$$0 \to H^1(L/\mathbb{Q}, E(L)) \to H^1(\mathbb{Q}, E(\overline{\mathbb{Q}})) \to H^1(L, E(\overline{\mathbb{Q}})).$$

Let p be an odd prime. Since $H^1(L/\mathbb{Q}, E(L))$ is a 2-group, the above sequence leads to an injective map

$$H^1(\mathbb{Q}, E(\overline{\mathbb{Q}}))[p] \hookrightarrow H^1(L, E(\overline{\mathbb{Q}}))[p],$$

which induces an inclusion

$$Ш(E/\mathbb{Q})[p] \hookrightarrow Ш(E/L)[p] = 0.$$

References

[AS02] A. Agashe and W. A. Stein, *Visibility of Shafarevich-Tate Groups of Abelian Varieties*, J. of Number Theory, **97** (2002), no. 1, 171–184.

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises*, J. Amer. Math. Soc. **15** (2001), no. 4, 843–939.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR **91i:**14034

[C97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[CrA] ———, *Elliptic Curve Data*, http://www.maths.nott.ac.uk/personal/jec/ftp/data/.

[CrB] ———, `mwrank` *(computer software)*, `http://www.maths.nott.ac.uk/personal/jec/ftp/progs/`.

[CM00] J. E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. **9** (2000), no. 1, 13–28. MR 1 758 797

[Fea01] J. Fearnley, *Vanishing and Non-Vanishing of L-series of Elliptic Curves Twisted by Dirichlet Characters*, Concordia Ph.D. thesis (2001).

[Fla90] M. Flach, *A generalisation of the Cassels-Tate pairing*, J. Reine Angew. Math. **412** (1990), 113–127. MR **92b:**11037

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR **87j:**11057

[JL99] B. W. Jordan and R. Livné, *On Atkin-Lehner quotients of Shimura curves*, Bull. London Math. Soc. **31** (1999), no. 6, 681–685. MR **2000j:**11090

[Kat] K. Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, Preprint, 244 pages.

[Kol90] V. A. Kolyvagin, *Euler systems*, The Grothendieck Festschrift, Vol. II, Birkhäuser Boston, Boston, MA, 1990, pp. 435–483. MR **92g:**11109

[Lan56] S. Lang, *Algebraic groups over finite fields*, Amer. J. Math. **78** (1956), 555–563. MR 19,174a

[Maz72] B. Mazur, *Rational points of abelian varieties with values in towers of number fields*, Invent. Math. **18** (1972), 183–266.

[McC86] W. G. McCallum, *Duality theorems for Néron models*, Duke Math. J. **53** (1986), no. 4, 1093–1124. MR **88c:**14062

[Mil72] J. S. Milne, *On the arithmetic of abelian varieties*, Invent. Math. **17** (1972), 177–190. MR 48 #8512

[Mil86] ______, *Arithmetic duality theorems*, Academic Press Inc., Boston, Mass., 1986.

[PS97] B. Poonen and E. F. Schaefer, *Explicit descent for Jacobians of cyclic covers of the projective line*, J. Reine Angew. Math. **488** (1997), 141–188. MR **98k:**11087

[PS99] B. Poonen and M. Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2) **150** (1999), no. 3, 1109–1149. MR **2000m:**11048

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 143–232. MR **2002h:**11047

[Rub98] K. Rubin, *Euler systems and modular elliptic curves*, Galois representations in arithmetic algebraic geometry (Durham, 1996), Cambridge Univ. Press, Cambridge, 1998, pp. 351–367. MR **2001a:**11106

[Sch83] P. Schneider, *Iwasawa L-functions of varieties over algebraic number fields. A first approach*, Invent. Math. **71** (1983), no. 2, 251–293. MR **85d:**11063

[SD67] P. Swinnerton-Dyer, *The conjectures of Birch and Swinnerton-Dyer, and of Tate*, Proc. Conf. Local Fields (Driebergen, 1966), Springer, Berlin, 1967, pp. 132–157. MR 37 #6287

[Ser79] J-P. Serre, *Local fields*, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

[Ser88] ______, *Algebraic groups and class fields*, Springer-Verlag, New York, 1988, Translated from the French.

[ST68] J-P. Serre and J. T. Tate, *Good reduction of abelian varieties*, Ann. of Math. (2) **88** (1968), 492–517.

[Tat63] J. Tate, *Duality theorems in Galois cohomology over number fields*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 288–295. MR 31 #168

515 Science Center, Department of Mathematics, Harvard University,
E-mail address: was@math.harvard.edu

17 Constructing Elements In Shafarevich-Tate Groups Of Modular Motives, with N. Dummigan and M. Watkins

CONSTRUCTING ELEMENTS IN SHAFAREVICH-TATE GROUPS OF MODULAR MOTIVES

NEIL DUMMIGAN, WILLIAM STEIN, AND MARK WATKINS

Abstract. We study Shafarevich-Tate groups of motives attached to modular forms on $\Gamma_0(N)$ of weight bigger than 2. We deduce a criterion for the existence of nontrivial elements of these Shafarevich-Tate groups, and give 16 examples in which a strong form of the Beilinson-Bloch conjecture implies the existence of such elements. We also use modular symbols and observations about Tamagawa numbers to compute nontrivial conjectural lower bounds on the orders of the Shafarevich-Tate groups of modular motives of low level and weight at most 12. Our methods build upon the idea of visibility due to Cremona and Mazur, but in the context of motives instead of abelian varieties.

1. Introduction

Let E be an elliptic curve defined over $\mathbb{Q}$ and let $L(E,s)$ be the associated L-function. The conjecture of Birch and Swinnerton-Dyer [BS-D] predicts that the order of vanishing of $L(E,s)$ at $s = 1$ is the rank of the group $E(\mathbb{Q})$ of rational points, and also gives an interpretation of the leading term in the Taylor expansion in terms of various quantities, including the order of the Shafarevich-Tate group of E.

Cremona and Mazur [CM1] look, among all strong Weil elliptic curves over $\mathbb{Q}$ of conductor $N \leq 5500$, at those with nontrivial Shafarevich-Tate group (according to the Birch and Swinnerton-Dyer conjecture). Suppose that the Shafarevich-Tate group has predicted elements of prime order p. In most cases they find another elliptic curve, often of the same conductor, whose p-torsion is Galois-isomorphic to that of the first one, and which has positive rank. The rational points on the second elliptic curve produce classes in the common $H^1(\mathbb{Q}, E[p])$. They show [CM2] that these lie in the Shafarevich-Tate group of the first curve, so rational points on one curve explain elements of the Shafarevich-Tate group of the other curve.

The Bloch-Kato conjecture [BK] is the generalisation to arbitrary motives of the leading term part of the Birch and Swinnerton-Dyer conjecture. The Beilinson-Bloch conjecture [B, Be] generalises the part about the order of vanishing at the central point, identifying it with the rank of a certain Chow group.

This paper is a partial generalisation of [CM1] and [AS] from abelian varieties over $\mathbb{Q}$ associated to modular forms of weight 2 to the motives attached to modular forms of higher weight. It also does for congruences between modular forms of equal weight what [Du2] did for congruences between modular forms of different weights.

Date: 28 January 2003.

1991 *Mathematics Subject Classification.* 11F33, 11F67, 11G40.

Key words and phrases. modular form, L-function, visibility, Bloch-Kato conjecture, Shafarevich-Tate group.

We consider the situation where two newforms f and g, both of even weight $k > 2$ and level N, are congruent modulo a maximal ideal $\mathfrak{q}$ of odd residue characteristic, and $L(g, k/2) = 0$ but $L(f, k/2) \neq 0$. It turns out that this forces $L(g, s)$ to vanish to order at least 2 at $s = k/2$. In Section 7, we give sixteen such examples (all with $k = 4$ and $k = 6$), and in each example, we find that $\mathfrak{q}$ divides the numerator of the algebraic number $L(f, k/2)/\operatorname{vol}_\infty$, where $\operatorname{vol}_\infty$ is a certain canonical period.

In fact, we show how this divisibility may be deduced from the vanishing of $L(g, k/2)$ using recent work of Vatsal [V]. The point is, the congruence between$f$ and g leads to a congruence between suitable "algebraic parts" of the special values $L(f, k/2)$ and $L(g, k/2)$. In slightly more detail, a multiplicity one result of Faltings and Jordan shows that the congruence of Fourier expansions leads to a congruence of certain associated cohomology classes. These are then identified with the modular symbols which give rise to the algebraic parts of special values. If $L(g, k/2)$ vanishes then the congruence implies that $L(f, k/2)/\operatorname{vol}_\infty$ must be divisible by $\mathfrak{q}$.

The Bloch-Kato conjecture sometimes then implies that the Shafarevich-Tate group Ш attached to f has nonzero $\mathfrak{q}$-torsion. Under certain hypotheses and assumptions, the most substantial of which is the Beilinson-Bloch conjecture relating the vanishing of $L(g, k/2)$ to the existence of algebraic cycles, we are able to construct some of the predicted elements of Ш using the Galois-theoretic interpretation of the congruence to transfer elements from a Selmer group for g to a Selmer group for f. One might say that algebraic cycles for one motive explain elements of Ш for the other, or that we use the congruence to link the Beilinson-Bloch conjecture for one motive with the Bloch-Kato conjecture for the other.

We also compute data which, assuming the Bloch-Kato conjecture, provides lower bounds for the orders of numerous Shafarevich-Tate groups (see Section 7.3). We thank the referee for many constructive comments.

2. Motives and Galois representations

This section and the next provide definitions of some of the quantities appearing later in the Bloch-Kato conjecture. Let $f = \sum a_n q^n$ be a newform of weight $k \geq 2$ for $\Gamma_0(N)$, with coefficients in an algebraic number field E, which is necessarily totally real. Let λ be any finite prime of E, and let ℓ denote its residue characteristic. A theorem of Deligne [De1] implies the existence of a two-dimensional vector space V_λ over E_λ, and a continuous representation

$$\rho_\lambda : \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \operatorname{Aut}(V_\lambda),$$

such that

(1) ρ_λ is unramified at p for all primes p not dividing ℓN, and
(2) if Frob_p is an arithmetic Frobenius element at such a p then the characteristic polynomial of $\operatorname{Frob}_p^{-1}$ acting on V_λ is $x^2 - a_p x + p^{k-1}$.

Following Scholl [Sc], V_λ may be constructed as the λ-adic realisation of a Grothendieck motive M_f. There are also Betti and de Rham realisations V_B and V_{dR}, both 2-dimensional E-vector spaces. For details of the construction see [Sc]. The de Rham realisation has a Hodge filtration $V_{\mathrm{dR}} = F^0 \supset F^1 = \cdots = F^{k-1} \supset F^k = \{0\}$. The Betti realisation V_B comes from singular cohomology, while V_λ comes from étale ℓ-adic cohomology. For each prime λ, there is a natural isomorphism $V_B \otimes E_\lambda \simeq V_\lambda$. We may choose a $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-stable O_λ-module T_λ inside each V_λ. Define $A_\lambda = V_\lambda/T_\lambda$. Let $A[\lambda]$ denote the λ-torsion in A_λ. There is the Tate twist $V_\lambda(j)$ (for any integer j), which amounts to multiplying the action of Frob_p by p^j.

Following [BK] (Section 3), for $p \neq \ell$ (including $p = \infty$) let

$$H^1_f(\mathbb{Q}_p, V_\lambda(j)) = \ker(H^1(D_p, V_\lambda(j)) \to H^1(I_p, V_\lambda(j))).$$

The subscript f stands for "finite part", D_p is a decomposition subgroup at a prime above p, I_p is the inertia subgroup, and the cohomology is for continuous cocycles and coboundaries. For $p = \ell$ let

$$H^1_f(\mathbb{Q}_\ell, V_\lambda(j)) = \ker(H^1(D_\ell, V_\lambda(j)) \to H^1(D_\ell, V_\lambda(j) \otimes_{\mathbb{Q}_\ell} B_{\mathrm{cris}}))$$

(see Section 1 of [BK] for definitions of Fontaine's rings B_{cris} and B_{dR}). Let $H^1_f(\mathbb{Q}, V_\lambda(j))$ be the subspace of elements of $H^1(\mathbb{Q}, V_\lambda(j))$ whose local restrictions lie in $H^1_f(\mathbb{Q}_p, V_\lambda(j))$ for all primes p.

There is a natural exact sequence

$$\begin{tikzcd} 0 \arrow[r] & T_\lambda(j) \arrow[r] & V_\lambda(j) \arrow[r, "\pi"] & A_\lambda(j) \arrow[r] & 0. \end{tikzcd}$$

Let $H^1_f(\mathbb{Q}_p, A_\lambda(j)) = \pi_* H^1_f(\mathbb{Q}_p, V_\lambda(j))$. Define the λ-Selmer group $H^1_f(\mathbb{Q}, A_\lambda(j))$ to be the subgroup of elements of $H^1(\mathbb{Q}, A_\lambda(j))$ whose local restrictions lie in $H^1_f(\mathbb{Q}_p, A_\lambda(j))$ for all primes p. Note that the condition at $p = \infty$ is superfluous unless $\ell = 2$. Define the Shafarevich-Tate group

$$Ш(j) = \oplus_\lambda H^1_f(\mathbb{Q}, A_\lambda(j))/\pi_* H^1_f(\mathbb{Q}, V_\lambda(j)).$$

Define an ideal $\#Ш(j)$ of O_E, in which the exponent of any prime ideal λ is the length of the λ-component of $Ш(j)$. We shall only concern ourselves with the case $j = k/2$, and write Ш for $Ш(k/2)$. It depends on the choice of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-stable O_λ-module T_λ inside each V_λ. But if $A[\lambda]$ is irreducible then T_λ is unique up to scaling and the λ-part of Ш is independent of choices.

In the case $k = 2$ the motive comes from a (self-dual) isogeny class of abelian varieties over $\mathbb{Q}$, with endomorphism algebra containing E. Choose an abelian variety B in the isogeny class in such a way that the endomorphism ring of B contains the full ring of integers O_E. If one takes all the $T_\lambda(1)$ to be λ-adic Tate modules, then what we have defined above coincides with the usual Shafarevich-Tate group of B (assuming finiteness of the latter, or just taking the quotient by its maximal divisible subgroup). To see this one uses 3.11 of [BK], for $\ell = p$. For $\ell \neq p$, $H^1_f(\mathbb{Q}_p, V_\ell) = 0$. Considering the formal group, every class in $B(\mathbb{Q}_p)/\ell B(\mathbb{Q}_p)$ is represented by an ℓ-power torsion point in $B(\mathbb{Q}_p)$, so maps to zero in $H^1(\mathbb{Q}_p, A_\ell)$.

Define the group of global torsion points

$$\Gamma_{\mathbb{Q}} = \oplus_\lambda H^0(\mathbb{Q}, A_\lambda(k/2)).$$

This is analogous to the group of rational torsion points on an elliptic curve. Define an ideal $\#\Gamma_{\mathbb{Q}}$ of O_E, in which the exponent of any prime ideal λ is the length of the λ-component of $\Gamma_{\mathbb{Q}}$.

3. Canonical periods

We assume from now on for convenience that $N \geq 3$. We need to choose convenient O_E-lattices T_B and T_{dR} in the Betti and de Rham realisations V_B and V_{dR} of M_f. We do this in a way such that T_B and $T_{\mathrm{dR}} \otimes_{O_E} O_E[1/Nk!]$ agree with (respectively) the O_E-lattice $\mathfrak{M}_{f,B}$ and the $O_E[1/Nk!]$-lattice $\mathfrak{M}_{f,\mathrm{dR}}$ defined in [DFG1] using cohomology, with non-constant coefficients, of modular curves. (In [DFG1], see especially Sections 2.2 and 5.4, and the paragraph preceding Lemma 2.3.)

For any finite prime λ of O_E define the O_λ module T_λ inside V_λ to be the image of $T_B \otimes O_\lambda$ under the natural isomorphism $V_B \otimes E_\lambda \simeq V_\lambda$. Then the O_λ-module T_λ is $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-stable.

Let $M(N)$ be the modular curve over $\mathbb{Z}[1/N]$ parametrising generalised elliptic curves with full level-N structure. Let $\mathfrak{E}$ be the universal generalised elliptic curve over $M(N)$. Let $\mathfrak{E}^{k-2}$ be the $(k-2)$-fold fibre product of $\mathfrak{E}$ over $M(N)$. (The motive M_f is constructed using a projector on the cohomology of a desingularisation of $\mathfrak{E}^{k-2}$). Realising $M(N)(\mathbb{C})$ as the disjoint union of $\phi(N)$ copies of the quotient $\Gamma(N)\backslash\mathfrak{H}^*$ (where $\mathfrak{H}^*$ is the completed upper half plane), and letting τ be a variable on $\mathfrak{H}$, the fibre $\mathfrak{E}_\tau$ is isomorphic to the elliptic curve with period lattice generated by 1 and τ. Let $z_i \in \mathbb{C}/\langle 1, \tau\rangle$ be a variable on the i^{th} copy of $\mathfrak{E}_\tau$ in the fibre product. Then $2\pi i f(\tau)\, d\tau \wedge dz_1 \wedge \ldots \wedge dz_{k-2}$ is a well-defined differential form on (a desingularisation of) $\mathfrak{E}^{k-2}$ and naturally represents a generating element of $F^{k-1}T_{\mathrm{dR}}$. (At least we can make our choices locally at primes dividing $Nk!$ so that this is the case.) We shall call this element $e(f)$.

Under the de Rham isomorphism between $V_{\mathrm{dR}} \otimes \mathbb{C}$ and $V_B \otimes \mathbb{C}$, $e(f)$ maps to some element ω_f. There is a natural action of complex conjugation on V_B, breaking it up into one-dimensional E-vector spaces V_B^+ and V_B^-. Let ω_f^+ and ω_f^- be the projections of ω_f to $V_B^+ \otimes \mathbb{C}$ and $V_B^- \otimes \mathbb{C}$, respectively. Let $T_B^\pm$ be the intersections of $V_B^\pm$ with T_B. These are rank one O_E-modules, but not necessarily free, since the class number of O_E may be greater than one. Choose nonzero elements $\delta_f^\pm$ of $T_B^\pm$ and let $\mathfrak{a}^\pm$ be the ideals $[T_B^\pm : O_E\delta_f^\pm]$. Define complex numbers $\Omega_f^\pm$ by $\omega_f^\pm = \Omega_f^\pm \delta_f^\pm$.

4. The Bloch-Kato conjecture

In this section we extract from the Bloch-Kato conjecture for $L(f, k/2)$ a prediction about the order of the Shafarevich-Tate group, by analysing the other terms in the formula.

Let $L(f,s)$ be the L-function attached to f. For $\Re(s) > \frac{k+1}{2}$ it is defined by the Dirichlet series with Euler product $\sum_{n=1}^{\infty} a_n n^{-s} = \prod_p (P_p(p^{-s}))^{-1}$, but there is an analytic continuation given by an integral, as described in the next section. Suppose that $L(f, k/2) \neq 0$. The Bloch-Kato conjecture for the motive $M_f(k/2)$ predicts the following equality of fractional ideals of E:

$$\frac{L(f,k/2)}{\mathrm{vol}_\infty} = \left(\prod_p c_p(k/2)\right) \frac{\#\text{Ш}}{\mathfrak{a}^\pm(\#\Gamma_\mathbb{Q})^2}.$$

Here, **and from this point onwards,** ± represents the parity of $(k/2) - 1$. The quantity vol_∞ is equal to $(2\pi i)^{k/2}$ multiplied by the determinant of the isomorphism $V_B^\pm \otimes \mathbb{C} \simeq (V_{\mathrm{dR}}/F^{k/2}) \otimes \mathbb{C}$, calculated with respect to the lattices $O_E\delta_f^\pm$ and the image of T_{dR}. For $l \neq p$, $\mathrm{ord}_\lambda(c_p(j))$ is defined to be

$$\begin{aligned} &\text{length } H^1_f(\mathbb{Q}_p, T_\lambda(j))_{\mathrm{tors}} - \mathrm{ord}_\lambda(P_p(p^{-j})) \\ &\quad = \text{length } \left(H^0(\mathbb{Q}_p, A_\lambda(j))/H^0\left(\mathbb{Q}_p, V_\lambda(j)^{I_p}/T_\lambda(j)^{I_p}\right)\right). \end{aligned}$$

We omit the definition of $\mathrm{ord}_\lambda(c_p(j))$ for $\lambda \mid p$, which requires one to assume Fontaine's de Rham conjecture ([Fo1], Appendix A6), and depends on the choices of T_{dR} and T_B, locally at λ. (We shall mainly be concerned with the q-part of the

Bloch-Kato conjecture, where q is a prime of good reduction. For such primes, the de Rham conjecture follows from Theorem 5.6 of [Fa].)

Strictly speaking, the conjecture in [BK] is only given for $E = \mathbb{Q}$. We have taken here the obvious generalisation of a slight rearrangement of (5.15.1) of [BK]. The Bloch-Kato conjecture has been reformulated and generalised by Fontaine and Perrin-Riou, who work with general E, though that is not really the point of their work. In Section 11 of [Fo2] it is sketched how to deduce the original conjecture from theirs, in the case $E = \mathbb{Q}$.

Lemma 4.1. $\mathrm{vol}_\infty / \mathfrak{a}^\pm = c(2\pi i)^{k/2}\mathfrak{a}^\pm\Omega_\pm$, *with* $c \in E$ *and* $\mathrm{ord}_\lambda(c) = 0$ *for* $\lambda \nmid Nk!$.

Proof. We note that vol_∞ is equal to $(2\pi i)^{k/2}$ times the determinant of the period map from $F^{k/2}V_{\mathrm{dR}} \otimes \mathbb{C}$ to $V_B^\pm \otimes \mathbb{C}$, with respect to lattices dual to those we used above in the definition of vol_∞ (c.f. the last paragraph of 1.7 of [De2]). We are using here natural pairings. Meanwhile, $\Omega_\pm$ is the determinant of the same map with respect to the lattices $F^{k/2}T_{\mathrm{dR}}$ and $O_E\delta_f^\pm$. Recall that the index of $O_E\delta_f^\pm$ in $T_B^\pm$ is the ideal $\mathfrak{a}^\pm$. Then the proof is completed by noting that, locally away from primes dividing $Nk!$, the index of T_{dR} in its dual is equal to the index of T_B in its dual, both being equal to the ideal denoted η in [DFG2]. □

Remark 4.2. Note that the "quantities" $\mathfrak{a}^\pm\Omega_\pm$ and $\mathrm{vol}_\infty / \mathfrak{a}^\pm$ are independent of the choice of $\delta_f^\pm$.

Lemma 4.3. *Let* $p \nmid N$ *be a prime and* j *an integer. Then the fractional ideal* $c_p(j)$ *is supported at most on divisors of* p.

Proof. As on p. 30 of [Fl2], for odd $l \neq p$, $\mathrm{ord}_\lambda(c_p(j))$ is the length of the finite O_λ-module $H^0(\mathbb{Q}_p, H^1(I_p, T_\lambda(j))_{\mathrm{tors}})$, where I_p is an inertia group at p. But $T_\lambda(j)$ is a trivial I_p-module, so $H^1(I_p, T_\lambda(j))$ is torsion free. □

Lemma 4.4. *Let* $q \nmid N$ *be a prime satisfying* $q > k$. *Suppose that* $A[\mathfrak{q}]$ *is an irreducible representation of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, *where* $\mathfrak{q} \mid q$. *Let* $p \mid N$ *be a prime, and if* $p^2 \mid N$ *suppose that* $p \not\equiv -1 \pmod{q}$. *Suppose also that* f *is not congruent modulo* $\mathfrak{q}$ *(for Fourier coefficients of index coprime to* Nq*) to any newform of weight* k*, trivial character, and level dividing* N/p. *Then* $\mathrm{ord}_\mathfrak{q}(c_p(j)) = 0$ *for all integers* j.

Proof. There is a natural injective map from $V_\mathfrak{q}(j)^{I_p}/T_\mathfrak{q}(j)^{I_p}$ to $H^0(I_p, A_\mathfrak{q}(j))$ (i.e., $A_\mathfrak{q}(j)^{I_p}$). Consideration of $\mathfrak{q}$-torsion shows that

$$\dim_{O_E/\mathfrak{q}} H^0(I_p, A[\mathfrak{q}](j)) \geq \dim_{E_\mathfrak{q}} H^0(I_p, V_\mathfrak{q}(j)).$$

To prove the lemma it suffices to show that

$$\dim_{O_E/\mathfrak{q}} H^0(I_p, A[\mathfrak{q}](j)) = \dim_{E_\mathfrak{q}} H^0(I_p, V_\mathfrak{q}(j)),$$

since this ensures that $H^0(I_p, A_\mathfrak{q}(j)) = V_\mathfrak{q}(j)^{I_p}/T_\mathfrak{q}(j)^{I_p}$, hence that $H^0(\mathbb{Q}_p, A_\mathfrak{q}(j)) = H^0(\mathbb{Q}_p, V_\mathfrak{q}(j)^{I_p}/T_\mathfrak{q}(j)^{I_p})$. If the dimensions differ then, given that f is not congruent modulo $\mathfrak{q}$ to a newform of level dividing N/p, Condition (b) of Proposition 2.3 of [L] is satisfied. If Condition (a) was not satisfied then Proposition 2.2 of [L] would imply that f was congruent modulo $\mathfrak{q}$ to a twist of level dividing N/p. Since Condition (c) is clearly also satisfied, we are in a situation covered by one of the three cases in Proposition 2.3 of [L]. Since $p \not\equiv -1 \pmod{q}$ if $p^2 \mid N$, Case 3 is excluded, so $A[\mathfrak{q}](j)$ is unramified at p and $\mathrm{ord}_p(N) = 1$. (Here we are using Carayol's result that N is the prime-to-q part of the conductor of $V_\mathfrak{q}$ [Ca1].) But then Theorem 1 of [JL] (which uses the condition $q > k$) implies the existence of a newform of weight k, trivial character and level dividing N/p, congruent to g modulo $\mathfrak{q}$, for Fourier coefficients of index coprime to Nq. This contradicts our hypotheses. □

Remark 4.5. For an example of what can be done when f is congruent to a form of lower level, see the first example in Section 7.4 below.

Lemma 4.6. *If* $\mathfrak{q} \mid q$ *is a prime of* E *such that* $q \nmid Nk!$*, then* $\mathrm{ord}_{\mathfrak{q}}(c_q) = 0$.

Proof. It follows from Lemma 5.7 of [DFG1] (whose proof relies on an application, at the end of Section 2.2, of the results of [Fa]) that $T_{\mathfrak{q}}$ is the $O_{\mathfrak{q}}[\mathrm{Gal}(\overline{\mathbb{Q}}_q/\mathbb{Q}_q)]$-module associated to the filtered module $T_{\mathrm{dR}} \otimes O_{\mathfrak{q}}$ by the functor they call $\mathbb{V}$. (This property is part of the definition of an S-integral premotivic structure given in Section 1.2 of [DFG1].) Given this, the lemma follows from Theorem 4.1(iii) of [BK]. (That $\mathbb{V}$ is the same as the functor used in Theorem 4.1 of [BK] follows from the first paragraph of 2(h) of [Fa].) □

Lemma 4.7. *If* $A[\lambda]$ *is an irreducible representation of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$*, then*

$$\mathrm{ord}_{\lambda}(\#\Gamma_{\mathbb{Q}}) = 0.$$

Proof. This follows trivially from the definition. □

Putting together the above lemmas we arrive at the following:

Proposition 4.8. *Let* $q \nmid N$ *be a prime satisfying* $q > k$ *and suppose that* $A[\mathfrak{q}]$ *is an irreducible representation of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$*, where* $\mathfrak{q} \mid q$*. Assume the same hypotheses as in Lemma 4.4 for all* $p \mid N$*. Choose* T_{dR} *and* T_B *which locally at* $\mathfrak{q}$ *are as in the previous section. If* $L(f, k/2)\mathfrak{a}^{\pm}/\mathrm{vol}_{\infty} \neq 0$ *then the Bloch-Kato conjecture predicts that*

$$\mathrm{ord}_{\mathfrak{q}}(\#Ш) = \mathrm{ord}_{\mathfrak{q}}(L(f, k/2)\mathfrak{a}^{\pm}/\mathrm{vol}_{\infty}).$$

5. Congruences of special values

Let $f = \sum a_n q^n$ and $g = \sum b_n q^n$ be newforms of equal weight $k \geq 2$ for $\Gamma_0(N)$. Let E be a number field large enough to contain all the coefficients a_n and b_n. Suppose that $\mathfrak{q} \mid q$ is a prime of E such that $f \equiv g \pmod{\mathfrak{q}}$, i.e. $a_n \equiv b_n \pmod{\mathfrak{q}}$ for all n. Assume that $A[\mathfrak{q}]$ is an irreducible representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and that $q \nmid N\phi(N)k!$. Choose $\delta_f^{\pm} \in T_B^{\pm}$ in such a way that $\mathrm{ord}_{\mathfrak{q}}(\mathfrak{a}^{\pm}) = 0$, i.e., $\delta_f^{\pm}$ generates $T_B^{\pm}$ locally at $\mathfrak{q}$. Make two further assumptions:

$$L(f, k/2) \neq 0 \qquad \text{and} \qquad L(g, k/2) = 0.$$

Proposition 5.1. *With assumptions as above,* $\mathrm{ord}_{\mathfrak{q}}(L(f, k/2)/\mathrm{vol}_{\infty}) > 0$.

Proof. This is based on some of the ideas used in Section 1 of [V]. Note the apparent typo in Theorem 1.13 of [V], which presumably should refer to “Condition 2”. Since $\mathrm{ord}_{\mathfrak{q}}(\mathfrak{a}^{\pm}) = 0$, we just need to show that $\mathrm{ord}_{\mathfrak{q}}(L(f, k/2)/((2\pi i)^{k/2}\Omega_{\pm})) > 0$, where $\pm 1 = (-1)^{(k/2)-1}$. It is well known, and easy to prove, that

$$\int_0^{\infty} f(iy) y^{s-1} dy = (2\pi)^{-s}\Gamma(s)L(f, s).$$

Hence, if for $0 \leq j \leq k-2$ we define the j^{th} period

$$r_j(f) = \int_0^{i\infty} f(z) z^j dz,$$

where the integral is taken along the positive imaginary axis, then

$$r_j(f) = j!(-2\pi i)^{-(j+1)} L_f(j+1).$$

Thus we are reduced to showing that $\mathrm{ord}_{\mathfrak{q}}(r_{(k/2)-1}(f)/\Omega_{\pm}) > 0$.

Let $\mathcal{D}_0$ be the group of divisors of degree zero supported on $\mathbb{P}^1(\mathbb{Q})$. For a $\mathbb{Z}$-algebra R and integer $r \geq 0$, let $P_r(R)$ be the additive group of homogeneous polynomials of degree r in $R[X,Y]$. Both these groups have a natural action of $\Gamma_1(N)$. Let $S_{\Gamma_1(N)}(k,R) := \mathrm{Hom}_{\Gamma_1(N)}(\mathcal{D}_0, P_{k-2}(R))$ be the R-module of weight k modular symbols for $\Gamma_1(N)$.

Via the isomorphism (8) in Section 1.5 of [V], combined with the argument in 1.7 of [V], the cohomology class $\omega_f^{\pm}$ corresponds to a modular symbol $\Phi_f^{\pm} \in S_{\Gamma_1(N)}(k,\mathbb{C})$, and $\delta_f^{\pm}$ corresponds to an element $\Delta_f^{\pm} \in S_{\Gamma_1(N)}(k, O_{E,\mathfrak{q}})$. We are now dealing with cohomology over $X_1(N)$ rather than $M(N)$, which is why we insist that $q \nmid \phi(N)$. It follows from the last line of Section 4.2 of [St] that, up to some small factorials which do not matter locally at $\mathfrak{q}$,

$$\Phi_f^{\pm}([\infty]-[0]) = \sum_{j=0, j\equiv (k/2)-1 \pmod 2}^{k-2} r_f(j) X^j Y^{k-2-j}.$$

Since $\omega_f^{\pm} = \Omega_f^{\pm}\delta_f^{\pm}$, we see that

$$\Delta_f^{\pm}([\infty]-[0]) = \sum_{j=0, j\equiv (k/2)-1 \pmod 2}^{k-2} (r_f(j)/\Omega_f^{\pm}) X^j Y^{k-2-j}.$$

The coefficient of $X^{(k/2)-1}Y^{(k/2)-1}$ is what we would like to show is divisible by $\mathfrak{q}$. Similarly

$$\Phi_g^{\pm}([\infty]-[0]) = \sum_{j=0, j\equiv (k/2)-1 \pmod 2}^{k-2} r_g(j) X^j Y^{k-2-j}.$$

The coefficient of $X^{(k/2)-1}Y^{(k/2)-1}$ in this is 0, since $L(g,k/2)=0$. Therefore it would suffice to show that, for some $\mu \in O_E$, the element $\Delta_f^{\pm} - \mu\Delta_g^{\pm}$ is divisible by $\mathfrak{q}$ in $S_{\Gamma_1(N)}(k, O_{E,\mathfrak{q}})$. It suffices to show that, for some $\mu \in O_E$, the element $\delta_f^{\pm} - \mu\delta_g^{\pm}$ is divisible by $\mathfrak{q}$, considered as an element of $\mathfrak{q}$-adic cohomology of $X_1(N)$ with non-constant coefficients. This would be the case if $\delta_f^{\pm}$ and $\delta_g^{\pm}$ generate the same one-dimensional subspace upon reduction modulo $\mathfrak{q}$. But this is a consequence of Theorem 2.1(1) of [FJ] (for which we need the irreducibility of $A[\mathfrak{q}]$). □

Remark 5.2. The signs in the functional equations of $L(f,s)$ and $L(g,s)$ are equal. They are determined by the eigenvalue of the Atkin-Lehner involution W_N, which is determined by a_N and b_N modulo $\mathfrak{q}$, because a_N and b_N are each $N^{k/2-1}$ times this sign and $\mathfrak{q}$ has residue characteristic coprime to $2N$. The common sign in the functional equation is $(-1)^{k/2}w_N$, where w_N is the common eigenvalue of W_N acting on f and g.

This is analogous to the remark at the end of Section 3 of [CM1], which shows that if $\mathfrak{q}$ has odd residue characteristic and $L(f,k/2)\neq 0$ but $L(g,k/2)=0$ then $L(g,s)$ must vanish to order at least two at $s=k/2$. Note that Maeda's conjecture implies that there are no examples of g of level one with positive sign in their functional equation such that $L(g,k/2)=0$ (see [CF]).

6. Constructing elements of the Shafarevich-Tate group

Let f, g and $\mathfrak{q}$ be as in the first paragraph of the previous section. In the previous section we showed how the congruence between f and g relates the vanishing of $L(g,k/2)$ to the divisibility by $\mathfrak{q}$ of an "algebraic part" of $L(f,k/2)$. Conjecturally the former is associated with the existence of certain algebraic cycles (for M_g) while the latter is associated with the existence of certain elements of the Shafarevich-Tate group (for M_f, as we saw in §4). In this section we show how the congruence, interpreted in terms of Galois representations, provides a direct link between algebraic cycles and the Shafarevich-Tate group.

For f we have defined V_λ, T_λ and A_λ. Let V'_λ, T'_λ and A'_λ be the corresponding objects for g. Since a_p is the trace of Frob_p^{-1} on V_λ, it follows from the Cebotarev Density Theorem that $A[\mathfrak{q}]$ and $A'[\mathfrak{q}]$, if irreducible, are isomorphic as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules.

Recall that $L(g,k/2)=0$ and $L(f,k/2)\neq 0$. Since the sign in the functional equation for $L(g,s)$ is positive (this follows from $L(f,k/2)\neq 0$, see Remark 5.2), the order of vanishing of $L(g,s)$ at $s=k/2$ is at least 2. According to the Beilinson-Bloch conjecture [B, Be], the order of vanishing of $L(g,s)$ at $s=k/2$ is the rank of the group $\mathrm{CH}_0^{k/2}(M_g)(\mathbb{Q})$ of $\mathbb{Q}$-rational rational equivalence classes of null-homologous, algebraic cycles of codimension $k/2$ on the motive M_g. (This generalises the part of the Birch–Swinnerton-Dyer conjecture which says that for an elliptic curve $E/\mathbb{Q}$, the order of vanishing of $L(E,s)$ at $s=1$ is equal to the rank of the Mordell-Weil group $E(\mathbb{Q})$.)

Via the $\mathfrak{q}$-adic Abel-Jacobi map, $\mathrm{CH}_0^{k/2}(M_g)(\mathbb{Q})$ maps to $H^1(\mathbb{Q},V'_{\mathfrak{q}}(k/2))$, and its image is contained in the subspace $H^1_f(\mathbb{Q},V'_{\mathfrak{q}}(k/2))$, by 3.1 and 3.2 of [Ne]. If, as expected, the $\mathfrak{q}$-adic Abel-Jacobi map is injective, we get (assuming also the Beilinson-Bloch conjecture) a subspace of $H^1_f(\mathbb{Q},V'_{\mathfrak{q}}(k/2))$ of dimension equal to the order of vanishing of $L(g,s)$ at $s=k/2$. In fact, one could simply conjecture that the dimension of $H^1_f(\mathbb{Q},V'_{\mathfrak{q}}(k/2))$ is equal to the order of vanishing of $L(g,s)$ at $s=k/2$. This would follow from the "conjectures" $C_r(M)$ and $C^i_\lambda(M)$ in Sections 1 and 6.5 of [Fo2]. We shall call it the "strong" Beilinson-Bloch conjecture.

Similarly, if $L(f,k/2)\neq 0$ then we expect that $H^1_f(\mathbb{Q},V_{\mathfrak{q}}(k/2))=0$, so that $H^1_f(\mathbb{Q},A_{\mathfrak{q}}(k/2))$ coincides with the $\mathfrak{q}$-part of Ш.

Theorem 6.1. *Let $q\nmid N$ be a prime satisfying $q>k$. Let r be the dimension of $H^1_f(\mathbb{Q},V'_{\mathfrak{q}}(k/2))$. Suppose that $A[\mathfrak{q}]$ is an irreducible representation of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *and that for no prime $p\mid N$ is f congruent modulo $\mathfrak{q}$ (for Fourier coefficients of index coprime to Nq) to a newform of weight k, trivial character and level dividing N/p. Suppose that, for all primes $p\mid N$, $p\not\equiv -w_p \pmod{q}$, with $p\not\equiv -1 \pmod{q}$ if $p^2\mid N$. (Here w_p is the common eigenvalue of the Atkin-Lehner involution W_p acting on f and g.) Then the $\mathfrak{q}$-torsion subgroup of $H^1_f(\mathbb{Q},A_{\mathfrak{q}}(k/2))$ has $\mathbb{F}_{\mathfrak{q}}$-rank at least r.*

Proof. The theorem is trivially true if $r = 0$, so we assume that $r > 0$. It follows easily from our hypothesis that the rank of the free part of $H^1_f(\mathbb{Q}, T'_{\mathfrak{q}}(k/2))$ is r. The natural map from $H^1_f(\mathbb{Q}, T'_{\mathfrak{q}}(k/2))/\mathfrak{q}H^1_f(\mathbb{Q}, T'_{\mathfrak{q}}(k/2))$ to $H^1(\mathbb{Q}, A'[\mathfrak{q}](k/2))$ is injective. Take a nonzero class c in the image, which has $\mathbb{F}_{\mathfrak{q}}$-rank r. Choose $d \in H^1_f(\mathbb{Q}, T'_{\mathfrak{q}}(k/2))$ mapping to c. Consider the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-cohomology of the short exact sequence

$$0 \longrightarrow A'[\mathfrak{q}](k/2) \longrightarrow A'_{\mathfrak{q}}(k/2) \xrightarrow{\pi} A'_{\mathfrak{q}}(k/2) \longrightarrow 0,$$

where π is multiplication by a uniformising element of $O_{\mathfrak{q}}$. By irreducibility, $H^0(\mathbb{Q}, A[\mathfrak{q}](k/2))$ is trivial. Hence $H^0(\mathbb{Q}, A_{\mathfrak{q}}(k/2))$ is trivial, so $H^1(\mathbb{Q}, A[\mathfrak{q}](k/2))$ injects into $H^1(\mathbb{Q}, A_{\mathfrak{q}}(k/2))$, and we get a nonzero, $\mathfrak{q}$-torsion class $\gamma \in H^1(\mathbb{Q}, A_{\mathfrak{q}}(k/2))$.

Our aim is to show that $\mathrm{res}_p(\gamma) \in H^1_f(\mathbb{Q}_p, A_{\mathfrak{q}}(k/2))$, for all (finite) primes p. We consider separately the cases $p \nmid qN$, $p \mid N$ and $p = q$.

Case (1) $p \nmid qN$:

Consider the I_p-cohomology of the short exact sequence above. Since in this case $A'_{\mathfrak{q}}(k/2)$ is unramified at p, $H^0(I_p, A'_{\mathfrak{q}}(k/2)) = A'_{\mathfrak{q}}(k/2)$, which is $\mathfrak{q}$-divisible. Therefore $H^1(I_p, A'[\mathfrak{q}](k/2))$ (which, remember, is the same as $H^1(I_p, A[\mathfrak{q}](k/2))$) injects into $H^1(I_p, A'_{\mathfrak{q}}(k/2))$. It follows from the fact that $d \in H^1_f(\mathbb{Q}, T'_{\mathfrak{q}}(k/2))$ that the image in $H^1(I_p, A'_{\mathfrak{q}}(k/2))$ of the restriction of c is zero, hence that the restriction of c to $H^1(I_p, A'[\mathfrak{q}](k/2)) \simeq H^1(I_p, A[\mathfrak{q}](k/2))$ is zero. Hence the restriction of γ to $H^1(I_p, A_{\mathfrak{q}}(k/2))$ is also zero. By line 3 of p. 125 of [Fl1], $H^1_f(\mathbb{Q}_p, A_{\mathfrak{q}}(k/2))$ is equal to (not just contained in) the kernel of the map from $H^1(\mathbb{Q}_p, A_{\mathfrak{q}}(k/2))$ to $H^1(I_p, A_{\mathfrak{q}}(k/2))$, so we have shown that $\mathrm{res}_p(\gamma) \in H^1_f(\mathbb{Q}_p, A_{\mathfrak{q}}(k/2))$.

Case (2) $p \mid N$:

First we show that $H^0(I_p, A'_{\mathfrak{q}}(k/2))$ is $\mathfrak{q}$-divisible. It suffices to show that

$$\dim H^0(I_p, A'[\mathfrak{q}](k/2)) = \dim H^0(I_p, V'_{\mathfrak{q}}(k/2)),$$

since then the natural map from $H^0(I_p, V'_{\mathfrak{q}}(k/2))$ to $H^0(I_p, A'_{\mathfrak{q}}(k/2))$ is surjective; this may be done as in the proof of Lemma 4.4. It follows as above that the image of $c \in H^1(\mathbb{Q}, A[\mathfrak{q}](k/2))$ in $H^1(I_p, A[\mathfrak{q}](k/2))$ is zero. Then $\mathrm{res}_p(c)$ comes from $H^1(D_p/I_p, H^0(I_p, A[\mathfrak{q}](k/2)))$, by inflation-restriction. The order of this group is the same as the order of the group $H^0(\mathbb{Q}_p, A[\mathfrak{q}](k/2))$ (this is Lemma 1 of [W]), which we claim is trivial. By the work of Carayol [Ca1], the level N is the conductor of $V_{\mathfrak{q}}(k/2)$, so $p \mid N$ implies that $V_{\mathfrak{q}}(k/2)$ is ramified at p, hence $\dim H^0(I_p, V_{\mathfrak{q}}(k/2)) = 0$ or 1. As above, we see that $\dim H^0(I_p, V_{\mathfrak{q}}(k/2)) = \dim H^0(I_p, A[\mathfrak{q}](k/2))$, so we need only consider the case where this common dimension is 1. The (motivic) Euler factor at p for M_f is $(1 - \alpha p^{-s})^{-1}$, where Frob_p^{-1} acts as multiplication by α on the one-dimensional space $H^0(I_p, V_{\mathfrak{q}})$. It follows from Theoréme A of [Ca1] that this is the same as the Euler factor at p of $L(f, s)$. By Theorems 3(ii) and 5 of [AL], it then follows that $p^2 \nmid N$ and $\alpha = -w_p p^{(k/2)-1}$, where $w_p = \pm 1$ is such that $W_p f = w_p f$. Twisting by $k/2$, Frob_p^{-1} acts on $H^0(I_p, V_{\mathfrak{q}}(k/2))$ (hence also on $H^0(I_p, A[\mathfrak{q}](k/2))$) as $-w_p p^{-1}$. Since $p \not\equiv -w_p \pmod{q}$, we see that $H^0(\mathbb{Q}_p, A[\mathfrak{q}](k/2))$ is trivial. Hence $\mathrm{res}_p(c) = 0$ so $\mathrm{res}_p(\gamma) = 0$ and certainly lies in $H^1_f(\mathbb{Q}_p, A_{\mathfrak{q}}(k/2))$.

Case (3) $p = q$:

Since $q \nmid N$ is a prime of good reduction for the motive M_g, $V'_{\mathfrak{q}}$ is a crystalline representation of $\mathrm{Gal}(\overline{\mathbb{Q}}_q/\mathbb{Q}_q)$, meaning $D_{\mathrm{cris}}(V'_{\mathfrak{q}})$ and $V'_{\mathfrak{q}}$ have the same dimension, where $D_{\mathrm{cris}}(V'_{\mathfrak{q}}) := H^0(\mathbb{Q}_q, V'_{\mathfrak{q}} \otimes_{\mathbb{Q}_q} B_{\mathrm{cris}})$. (This is a consequence of Theorem 5.6 of [Fa].) As already noted in the proof of Lemma 4.6, $T_{\mathfrak{q}}$ is the $O_{\mathfrak{q}}[\mathrm{Gal}(\overline{\mathbb{Q}}_q/\mathbb{Q}_q)]$-module associated to the filtered module $T_{\mathrm{dR}} \otimes O_{\mathfrak{q}}$. Since also $q > k$, we may now prove, in the same manner as Proposition 9.2 of [Du1], that $\mathrm{res}_q(\gamma) \in H^1_f(\mathbb{Q}_q, A_{\mathfrak{q}}(k/2))$. For the convenience of the reader, we give some details.

In Lemma 4.4 of [BK], a cohomological functor $\{h^i\}_{i\geq 0}$ is constructed on the Fontaine-Lafaille category of filtered Dieudonné modules over $\mathbb{Z}_q$. $h^i(D) = 0$ for all $i \geq 2$ and all D, and $h^i(D) = \mathrm{Ext}^i(1_{FD}, D)$ for all i and D, where 1_{FD} is the "unit" filtered Dieudonné module.

Now let $D = T_{\mathrm{dR}} \otimes O_{\mathfrak{q}}$ and $D' = T'_{\mathrm{dR}} \otimes O_{\mathfrak{q}}$. By Lemma 4.5 (c) of [BK],

$$h^1(D) \simeq H^1_e(\mathbb{Q}_q, T_{\mathfrak{q}}),$$

where

$$H^1_e(\mathbb{Q}_q, T_{\mathfrak{q}}) = \ker(H^1(\mathbb{Q}_q, T_{\mathfrak{q}}) \to H^1(\mathbb{Q}_q, V_{\mathfrak{q}})/H^1_e(\mathbb{Q}_q, V_{\mathfrak{q}}))$$

and

$$H^1_e(\mathbb{Q}_q, V_{\mathfrak{q}}) = \ker(H^1(\mathbb{Q}_q, V_{\mathfrak{q}}) \to H^1(\mathbb{Q}_q, B^{f=1}_{\mathrm{cris}} \otimes_{\mathbb{Q}_q} V_{\mathfrak{q}})).$$

Likewise $h^1(D') \simeq H^1_e(\mathbb{Q}_q, T'_{\mathfrak{q}})$. When applying results of [BK] we view D, $T_{\mathfrak{q}}$ etc. simply as $\mathbb{Z}_q$-modules, forgetting the $O_{\mathfrak{q}}$-structure.

For an integer j let $D(j)$ be D with the Hodge filtration shifted by j. Then

$$h^1(D(j)) \simeq H^1_e(\mathbb{Q}_q, T_{\mathfrak{q}}(j))$$

(as long as $k - p + 1 < j < p - 1$, so that $D(j)$ satisfies the hypotheses of Lemma 4.5 of [BK]). By Corollary 3.8.4 of [BK],

$$H^1_f(\mathbb{Q}_q, V_{\mathfrak{q}}(j))/H^1_e(\mathbb{Q}_q, V_{\mathfrak{q}}(j)) \simeq (D(j) \otimes_{\mathbb{Z}_q} \mathbb{Q}_q)/(1-f)(D(j) \otimes_{\mathbb{Z}_q} \mathbb{Q}_q),$$

where f is the Frobenius operator on crystalline cohomology. By 1.2.4(ii) of [Sc], and the Weil conjectures, $H^1_e(\mathbb{Q}_q, V_{\mathfrak{q}}(j)) = H^1_f(\mathbb{Q}_q, V_{\mathfrak{q}}(j))$, since $j \neq (k-1)/2$. Similarly $H^1_e(\mathbb{Q}_q, V'_{\mathfrak{q}}(j)) = H^1_f(\mathbb{Q}_q, V'_{\mathfrak{q}}(j))$.

We have

$$h^1(D(k/2)) \simeq H^1_f(\mathbb{Q}_q, T_{\mathfrak{q}}(k/2)) \quad \text{and} \quad h^1(D'(k/2)) \simeq H^1_f(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2)).$$

The exact sequence in the middle of page 366 of [BK] gives us a commutative diagram.

$$\begin{tikzcd}
h^1(D'(k/2)) \arrow[r, "\pi"] \arrow[d] & h^1(D'(k/2)) \arrow[r] \arrow[d] & h^1(D'(k/2)/\mathfrak{q}D'(k/2)) \arrow[d] \\
H^1(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2)) \arrow[r, "\pi"] & H^1(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2)) \arrow[r] & H^1(\mathbb{Q}_q, A'[\mathfrak{q}](k/2)).
\end{tikzcd}$$

The vertical arrows are all inclusions and we know that the image of $h^1(D'(k/2))$ in $H^1(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2))$ is exactly $H^1_f(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2))$. The top right horizontal map is surjective since $h^2(D'(k/2)) = 0$.

The class $\mathrm{res}_q(c) \in H^1(\mathbb{Q}_q, A'[\mathfrak{q}](k/2))$ is in the image of $H^1_f(\mathbb{Q}_q, T'_{\mathfrak{q}}(k/2))$, by construction, and therefore is in the image of $h^1(D'(k/2)/\mathfrak{q}D'(k/2))$. By the fullness and exactness of the Fontaine-Lafaille functor [FL] (see Theorem 4.3 of [BK]), $D'(k/2)/\mathfrak{q}D'(k/2)$ is isomorphic to $D(k/2)/\mathfrak{q}D(k/2)$.

It follows that the class $\mathrm{res}_q(c) \in H^1(\mathbb{Q}_q, A[\mathfrak{q}](k/2))$ is in the image of $h^1(D(k/2)/\mathfrak{q}D(k/2))$ by the vertical map in the exact sequence analogous to the above. Since the map from $h^1(D(k/2))$ to $h^1(D(k/2)/\mathfrak{q}D(k/2))$ is surjective, $\mathrm{res}_q(c)$ lies in the image of $H^1_f(\mathbb{Q}_q, T_{\mathfrak{q}}(k/2))$. From this it follows that $\mathrm{res}_q(\gamma) \in H^1_f(\mathbb{Q}_q, A_{\mathfrak{q}}(k/2))$, as desired. □

Theorem 2.7 of [AS] is concerned with verifying local conditions in the case $k = 2$, where f and g are associated with abelian varieties A and B. (Their theorem also applies to abelian varieties over number fields.) Our restriction outlawing congruences modulo $\mathfrak{q}$ with cusp forms of lower level is analogous to theirs forbidding q from dividing Tamagawa factors $c_{A,l}$ and $c_{B,l}$. (In the case where A is an elliptic curve with $\mathrm{ord}_l(j(A)) < 0$, consideration of a Tate parametrisation shows that if $q \mid c_{A,l}$, i.e., if $q \mid \mathrm{ord}_l(j(A))$, then it is possible that $A[q]$ is unramified at l.)

In this paper we have encountered two technical problems which we dealt with in quite similar ways:

(1) dealing with the $\mathfrak{q}$-part of c_p for $p \mid N$;
(2) proving local conditions at primes $p \mid N$, for an element of $\mathfrak{q}$-torsion.

If our only interest was in testing the Bloch-Kato conjecture at $\mathfrak{q}$, we could have made these problems cancel out, as in Lemma 8.11 of [DFG1], by weakening the local conditions. However, we have chosen not to do so, since we are also interested in the Shafarevich-Tate group, and since the hypotheses we had to assume are not particularly strong. Note that, since $A[\mathfrak{q}]$ is irreducible, the $\mathfrak{q}$-part of Ш does not depend on the choice of $T_{\mathfrak{q}}$.

7. Examples and Experiments

This section contains tables and numerical examples that illustrate the main themes of this paper. In Section 7.1, we explain Table 1, which contains 16 examples of pairs f, g such that the strong Beilinson-Bloch conjecture and Theorem 6.1 together imply the existence of nontrivial elements of the Shafarevich-Tate group of the motive attached to f. Section 7.2 outlines the higher-weight modular symbol computations that were used in making Table 1. Section 7.3 discusses Table 2, which summarizes the results of an extensive computation of conjectural orders of Shafarevich-Tate groups for modular motives of low level and weight. Section 7.4 gives specific examples in which various hypotheses fail. Note that in §7 "modular symbol" has a different meaning from in §5, being related to homology rather than cohomology. For precise definitions see [SV].

7.1. **Visible Ш Table 1.** Table 1 on page 11 lists sixteen pairs of newforms f and g (of equal weights and levels) along with at least one prime q such that there is a prime $\mathfrak{q} \mid q$ with $f \equiv g \pmod{\mathfrak{q}}$. In each case, $\mathrm{ord}_{s=k/2} L(g, k/2) \geq 2$ while $L(f, k/2) \neq 0$. The notation is as follows. The first column contains a label whose structure is

[Level]k[Weight][GaloisOrbit]

This label determines a newform $g = \sum a_n q^n$, up to Galois conjugacy. For example, **127k4C** denotes a newform in the third Galois orbit of newforms in $S_4(\Gamma_0(127))$. The Galois orbits are ordered first by the degree of $\mathbb{Q}(\ldots, a_n, \ldots)$, then by the sequence of absolute values $|\mathrm{Tr}(a_p(g))|$ for p not dividing the level, with positive trace being first in the event that the two absolute values are equal, and the first

Table 1. Visible Ш

g	$\deg(g)$	f	$\deg(f)$	q's
127k4A	1	**127k4C**	17	43
159k4B	1	**159k4E**	16	$5, 23$
365k4A	1	**365k4E**	18	29
369k4B	1	**369k4I**	9	13
453k4A	1	**453k4E**	23	17
465k4B	1	**465k4I**	7	11
477k4B	1	**477k4L**	12	73
567k4B	1	**567k4H**	8	23
581k4A	1	**581k4E**	34	19^2
657k4A	1	**657k4C**	7	5
657k4A	1	**657k4G**	12	5
681k4A	1	**681k4D**	30	59
684k4C	1	**684k4K**	4	7^2
95k6A	1	**95k6D**	9	$31, 59$
122k6A	1	**122k6D**	6	73
260k6A	1	**260k6E**	4	17

Galois orbit is denoted **A**, the second **B**, and so on. The second column contains the degree of the field $\mathbb{Q}(\ldots, a_n, \ldots)$. The third and fourth columns contain f and its degree, respectively. The fifth column contains at least one prime q such that there is a prime $\mathfrak{q} \mid q$ with $f \equiv g \pmod{\mathfrak{q}}$, and such that the hypotheses of Theorem 6.1 (except possibly $r > 0$) are satisfied for f, g, and $\mathfrak{q}$.

For the two examples **581k4E** and **684k4K**, the square of a prime q appears in the q-column, meaning q^2 divides the order of the group $S_k(\Gamma_0(N), \mathbb{Z})/(W + W^{\perp})$, defined at the end of 7.3 below.

We describe the first line of Table 1 in more detail. See the next section for further details on how the computations were performed.

Using modular symbols, we find that there is a newform

$$g = q - q^2 - 8q^3 - 7q^4 - 15q^5 + 8q^6 - 25q^7 + \cdots \in S_4(\Gamma_0(127))$$

with $L(g, 2) = 0$. Because $W_{127}(g) = g$, the functional equation has sign $+1$, so $L'(g, 2) = 0$ as well. We also find a newform $f \in S_4(\Gamma_0(127))$ whose Fourier coefficients generate a number field K of degree 17, and by computing the image of the modular symbol $XY\{0, \infty\}$ under the period mapping, we find that $L(f, 2) \neq 0$. The newforms f and g are congruent modulo a prime $\mathfrak{q}$ of K of residue characteristic 43. The mod $\mathfrak{q}$ reductions of f and g are both equal to

$$\overline{f} = q + 42q^2 + 35q^3 + 36q^4 + 28q^5 + 8q^6 + 18q^7 + \cdots \in \mathbb{F}_{43}[[q]].$$

There is no form in the Eisenstein subspaces of $M_4(\Gamma_0(127))$ whose Fourier coefficients of index n, with $(n, 127) = 1$, are congruent modulo 43 to those of $\overline{f}$, so $\rho_{f,\mathfrak{q}} \approx \rho_{g,\mathfrak{q}}$ is irreducible. Since 127 is prime and $S_4(\mathrm{SL}_2(\mathbb{Z})) = 0$, $\overline{f}$ does not arise from a level 1 form of weight 4. Thus we have checked the hypotheses of Theorem 6.1, so if r is the dimension of $H^1_f(\mathbb{Q}, V'_{\mathfrak{q}}(k/2))$ then the $\mathfrak{q}$-torsion subgroup of $H^1_f(\mathbb{Q}, A_{\mathfrak{q}}(k/2))$ has $\mathbb{F}_{\mathfrak{q}}$-rank at least r.

Recall that since $\operatorname{ord}_{s=k/2} L(g,s) \geq 2$, we expect that $r \geq 2$. Then, since $L(f,k/2) \neq 0$, we expect that the $\mathfrak{q}$-torsion subgroup of $H^1_f(\mathbb{Q}, A_{\mathfrak{q}}(k/2))$ is equal to the $\mathfrak{q}$-torsion subgroup of Ш. Admitting these assumptions, we have constructed the $\mathfrak{q}$-torsion in Ш predicted by the Bloch-Kato conjecture.

For particular examples of elliptic curves one can often find and write down rational points predicted by the Birch and Swinnerton-Dyer conjecture. It would be nice if likewise one could explicitly produce algebraic cycles predicted by the Beilinson-Bloch conjecture in the above examples. Since $L'(g,k/2) = 0$, Heegner cycles have height zero (see Corollary 0.3.2 of [Z]), so ought to be trivial in $\mathrm{CH}_0^{k/2}(M_g) \otimes \mathbb{Q}$.

7.2. **How the computation was performed.** We give a brief summary of how the computation was performed. The algorithms that we used were implemented by the second author, and most are a standard part of MAGMA (see [BCP]).

Let g, f, and q be some data from a line of Table 1 and let N denote the level of g. We verified the existence of a congruence modulo q, that $L(g, \frac{k}{2}) = L'(g, \frac{k}{2}) = 0$ and $L(f, \frac{k}{2}) \neq 0$, and that $\rho_{f,\mathfrak{q}} = \rho_{g,\mathfrak{q}}$ is irreducible and does not arise from any $S_k(\Gamma_0(N/p))$, as follows:

To prove there is a congruence, we showed that the corresponding *integral* spaces of modular symbols satisfy an appropriate congruence, which forces the existence of a congruence on the level of Fourier expansions. We showed that $\rho_{g,\mathfrak{q}}$ is irreducible by computing a set that contains all possible residue characteristics of congruences between g and any Eisenstein series of level dividing N, where by congruence, we mean a congruence for all Fourier coefficients of index n with $(n,N) = 1$. Similarly, we checked that g is not congruent to any form h of level N/p for any p that exactly divides N by listing a basis of such h and finding the possible congruences, where again we disregard the Fourier coefficients of index not coprime to N.

To verify that $L(g, \frac{k}{2}) = 0$, we computed the image of the modular symbol $\mathbf{e} = X^{\frac{k}{2}-1}Y^{\frac{k}{2}-1}\{0,\infty\}$ under a map with the same kernel as the period mapping, and found that the image was 0. The period mapping sends the modular symbol $\mathbf{e}$ to a nonzero multiple of $L(g, \frac{k}{2})$, so that $\mathbf{e}$ maps to 0 implies that $L(g, \frac{k}{2}) = 0$. In a similar way, we verified that $L(f, \frac{k}{2}) \neq 0$. Next, we checked that $W_N(g) = (-1)^{k/2} g$ which, because of the functional equation, implies that $L'(g, \frac{k}{2}) = 0$. Table 1 is of independent interest because it includes examples of modular forms of even weight > 2 with a zero at $\frac{k}{2}$ that is not forced by the functional equation. We found no such examples of weights ≥ 8.

7.3. **Conjecturally nontrivial Ш.** In this section we apply some of the results of Section 4 to compute lower bounds on conjectural orders of Shafarevich-Tate groups of many modular motives. The results of this section suggest that Ш of a modular motive is usually not "visible at level N", i.e., explained by congruences at level N, which agrees with the observations of [CM1] and [AS]. For example, when $k > 6$ we find many examples of conjecturally nontrivial Ш but no examples of nontrivial visible Ш.

For any newform f, let $L(M_f/\mathbb{Q}, s) = \prod_{i=1}^{d} L(f^{(i)}, s)$ where $f^{(i)}$ runs over the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f. Let T be the complex torus $\mathbb{C}^d/(2\pi i)^{k/2}\mathcal{L}$, where the lattice $\mathcal{L}$ is defined by integrating integral cuspidal modular symbols (for $\Gamma_0(N)$) against the conjugates of f. Let $\Omega_{M_f/\mathbb{Q}}$ denote the volume of the $(-1)^{(k/2)-1}$ eigenspace $T^{\pm} = \{z \in T : \overline{z} = (-1)^{(k/2)-1} z\}$ for complex conjugation on T.

Lemma 7.1. *Suppose that $p \nmid Nk!$ is such that f is not congruent to any of its Galois conjugates modulo a prime dividing p. Then the p-parts of*

$$\frac{L(M_f/\mathbb{Q}, k/2)}{\Omega_{M_f/\mathbb{Q}}} \qquad \textit{and} \qquad \mathrm{Norm}\left(\frac{L(f,k/2)}{\mathrm{vol}_\infty}\mathfrak{a}^{\pm}\right)$$

are equal, where vol_∞ *is as in Section 4.*

Proof. Let H be the $\mathbb{Z}$-module of all integral cuspidal modular symbols for $\Gamma_0(N)$. Let I be the image of H under projection into the submodule of $H \otimes \mathbb{Q}$ corresponding to f and its Galois conjugates. Note that I is not necessarily contained in H, but it is contained in $H \otimes \mathbb{Z}[\frac{1}{m}]$ where m is divisible by the residue characteristics of any primes of congruence between f and cuspforms of weight k for $\Gamma_0(N)$ which are not Galois conjugate to f.

The lattice $\mathcal{L}$ defined in the paragraph before the lemma is (up to divisors of $Nk!$) obtained by pairing the cohomology modular symbols $\Phi^{\pm}_{f^{(i)}}$ (as in §5) with the homology modular symbols in H; equivalently, since the pairing factors through the map $H \to I$, the lattice $\mathcal{L}$ is obtained by pairing with the elements of I. For $1 \leq i \leq d$ let I_i be the O_E-module generated by the image of the projection of I into $I \otimes E$ corresponding to $f^{(i)}$. The finite index of $I \otimes O_E$ in $\oplus_{i=1}^{d} I_i$ is divisible only by primes of congruence between f and its Galois conjugates. Up to these primes, $\Omega_{M_f/\mathbb{Q}}/(2\pi i)^{((k/2)-1)d}$ is then a product of the d quantities obtained by pairing $\Phi^{\pm}_{f^{(i)}}$ with I_i, for $1 \leq i \leq d$. (These quantities inhabit a kind of tensor product of $\mathbb{C}^*$ over E^* with the group of fractional ideals of E.) Bearing in mind the last line of §3, we see that these quantities are the $\mathfrak{a}^{\pm}\Omega^{\pm}_{f^{(i)}}$, up to divisors of $Nk!$. Now we may apply Lemma 4.1. We have then a factorisation of the left hand side which shows it to be equal to the right hand side, to the extent claimed by the lemma. Note that $\frac{L(f,k/2)}{\mathrm{vol}_\infty}\mathfrak{a}^{\pm}$ has an interpretation in terms of integral modular symbols, as in §5, and just gets Galois-conjugated when one replaces f by some $f^{(i)}$. □

Remark 7.2. The newform $f =$ **319k4C** is congruent to one of its Galois conjugates modulo 17 and $17 \mid \frac{L(M_f/\mathbb{Q},k/2)}{\Omega_{M_f/\mathbb{Q}}}$ so the lemma and our computations say nothing about whether or not 17 divides $\mathrm{Norm}\left(\frac{L(f,k/2)}{\mathrm{vol}_\infty}\mathfrak{a}^{\pm}\right)$.

Let $\mathcal{S}$ be the set of newforms with level N and weight k satisfying either $k = 4$ and $N \leq 321$, or $k = 6$ and $N \leq 199$, or $k = 8$ and $N \leq 149$, or $k = 10$ and $N \leq 72$, or $k = 12$ and $N \leq 49$. Given $f \in \mathcal{S}$, let B be defined as follows:

1. Let L_1 be the numerator of the rational number $L(M_f/\mathbb{Q}, k/2)/\Omega_{M_f/\mathbb{Q}}$. If $L_1 = 0$ let $B = 1$ and terminate.
2. Let L_2 be the part of L_1 that is coprime to $Nk!$.
3. Let L_3 be the part of L_2 that is coprime to $p + 1$ for every prime p such that $p^2 \mid N$.
4. Let L_4 be the part of L_3 coprime to the residue characteristic of any prime of congruence between f and a form of weight k and lower level. (By congruence here, we mean a congruence for coefficients a_n with n coprime to the level of f.)
5. Let L_5 be the part of L_4 coprime to the residue characteristic of any prime of congruence between f and an Eisenstein series. (This eliminates residue characteristics of reducible representations.)

Table 2. Conjecturally nontrivial Ш (mostly invisible)

f	$\deg(f)$	B (Ш bound)	all odd congruence primes
127k4C$*$	17	43^2	$43, 127$
159k4E$*$	8	23^2	$3, 5, 11, 23, 53, 13605689$
263k4B	39	41^2	263
269k4C	39	23^2	269
271k4B	39	29^2	271
281k4B	40	29^2	281
295k4C	16	7^2	$3, 5, 11, 59, 101, 659, 70791023$
299k4C	20	29^2	$13, 23, 103, 20063, 21961$
321k4C	16	13^2	$3, 5, 107, 157, 12782373452377$
95k6D$*$	9	$31^2{\cdot}59^2$	$3, 5, 17, 19, 31, 59, 113, 26701$
101k6B	24	17^2	101
103k6B	24	23^2	103
111k6C	9	11^2	$3, 37, 2796169609$
122k6D$*$	6	73^2	$3, 5, 61, 73, 1303196179$
153k6G	5	7^2	$3, 17, 61, 227$
157k6B	34	251^2	157
167k6B	40	41^2	167
172k6B	9	7^2	$3, 11, 43, 787$
173k6B	39	71^2	173
181k6B	40	107^2	181
191k6B	46	85091^2	191
193k6B	41	31^2	193
199k6B	46	200329^2	199
47k8B	16	19^2	47
59k8B	20	29^2	59
67k8B	20	29^2	67
71k8B	24	379^2	71
73k8B	22	197^2	73
74k8C	6	23^2	$37, 127, 821, 8327168869$
79k8B	25	307^2	79
83k8B	27	1019^2	83
87k8C	9	11^2	$3, 5, 7, 29, 31, 59, 947, 22877, 3549902897$
89k8B	29	44491^2	89
97k8B	29	$11^2{\cdot}277^2$	97
101k8B	33	$19^2{\cdot}11503^2$	101
103k8B	32	75367^2	103
107k8B	34	$17^2{\cdot}491^2$	107
109k8B	33	$23^2{\cdot}229^2$	109
111k8C	12	127^2	$3, 7, 11, 13, 17, 23, 37, 6451, 18583, 51162187$
113k8B	35	$67^2{\cdot}641^2$	113
115k8B	12	37^2	$3, 5, 19, 23, 572437, 5168196102449$
117k8I	8	19^2	$3, 13, 181$
118k8C	8	37^2	$5, 13, 17, 59, 163, 3923085859759909$
119k8C	16	1283^2	$3, 7, 13, 17, 109, 883, 5324191, 91528147213$

f	$\deg(f)$	B (Ш bound)	all odd congruence primes
121k8F	6	71^2	$3, 11, 17, 41$
121k8G	12	13^2	$3, 11$
121k8H	12	19^2	$5, 11$
125k8D	16	179^2	5
127k8B	39	59^2	127
128k8F	4	11^2	1
131k8B	43	$241^2 \cdot 817838201^2$	131
134k8C	11	61^2	$11, 17, 41, 67, 71, 421, 2356138931854759$
137k8B	42	$71^2 \cdot 749093^2$	137
139k8B	43	$47^2 \cdot 89^2 \cdot 1021^2$	139
141k8C	14	13^2	$3, 5, 7, 47, 4639, 43831013, 4047347102598757$
142k8B	10	11^2	$3, 53, 71, 56377, 1965431024315921873$
143k8C	19	307^2	$3, 11, 13, 89, 199, 409, 178397, 639259, 1744053597287$
143k8D	21	109^2	$3, 7, 11, 13, 61, 79, 103, 173, 241, 769, 36583$
145k8C	17	29587^2	$5, 11, 29, 107, 251623, 393577, 518737, 9837145699$
146k8C	12	3691^2	$11, 73, 269, 503, 1673540153, 11374452082219$
148k8B	11	19^2	$3, 37$
149k8B	47	$11^4 \cdot 40996789^2$	149
43k10B	17	449^2	43
47k10B	20	2213^2	47
53k10B	21	673^2	53
55k10D	9	71^2	$3, 5, 11, 251, 317, 61339, 19869191$
59k10B	25	37^2	59
62k10E	7	23^2	$3, 31, 101, 523, 617, 41192083$
64k10K	2	19^2	3
67k10B	26	$191^2 \cdot 617^2$	67
68k10B	7	83^2	$3, 7, 17, 8311$
71k10B	30	1103^2	71
19k12B	9	67^2	$5, 17, 19, 31, 571$
31k12B	15	$67^2 \cdot 71^2$	$31, 13488901$
35k12C	6	17^2	$5, 7, 23, 29, 107, 8609, 1307051$
39k12C	6	73^2	$3, 13, 1491079, 3719832979693$
41k12B	20	54347^2	$7, 41, 3271, 6277$
43k12B	20	212969^2	$43, 1669, 483167$
47k12B	23	24469^2	$17, 47, 59, 2789$
49k12H	12	271^2	7

(6) Let B be the part of L_5 coprime to the residue characteristic of any prime of congruence between f and any one of its Galois conjugates.

Proposition 4.8 and Lemma 7.1 imply that if $\mathrm{ord}_p(B) > 0$ then, according to the Bloch-Kato conjecture, $\mathrm{ord}_p(\#Ш) = \mathrm{ord}_p(B) > 0$.

We computed B for every newform in $\mathcal{S}$. There are many examples in which L_3 is large, but B is not, and this is because of Tamagawa factors. For example, **39k4C** has $L_3 = 19$, but $B = 1$ because of a 19-congruence with a form of level 13; in this case we must have $19 \mid c_3(2)$, where $c_3(2)$ is as in Section 4. See Section 7.4 for more details. Also note that in every example B is a perfect square, which, away from congruence primes, is as predicted by the existence of Flach's generalised Cassels-Tate pairing [Fl1]. (Note that if $A[\lambda]$ is irreducible then the lattice T_λ is at worst a scalar multiple of its dual, so the pairing shows that the order of the λ-part of Ш, if finite, is a square.) That our computed value of B should be a square is not *a priori* obvious.

For simplicity, we discard residue characteristics instead of primes of rings of integers, so our definition of B is overly conservative. For example, 5 occurs in row 2 of Table 1 but not in Table 2, because **159k4E** is Eisenstein at some prime above 5, but the prime of congruences of characteristic 5 between **159k4B** and **159k4E** is not Eisenstein.

The newforms for which $B > 1$ are given in Table 2. The second column of the table records the degree of the field generated by the Fourier coefficients of f. The third contains B. Let W be the intersection of the span of all conjugates of f with $S_k(\Gamma_0(N), \mathbb{Z})$ and $W^\perp$ the Petersson orthogonal complement of W in $S_k(\Gamma_0(N), \mathbb{Z})$. The fourth column contains the odd prime divisors of $\#(S_k(\Gamma_0(N), \mathbb{Z})/(W + W^\perp))$, which are exactly the possible primes of congruence between f and non-conjugate cusp forms of the same weight and level. We place a $*$ next to the four entries of Table 2 that also occur in Table 1.

7.4. **Examples in which hypotheses fail.** We have some other examples where forms of different levels are congruent (for Fourier coefficients of index coprime to the levels). However, Remark 5.2 does not apply, so that one of the forms could have an odd functional equation, and the other could have an even functional equation. For instance, we have a 19-congruence between the newforms g = **13k4A** and f = **39k4C** of Fourier coefficients of index coprime to 39. Here $L(f, 2) \neq 0$, while $L(g, 2) = 0$ since $L(g, s)$ has *odd* functional equation. Here f fails the condition about not being congruent to a form of lower level, so in Lemma 4.4 it is possible that $\mathrm{ord}_{\mathfrak{q}}(c_3(2)) > 0$. In fact this does happen. Because $V'_{\mathfrak{q}}$ (attached to g of level 13) is unramified at $p = 3$, $H^0(I_p, A[\mathfrak{q}])$ (the same as $H^0(I_p, A'[\mathfrak{q}])$) is two-dimensional. As in (2) of the proof of Theorem 6.1, one of the eigenvalues of Frob_p^{-1} acting on this two-dimensional space is $\alpha = -w_p p^{(k/2)-1}$, where $W_p f = w_p f$. The other must be $\beta = -w_p p^{k/2}$, so that $\alpha\beta = p^{k-1}$. Twisting by $k/2$, we see that Frob_p^{-1} acts as $-w_p$ on the quotient of $H^0(I_p, A[\mathfrak{q}](k/2))$ by the image of $H^0(I_p, V_{\mathfrak{q}}(k/2))$. Hence $\mathrm{ord}_{\mathfrak{q}}(c_p(k/2)) > 0$ when $w_p = -1$, which is the case in our example here with $p = 3$. Likewise $H^0(\mathbb{Q}_p, A[\mathfrak{q}](k/2))$ is nontrivial when $w_p = -1$, so (2) of the proof of Theorem 6.1 does not work. This is just as well, since had it worked we would have expected $\mathrm{ord}_{\mathfrak{q}}(L(f, k/2)/\mathrm{vol}_\infty) \geq 3$, which computation shows not to be the case.

In the following example, the divisibility between the levels is the other way round. There is a 7-congruence between g = **122k6A** and f = **61k6B**, both L-functions have even functional equation, and $L(g, 3) = 0$. In the proof of Theorem 6.1, there is a problem with the local condition at $p = 2$. The map from $H^1(I_2, A'[\mathfrak{q}](3))$ to $H^1(I_2, A'_{\mathfrak{q}}(3))$ is not necessarily injective, but its kernel is at most one dimensional, so we still get the $\mathfrak{q}$-torsion subgroup of $H^1_f(\mathbb{Q}, A_{\mathfrak{q}}(2))$ having $\mathbb{F}_{\mathfrak{q}}$-rank at least 1 (assuming $r \geq 2$), and thus get elements of Ш for **61k6B** (assuming all along the strong Beilinson-Bloch conjecture). In particular, these elements of Ш are *invisible* at level 61. When the levels are different we are no longer able to apply Theorem 2.1 of [FJ]. However, we still have the congruences of integral modular symbols required to make the proof of Proposition 5.1 go through. Indeed, as noted above, the congruences of modular forms were found by producing congruences of modular symbols. Despite these congruences of modular symbols, Remark 5.2 does not apply, since there is no reason to suppose that $w_N = w_{N'}$, where N and N' are the distinct levels.

Finally, there are two examples where we have a form g with even functional equation such that $L(g, k/2) = 0$, and a congruent form f which has odd functional equation; these are a 23-congruence between $g =$ **453k4A** and $f =$ **151k4A**, and a 43-congruence between $g =$ **681k4A** and $f =$ **227k4A**. If $\mathrm{ord}_{s=2} L(f, s) = 1$, it ought to be the case that $\dim(H^1_f(\mathbb{Q}, V_{\mathfrak{q}}(2))) = 1$. If we assume this is so, and similarly that $r = \mathrm{ord}_{s=2}(L(g, s)) \geq 2$, then unfortunately the appropriate modification of Theorem 6.1 (with strong Beilinson-Bloch conjecture) does not necessarily provide us with nontrivial $\mathfrak{q}$-torsion in Ш. It only tells us that the $\mathfrak{q}$-torsion subgroup of $H^1_f(\mathbb{Q}, A_{\mathfrak{q}}(2))$ has $\mathbb{F}_{\mathfrak{q}}$-rank at least 1. It could all be in the image of $H^1_f(\mathbb{Q}, V_{\mathfrak{q}}(2))$. Ш appears in the conjectural formula for the first derivative of the complex L function, evaluated at $s = k/2$, but in combination with a regulator that we have no way of calculating.

Let $L_q(f, s)$ and $L_q(g, s)$ be the q-adic L functions associated with f and g by the construction of Mazur, Tate and Teitelbaum [MTT], each divided by a suitable canonical period. We may show that $\mathfrak{q} \mid L'_q(f, k/2)$, though it is not quite clear what to make of this. This divisibility may be proved as follows. The measures $d\mu_{f,\alpha}$ and (a q-adic unit times) $d\mu_{g,\alpha'}$ in [MTT] (again, suitably normalised) are congruent mod $\mathfrak{q}$, as a result of the congruence between the modular symbols out of which they are constructed. Integrating an appropriate function against these measures, we find that $L'_q(f, k/2)$ is congruent mod $\mathfrak{q}$ to $L'_q(g, k/2)$. It remains to observe that $L'_q(g, k/2) = 0$, since $L(g, k/2) = 0$ forces $L_q(g, k/2) = 0$, but we are in a case where the signs in the functional equations of $L(g, s)$ and $L_q(g, s)$ are the same, positive in this instance. (According to the proposition in Section 18 of [MTT], the signs differ precisely when $L_q(g, s)$ has a "trivial zero" at $s = k/2$.)

We also found some examples for which the conditions of Theorem 6.1 were not met. For example, we have a 7-congruence between **639k4B** and **639k4H**, but $w_{71} = -1$, so that $71 \equiv -w_{71} \pmod 7$. There is a similar problem with a 7-congruence between **260k6A** and **260k6E** — here $w_{13} = 1$ so that $13 \equiv -w_{13} \pmod 7$. According to Propositions 5.1 and 4.8, Bloch-Kato still predicts that the $\mathfrak{q}$-part of Ш is non-trivial in these examples. Finally, there is a 5-congruence between **116k6A** and **116k6D**, but here the prime 5 is less than the weight 6 so Propositions 5.1 and 4.8 (and even Lemma 7.1) do not apply.

References

[AL] A. O. L. Atkin, J. Lehner, Hecke operators on $\Gamma_0(m)$, *Math. Ann.* **185** (1970), 135–160.

[AS] A. Agashe, W. Stein, Visibility of Shafarevich-Tate groups of abelian varieties, preprint.

[BS-D] B. J. Birch, H. P. F. Swinnerton-Dyer, Notes on elliptic curves. I and II. *J. reine angew. Math.* **212** (1963), 7–25, **218** (1965), 79–108.

[B] S. Bloch, Algebraic cycles and values of L-functions, *J. reine angew. Math.* **350** (1984), 94–108.

[BCP] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[Be] A. Beilinson, Height pairing between algebraic cycles, *in* Current trends in arithmetical algebraic geometry (K. Ribet, ed.) *Contemp. Math.* **67** (1987), 1–24.

[BK] S. Bloch, K. Kato, L-functions and Tamagawa numbers of motives, The Grothendieck Festschrift Volume I, 333–400, Progress in Mathematics, 86, Birkhäuser, Boston, 1990.

[Ca1] H. Carayol, Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert, *Ann. Sci. École Norm. Sup. (4)***19** (1986), 409–468.

[Ca2] H. Carayol, Sur les représentations Galoisiennes modulo ℓ attachées aux formes modulaires, *Duke Math. J.* **59** (1989), 785–801.

[CM1] J. E. Cremona, B. Mazur, Visualizing elements in the Shafarevich-Tate group, *Experiment. Math.* **9** (2000), 13–28.

[CM2] J. E. Cremona, B. Mazur, Appendix to A. Agashe, W. Stein, Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of rank zero, preprint.

[CF] B. Conrey, D. Farmer, On the non-vanishing of $L_f(s)$ at the center of the critical strip, preprint.

[De1] P. Deligne, Formes modulaires et représentations ℓ-adiques. Sém. Bourbaki, éxp. 355, Lect. Notes Math. **179**, 139–172, Springer, 1969.

[De2] P. Deligne, Valeurs de Fonctions L et Périodes d'Intégrales, *AMS Proc. Symp. Pure Math.,* Vol. 33 (1979), part 2, 313–346.

[DFG1] F. Diamond, M. Flach, L. Guo, Adjoint motives of modular forms and the Tamagawa number conjecture, preprint. http://www.andromeda.rutgers.edu/~liguo/lgpapers.html

[DFG2] F. Diamond, M. Flach, L. Guo, The Bloch-Kato conjecture for adjoint motives of modular forms, *Math. Res. Lett.* **8** (2001), 437–442.

[Du1] N. Dummigan, Symmetric square L-functions and Shafarevich-Tate groups, *Experiment. Math.* **10** (2001), 383–400.

[Du2] N. Dummigan, Congruences of modular forms and Selmer groups, *Math. Res. Lett.* **8** (2001), 479–494.

[Fa] G. Faltings, Crystalline cohomology and p-adic Galois representations, *in* Algebraic analysis, geometry and number theory (J. Igusa, ed.), 25–80, Johns Hopkins University Press, Baltimore, 1989.

[FJ] G. Faltings, B. Jordan, Crystalline cohomology and $\mathrm{GL}(2,\mathbb{Q})$, *Israel J. Math.* **90** (1995), 1–66.

[Fl1] M. Flach, A generalisation of the Cassels-Tate pairing, *J. reine angew. Math.* **412** (1990), 113–127.

[Fl2] M. Flach, On the degree of modular parametrisations, Séminaire de Théorie des Nombres, Paris 1991-92 (S. David, ed.), 23–36, Progress in mathematics, 116, Birkhäuser, Basel Boston Berlin, 1993.

[Fo1] J.-M. Fontaine, Sur certains types de représentations p-adiques du groupe de Galois d'un corps local, construction d'un anneau de Barsotti-Tate, *Ann. Math.* **115** (1982), 529–577.

[Fo2] J.-M. Fontaine, Valeurs spéciales des fonctions L des motifs, Séminaire Bourbaki, Vol. 1991/92. *Astérisque* **206** (1992), Exp. No. 751, 4, 205–249.

[FL] J.-M. Fontaine, G. Lafaille, Construction de représentations p-adiques, *Ann. Sci. E.N.S.* **15** (1982), 547–608.

[JL] B. W. Jordan, R. Livné, Conjecture "epsilon" for weight $k > 2$, *Bull. Amer. Math. Soc.* **21** (1989), 51–56.

[L] R. Livné, On the conductors of mod ℓ Galois representations coming from modular forms, *J. Number Theory* **31** (1989), 133–141.

[MTT] B. Mazur, J. Tate, J. Teitelbaum, On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer, *Invent. Math.* **84** (1986), 1–48.

[Ne] J. Nekovár, p-adic Abel-Jacobi maps and p-adic heights. The arithmetic and geometry of algebraic cycles (Banff, AB, 1998), 367–379, CRM Proc. Lecture Notes, 24, Amer. Math. Soc., Providence, RI, 2000.

[Sc] A. J. Scholl, Motives for modular forms, *Invent. Math.* **100** (1990), 419–430.

[SV] W. A. Stein, H. A. Verrill, Cuspidal modular symbols are transportable, *L.M.S. Journal of Computational Mathematics* **4** (2001), 170–181.

[St] G. Stevens, Λ-adic modular forms of half-integral weight and a Λ-adic Shintani lifting. Arithmetic geometry (Tempe, AZ, 1993), 129–151, Contemp. Math., **174**, Amer. Math. Soc., Providence, RI, 1994.

[V] V. Vatsal, Canonical periods and congruence formulae, *Duke Math. J.* **98** (1999), 397–419.

[W] L. C. Washington, Galois cohomology, *in* Modular Forms and Fermat's Last Theorem, (G. Cornell, J. H. Silverman, G. Stevens, eds.),101–120, Springer-Verlag, New York, 1997.

[Z] S. Zhang, Heights of Heegner cycles and derivatives of L-series, *Invent. Math.* **130** (1997), 99–152.

University of Sheffield, Department of Pure Mathematics, Hicks Building, Hounsfield Road, Sheffield, S3 7RH, U.K.

Harvard University, Department of Mathematics, One Oxford Street, Cambridge, MA 02138, U.S.A.

Penn State Mathematics Department, University Park, State College, PA 16802, U.S.A.

E-mail address: `n.p.dummigan@shef.ac.uk`
E-mail address: `was@math.harvard.edu`
E-mail address: `watkins@math.psu.edu`

18 A Database of Elliptic CurvesFirst Report, with M. Watkins

A Database of Elliptic Curves—First Report

William A. Stein[1] and Mark Watkins[2]

[1] Harvard University
was@math.harvard.edu
http://modular.fas.harvard.edu
[2] The Pennsylvania State Univeristy
watkins@math.psu.edu
http://www.math.psu.edu/watkins

1 Introduction

In the late 1980s, Brumer and McGuinness [2] undertook the construction of a database of elliptic curves whose discriminant Δ was both prime and satisfied $|\Delta| \leq 10^8$. While the restriction to primality was nice for many reasons, there are still many curves of interest lacking this property. As ten years have passed since the original experiment, we decided to undertake an extension of it, simultaneously extending the range for the type of curves they considered, and also including curves with composite discriminant. Our database can be crudely described as being the curves with $|\Delta| \leq 10^{12}$ which either have conductor smaller than 10^8 or have prime conductor less than 10^{10}—but there are a few caveats concerning issues like quadratic twists and isogenous curves. For each curve in our database, we have undertaken to compute various invariants (as did Brumer and McGuinness), such as the Birch–Swinnerton-Dyer L-ratio, generators, and the modular degree. We did not compute the latter two of these for every curve. The database currently contains about 44 million curves; the end goal is find as many curves with conductor less than 10^8 as possible, and we comment on this direction of growth of the database below. Of these 44 million curves, we have started a first stage of processing (computation of analytic rank data), with point searching to be carried out in a later second stage of computation.

Our general frame of mind is that computation of many of the invariants is rather trivial, for instance, the discriminant, conductor, and even the isogeny structure. We do not even save these data, expecting them to be recomputable quite easily in real time. For instance, for each isogeny class, we store only one representative (the one of minimal Faltings height), as we view the construction of isogenous curves as a “fast” process. It is only information like analytic ranks, modular degrees (both of which use computation of the Frobenius traces a_p), and coordinates of generators that we save; saving the a_p would take too much storage space. It might be seen that our database could be used a "seed" for other more specialised databases, as we can quickly calculate the less time-consuming information and append it to the saved data.

2 Generating the Curves.

While Brumer and McGuinness fixed the a_1, a_2, a_3 invariants of the elliptic curve (12 total possibilities) and then searched for a_4 and a_6 which made $|\Delta|$ small, we instead decided to break the c_4 and c_6 invariants into congruence classes, and then find small solutions to $c_4^3 - c_6^2 = 1728\Delta$. We write $c_4^\star$ for the least nonnegative residue of c_4 modulo 576, and $c_6^\star$ for the least nonnegative residue of c_6 modulo 1728. The work of Connell [3] gives necessary and sufficient conditions on c_4 and c_6 for an elliptic curve with such invariants to exist. We first need that $c_6 \equiv 3 \pmod 4$ (when it follows that c_4 is odd), or $2^4 \mid c_4$ and $c_6 \equiv 0, 8 \pmod{32}$, and secondly we require a local condition at the prime 3, namely that $c_6 \not\equiv \pm 9 \pmod{27}$. Using this information and the fact that $1728 \mid (c_4^3 - c_6^2)$, this leads to 288 possible $(c_4^\star, c_6^\star)$ pairs.

For each fixed such $(c_4^\star, c_6^\star)$ pair, we can simply loop over c_4 and c_6, finding the curves with $|\Delta| \le 10^{12}$. Of course, it is only under the ABC-conjecture that we would have an upper bound on c_4 to ensure that we would have found all such curves, and even then the bound would be too large. Our method was to take $c_4 \le 1.44 \cdot 10^{12}$ in this first step; in any case, curves with larger c_4 are most likely found more easily using the method of Elkies [5].

2.1 Minimal Twists

In the sequel, we shall write E_d for the quadratic twist of E by d. For each (c_4, c_6) pair (again with $c_4 \le 1.44 \cdot 10^{12}$) which satisfies the $|\Delta| \le 10^{12}$ condition, we then determine whether this curve is minimal—not only in the traditional sense for its minimal discriminant, but also whether it is has the minimal discriminant in its family of quadratic twists. For $p \ge 5$, this is rather easy to determine; unless $p^6 \mid \Delta$ and $p \mid c_4$, the curve is minimal for quadratic twists (the only difference between this and the standard notion of minimality is that the exponent here is 6 instead of 12). If both the above conditions hold, then we throw the curve out, as $E_{\tilde{p}}$, where $\tilde{p} = \left(\frac{-1}{p}\right)p$, is a curve with lesser discriminant (which will be found by our search procedure). Given that the curve is minimal at a prime divisor $p \geq 5$ of $\varDelta$, the local conductor at p is p^2 if $p \mid c_4$ and p^1 otherwise.

The case with $p = 3$ is a bit harder. By Connell's conditions, we see that if $3^9 \mid (c_4^3 - c_6^2)$ while $3 \mid c_6$ but 3^5 does not exactly divide c_6, then E_{-3} is a curve with invariants $(c_4/9, -c_6/27)$ which has the discriminant reduced by 3^6. This is the only prohibition against the curve being the minimal twist at 3. If $3 \mid\mid c_4$, the curve has good reduction (at 3), while if c_4 is not divisible by 3, the curve has either good or multiplicative reduction. In both cases, the local conductor can be computed readily, it being 3^0 for good reduction and 3^1 for multiplicative. To compute the conductor in the remaining cases, let $\tilde{c}_4$ be the the least nonnegative residue of $(c_4/9)$ modulo 3, and $\tilde{c}_6$ be the the least nonnegative residue of $(c_6/27)$ modulo 9. Table 1 then gives us the exponent of the local conductor. Here $e = 5$ if $3^4 \mid c_4$ and $e = 4$ if $3^3 \mid\mid c_4$ (note that we must have $3^5 \mid\mid c_6$ in this case for the curve to be twist-minimal).

Table 1. Local Conductors at 3

$\tilde{c}_4 \backslash \tilde{c}_6$	0	1	2	3	4	5	6	7	8
0	e	3	3	5	2	2	5	3	3
1	2	3	4	3	4	4	3	4	3
2	2	3	2	3	3	3	3	2	3

For $p = 2$, the minimality test and conductor computation is much more complicated. We include the prime at infinity (twisting by -1) in the test for $p = 2$. By Connell's conditions, if $2^6 \mid c_4$ and $2^8 \mid c_6$, we see that E_2 is a curve with invariants $(c_4/4, c_6/8)$, and has a lesser discriminant. Also if $2^6 \mid c_4$ and $2^6 \mid\mid c_6$, then one of the twists $E_{\pm 2}$ (the sign depending on whether $c_6/8$ is 8 mod 32) has lesser discriminant. And finally if we have $2^4 \mid\mid c_4$ and $2^6 \mid\mid c_6$ and $2^{18} \mid (c_4^3 - c_6^2)$, then one of $E_{\pm 1}$ (depending on whether $c_6/64$ is 3 mod 4) is nonminimal (in the standard sense) at 2, and hence can be ignored. If none of these events happens, then the curve is twist-minimal at $p = 2$ and the infinite prime. We next describe how to compute the local conductor at $p = 2$ in terms of congruence conditions. If c_4 is odd, then the local conductor is 2^0 or 2^1, depending on whether 2 divides $\varDelta$. If c_4 is even, then it is divisible by 16. In this case, if c_6 is 8 mod 32, there is good reduction at 2, and again the local conductor is 2^0. So we are left to consider the cases of additive reduction where $2^4 \mid c_4$ and $2^5 \mid c_6$. Let $\tilde{c}_4$ be the the least nonnegative residue of

$(c_4/16)$ modulo 8, and $\tilde{c}_6$ be the the least nonnegative residue of $(c_6/32)$ modulo 8. Table 2 then gives the exponent of the local conductor at 2. In this, the dashed entries simply do not occur. For the entries marked by e, let $\tilde{c}_4$ be the the least nonnegative residue of $(c_4/16)$ modulo 16, and $\tilde{c}_6$ be the the least nonnegative residue of $(c_6/32)$ modulo 16. We then use the further Table 3. All the conductor computations are exercises with Tate's algorithm [9].

Table 2. Local Conductors at 2

$\tilde{c}_4 \backslash \tilde{c}_6$	0	1	2	3	4	5	6	7
1,5	6	4	e	3	6	4	e	3
2,6	8	3	6	4	7	3	6	4
3,7	5	2	7	2	5	2	7	4
4	6	2	-	4	3	2	-	4
0	6	2	-	4	2	2	-	4

Table 3. More of the Same

$\tilde{c}_4 \backslash \tilde{c}_6$	2	6	10	14
1	4	5	5	3
5	3	2	4	4
9	5	3	4	5
13	4	4	3	2

A curve which has minimal discriminant at $p = 2$ will be of minimal conductor at $p = 2$ unless $2^4 \mid\mid N$ or $2^6 \mid\mid N$; we can throw out the curve in the first case, since E_{-1} will be found in the search process (and it has lesser conductor). But in the latter case, we cannot immediately discard the curve, as E_2 will have conductor smaller by a factor of 2, but the discriminant rises by a factor of 64. So only if $|\Delta| \leq 10^{12}/64$ do we discard the curve; in the alternative case we replace the curve by E_2, so that we have the twist of minimal conductor. Finally, if we have $2^5 \mid\mid N$ (possibly after the above twisting by 2), or $2^7 \mid N$, we make the arbitrary decision to discard the curve if $c_6 < 0$, as we will also find E_{-1} in the search, which will have the same conductor and discriminant.

Using the above method, we can rid ourselves of all curves which are not minimal twists, and simultaneously compute the conductor. If $N > 10^{10}$, we simply ignore the curve; if $N > 10^8$ (and $N \leq 10^{10}$), we check whether N is a strong pseudoprime for 2, 13, 23, and 1662803, this being sufficient to prove primality [6]. At this point, we have a list of curves which meet our size conditions on the discriminant, and which have the minimal conductor in a family of quadratic twists, and minimal discriminant at primes other than $p = 2$.

2.2 Isogenous Curves

The next step will be to get rid of isogenous curves. The process of finding all curves isogenous to a given one is described in [4]. This is a fairly fast process, as most curves will have no nontrivial isogenies. Amongst the isogenous curves, we then take the curve of largest fundamental volume, that is, minimal Faltings height (which is unique by [8], as our representative. Note that this curve might not have the minimal discriminant in the isogeny class. Our final set of curves is then: the set of elliptic curves E such that E has minimal height in its isogeny class, and has some isogenous curve F for which we have $c_4 \leq 1.44 \cdot 10^{12}$ and either $N \leq 10^{10}$ with $|\Delta|$ prime, or $N \leq 10^8$ with $|\Delta| \leq 10^{12}$ for either the curve F or F_2.

2.3 Future Extension of the Database

As stated above, we would desire to have all minimal twists which have conductor less than 10^8. There are three ways of enlarging the database. The first is extending the range on c_4 by using the algorithm of [5]. The second is to incorporate the data from the exhaustive methods of Cremona. The third is to find families in which we expect the conductor to be substantially less than the discriminant; for instance, curves with a rational point of order 5 often have some prime to the 5th power dividing the discriminant. In the same vein, curves with (say) a 5-isogeny are parametrised from $X_0(5)$, and in such a parametrised family we again expect a large difference between the conductor and discriminant. We could also extend the discriminant limit to (say) 10^{13} for certain $(c_4^\star, c_6^\star)$ pairs, especially those for which we know ahead of time that we will save significant powers of 2 and 3 in the conductor compared to the discriminant.

3 Data Computed for Each Curve

One object of interest for an elliptic curve is its algebraic rank. This is hard to compute; indeed, there is no known algorithm to do this, only ones which work conditionally. By the process given in [4], we can try to determine the **analytic rank** of the curve, which is the degree of vanishing of its L-series at the central point. Of course, as there is no way to determine if a computed number is exactly zero, we can only give a good guess as to the analytic rank. The conjecture of Birch and Swinnerton-Dyer asserts that the algebraic rank and the analytic rank are equal, and that the first nonzero derivative of the L-function at the central point has arithmetic significance. For each curve in the database, we computed the suspected analytic rank and first nonzero derivative for both the curve itself and some of its quadratic twists.

Each curve in our database is the curve of minimal Faltings height in its isogeny class. A conjecture of Stevens [8] asserts that this curve should be the **optimal** curve for parametrisations from $X_1(N)$, in the sense that the parametrisations to the isogenous curves factor through the parametrisation to the strong curve (the existence of a modular parametrisation from $X_1(N)$ was proved in [1] following the methods initiated by Wiles [11]). It is sometimes the case that the optimal curve for parametrisations from $X_0(N)$ differs from the curve we find; in [10], a process is given to find the $X_0(N)$-optimal curve, assuming a technical condition, namely that the Manin constant of the optimal curve is 1 (this is similar to the Stevens conjecture). As many of the Frobenius traces were already computed for the analytic rank computation, these can be re-used at this stage. In a section below, we discuss the data obtained.

In the aforementioned paper [10], a process is given to compute the modular degree of an elliptic curve, again assuming that the Manin constant is 1. Compared to the computation of the analytic rank, which requires about the first $\sqrt{N}$ of the Frobenius traces, this method requires on the order of N of these (actually $\tilde{N}$, the symmetric-square conductor; see below). Thus for $N \geq 300000$ or so, it becomes rather time-consuming to compute the modular degree. We therefore compromised, computing the modular degree only if the symmetric-square conductor of the elliptic curve was sufficiently small (if we write $N = \prod_p p^{f_p}$ as a product of local conductors, then the symmetric-square conductor is simply $\tilde{N} = \prod_p p^{\lceil f_p/2 \rceil}$, except possibly when $f_2 = 8$, when the local symmetric-square conductor at 2 might be either 2^3 or 2^4; see [10] for details). We also computed the modular degree in some other interesting cases, for instance, when the rank is large.

4 Differing Optimal Curves

Here we discuss the question of differing optimal curves for parametrisations from $X_0(N)$ and $X_1(N)$. Note that we do not compute the actual optimal curve for the latter, relying instead on the Stevens conjecture, and compute the optimal curve for $X_0(N)$ only under the assumption that the Manin constant is 1. But the results are still interesting.

There appear to be three major cases when the optimal curves differ by a 2-isogeny. One of these, the so-called Setzer-Neumann curves, was considered in [7]. These curves are parametrised by $c_4 = u^2 + 48$ and $c_6 = -u\left(u^2 + 72\right)$, with the discriminant $u^2 + 64$ being a prime and u being taken to be congruent to 1 mod 4. The second family corresponds to taking $c_4 = 16\left(u^2 + 3\right)$ and $c_6 = -32u\left(2u^2 + 9\right)$ with u again being 1 mod 4 and $p = u^2 + 4$ being prime. Here the conductor is $4p$ and the discriminant is $16p$; the differing optimal curves property appears to be preserved upon twisting by -1.

The third family we found is obtained by taking $c_4 = p\,(p + 16) + 16$ and $c_6 = (p + 8)\left(p^2 + 16p - 8\right)$ of discriminant $p\,(p + 16)$ with both p and $p + 16$ primes congruent to 3 mod 4. A similar thing occurs if p and $p + 16$ are more generally powers of primes, but at least one of the two must be a power of a prime which is congruent to 3 mod 4 (i.e. $p = 11$ or $p = 2401$ works, but $p = 625$ does not). If p is congruent to 1 mod 4, then the sign of c_6 must be switched. Finally, p can be taken to be negative, for instance $p = -5$. Note that $p = 9$ leads to 15A, in which the optimal curves differ by a 4-isogeny; also, 17A might be thrown into consideration here with $p = 1$, which also has the optimal curves differing by a 4-isogeny.

With these considerations, there are but a couple of outstanding cases of optimal curves differing by a 2-isogeny (though proofs of this classification are lacking), those being the isogeny classes 24A/48A, 40A/80A, 32A/64A, and 128B/128D, though this last case can be seen as the $p = 8$ case of the second family. Ignoring the 5-isogeny example of 11A as being spurious, this leaves just the occasions of the optimal curves differing by a 3-isogeny. Here, all known examples are parametrised by

$$c_4 = (n + 3)\left(n^3 + 9n^2 + 27n + 3\right)$$

and

$$c_6 = -\left(n^6 + 18n^5 + 135n^4 + 504n^3 + 891n^2 + 486n - 27\right)$$

with the discriminant being $n\left(n^2 + 9n + 27\right)$. The n's for which the optimal curves differ are (experimentally) precisely those for which $n^2{+}9n{+}27$

is a prime power and n has no prime factors congruent to 1 mod 6; else the optimal curves are the same. We have no theoretical justification of this observation.

5 Data Obtained

This may seem strange for a comprehensive database project, but we do not dwell on large-scale phemonemon; indeed, the Brumer–McGuinness work is probably already sufficient in this manner, at least for prime conductor. As noted there, telling the difference between a small power of 10^8 (or whatever the upper limit of consideration may be) and a large power of its logarithm is rather hopeless—extending their data by a factor of 5/4 on the logarithmic scale does not help matters much. We mention that there are 11386955 isogeny classes of curves with prime conductor less than 10^{10} in our database (this should grow slightly when curves with $c_4 \geq 1.44 \cdot 10^{12}$ are added). Of these curves with prime conductor, of the ones we have processed, we have that 62.5% of the curves with even functional equation possess rank 0, compared to about 60% for Brumer–McGuinness. It is conjectured that asymptotically this percentage should be 100%. Similarly, 92.5% of the curves with odd functional equation have rank 1, slightly more than the previous results; there is no real reason to think that our numbers will change drastically upon extending the rank computation to all the prime conductor curves we have. The least conductor for a rank 5 curve we have found is 48012824 for $[0, 1, 0, -625, 6099]$, and for rank 6 we have $[0, 0, 1, -277, 4566]$ of conductor 7647224363. These respectively fall short to the best-known (to the authors) examples of $[0, 0, 1, -79, 342]$ of conductor 19047851 and $[0, 0, 1, -7077, 235516]$ of conductor 5258110041.

Instead of concentrating on large-scale behavior, we see our database as more of a tool to be used by other mathematicians. For instance, Neil Dummigan queried us concerning examples of strong Weil curves with rank 2 and a rational point of order 5 for which the conductor is not divisible by 5, and we were able to provide him with the example $[0, 1, 1, -840, 39800]$ of conductor 13881 (and modular degree 52000), among other examples which were beyond the range of Cremona's tables (which include $[1, 1, 1, -2365, 43251]$ of conductor 5302). Though we would likely be better able to answer the question after extending our database with parametrisations from $X_0(5)$, the efficacy of our database was evinced. As another example, the second author has conjectured in [10] that 2^r divides the modular degree for any curve (where r is the rank), and perhaps higher powers of 2 should divide the modular degree when the conductor is composite, due to factorisation through Atkin–Lehner involutions. For many large-rank curves in the Brumer–McGuinness database, we verified this. With our extension to curves of composite conductor, we are able to give more evidence for this conjecture. Also, the third 2-isogeny family in the previous section was discovered after looking at our data, as was the parametrisation of the 3-isogeny family, and finally our analytic rank data concerning quadratic twists could be of use.

6 Acknowledgements

The authors would like to thank Neil Dummigan for the question mentioned in Section V, Noam Elkies for many useful observations, especially concerning Section IV, and Blair Kelly III and Wayne Whitney for providing computing power.

References

1. C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: Wild 3-adic exercises.* J. Amer. Math. Soc. **14** (2001), 843–939.
2. A. Brumer, O. McGuinness, *The behavior of the Mordell-Weil group of elliptic curves.* Bull. Amer. Math. Soc. (N.S.) **23** (1990), no. 2, 375–382.
3. I. Connell, Lecture Notes from class at McGill University, 1991.
4. J. Cremona, *Algorithms for modular elliptic curves.* Cambridge University Press, Cambridge, 1992. Second edition 1997.
5. N. Elkies, *Rational points near curves and small nonzero* $|x^3 - y^2|$ *via lattice reduction.* In *Algorithmic number theory* (Leiden 2000), 33–63, Lecture Notes in Comput. Sci., 1838, Springer, Berlin, 2000.
6. G. Jaeschke, *On strong pseudoprimes to several bases.* Math. Comp. **61** (1993), no. 204, 915–926.
7. J.-F. Mestre, J. Oesterlé, *Courbes de Weil semi-stables de discriminant une puissance* m*-ième.* (French) J. Reine Angew. Math. **400** (1989), 173–184.
8. G. Stevens, *Stickelberger elements and modular parametrizations of elliptic curves.* Invent. Math. **98** (1989), no. 1, 75–106.
9. J. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil.* In *Modular functions of one variable IV*, edited by B. Birch and W. Kuyk, 33–52, Lecture Notes in Math., Vol. 476, Springer, Berlin, 1975.
10. M. Watkins, *Computing the modular degree of an elliptic curve,* preprint, 2001.
11. A. Wiles, *Modular elliptic curves and Fermat's last theorem.* Ann. of Math. (2) **141** (1995), no. 3, 443–551.

19 Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using Magma

Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using MAGMA

William Stein

Harvard University, Cambridge, Massachusetts, USA

1 Introduction

In this paper we describe the Birch and Swinnerton-Dyer conjecture in the case of modular abelian varieties and how to use MAGMA to do computations with some of the quantities that appear in the conjecture. We assume the reader has some experience with algebraic varieties and number theory, but do not assume the reader has proficiency working with elliptic curves, abelian varieties, modular forms, or modular symbols.

In Section 2 we quickly survey abelian varieties, modular forms, Hecke algebras, modular curves, and modular Jacobians, then discuss Shimura's construction of abelian varieties attached to modular forms. In Section 3 we survey many quantities associated to an abelian variety, including the Mordell-Weil group, torsion subgroup, regulator, Tamagawa numbers, real volume, and Shafarevich-Tate group, and use these to state the full Birch and Swinnerton-Dyer conjecture for modular abelian varieties. Section 4 contains some computational results from other papers about the Birch and Swinnerton-Dyer conjecture.

The rest of the paper is about how to use the package that I wrote for MAGMA to carry out an explicit computational study of modular abelian varieties. Section 5 is about modular symbols and how to compute with them in MAGMA. In Section 6 we state a theorem that allows us to use MAGMA to compute subgroups of Shafarevich-Tate groups of abelian varieties. In Section 7 we discuss computation of special values of L-functions. Section 8 is about computing Tamagawa numbers, and in Section 9 we describe how to compute a divisor and multiple of the order of the torsion subgroup. All these computations are pulled together in Section 10 to obtain a conjectural divisor and multiple of the order of the Shafarevich-Tate group of a modular abelian variety of dimension 20. We finish with Section 11, which contains an example in which the level is composite and elements of the Shafarevich-Tate group only becomes "visible" at higher level.

Taken together, these computations give evidence for the Birch and Swinnerton-Dyer conjecture and increase our explicit understanding of modular abelian varieties.

2 Modular Abelian Varieties

An elliptic curve E over the rational numbers $\mathbf{Q}$ is a one-dimensional commutative compact algebraic group. Such a curve is usually given as the projective closure of an affine curve $y^2 = x^3 + ax + b$, with a and b in $\mathbf{Q}$. The points over the real numbers $\mathbf{R}$ of $y^2 = x^3 - x + 1$ are illustrated in Figure 1. If P and Q are two distinct points on E, we find their sum as follows: draw the unique line through them and let (x, y) be the third point of intersection of this line with E. Then the sum of P and Q is $R = (x, -y)$, as illustrated in Figure 1. For more about elliptic curves, see [33, 34].

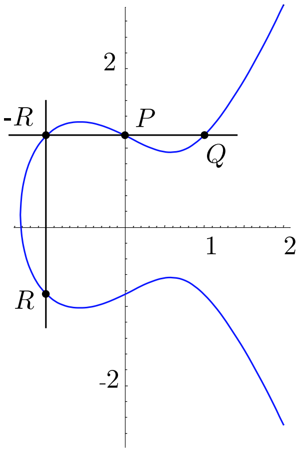


Fig. 1. Adding $P = (0, 1)$ to $Q = (1, 1)$ to get $R = (-1, -1)$ on $y^2 = x^3 - x + 1$

This paper is about abelian varieties, which are compact (commutative) algebraic groups of dimension possibly greater than 1. For example, the Cartesian product of two elliptic curves is an abelian variety of dimension 2.

Explicit equations for abelian varieties are vastly more complicated than for elliptic curves, so algorithms for computing with abelian varieties without recourse to explicit algebraic equations are of great value. In this paper we focus on such algorithms in the case when the abelian variety is endowed with extra structure coming from modular forms.

A cuspidal modular form of weight 2 for

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : N \mid c \right\}$$

is a holomorphic function $f(z)$ on the upper half plane such that for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ we have

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z),$$

and which satisfies certain vanishing conditions at the cusps (see [13, pg. 42] for a precise definition). We denote the finite dimensional complex vector space of all cuspidal modular forms of weight 2 for $\Gamma_0(N)$ by $S_2(\Gamma_0(N))$. Because $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$, cuspidal modular forms have a Fourier series representation

$$f(z) = \sum_{n=1}^{\infty} a_n q^n = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}.$$

The Hecke algebra

$$\mathbf{T} = \mathbf{Z}[T_1, T_2, T_3, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N)))$$

is a commutative ring that is free of rank equal to $\dim_{\mathbf{C}} S_2(\Gamma_0(N))$ (for the definition and basic properties of the Hecke operators T_n, see [13, §3] and the references therein).

A newform is a modular form

$$f = q + \sum_{n \geq 2} a_n q^n$$

that is a simultaneous eigenvector for every element of the Hecke algebra and such that the coefficients $\{a_p : p \nmid N\}$ are not the prime-index coefficients of another eigenform of some level that strictly divides N.

The group $\Gamma_0(N)$ acts as a discrete group of linear fractional transformations on the upper half plane; the quotient of the upper half plane by this action is a non-compact Riemann surface. Its compactification has the structure of algebraic curve over $\mathbf{Q}$, i.e., the compactification is the set of complex points of an algebraic curve $X_0(N)$ defined by polynomial equations with coefficients in $\mathbf{Q}$.

A divisor on an algebraic curve X is an element of the free abelian group generated by the points of X. For example, if f is a rational function on X then

$$(f) = (\text{formal sum of poles of } f) - (\text{formal sum of zeros of } f)$$

is a divisor on X, where the sums are with multiplicity. Two divisors D_1 and D_2 are linearly equivalent if there is a rational function f on X such that

$D_1 - D_2 = (f)$. The Jacobian J of an algebraic curve X is an abelian variety of dimension equal to the genus (number of holes in the Riemann surface $X(\mathbf{C})$) of X such that the underlying group of J is naturally isomorphic to the group of divisor classes of degree 0 on X. Let $J_0(N)$ denote the Jacobian of $X_0(N)$.

Similarly, let

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : N \mid c \text{ and } a \equiv 1 \pmod{N} \right\},$$

define $X_1(N)$ similarly, and let $J_1(N)$ be the Jacobian of $X_1(N)$.

A modular abelian variety is an abelian variety A for which there exists a surjective morphism $J_1(N) \to A$. Modular abelian varieties are appealing objects to study. For example, it is a deep theorem that every elliptic curve over $\mathbf{Q}$ is modular (see [7, 40, 41]), and this implies Fermat's Last Theorem (see [25, Cor. 1.2]). In [27], Ken Ribet conjectured that the simple abelian varieties over $\mathbf{Q}$ of "GL_2-type" are exactly the simple modular abelian varieties. A closely related conjecture of Serre (see [29, pg. 179] and [28]) asserts that every odd irreducible Galois representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

is "modular"; this conjecture is equivalent to the assertion that ρ can be realized (up to twist) as the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on a subgroup of the points on some $J_1(N)$ (see [28, §3.3.1] for a partial explanation). Though Serre's conjecture is still far from proved, it implies Ribet's conjecture (see [27, Thm. 4.4]).

We now return to considering $\Gamma_0(N)$, though we could consider $\Gamma_1(N)$ for everything in the rest of this section. The Hecke algebra $\mathbf{T}$, which we introduced above as a ring of linear transformations on $S_2(\Gamma_0(N))$, also acts via endomophisms on $J_0(N)$.

In order to construct Galois representations attached to modular forms, Goro Shimura (see [31, §1] and [32, §7.14]) associated to each newform $f = \sum a_n q^n$ a simple abelian variety A_f defined over $\mathbf{Q}$. Let I_f be the ideal of elements of $\mathbf{T}$ that annihilate f. Then

$$A_f = J_0(N)/I_f J_0(N).$$

The dimension of A_f equals the degree of the field generated over $\mathbf{Q}$ by the coefficients a_n of f. Note that A_f need not be simple over $\overline{\mathbf{Q}}$.

We will frequently mention the dual $A_f^\vee$ below. The dual can be considered as an abelian subvariety of $J_0(N)$, by using that Jacobians are canonically self dual and the dual of the quotient map $J_0(N) \to A_f$ is an inclusion $A_f^\vee \hookrightarrow J_0(N)$. Note that $A_f^\vee$ is the connected component of the intersection of the kernels of all elements of I_f.

We say that a newform g is a Galois conjugate of f if there is σ in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ such that $g = \sum \sigma(a_n) q^n$. If g is a Galois conjugate of f, then

$A_f = A_g$; if g is not a conjugate of f then the only homomorphism from A_f to A_g is the zero map. (A nonzero homomorphism $A_f \to A_g$ would induce an isogeny of Tate modules, from which one could deduce that f and g are Galois conjugate.)

We will concern ourselves almost entirely with these modular abelian varieties attached to newforms, because, as mentioned above, there are a number of algorithms for computing with them that do not require explicit algebraic equations (see [2, 3, 9, 10, 15, 19, 37, 35]). Also, it follows from standard results about constructing spaces of cusp forms from newforms, which can be found in [4, 22], that every modular abelian variety is isogenous to a product of abelian varieties of the form A_f. (An isogeny of abelian varieties is a surjective homomorphism with finite kernel.)

3 The Birch and Swinnerton-Dyer Conjecture

In the 1960s Bryan Birch and Peter Swinnerton-Dyer did computations with elliptic curves at Cambridge University on the EDSAC computer (see, e.g., [5]). These computations led to earth-shattering conjectures about the arithmetic of elliptic curves over $\mathbf{Q}$. Tate [39] formulated their conjectures in a more functorial way that generalized them to abelian varieties over global fields (such as the rational numbers). We now state their conjectures below for modular abelian varieties over $\mathbf{Q}$.

Let A_f be a modular abelian variety. Mordell and Weil proved that the abelian group $A_f(\mathbf{Q})$ of rational points on A_f is finitely generated, so it is isomorphic to $\mathbf{Z}^r \times T$ where T is the finite group $A_f(\mathbf{Q})_{\mathrm{tor}}$ of all elements of finite order in $A_f(\mathbf{Q})$. The exponent r is called the Mordell-Weil rank of A_f.

If f is a newform, the L-function of f is defined by the Dirichlet series $L(f,s) = \sum_{n\geq 1} a_n n^{-s}$. Hasse showed that $L(f,s)$ has an analytic continuation to a holomorphic function on the whole complex plane. The Hasse-Weil L-function of A_f is

$$L(A_f, s) = \prod L(g, s)$$

where the product is over the Galois conjugates g of f. The analytic rank of A_f is $\mathrm{ord}_{s=1} L(A_f, s)$.

We are now ready to state the first part of the conjecture.

Conjecture 3.1 (Birch and Swinnerton-Dyer) *The analytic rank of A_f is equal to the Mordell-Weil rank of A_f.*

Remark 1.

1. It is an open problem to give, with proof, an example of an elliptic curve with analytic rank at least 4. No examples with analytic rank at least 3 were known until the deep theorem of [16, Prop. 7.4].

2. When A_f is an elliptic curve, Conjecture 3.1 is the Clay Mathematics Institute Millennium Prize Problem from arithmetic geometry [17], so it has received much publicity.

In order to explain the conjecture of Birch and Swinnerton-Dyer about the leading coefficient of $L(A_f, s)$ at $s = 1$, we introduce the regulator, real volume, Tamagawa numbers, and Shafarevich-Tate group of A_f. Most of what we say below is true for a general abelian variety over a global field; the notable exceptions are that we do not know that the L-function is defined on the whole complex plane, and there are hardly any cases in general when the Shafarevich-Tate group is known to be finite.

Let $A_f(\mathbf{Q})_{/\mathrm{tor}}$ denote the quotient of $A_f(\mathbf{Q})$ by its torsion subgroup, so $A_f(\mathbf{Q})_{/\mathrm{tor}}$ is isomorphic to $\mathbf{Z}^r$, where r is the Mordell-Weil rank of A_f. The height pairing is a nondegenerate bilinear pairing h on $A_f(\mathbf{Q})_{/\mathrm{tor}}$. The regulator Reg_{A_f} of A_f is the absolute value of the determinant of a matrix who entries are $h(P_i, P_j)$, where $P_1, \ldots, P_r$ are a basis for $A_f(\mathbf{Q})_{/\mathrm{tor}}$. When $A_f(\mathbf{Q})$ has rank zero, the regulator is 1.

We use a certain integral model of A_f to define the real volume and Tamagawa numbers of A_f. The Néron model $\mathcal{A}$ of A_f, whose existence was established by Néron in [24] (see also [6, Ch. 1]), is a canonical object associated to A_f that is defined over $\mathbf{Z}$. The Néron model can be reduced modulo p for every prime p, and when base extended to $\mathbf{Q}$, the Néron model is isomorphic to A_f. The Néron model $\mathcal{A}$ is determined, up to unique isomorphism, by the following properties, which the reader unfamiliar with schemes can safely ignore: $\mathcal{A}$ is a smooth commutative group scheme over $\mathbf{Z}$ such that whenever S is a smooth scheme over $\mathbf{Z}$ the restriction map

$$\mathrm{Hom}(S, \mathcal{A}) \to \mathrm{Hom}_{\mathbf{Q}}(S_{\mathbf{Q}}, A)$$

is a bijection.

The real volume Ω_{A_f} of A_f is the absolute value of the integral over $A_f(\mathbf{R})$ of $h_1 \wedge \cdots \wedge h_d$ where $h_1, \ldots, h_d$ are a basis for the holomorphic 1-forms on $\mathcal{A}$. Using various identifications as in [1, §2.2.2] one sees that the $\mathbf{Z}$-span M of $h_1, \ldots, h_d$ can be viewed as a submodule of

$$W = S_2(\Gamma_0(N), \mathbf{Z}) \cap (\mathbf{C}f_1 \oplus \cdots \oplus \mathbf{C}f_d)$$

where $f_1, \ldots, f_d$ are the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ conjugates of f. We call the index of M in W the Manin constant of A_f, and conjecture (see [1]) that the Manin constant is 1. This conjecture would imply that a basis $h_1, \ldots, h_d$ can be computed, since W can be computed.

The reduction modulo p of $\mathcal{A}$ is an algebraic group $\mathcal{A}_{\mathbf{F}_p}$ over the finite field $\mathbf{F}_p$ with p elements. If p does not divide N, then this group is connected, but when p divides N, the reduction $\mathcal{A}_{\mathbf{F}_p}$ need not be connected. Let

$$\Phi_{A,p} = \mathcal{A}_{\mathbf{F}_p} / \mathcal{A}^0_{\mathbf{F}_p}$$

be the finite group of components. The Tamagawa number of A_f at p, denoted c_p, is the number of $\mathbf{F}_p$-rational components of the reduction of $\mathcal{A}$ modulo p, so $c_p = \#\Phi_{A,p}(\mathbf{F}_p)$.

The only object left to define before we state the second part of the Birch and Swinnerton-Dyer conjecture is the Shafarevich-Tate group of A_f. This is a group that measures the failure of a certain local-to-global principle for A_f. To give an exact description, we let $\mathrm{H}^1(\mathbf{Q}, A_f)$ be the first Galois cohomology group of A_f, which is a torsion group with infinitely many elements of any order bigger than 1 (see [30] for a proof in the case when A_f is an elliptic curve; the top of page 278 of [8] also purports to contain a proof). More precisely, $\mathrm{H}^1(\mathbf{Q}, A_f)$ is the set of equivalence classes of maps $c : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to A_f(\overline{\mathbf{Q}})$, with finite image, such that $c(\sigma\tau) = c(\sigma) + \sigma c(\tau)$, and two classes c_1 and c_2 are equivalent if there exists P in $A_f(\overline{\mathbf{Q}})$ such that $c_1(\sigma) - c_2(\sigma) = \sigma(P) - P$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. For each prime p we define $\mathrm{H}^1(\mathbf{Q}_p, A_f)$ analogously, but with the rational numbers $\mathbf{Q}$ replaced by the p-adic numbers $\mathbf{Q}_p$. Also, we allow $p = \infty$, in which case $\mathbf{Q}_p = \mathbf{R}$. Then

$$Ш(A_f) = \ker\left(\mathrm{H}^1(\mathbf{Q}, A_f) \longrightarrow \bigoplus_{\text{primes } p \leq \infty} \mathrm{H}^1(\mathbf{Q}_p, A_f)\right).$$

We are now ready to state the full Birch and Swinnerton-Dyer conjecture for modular abelian varieties A_f.

Conjecture 3.2 *Let $A = A_f$ be a modular abelian variety attached to a newform, and let $r = \mathrm{ord}_{s=1}L(A, s)$ be the analytic rank of A. Then*

$$\frac{L^{(r)}(A,1)}{r!} = \frac{\prod c_p \cdot \Omega_A \cdot \mathrm{Reg}_A}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^\vee(\mathbf{Q})_{\mathrm{tor}}} \cdot \#Ш(A).$$

Recall that $L^{(r)}(A_f, 1)$ makes sense at $s = 1$ because A_f is attached to a modular form. Also Kato established in [18, Cor. 14.3] that if $L(A_f, 1) \neq 0$ then $Ш(A_f)$ is finite, and Kolyvagin-Logachev ([20, Thm. 0.3]) proved that if f is a modular form in $S_2(\Gamma_0(N))$ and $\mathrm{ord}_{s=1}L(f, s) \leq 1$, then $Ш(A_f)$ is finite. When the theorems of Kato, Kolyvagin, and Logachev do not apply, we do not know even one example of a modular abelian variety A_f for which $Ш(A_f)$ is provably finite. John Tate once remarked that Conjecture 3.2 (for arbitrary abelian varieties) relates the value of a function where it is not known to be defined to the order of a group that is not known to be finite.

The rest of this paper is about how to use Magma to gather computational evidence for Conjecture 3.2, a task well worth pursuing. Elliptic curves are naturally surrounded by modular abelian varieties, so we want to understand modular abelian varieties well in order to say something about Conjectures 3.1–3.2 for elliptic curves. Doing explicit computations about these conjectures results in stimulating tables of data about modular abelian varieties, which could never be obtained except by direct computation. Until [2, 15] there were very few nontrivial computational examples of Conjecture 3.2 for abelian varieties in the literature, so it is important to test the conjecture since we might find a counterexample. Trying to compute information about a conjecture stimulates development of algorithms and theorems about that conjecture. Finally, our computations may lead to refinements of Conjecture 3.2 in the special case of modular abelian varieties; for example, most objects in Conjecture 3.2 are modules over the Hecke algebra so there should be more precise module-theoretic versions of the conjecture.

4 Some Computational Results

In [2] we use Magma to compute some of the arithmetic invariants of the 19608 abelian variety quotients A_f of $J_0(N)$ with $N \leq 2333$. Over half of these A_f have analytic rank 0, and for these we compute a divisor and a multiple of the order of $Ш(A_f)$ predicted by Conjecture 3.2. We find that there are at least 168 abelian varieties A_f such that the Birch and Swinnerton-Dyer Conjecture implies that $\#Ш(A_f)$ is divisible by an odd prime, and we use Magma to show that for 37 of these the odd part of the conjectural order of $Ш(A_f)$ divides $\#Ш(A_f)$ by constructing nontrivial elements of $Ш(A_f)$ using visibility theory. The challenge remains to show that the remaining 131 abelian varieties A_f have odd part of $Ш(A_f)$ divisible by the odd part of the conjectural order of $Ш(A_f)$ (we successfully take up this challenge for one example of level 551 in Section 11 of the present paper).

In [9, §2 and §7] we investigate Conjecture 3.1–3.2 when A_f is a quotient of $J_1(p)$ with p prime. In particular, we compute some of the invariants of every A_f for $p \leq 71$.

It was once thought by some mathematicians that Shafarevich-Tate groups of abelian varieties would have order a perfect square (or at least twice a perfect square). This is false, as we showed in the paper [36], where we use Magma to prove that for every odd prime $p < 25000$ there is an abelian variety whose Shafarevich-Tate group has order pn^2 with n an integer.

Much of the data mentioned above is of interest even if the full Birch and Swinnerton-Dyer conjecture were known since this data could probably never be discovered without considerable computation, even assuming the conjectures were true.

The rest of this paper is about how to use Magma to do computations with newform quotients A_f of $J_0(N)$ as in [2]. These computations involve modular symbols, which underly most algorithms for working with modular abelian varieties. (I hope to add functionality to a future release of Magma for computing directly with modular abelian varieties, so that no explicit mention of modular symbols is required.)

Remark 2. From a computational point of view, it is difficult to give evidence for Conjecture 3.1 when the dimension is greater than 1 in cases not covered by the general theorems of Kato, Kolyvagin, and Logachev. To give new evidence we would have to consider a modular abelian variety A_f with either f a newform in $S_2(\Gamma_0(N))$ and $\mathrm{ord}_{s=1}L(f,s) > 1$, or f a newform in $S_2(\Gamma_1(N))$ but not in $S_2(\Gamma_0(N))$ and $\mathrm{ord}_{s=1}L(f,s) > 0$. We would then show that $A_f(\mathbf{Q})$ is infinite, and more precisely that it has the rank predicted by Conjecture 3.1. In the above 2 cases the only known way to show that $A_f(\mathbf{Q})$ is infinite is to exhibit a point of infinite order in $A_f(\mathbf{Q})$, and this seems to require knowing equations for A_f. Also when $L(A_f,1) = 0$, Conjecture 3.2 involves a regulator term, which we do not know how to compute without explicitly finding the points on a model for A_f. Thus we will focus on giving evidence for Conjecture 3.2 in the case when $L(f,1) \neq 0$.

5 Modular Symbols

In this section we describe how modular symbols are related to homology of modular curves, and illustrate how to compute with modular symbols in Magma. We also discuss computing decomposition of modular symbols spaces and, for efficiency reasons, computing in the +1 quotient.

Let N be a positive integer. The integral homology $\mathrm{H}_1(X_0(N),\mathbf{Z})$ of the modular curve $X_0(N)$ is a free abelian group of rank equal to the genus of $X_0(N)$. The Hecke algebra $\mathbf{T} = \mathbf{Z}[T_1, T_2, T_3, \ldots]$ acts on a $\mathrm{H}_1(X_0(N),\mathbf{Z})$ as a ring of homomorphisms and makes $\mathrm{H}_1(X_0(N),\mathbf{Z})$ into a $\mathbf{T}$-module. This section is concerned with how to compute with this module using Magma. Section 12 contains a complete log of all Magma computations given below.

Modular symbols provide a finite computable presentation for the homology of $X_0(N)$ along with the action of the Hecke algebra $\mathbf{T}$ on this homology. The relative rational homology $\mathrm{H}_1(X_0(N),\mathbf{Q},\text{cusps})$ is the rational homology of $X_0(N)$ relative to the cusps; it is the finitely generated free abelian group of homology equivalence classes of geodesic paths from α to β, where α and β lie in $\mathbf{P}^1(\mathbf{Q}) = \mathbf{Q} \cup \{\infty\}$. A finite presentation for $\mathrm{H}_1(X_0(N),\mathbf{Q},\text{cusps})$ can be found in [23]. For simplicity, we typically compute $\mathrm{H}_1(X_0(N),\mathbf{Q},\text{cusps})$ first, then find $\mathrm{H}_1(X_0(N),\mathbf{Z})$ inside $\mathrm{H}_1(X_0(N),\mathbf{Q},\text{cusps})$ if it is needed. By definition, We now illustrate how Magma can compute a basis for $\mathrm{H}_1(X_0(N),\mathbf{Q},\text{cusps})$, and, given arbitrary α and β in $\mathbf{P}^1(\mathbf{Q})$, find an equivalent linear combination of basis elements.

```
M := ModularSymbols(389);
Basis(M);
```

The output of Basis(M) begins with the symbol $\{-1/337, 0\}$. Figure 2 on page 10 illustrates how the expression $\{-1/337, 0\}$ represents the relative rational homology class determined by a geodesic path from $-1/337$ to 0 in the upper half plane. The cusps determined by $-1/337$ and 0 are equivalent by an element of $\Gamma_0(389)$, so the image of the geodesic path in the 32 holed torus $X_0(389)(\mathbf{C})$ is a closed loop.

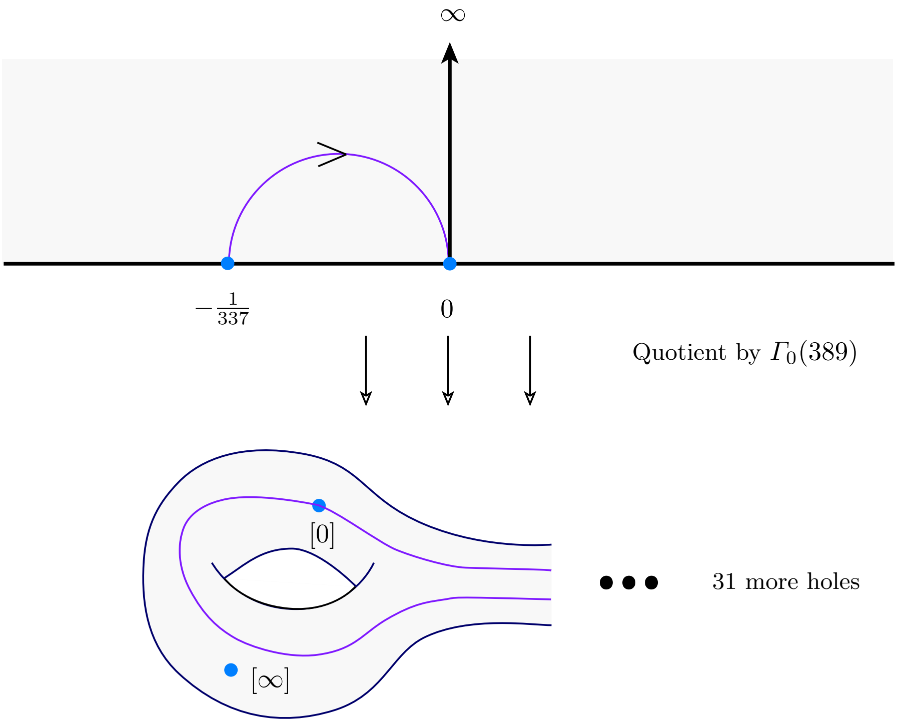


Fig. 2. The Modular Symbol $\{-1/337, 0\}$

The following Magma code illustrates how to find the image in the relative homology of an arbitrary path between cusps. The extra < and > are needed because we are considering modular symbols of weight $k = 2$; in general there is a coefficient which is a homogenous polynomial of degree $k - 2$, which is the first argument to the coercion. The Cusps()| part of the expression is needed so that the sequence is a sequence of cusps (this is not required if both cusps are rational numbers).

```
M ! <1, [Cusps() | −1/337, Infinity()] > ;
```

For more about computing with modular symbols, see [11, 12, 23, 37, 35].

Precise relationships between $\mathrm{H}_1(X_0(N), \mathbf{Q})$ and $S_2(\Gamma_0(N))$, along with some linear algebra, make it possible for us to compute a basis of $S_2(\Gamma_0(N))$

from knowledge about $\mathrm{H}_1(X_0(N), \mathbf{Q})$ as a $\mathbf{T}$-module. The following code, which computes a basis for $S_2(\Gamma_0(389))$, computes $\mathrm{H}_1(X_0(389), \mathbf{Q})$ and uses it to deduce the basis.

```
S := CuspForms(389);
SetPrecision(S,40);
Basis(S);
```

The SetPrecision command sets the output precision for q-expansions. The computed basis consists of q-expansions with coefficients in $\mathbf{Z}$.

Using NewformDecomposition, we find the submodules of $\mathrm{H}_1(X_0(389), \mathbf{Q})$ that correspond to Galois-conjugacy classes of newforms. These in turn correspond to the modular abelian varieties A_f attached to newforms. Magma excels at dense linear algebra over $\mathbf{Q}$ and is highly optimized for computing these decompositions. The following commands compute a decomposition of the new subspace of $\mathrm{H}_1(X_0(389), \mathbf{Q})$ corresponding to newforms.

```
M := ModularSymbols(389);
N := NewSubspace(CuspidalSubspace(M));
NewformDecomposition(N);
```

Since 389 is prime, the NewSubspace command is not necessary since everything is automatically new (there are no nonzero cusp forms of level 1 and weight 2). The decomposition consists of five factors of dimensions 2, 4, 6, 12, and 40; these correspond to newforms defined over fields of degrees 1, 2, 3, 6, and 20, respectively, which in turn correspond to abelian varieties over $\mathbf{Q}$ of dimensions 1, 2, 3, 6, and 20, respectively.

Remark 3. When information about the powers of 2 appearing in Conjecture 3.2 is not needed, we can instead do all computations in the "+1 quotient" of the space of modular symbols, which has half the dimension.

```
M := ModularSymbols(389,2,+1);    // the plus one quotient
```

6 Visibility Theory

Mazur introduced the notion of visibility to unify diverse ideas for constructing elements of Shafarevich-Tate groups. In this section we define what it means for an element of the Shafarevich-Tate group to be visible, state a theorem that allows us to compute pieces of this visible subgroup in some cases, and illustrate the theorem with a 20 dimensional abelian variety of level 389.

Suppose $i : A \to J$ is an injective morphism of abelian varieties over $\mathbf{Q}$. Then the visible subgroup of $\text{Ш}(A)$ is the kernel of the induced map $\text{Ш}(A) \to \text{Ш}(J)$.

Our interest in visibility in the present paper is that it allows us to obtain a provable divisor of $\#\text{Ш}(A)$, which is useful in giving evidence for Conjecture 3.2. The following theorem is proved in [3, Thm. 3.1] for abelian varieties over number fields.

Theorem 6.1 *Let A and B be abelian subvarieties of an abelian variety J over $\mathbf{Q}$ such that $A(\overline{\mathbf{Q}}) \cap B(\overline{\mathbf{Q}})$ is finite. (Note that J need not be a Jacobian.) Let N be an integer divisible by the residue characteristics of primes of bad reduction for B (so if A and B are modular then N is the level). Suppose p is an odd prime and that*

$$p \nmid N \cdot \#(J/B)(\mathbf{Q})_{\mathrm{tor}} \cdot \#B(\mathbf{Q})_{\mathrm{tor}} \cdot \prod_{p} c_{A,p} \cdot c_{B,p},$$

where $c_{A,p} = \#\Phi_{A,p}(\mathbf{F}_p)$ (resp., $c_{B,p}$) is the Tamagawa number of A (resp., B) at p. Suppose furthermore that $B(\overline{\mathbf{Q}})[p] \subset A(\overline{\mathbf{Q}})$, where both are viewed as subgroups of $J(\overline{\mathbf{Q}})$. Then there is a natural map

$$\varphi : B(\mathbf{Q})/pB(\mathbf{Q}) \to \text{Ш}(A)[p]$$

such that

$$\dim_{\mathbf{F}_p} \ker(\varphi) \leq \dim_{\mathbf{Q}} A(\mathbf{Q}) \otimes \mathbf{Q}.$$

In particular, if A has Mordell-Weil rank 0, then φ is injective.

Let A be the 20 dimensional quotient of $J_0(389)$ attached to a newform and B the elliptic curve quotient of $J_0(389)$. We use Magma to verify the hypothesis of Theorem 6.1 for $J = A^\vee + B^\vee \subset J_0(389)$ with $p = 5$, and hence deduce that $B(\mathbf{Q})/5B(\mathbf{Q}) = (\mathbf{Z}/5\mathbf{Z}) \times (\mathbf{Z}/5\mathbf{Z})$ injects into $\text{Ш}(A)$.

Since A and B are quotients of $J_0(389)$, we have $N = 389$. Next we construct the corresponding spaces A and B of modular symbols.

```
M := ModularSymbols(389);
N := NewSubspace(CuspidalSubspace(M));
D := SortDecomposition(NewformDecomposition(N));
A := D[5]; B := D[1];
```

The command IntersectionGroup computes the group structure of the intersection of two abelian subvarieties. In our case these are the abelian varieties $A^\vee$ and $B^\vee$, and we find that $A^\vee \cap B^\vee = (\mathbf{Z}/20\mathbf{Z}) \times (\mathbf{Z}/20\mathbf{Z})$. In particular, $B^\vee[5] = (\mathbf{Z}/5\mathbf{Z}) \times (\mathbf{Z}/5\mathbf{Z})$ is contained in $A^\vee$ as abelian subvarieties of $J_0(389)$.

```
IntersectionGroup(A,B);
```

Using the TorsionBound command (see Section 9 below), we obtain a multiple of the order of the torsion subgroup of B (it is 1) and of J/B (it is 97).

```
TorsionBound(A,7);
TorsionBound(B,7);
```

Neither torsion subgroup has order divisible by 5, as required to apply Theorem 6.1. The reason that TorsionBound(A,7) is a multiple of the order of the torsion subgroup of J/B is because TorsionBound is an isogeny invariant and A is isogenous to J/B. (The kernel of the natural map from A to J/B is $A \cap B = (\mathbf{Z}/20\mathbf{Z}) \times (\mathbf{Z}/20\mathbf{Z})$, which is finite.)

Finally, we compute the Tamagawa numbers of A and B and obtain 97 and 1, respectively (see Section 8 below).

```
TamagawaNumber(A, 389);
TamagawaNumber(B, 389);
```

Putting everything together we see that $B(\mathbf{Q})/5B(\mathbf{Q})$ is a subgroup of $\text{Ш}(A)$. Finally, using the Rank command on the elliptic curve attached to B, we see that $B(\mathbf{Q})/5B(\mathbf{Q}) = (\mathbf{Z}/5\mathbf{Z}) \times (\mathbf{Z}/5\mathbf{Z})$.

```
E := EllipticCurve(B);
Rank(E);
```

Thus 25 divides $\#\text{Ш}(A)$, which gives evidence for Conjecture 3.2, as we will see in Section 10.

Frequently not all of $\text{Ш}(A)$ can be constructed using Theorem 6.1 and abelian subvarieties B of $J_0(N)$. One obstruction to visibility arises from a canonical homomorphism from $A^\vee$ to A. Jacobians of curves are canonically isomorphic to their dual abelian variety and the composition $A^\vee \to J_0(389)^\vee \cong J_0(389) \to A$ defines a homomorphism from $A^\vee$ to A. According to [3, §5.3], if p does not divide the kernel of $A^\vee \to A$, then no element of order p in $\text{Ш}(A)$ is visible in $J_0(N)$. The command ModularKernel computes the group structure of the kernel of $A^\vee \to A$.

```
G := ModularKernel(A);
Factorization(#G);
```

We find that the modular kernel has order $2^{24}5^2$, so any element of $\text{Ш}(A^\vee)$ that is visible in $J_0(389)$ has order divisible only by 2 and 5.

7 Computing Special Values of Modular L-function

This section is about computing the quotient $L(A_f, 1)/\Omega_{A_f}$. We discuss the Manin constant and the LRatio command.

Let $A = A_f$ for some newform f and assume that $L(A, 1) \neq 0$. We can then rewrite Conjecture 3.2 as follows:

$$\frac{L(A,1)}{\Omega_A} = \frac{\prod c_p \cdot \#\text{Ш}(A)}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^\vee(\mathbf{Q})_{\mathrm{tor}}}.$$

We do not know an algorithm, in general, to compute $L(A, 1)/\Omega_A$. However, we can compute $c_A \cdot L(A, 1)/\Omega_A$, where c_A is the Manin constant, which is defined in [1, §2.2]. We conjecture that $c_A = 1$, and prove in [1, §2.2.2] that if f is a newform on $\Gamma_0(N)$ then c_A is an integer divisible only by primes whose square divides $4N$. Moreover, if N is odd then $2^{\dim A}$ is the largest power of 2 that can divide c_A. See also [14] for results when A has dimension 1, and [9, §6.1.2] for a proof that c_A is an integer when $\Gamma_0(N)$ is replaced by $\Gamma_1(N)$.

The algorithm described in [9, §2.1.3], [2, §4] and [37, §3.10] to compute $c_A \cdot L(A,1)/\mathit{\Omega}_A$ is implemented in MAGMA via the LRATIO command. For example, if A is as in Section 6, then $c_A \cdot L(A,1)/\mathit{\Omega}_A = 2^{11} \cdot 5^2/97$.

```
LRatio(A, 1)
```

8 Computing Tamagawa Numbers

In this section we discuss computing Tamagawa numbers when $p \mid\mid N$ and some bounds when $p^2 \mid N$. We also discuss issues that arise in going from the order of the component group to the Tamagawa number when $p \mid\mid N$.

Let $A = A_f$ be a modular abelian variety attached to a newform $f \in S_2(\mathit{\Gamma}_0(N))$. When $p \mid\mid N$, [10, §2.1] contains a computable formula for $\#\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$ and for $c_p = \#\mathit{\Phi}_{A,p}(\mathbf{F}_p)$, where the latter formula is in some cases only valid up to a bounded power of 2. Also [19] is about how to compute these orders. Note that the Tamagawa number of A at p is the same as the Tamagawa number of $A^\vee$ at p.

When $p^2 \mid N$ the authors do not know an algorithm to compute c_p. However, in this case Lenstra and Oort proved in [21, Cor. 1.15] that

$$\sum_{\ell \neq p} (\ell - 1)\mathrm{ord}_\ell(\#\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)) \leq 2\dim(A_f),$$

so if $\ell \mid \#\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$ then $\ell \leq 2 \cdot \dim(A_f) + 1$ or $\ell = p$. (Here $\mathrm{ord}_\ell(x)$ denotes the exponent of the largest power of ℓ that divides x.)

Using [10], when $p \mid\mid N$ we know how to compute the order of the component group over the algebraic closure, but not its structure as a group. The command COMPONENTGROUPORDER computes the order of $\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$. The command TAMAGAWANUMBER computes $c_p = \#\mathit{\Phi}_{A,p}(\mathbf{F}_p)$ when the subgroup of elements of $\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$ fixed by the Galois group has order that does not depend on the underlying group structure. By computing the Atkin-Lehner involution on modular symbols, we can decide whether the Galois group acts trivially or by -1 on $\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$ since the Atkin-Lehner involution acts as the negative of the canonical generator Frobenius of $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. We can thus compute $\#\mathit{\Phi}_{A,p}(\mathbf{F}_p)$ when the Galois group acts trivially. When the Galois group acts nontrivially, $\mathit{\Phi}_{A,p}(\mathbf{F}_p)$ is the 2-torsion subgroup of $\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$, whose order we know as long as 4 does not divide $\#\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$. It is an open problem to given an algorithm to compute the group structure of $\mathit{\Phi}_{A,p}(\overline{\mathbf{F}}_p)$ or the order of $\mathit{\Phi}_{A,p}(\mathbf{F}_p)$ in general.

Section 11 contains an example of an abelian variety of dimension 18 in which the author is only able to find the Tamagawa number up to a controlled power of 2.

9 Computing the Torsion Subgroup

In this section we describe how to compute a divisor and multiple of the order of the torsion subgroup and explain how knowing a divisor of $\#A_f(\mathbf{Q})_{\mathrm{tor}}$ yields a divisor of $\#A_f^\vee(\mathbf{Q})_{\mathrm{tor}}$.

The papers [2, §3.5–3.6] and [9, §2.1.1] contain discussions of how to compute a divisor and multiple of the order of the torsion subgroup $A_f(\mathbf{Q})_{\mathrm{tor}}$ of $A_f(\mathbf{Q})$, and likewise for $A_f^\vee(\mathbf{Q})_{\mathrm{tor}}$. (The multiple of $\#A_f^\vee(\mathbf{Q})_{\mathrm{tor}}$ is the same as for $A_f(\mathbf{Q})_{\mathrm{tor}}$, and the divisor can be computed as described below.) We compute the multiple by using that $A_f(\mathbf{Q})_{\mathrm{tor}}$ injects into $A_f(\mathbf{F}_p)$ for all p not dividing $2N$, and that $\#A_f(\mathbf{F}_p)$ is fairly easy to compute, though we do not know how to compute the group structure. We compute the lower bound by considering the subgroup of elements of $J_0(N)(\mathbf{Q})_{\mathrm{tor}}$ generated by rational cusps on $X_0(N)$ (see [38, §1.3]), and taking its image in $A_f(\mathbf{Q})_{\mathrm{tor}}$ or intersecting its image with $A_f^\vee(\mathbf{Q})_{\mathrm{tor}} \subset J_0(N)(\mathbf{Q})_{\mathrm{tor}}$. Note that there is no reason for the subgroup generated by rational cusps to equal the rational subgroup of the group generated by all cusps, and one might want to compute and work with this possibly larger group instead.

Let A and B be as in Section 6, where we showed that the torsion subgroup of B is trivial and the order of $B(\mathbf{Q})$ and $B^\vee(\mathbf{Q})$ divides 97. In Section 10, we give an example in which the divisor and multiple of the order of the torsion subgroup differ by a power of 2.

```
RationalCuspidalSubgroup(A);    // subgroup of $A(\mathbf{Q})$
```

As mentioned in Section 6, there is a homomorphism $A^\vee \to A$ of degree $2^{24} \cdot 5^2$, which implies that 97 also divides $\#A_f^\vee(\mathbf{Q})_{\mathrm{tor}}$. Thus $\#A_f(\mathbf{Q})_{\mathrm{tor}} = \#A_f^\vee(\mathbf{Q})_{\mathrm{tor}} = 97$.

Remark 4. Computation of a nontrivial divisor of $\#A_f^\vee(\mathbf{Q})_{\mathrm{tor}}$ directly using rational cusps is not yet implemented in Magma, though in principle this should not be difficult to implement.

10 A Divisor and Multiple of the Order of the Shafarevich-Tate Group

In this section we substitute the values computed above into Conjecture 3.2 to obtain a conjectural divisor and multiple of the order of a Shafarevich-Tate group. We then remark that the visibility computation of Section 6 gives evidence for Conjecture 3.2. This example is also discussed in [3, §4.2].

To obtain evidence for Conjecture 3.2, we consider an abelian variety A_f with $L(A_f, 1) \neq 0$ and combine the invariants whose computation is described above with Conjecture 3.2 to obtain a conjectural divisor and multiple of the order of $\text{Ш}(A_f)$. We then observe that this divisor and multiple is consistent with Conjecture 3.2.

We now combine the computations from the previous sections for the 20 dimensional quotient A of $J_0(389)$. Recall that Conjecture 3.2 asserts that

$$\frac{L(A,1)}{\Omega_A} = \frac{\prod c_p \cdot \#Ш(A)}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^{\vee}(\mathbf{Q})_{\mathrm{tor}}}.$$

This equation becomes

$$\frac{2^n \cdot 2^{11} \cdot 5^2}{97} = \frac{97 \cdot \#Ш(A)}{97^2}$$

where $0 \leq n \leq 20$ (using the bound from [1, Thm. 2.7]). Thus the conjecture asserts that $\#Ш(A) = 5^2 \cdot 2^{11+n}$, and we have computed a conjectural divisor $5^2 \cdot 2^{11}$ and a conjectural multiple $5^2 \cdot 2^{31}$ of $\#Ш(A)$. Using visibility theory from Section 6 we have proved that $5^2 \mid \#Ш(A)$, which provides evidence for Conjecture 3.2.

11 An Element of the Shafarevich-Tate Group that Becomes Visible at Higher Level

We finish this paper by considering the 18-dimensional newform quotient A of $J_0(551)$. In this example, the level $551 = 19 \cdot 29$ is composite, the Shafarevich-Tate group is conjecturally nontrivial, and the methods of Section 6 do not produce nontrivial elements of the Shafarevich-Tate group at level 551.

This example is striking because it is, in some sense, the simplest known example of "visibility only at a higher level"; more precisely, the methods of Section 6 do produce a nontrivial element at the rather small level 1102. For a similar example, see [3, §4.3], where the levels involved are much larger.

We first compute the space of modular symbols corresponding to A:

```
M := ModularSymbols(551);
N := NewSubspace(CuspidalSubspace(M));
D := SortDecomposition(NewformDecomposition(N));
A := D[8];
```

Next we compute a divisor and a multiple of the order of the torsion subgroup of $A(\mathbf{Q})$ and $A^{\vee}(\mathbf{Q})$. Using odd primes $p \leq 7$ we obtain the multiple 160, and using the rational cuspidal subgroup we obtain the divisor 40.

```
TorsionBound(A, 7);
RationalCuspidalSubgroup(A);
```

Since the divisor and multiple are different, we try more finite fields. For $p \leq 29$ the multiple we obtain is still 160; however, for $p = 31$ the multiple is 80, which is where it appears to stabilize.

```
TorsionBound(A, 31);
```

We conclude that $40 \mid \#A(\mathbf{Q})_{\mathrm{tor}} \mid 80$ and $5 \mid \#A^{\vee}(\mathbf{Q})_{\mathrm{tor}} \mid 80$. We know that 5 divides $\#A^{\vee}(\mathbf{Q})_{\mathrm{tor}}$ because, as we will see below, there is a homomorphism $A^{\vee} \to A$ of degree not divisible by 5.

Next we compute the modular kernel, which is of order $2^{44} \cdot 13^4$.

```
Factorization(#ModularKernel(A));
```

The only possible elements of $Ш(A)$ that we can construct using Theorem 6.1 at level 551 are of order 13.

The level 551 is not prime, so computation of the Tamagawa numbers involves certain relatively slow algorithms (a minute rather than seconds) that involve arithmetic in quaternion algebras. Also, in this example, we are unable to determine the exact power of 2 that divides the Tamagawa number at 19.

```
TamagawaNumber(A, 19);    // takes over a minute; gives an error
TamagawaNumber(A, 29);
```

We find that $c_{29} = 40$. We also deduce that $c_{19} = 2$ or 4 by noting that the component group over $\overline{\mathbf{F}}_{19}$ has order $2^2 \cdot 13^2$ by using the command

```
ComponentGroupOrder(A, 19);
```

and noting that the Galois generator Frobenius acts as -1 because

```
AtkinLehnerOperator(A, 19)[1, 1];
```

returns 1. Finally note that the 2 torsion in any group of order $2^2 \cdot 13^2$ is a subgroup of order either 2 or 4.

Next we find that $L(A, 1)/\Omega_A = 2^n \cdot 2^2 \cdot 3^2/5$, with $0 \leq n \leq 18$, using the command

```
LRatio(A, 1)
```

and the fact that the Manin constant divides $2^{\dim A}$ (see [1, Thm. 2.7]).

Putting these computations together we find that Conjecture 3.2 asserts that

$$\frac{2^n \cdot 2^2 \cdot 3^2}{5} = \frac{2^m \cdot 40 \cdot \#Ш(A)}{40 \cdot 2^r \cdot 5 \cdot 2^s},$$

where $0 \leq n \leq 18$, $1 \leq m \leq 2$, $0 \leq r \leq 1$, and $0 \leq s \leq 4$. Solving for $\#Ш(A)$, we see that Conjecture 3.2 predicts that

$$\#Ш(A) = 2^t \cdot 3^2$$

with $2 \leq t \leq 24$.

Theorem 6.1 does not construct elements of order 2 (yet), so we do not consider the factor 2^t further. As mentioned above, we cannot construct any elements of $Ш(A)$ of order 3 using visibility at level 551. We can, however, consider the images of A in $J_0(2 \cdot 551)$ under various natural maps. These natural maps are the degeneracy maps δ_1 and δ_2, which correspond to the maps $f(q) \mapsto f(q)$ and $f(q) \mapsto f(q^2)$ from $S_2(\Gamma_0(551))$ to $S_2(\Gamma_0(2 \cdot 551))$.

We next compute the space of modular symbols that corresponds to the sum $C = \delta_1(A)+\delta_2(A)$ of the images of A at level $2\cdot551$ by the two degeneracy maps δ_1 and δ_2.

```
M := ModularSymbols(2*551,2);
N := NewSubspace(CuspidalSubspace(M));
D := SortDecomposition(NewformDecomposition(N));
M551 := ModularSymbols(M,551);
N551 := NewSubspace(CuspidalSubspace(M551));
D551 := SortDecomposition(NewformDecomposition(N551));
A551 := D551[#D551];
C := M !! A551;    // sum of images under degeneracy maps
```

The sum C contains the 3-torsion of the rank 2 elliptic curve B defined by $y^2 + xy = x^3 + x^2 - 29x + 61$, as the following computation shows.

```
IntersectionGroup(C,D[1]);
B := EllipticCurve(D[1]); B;
```

It follows that $B[3]$ is contained in C. The following computation shows that the Tamagawa numbers of B are 2, 2, and 1 and $B(\mathbf{Q}) \equiv \mathbf{Z} \times \mathbf{Z}$:

```
TamagawaNumber(B,2);
TamagawaNumber(B,19);
TamagawaNumber(B,29);
MordellWeilGroup(B);
```

Theorem 6.1 implies that $B(\mathbf{Q})/3B(\mathbf{Q}) = \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}$ is a subgroup of $\text{Ш}(C)$. By [26, §2], there is an isogeny φ from $A \times A$ to C whose kernel is isomorphic to the intersection of A with the Shimura subgroup of $J_0(551)$. The Shimura subgroup Σ is a subgroup of $J_0(N)$ that, according to [26, Prop. 2], is annihilated by $T_p - (p + 1)$ for all primes $p \nmid 551$. Using Magma we find that $3 \nmid \det(T_3|_A - 4) = 12625812402998886400$, so the degree of φ is coprime to 3.

```
T3 := HeckeOperator(A,3);
d := Determinant(T3-4);
Valuation(d,3);
```

Since $3 \mid \#\text{Ш}(C)$ it follows that $3 \mid \#\text{Ш}(A)$. By [2, §5.3] the power of 3 that divides $\#\text{Ш}(A)$ is even, so $9 \mid \#\text{Ш}(A)$, as predicted by the Birch and Swinnerton-Dyer conjecture.

12 Complete Magma Log

This is a complete log of using MAGMA V2.10-6 to do all of the computations discussed in this paper. The output has been edited slightly to save space.

```
> M := ModularSymbols(389);
> Basis(M);
[
    {-1/337, 0},{-1/237, 0},{-1/342, 0},{-1/266, 0},{-1/170, 0},
    {-1/272, 0},{-1/333, 0},{-1/355, 0},{-1/270, 0},{-1/301, 0},
    {-1/293, 0},{-1/87, 0},{-1/306, 0},{-1/205, 0},{-1/209, 0},
    {-1/277, 0},{-1/383, 0},{-1/142, 0},{-1/178, 0},{-1/116, 0},
    {-1/61, 0},{-1/127, 0},{-1/235, 0},{-1/240, 0},{-1/93, 0},
    {-1/121, 0},{-1/221, 0},{-1/199, 0},{-1/213, 0},{-1/370, 0},
    {-1/282, 0},{-1/379, 0},{-1/100, 0},{-1/286, 0},{-1/165, 0},
    {-1/158, 0},{-1/376, 0},{-1/228, 0},{-1/125, 0},{-1/72, 0},
    {-1/374, 0},{-1/140, 0},{-1/81, 0},{-1/186, 0},{-1/53, 0},
    {-1/37, 0},{-1/175, 0},{-1/108, 0},{-1/183, 0},{-1/316, 0},
    {-1/363, 0},{-1/250, 0},{-1/359, 0},{-1/162, 0},{-1/106, 0},
    {-1/350, 0},{-1/216, 0},{-1/243, 0},{-1/111, 0},{-1/324, 0},
    {-1/311, 0},{-1/97, 0},{-1/259, 0},{-1/194, 0},{oo, 0}
]
> M ! <1, [Cusps() | -1/337, Infinity()]>;
{-1/337, 0} + -1*{oo, 0}
> S := CuspForms(389);
> SetPrecision(S,40);
> Basis(S);
[
    q + 474049571*q^32 + 480335856*q^33 + 984946270*q^34 +
    1338756227*q^35 + 1246938503*q^36 - 29119245*q^37 +
    1504020580*q^38 - 2463550751*q^39 + O(q^40),
    ...
]
```

```
> M := ModularSymbols(389);
> N := NewSubspace(CuspidalSubspace(M));
> NewformDecomposition(N);
[
    Modular symbols space for Gamma_0(389) of weight 2 and
    dimension 2 over Rational Field,
    Modular symbols space for Gamma_0(389) of weight 2 and
    dimension 4 over Rational Field,
    Modular symbols space for Gamma_0(389) of weight 2 and
    dimension 6 over Rational Field,
    Modular symbols space for Gamma_0(389) of weight 2 and
    dimension 12 over Rational Field,
    Modular symbols space for Gamma_0(389) of weight 2 and
    dimension 40 over Rational Field ]
```

```
> M := ModularSymbols(389,2,+1);
```

```
> M := ModularSymbols(389);
> N := NewSubspace(CuspidalSubspace(M));
> D := NewformDecomposition(N);
> A := D[5]; B := D[1];
> IntersectionGroup(A,B);
Abelian Group isomorphic to Z/20 + Z/20
> TorsionBound(A,7);
97
> TorsionBound(B,7);
1
> TamagawaNumber(A,389);
97
> TamagawaNumber(B,389);
1
> E := EllipticCurve(B);
> Rank(E);
2
> G := ModularKernel(A);
Abelian Group isomorphic to Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 +
Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 + Z/2 +
Z/2 + Z/40 + Z/40
> Factorization(#G);
[ <2, 24>, <5, 2> ]
> LRatio(A,1);
51200/97
> RationalCuspidalSubgroup(A);
Abelian Group isomorphic to Z/97
> M := ModularSymbols(551);
> N := NewSubspace(CuspidalSubspace(M));
> D := NewformDecomposition(N);
> A := D[8];
> TorsionBound(A,7);
160
> RationalCuspidalSubgroup(A);
Abelian Group isomorphic to Z/2 + Z/20
> TorsionBound(A,31);
80
> Factorization(#ModularKernel(A));
[ <2, 44>, <13, 4> ]
> TamagawaNumber(A,19);
No algorithm known to compute the Tamagawa number at 2. Use
ComponentGroupOrder instead.
> TamagawaNumber(A,29);
40
> ComponentGroupOrder(A,19);
676
> AtkinLehnerOperator(A,19)[1,1];
1
> LRatio(A,1);
36/5
> M := ModularSymbols(2*551,2);
> N := NewSubspace(CuspidalSubspace(M));
> D := SortDecomposition(NewformDecomposition(N));
> M551 := ModularSymbols(M,551);
> N551 := NewSubspace(CuspidalSubspace(M551));
> D551 := NewformDecomposition(N551);
> A551 := D551[#D551];
> C := M!!A551;
> IntersectionGroup(C,D[1]);
Abelian Group isomorphic to Z/32 + Z/32
> B := EllipticCurve(D[1]); B;
Elliptic Curve defined by y^2 + x*y = x^3 + x^2 - 29*x + 61
> TamagawaNumber(B,2);
2
> TamagawaNumber(B,19);
2
> TamagawaNumber(B,29);
1
> MordellWeilGroup(B);
Abelian Group isomorphic to Z + Z
> MordellWeilGroup(B);
Abelian Group isomorphic to Z + Z
> T3 := HeckeOperator(A,3);
> d := Determinant(T3-4);
> Valuation(d,3);
0
```

References

1. A. Agashe and W. A. Stein. The manin constant, congruence primes, and the modular degree. *Submitted*.
2. A. Agashe and W. A. Stein. Visible Evidence for the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties of Analytic Rank 0. *To appear in Mathematics of Computation*.
3. A. Agashe and W. A. Stein. Visibility of Shafarevich-Tate groups of abelian varieties. *J. Number Theory*, 97(1):171–185, 2002.
4. A. O. L. Atkin and J. Lehner. Hecke operators on $\Gamma_0(m)$. *Math. Ann.*, 185:134–160, 1970.
5. B. J. Birch. Conjectures concerning elliptic curves. In *Proceedings of Symposia in Pure Mathematics, VIII*, pages 106–112. Amer. Math. Soc., Providence, R.I., 1965.
6. S. Bosch, W. Lütkebohmert, and M. Raynaud. *Néron models*. Springer-Verlag, Berlin, 1990.
7. C. Breuil, B. Conrad, F. Diamond, and R. Taylor. On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises. *J. Amer. Math. Soc.*, 14(4):843–939 (electronic), 2001.
8. J. W. S. Cassels. Diophantine equations with special reference to elliptic curves. *J. London Math. Soc.*, 41:193–291, 1966.
9. B. Conrad, S. Edixhoven, and W. A. Stein. $J_1(p)$ Has Connected Fibers. *To appear in Documenta Mathematica*, 2003.

10. B. Conrad and W. A. Stein. Component Groups of Purely Toric Quotients. *To appear in Math Research Letters*, 2002.
11. J. E. Cremona. Modular symbols for $\Gamma_1(N)$ and elliptic curves with everywhere good reduction. *Math. Proc. Cambridge Philos. Soc.*, 111(2):199–218, 1992.
12. J. E. Cremona. *Algorithms for modular elliptic curves.* Cambridge University Press, Cambridge, second edition, 1997.
13. F. Diamond and J. Im. Modular forms and modular curves. In *Seminar on Fermat's Last Theorem*, pages 39–133. Providence, RI, 1995.
14. B. Edixhoven. On the Manin constants of modular elliptic curves. In *Arithmetic Algebraic Geometry*, pages 25–39 (G. van der Geer, F. Oort et al., eds.), Basel: Birkhäuser, Progress in Mathematics Volume 89, 1991.
15. E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell. Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves. *Math. Comp.*, 70(236):1675–1697 (electronic), 2001.
16. B. Gross and D. Zagier. Heegner points and derivatives of L-series. *Invent. Math.*, 84(2):225–320, 1986.
17. Clay Mathematics Institute. Millennium prize problems, `http://www.claymath.org/millennium_prize_problems/`.
18. K. Kato. p-adic Hodge theory and values of zeta functions of modular forms. *Preprint*, page 244 pages.
19. D. R. Kohel and W. A. Stein. Component Groups of Quotients of $J_0(N)$. In *Proceedings of the 4th International Symposium (ANTS-IV), Leiden, Netherlands, July 2–7, 2000*, Berlin, 2000. Springer.
20. V. A. Kolyvagin and D. Y. Logachev. Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties. *Algebra i Analiz*, 1(5):171–196, 1989.
21. H. W. Lenstra, Jr. and F. Oort. Abelian varieties having purely additive reduction. *J. Pure Appl. Algebra*, 36(3):281–298, 1985.
22. W-C. Li. Newforms and functional equations. *Math. Ann.*, 212:285–315, 1975.
23. J. I. Manin. Parabolic points and zeta functions of modular curves. *Izv. Akad. Nauk SSSR Ser. Mat.*, 36:19–66, 1972.
24. A. Néron. Modèles minimaux des variétés abéliennes sur les corps locaux et globaux. *Inst. Hautes Études Sci. Publ.Math. No.*, 21:128, 1964.
25. K. A. Ribet. On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms. *Invent. Math.*, 100(2):431–476, 1990.
26. K. A. Ribet. Raising the levels of modular representations. In *Séminaire de Théorie des Nombres, Paris 1987–88*, pages 259–271. Birkhäuser Boston, Boston, MA, 1990.
27. K. A. Ribet. Abelian varieties over $\mathbf{Q}$ and modular forms. In *Algebra and topology 1992 (Taejŏn)*, pages 53–79. Korea Adv. Inst. Sci. Tech., Taejŏn, 1992.
28. K. A. Ribet and W. A. Stein. Lectures on Serre's conjectures. In *Arithmetic algebraic geometry (Park City, UT, 1999)*, volume 9 of *IAS/Park City Math. Ser.*, pages 143–232. Amer. Math. Soc., Providence, RI, 2001.
29. J-P. Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987.
30. I. R. Shafarevich. Exponents of elliptic curves. *Dokl. Akad. Nauk SSSR (N.S.)*, 114:714–716, 1957.
31. G. Shimura. On the factors of the jacobian variety of a modular function field. *J. Math. Soc. Japan*, 25(3):523–544, 1973.

32. G. Shimura. *Introduction to the arithmetic theory of automorphic functions.* Princeton University Press, Princeton, NJ, 1994. Reprint of the 1971 original, Kan Memorial Lectures, 1.
33. J. H. Silverman. *The arithmetic of elliptic curves.* Springer-Verlag, New York, 1992. Corrected reprint of the 1986 original.
34. J. H. Silverman and J. Tate. *Rational points on elliptic curves.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1992.
35. W. A. Stein. An introduction to computing modular forms using modular symbols. *To appear in an MSRI Proceedings.*
36. W. A. Stein. Shafarevich-Tate groups of nonsquare order. *Proceedings of MCAV 2002, Progress of Mathematics (to appear).*
37. W. A. Stein. Explicit approaches to modular abelian varieties. *Ph.D. thesis, University of California, Berkeley*, 2000.
38. G. Stevens. *Arithmetic on modular curves.* Birkhäuser Boston Inc., Boston, Mass., 1982.
39. J. Tate. On the conjectures of Birch and Swinnerton-Dyer and a geometric analog. In *Séminaire Bourbaki, Vol. 9*, pages Exp. No. 306, 415–440. Soc. Math. France, Paris, 1995.
40. R. Taylor and A. J. Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995.
41. A. J. Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995.

20 Modular Parametrizations Of NeumannSetzer Elliptic Curves, with M. Watkins

MODULAR PARAMETRIZATIONS OF NEUMANN–SETZER ELLIPTIC CURVES

WILLIAM STEIN AND MARK WATKINS

Abstract. Suppose p is a prime of the form $u^2 + 64$ for some integer u, which we take to be 3 mod 4. Then there are two Neumann–Setzer elliptic curves E_0 and E_1 of prime conductor p, and both have Mordell–Weil group $\mathbf{Z}/2\mathbf{Z}$. There is a surjective map $X_0(p) \xrightarrow{\pi} E_0$ that does not factor through any other elliptic curve (i.e., π is optimal), where $X_0(p)$ is the modular curve of level p. Our main result is that the degree of π is odd if and only if $u \equiv 3 \pmod 8$. We also prove the prime-conductor case of a conjecture of Glenn Stevens, namely that that if E is an elliptic curve of prime conductor p then the optimal quotient of $X_1(p)$ in the isogeny class of E is the curve with minimal Faltings height. Finally we discuss some conjectures and data about modular degrees and orders of Shafarevich–Tate groups of Neumann–Setzer curves.

1. Introduction

Let p be a prime of the form $u^2 + 64$ for some integer u, which we take to be 3 modulo 4. Neumann and Setzer [Neu71, Set75] considered the following two elliptic curves of conductor p (note that Setzer chose $u \equiv 1 \pmod 4$ instead):

$$E_0: \quad y^2 + xy = x^3 - \frac{u+1}{4}x^2 + 4x - u, \tag{1.1}$$

$$E_1: \quad y^2 + xy = x^3 - \frac{u+1}{4}x^2 - x. \tag{1.2}$$

For E_1 we have $c_4 = p - 16$ and $c_6 = u(p+8)$ with $\Delta = p = u^2 + 64$, while for E_0 we have $c_4 = p - 256$ and $c_6 = u(p + 512)$ with $\Delta = -p^2$. Thus each E_i is isomorphic to a curve of the form $y^2 = x^3 - 27c_4x - 54c_6$ for the indicated values of c_4 and c_6. The curves E_0 and E_1 are 2-isogenous and one can show using Lutz-Nagell and descent via 2-isogeny that

$$E_0(\mathbf{Q}) = E_1(\mathbf{Q}) = \mathbf{Z}/2\mathbf{Z}.$$

Moreover, if E is *any* elliptic curve over $\mathbf{Q}$ of prime conductor with a rational point of order 2 then E is a Neumann–Setzer curve or has conductor 17 (see [Set75]).

Let $X_0(p)$ be the modular curve of level p. By [Wil95] there is a surjective map $\pi : X_0(p) \to E_0$, and by [MO89, §5, Lem. 3] we may choose π to be optimal, in the sense that π does not factor through any other elliptic curve. The *modular degree* of E_0 is $\deg(\pi)$.

We prove in Section 2 that the modular degree of E_0 is odd if and only if $u \equiv 3 \pmod 8$. Our proof relies mostly on results from [Maz77]. In Section 3 we show that E_1 is the curve of minimal Faltings height in the isogeny class $\{E_0, E_1\}$ of E_1 and prove that E_1 is an optimal quotient of $X_1(p)$, which is enough to prove the prime-conductor case of a conjecture of [Ste89] (this case is not covered by the results of [Vat03]). Finally, in Section 4 we give evidence for our conjecture that there are infinitely many elliptic curves with odd modular degree, and give a conjectural refinement of

Date: December 2003.

Theorem 2.1. We also present some data about p-divisibility of conjectural orders of Shafarevich–Tate groups of Neumann–Setzer curves.

1.1. Notation.

Let p be a prime and n be the numerator of $(p-1)/12$.

We use standard notation for modular forms, modular curves, and Hecke algebras, as in [DI95] and [Maz77]. In particular, let $X_0(p)$ be the compactified coarse moduli space of elliptic curves with a cyclic subgroup of order p. Then $X_0(p)$ is an algebraic curve defined over $\mathbf{Q}$. Let $J = J_0(p)$ be the Jacobian of $X_0(p)$, and let $\mathbf{T} = \mathbf{Z}[T_2, T_3, \ldots] \subset \mathrm{End}(J)$ be the Hecke algebra. Also, let $X_1(p)$ be the modular curve the classifies isomorphism classes of pairs (E, P), where $P \in E$ is a point of order p.

To each newform $f \in S_2(\Gamma_0(p))$, there is an associated abelian subvariety $A = A_f \subset J_0(p)$. We call the kernel Ψ_A of the natural map $A \hookrightarrow J \to A^\vee$ the *modular kernel*. For example, when A is an elliptic curve, this map is induced by pullback followed by push forward on divisors and Ψ_A is multiplication by $\deg(X_0(p) \to A)$. The *modular degree* of A is the square root of the degree of Ψ_A. This definition makes sense even when $\dim(A) > 1$, since the degree of a polarization is the square of its Euler characteristic, hence a perfect square (see [Mum70, §16, pg. 150]). If $I \subset \mathbf{T}$ is an ideal, let

$$A[I] = \{x \in A(\overline{\mathbf{Q}}) : Ix = 0\} \qquad \text{and} \qquad A[I^\infty] = \bigcup_{n>0} A[I^n].$$

1.2. Acknowledgements.

The authors would like to thank the American Institute of Mathematics for hospitality while they worked on this paper, the National Science Foundation for financial support, and Matt Baker, Frank Calegari, and Barry Mazur for helpful conversations.

2. Determination of the Parity of the Modular Degree

Let p, E_0, J and n be as in Section 1, and fix notation as in Section 1.1. In this section we prove the following theorem.

Theorem 2.1. *The modular degree of E_0 is odd if and only if $u \equiv 3 \pmod 8$.*

In order to prove the theorem we deduce seven lemmas using techniques and results from [Maz77].

Let m be the modular degree of E_0, and let

$$B = \ker(J \xrightarrow{\pi} E_0).$$

Lemma 2.2. *We have $m^2 = \#(B \cap E_0)$.*

Proof. As mentioned in Section 1.1, the composition $E_0 \to J \to E_0$ is multiplication by the degree of $X_0(p) \to E_0$, i.e., multiplication by the modular degree of E_0. The lemma follows since multiplication by m on E_0 has degree m^2. □

The *Eisenstein ideal* $\mathcal{I}$ of $\mathbf{T}$ is the ideal generated by $T_\ell - (\ell + 1)$ for $\ell \neq p$ and $T_p - 1$. By hypothesis, there is a Neumann–Setzer curve of conductor p, which implies that the numerator n of $(p-1)/12$ is even (we do the elementary verification that this numerator is even in the proof of Theorem 2.1 below). As discussed in [Maz77, Prop. II.9], the 2-*Eisenstein prime* $\mathfrak{m} = (2) + \mathcal{I}$ of $\mathbf{T}$ is a maximal ideal of $\mathbf{T}$, with $\mathbf{T}/\mathfrak{m} \cong \mathbf{Z}/2\mathbf{Z}$.

Lemma 2.3. *We have $E_0[\mathfrak{m}] = E_0[2]$.*

Proof. By [Maz77, Prop. II.11.1, Thm. III.1.2], the Eisenstein ideal $\mathcal{I}$ annihilates $J(\mathbf{Q})_{\mathrm{tor}}$, so $\mathfrak{m}$ annihilates $J(\mathbf{Q})_{\mathrm{tor}}[2]$. Since $J(\mathbf{Q})_{\mathrm{tor}}$ is cyclic of order n (by [Maz77, Thm. III.1.2]), $J(\mathbf{Q})_{\mathrm{tor}}[2]$ has order 2, so $J(\mathbf{Q})_{\mathrm{tor}}[2] = E_0(\mathbf{Q})_{\mathrm{tor}}[2]$, hence $E_0(\mathbf{Q})[\mathfrak{m}] \neq 0$. The Hecke algebra $\mathbf{T}$ acts on E_0 through $\mathrm{End}(E_0) \cong \mathbf{Z}$, so each element of $\mathbf{T}$ acts on E_0 as an integer; in particular, the elements of $\mathfrak{m}$ all act as multiples of 2 (since $E_0[\mathfrak{m}] \neq 0$ and $2 \in \mathfrak{m}$), so $E_0[\mathfrak{m}] = E_0[2]$ since $2 \in \mathfrak{m}$. □

Lemma 2.4. *Suppose $A \subset J_0(p)$ is a $\mathbf{T}$-stable abelian subvariety and $\wp \subset \mathbf{T}$ is a maximal ideal such that $A[\wp^\infty] \neq 0$. Then $A[\wp] \neq 0$. Also $A[\wp^\infty]$ is infinite.*

Proof. Arguing as in [Maz77, §II.14, pg. 112], we see that for any r, $A[\wp^r]/A[\wp^{r+1}]$ is isomorphic to a direct sum of copies of $A[\wp]$. If $A[\wp] = 0$, then since $A[\wp^\infty] \neq 0$, there must exist an r such that $A[\wp^r]/A[\wp^{r+1}] \neq 0$. But $A[\wp^r]/A[\wp^{r+1}]$ is contained in a direct sum of copies of $A[\wp] = 0$, which is a contradiction.

To see that $A[\wp^\infty]$ is infinite, note that if ℓ is the residue characteristic of $\wp$ and $\mathrm{Tate}_\ell(A)$ is the Tate module of A at ℓ, then

$$\mathrm{Tate}_\wp(A) = \varprojlim_r A[\wp^r] = \mathrm{Tate}_\ell(A) \otimes_{\mathbf{T}} \mathbf{T}_\wp$$

is infinite. (For more details, see the proof of [RS01, Prop. 3.2].) □

The analogues of Lemmas 2.5–2.7 below are true, with the same proofs, for $\mathfrak{m}$ any Eisenstein prime. We state and prove them for the 2-Eisenstein prime, since that is the main case of interest to us. Let $\tilde{J}^{(2)}$ be the 2-*Eisenstein quotient of* J, where $\tilde{J}^{(2)}$ is as defined in [Maz77, §II.10]. More precisely, we have the following:

Lemma 2.5. *The simple factors of $\tilde{J}^{(2)}$ correspond to the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugacy classes of newforms f such that $A_f[\mathfrak{m}] \neq 0$ (or equivalently, $A_f^\vee[\mathfrak{m}] \neq 0$).*

Proof. On page 97 of [Maz77] we find that the $\mathbf{C}$-simple factors of $\tilde{J}^{(2)}$ are in bijection with the irreducible components $\mathrm{Spec}(I_f)$ of $\mathrm{Spec}(\mathbf{T})$ which meet the support of the ideal $\mathfrak{m}$, so the I_f are the newform ideals contained in $\mathfrak{m}$. We have for any I_f,

$$I_f \subset \mathfrak{m} \iff \mathbf{T}_m/(I_f)_\mathfrak{m} \neq 0 \iff \mathrm{Tate}_\mathfrak{m}(A_f) \neq 0 \iff A_f[\mathfrak{m}] \neq 0.$$

Note that the same argument applies to $A_f^\vee$. □

Lemma 2.6. *Suppose A and B are abelian varieties equipped with an action of the Hecke ring $\mathbf{T}$ and that $\varphi : A \to B$ is a $\mathbf{T}$-module isogeny. If $\wp \subset \mathbf{T}$ is a maximal ideal and $B[\wp] \neq 0$, then also $A[\wp] \neq 0$.*

Proof. Let $\psi : B \to A$ be the isogeny complementary to φ, so ψ is the unique isogeny such that $\psi \circ \varphi$ is multiplication by $\deg(\varphi)$. Then ψ is also a $\mathbf{T}$-module homomorphism (one can see this in various ways; one way is to use the rational representation on homology to view the endomorphisms as matrices acting on lattices, and to note that if matrices M and N commute, then M^{-1} and N also commute). By Lemma 2.4, the union $B[\wp^\infty]$ is infinite, so $\psi(B[\wp^\infty]) \neq 0$. Since $\psi(B[\wp^\infty]) \subset A[\wp^\infty]$, Lemma 2.4 implies that $A[\wp] \neq 0$, as claimed. □

Lemma 2.7. *Suppose $B \subset J_0(p)$ is a sum of abelian subvarieties A_f attached to newforms. If $B[\mathfrak{m}] \neq 0$, then there is some $A_f \subset B$ such that $A_f[\mathfrak{m}] \neq 0$.*

Proof. There is something to be proved because if $x \in B[\mathfrak{m}]$ it could be the case that $x = y + z$ with $y \in A_f$ and $z \in A_g$, but $x \notin A_h$ for any h. Let $C = \oplus A_f$, where the $A_f \subset J_0(p)$ are simple abelian subvarieties of B corresponding to conjugacy classes of newforms. Then there is an isogeny $\varphi : C \to B$ given by

$$\varphi(x_1, \ldots, x_n) = x_1 + \cdots + x_n,$$

where the sum is in $B \subset J_0(p)$. By Lemma 2.6, $C[\mathfrak{m}] \neq 0$. Since $C[\mathfrak{m}] \cong \oplus A_f[\mathfrak{m}]$, it follows that $A_f[\mathfrak{m}] \neq 0$ for some $A_f \subset B$. □

Lemma 2.8. *If* $4 \mid n$, *then* $\dim \tilde{J}^{(2)} > 1$.

Proof. This follows from the remark on page 163 of [Maz77]. Since the proof is only sketched there, we give further details for the convenience of the reader. Because $4 \mid n$, the cuspidal subgroup C, which is generated in $J_0(p)$ by $(0) - (\infty)$ and is cyclic of order n, contains an element of order 4. Let $C(2)$ be the 2-primary part of C, and let $D = \ker(J_0(p) \to \tilde{J}^{(2)})$. If there is a nonzero element in the kernel of the homomorphism $C(2) \to \tilde{J}^{(2)}$, then $D[\mathfrak{m}] \neq 0$, where $\mathfrak{m}$ is the 2-Eisenstein prime. But then by Lemma 2.7, there is an $A_f \subset D$ such that $A_f[\mathfrak{m}] \neq 0$. By Lemma 2.5, $A_f^\vee$ is a quotient of $\tilde{J}^{(2)}$, so $A_f \subset (\tilde{J}^{(2)})^\vee$ so A_f cannot be in D. This contradiction shows that the map $C(2) \to \tilde{J}^{(2)}$ is injective, so $\tilde{J}^{(2)}$ contains a rational point of order 4. However, as mentioned in the introduction, $E_0(\mathbf{Q})$ has order 2, so $\tilde{J}^{(2)} \neq E_0$. Thus $\tilde{J}^{(2)}$ has dimension bigger than 1. □

Having established the above lemmas, we are now ready to deduce the theorem.

Proof of Theorem 2.1. It seems more straightforward to prove the equivalent statement that the modular degree is even if and only if $u \equiv 7 \pmod 8$, so we will prove this instead.

($\Longrightarrow$) $u \equiv 7 \pmod 8$ *implies that the modular degree is even:* Writing $u = 8k + 7$ we see that $p = (8k+7)^2 + 64 \equiv 1 \pmod{16}$, so $16 \mid (p-1)$ hence $4 \mid n$. By Lemma 2.3 and Lemma 2.5, E_0 is a factor of $\tilde{J}^{(2)}$. By Lemma 2.8, the dimension of $\tilde{J}^{(2)}$ is bigger than 1, so by Lemma 2.5 there is an A_f distinct from E_0 such that $A_f[\mathfrak{m}] \neq 0$. Since $A_f \subset B = \ker(J_0(p) \to E_0)$, it follows that $B[\mathfrak{m}] \neq 0$. As discussed on page 38 of [Maz77], $J[\mathfrak{m}]$ has dimension 2 over $\mathbf{F}_2$ so $E_0[\mathfrak{m}] = J[\mathfrak{m}]$, hence $B[\mathfrak{m}] \subset E_0[\mathfrak{m}]$. It follows that $2 \mid \#(B \cap E_0)$, so E_0 has even modular degree.

($\Longleftarrow$) *Modular degree even implies that* $u \equiv 7 \pmod 8$*:* Suppose that the modular degree m of E_0 is even. Letting $B = \ker(J_0(p) \to E_0)$, we have

$$E_0 \cap B \cong \ker(E_0 \to J_0(p) \to E_0),$$

so $\Psi := E_0 \cap B = E_0[m]$. Lemma 2.3 and our assumption that m is even imply that

$$E_0[\mathfrak{m}] = E_0[2] \subset E_0[m] = \Psi,$$

so $\Psi[\mathfrak{m}] \neq 0$. Since $\Psi[\mathfrak{m}] \neq 0$, and $\Psi \subset B$, we have $B[\mathfrak{m}] \neq 0$. By Lemma 2.7, there is some $A_f \subset B$ such that $A_f[\mathfrak{m}] \neq 0$. Then by Lemma 2.5 we see that A_f is an isogeny factor of $\tilde{J}^{(2)}$. Thus $\tilde{J}^{(2)}$ has dimension bigger than 1. If $u = 8k+3$, then $p = (8k+3)^2 + 64 \equiv 9 \pmod{16}$, so that $2 \mid\mid n$. However, when $2 \mid\mid n$, [Maz77, Prop. III.7.5] implies that $\tilde{J}^{(2)} = E_0$, which is false, so $u \equiv 7 \pmod 8$. □

Remark 2.9. Frank Calegari observed that Lemma 2.8 and its converse also follow from conditions (i) and (v) of Théorème 3 of [Mer96].

3. The Stevens Conjecture for Neumann–Setzer Curves is True

Let E be an arbitrary elliptic curve over $\mathbf{Q}$ of conductor N. Stevens conjectured in [Ste89] that the optimal quotient of $X_1(N)$ in the isogeny class of E is the curve in the isogeny class of E with minimal Faltings height. In this section we explain why this conjecture is true when N is prime.

Let $p = u^2 + 64$ be prime and E_1 and E_0 be as in Section 1. In this section we verify that the curve E_1 has smaller Faltings height than E_0, then show that E_1 is $X_1(p)$-optimal. The Stevens conjecture asserts that the $X_1(p)$-optimal curve is the curve of minimal Faltings height in an isogeny class, so our results verify the conjecture for Neumann–Setzer curves. In fact, the Stevens conjecture is true for all isogeny classes of elliptic curves of prime conductor. For if E is an elliptic curve of prime conductor, then by [Set75] there is only one curve in the isogeny class of E, unless E is a Neumann–Setzer curve or the conductor of E is 11, 17, 19, or 37. When the isogeny class of E contains only one curve, that curve is obviously both X_1-optimal and of minimal Faltings height. The conjecture is also well-known to be true for curves of conductor 11, 17, 19, or 37 (see [Ste89]). We note that Vatsal [Vat03] has recently extended results of Tang [Tan97] that make considerable progress toward the Stevens conjecture, but his work is not applicable to Neumann–Setzer curves.

Lemma 3.1. *The curve E_1 has smaller Faltings height than E_0.*

Proof. By [Ste89, Thm. 2.3, pg. 84] it is enough to exhibit an isogeny from E_1 to E_0 whose extension to Néron models is étale. Let φ be the isogeny $E_0 \to E_1$ of degree 2 whose kernel is the subgroup generated by the point whose coordinates are $(u/4, -u/8)$ in terms of the Weierstrass equation (1.1) for E_0, which is a global minimal model for E_0. The kernel of φ does not extend to an étale group scheme over $\mathbf{Z}$, since its special fiber at 2 is not étale (it has only one $\overline{\mathbf{F}}_2$-point), so the morphism on Néron models induced by $E_0 \to E_1$ cannot be étale, since kernels of étale morphisms are étale. By [Ste89, Lemma 2.5] the dual isogeny $E_1 \to E_0$ extends to an étale morphism of Néron models. □

Proposition 3.2. *The curve E_1 is $X_1(p)$-optimal.*

Proof. By [MO89, §5, Lem. 3], E_0 is an optimal quotient of $X_0(p)$, so we have an injection $E_0 \hookrightarrow J_0(p)$. As in [Maz77, pg. 100], let Σ be the kernel of the functorial map $J_0(p) \to J_1(p)$ induced by the cover $X_1(p) \to X_0(p)$. By [Maz77, Prop. II.11.6], Σ is the Cartier dual of the constant subgroup scheme U, which turns out to equal $J_0(p)(\mathbf{Q})_{\mathrm{tor}}$. Because $\#(E_0 \cap U) = 2$ and $E_0[2]$ is self dual, we have $\#(E_0 \cap \Sigma) = 2$. Thus the image of E_0 in $J_1(p)$ is the quotient of E_0 by the subgroup generated by the rational point of order 2 (note that the Cartier dual of $\mathbf{Z}/2\mathbf{Z}$ is $\mu_2 = \mathbf{Z}/2\mathbf{Z}$). This quotient is E_1, so $E_1 \subset J_1(p)$, which implies that E_1 is an optimal quotient of $X_1(p)$, as claimed. □

Remark 3.3. The above proposition could also be proved in a slightly different manner. The Faltings height of an elliptic curve is $\sqrt{2\pi/\Omega}$ where Ω is the volume of the fundamental parallelogram associated to the curve. When the conductor is prime, we have by [AL96] that the Manin constants for $X_0(p)$ and $X_1(p)$ are 1; this says that for a G-optimal curve E, the period lattice generated by G has covolume equal to Ω_E. Since the lattice generated by $\Gamma_1(p)$ is contained in the lattice generated by $\Gamma_0(p)$ (and thus has larger covolume), the Faltings height of the $X_1(p)$-optimal curve must be less than or equal to that of the $X_0(p)$-optimal curve. So if these two curves differ, the $X_1(p)$-optimal curve must have smaller Faltings height.

Remark 3.4. On page 12 of [Maz98], there is a "To be removed from the final draft" comment that asks (in our notation) whether E_0 is $X_0(p)$-optimal when $p \equiv 1 \pmod{16}$. This is already answered by [MO89], whereas here we go further and show additionally that E_1 is $X_1(p)$-optimal.

4. Conjectures

4.1. **Refinement of Theorem 2.1.** The following conjectural refinement of Theorem 2.1 is supported by the experimental data of [Wat02]. It is unclear whether the method of proof of Theorem 2.1 can be extended to prove this conjecture.

Conjecture 4.1. *If $u \equiv 7 \pmod 8$, then 2 exactly divides the modular degree of E_0 if and only if $u \equiv 7 \pmod{16}$.*

We can note that the pattern seems to end here; for curves with $u \equiv 15 \pmod{16}$ the data give no further information about the 2-valuation of the modular degree. For instance, with $u = -17$ we have that $[1, 1, 1, -2, 16]$ has modular degree $2^3 \cdot 3$, while with $u = 175$ the curve $[1, 1, 1, -634, -6484]$ has modular degree $2^2 \cdot 3^3 \cdot 5 \cdot 23$. Similarly, we have that $u = -33$ gives the curve $[1, -1, 1, -19, 68]$ with modular degree $2^5 \cdot 3$, while $u = 127$ gives the curve $[1, 1, 1, -332, -2594]$ of modular degree $2^2 \cdot 3^2 \cdot 5 \cdot 43$.

4.2. **The Parity of the Modular Degree.** According to Cremona's tables [Cre], of the 29755 new optimal elliptic curve quotients of $J_0(N)$ with $N < 8000$, a mere 89 have odd modular degree, which is less than 0.3%. There are 52878 non Neumann–Setzer curves in the database of [BM90] with prime conductor $N \leq 10^7$; of these curves 4592, or 8%, have odd modular degree (see [Wat02]). One reason that curves tend to have even modular degree is that for many curves the modular parametrization factors through an Atkin-Lehner quotient. Note that the method of [Wat02] used to compute the modular degree is rigourous when the level is prime because by [AL96] the Manin constant is 1 when the level is odd and square-free.

If $f(x) = (8x + 3)^2 + 64$, then it is a well-known conjecture (see [HL22] and e.g., [Guy94, §A1]) that there are infinitely many primes of the form $f(n)$ for some integer n, thus we make the following conjecture.

Conjecture 4.2. *There are infinitely many elliptic curves over* $\mathbf{Q}$ *with odd modular degree.*

Our data suggest the following conjecture:

Conjecture 4.3. *If E is an optimal elliptic curve quotient of $J_0(p)$ with $p \not\equiv 3 \pmod 8$ and E is not a Neumann–Setzer curve then the modular degree of E is even or $p = 17$.*

There are 23442 Brumer-McGuinness (see [BM90]) curves of conductor $37 \leq p \leq 10^7$ with $p \equiv 3 \pmod 8$, of which 11815 have even functional equation, of which 7322 have rank 0, and 4589 have odd modular degree. The significance of the data concerning the rank is that the second author has conjectured that 2^r divides the modular degree, where r is the rank.

Remark 4.4. Instead of asking about divisibility by 2, one could ask about divisibility by p. The first author and Frank Calegari make a conjecture about discriminants of Hecke algebras in [CS03] that implies that the modular degree of an elliptic curve of prime conductor p is not divisible by p. This conjecture agrees with our data.

4.3. **Shafarevich–Tate Groups of Neumann–Setzer Curves.** We consider the distribution of Ш in the Neumann–Setzer family (and note that similar phenomena occur in the related families listed in [SW02]). We look at u with $u^2 + 64$ prime and less than $2 \cdot 10^{12}$. We now take u to be positive, which thus replaces the restriction that u be 3 mod 4. The heuristics of [Del01] would seem to give us an idea of how

Table 1. Frequency of a prime dividing Ш

restriction	number	$p = 3$	$p = 5$	$p = 7$	$p = 11$
$u \equiv 1 \pmod 8$	25559	33.2%	16.9%	9.2%	3.0%
$u \equiv 3 \pmod 8$	25557	39.7%	20.3%	14.3%	8.4%
$u \equiv 5 \pmod 8$	25584	36.2%	18.5%	11.5%	5.0%
$u \equiv 7 \pmod 8$	25612	34.3%	20.3%	14.3%	8.2%
$u \equiv 0 \pmod 3$	34009	36.0%	18.7%	12.1%	6.0%
$u \equiv 1 \pmod 3$	34032	35.2%	18.6%	11.5%	5.6%
$u \equiv 2 \pmod 3$	34271	36.3%	19.7%	13.3%	6.9%
$u \equiv 0 \pmod 5$	34208	33.1%	18.0%	11.4%	5.4%
$u \equiv 2 \pmod 5$	33879	37.1%	19.5%	12.8%	6.5%
$u \equiv 3 \pmod 5$	34225	37.3%	19.5%	12.7%	6.5%
total	102312	35.8%	19.0%	12.3%	6.2%
Delaunay		36.1%	20.7%	14.5%	9.2%

often we expect a given prime to divide Ш. For instance, since Neumann–Setzer curves have rank 0, the prime 3 should divide Ш about 36.1% of the time. However, Table 1 gives a slightly different story with effects seen that depend on the various congruential properties of u.

References

[AL96] A. Abbes, E. Ullmo, *À propos de la conjecture de Manin pour les courbes elliptiques modulaires* (French). Compositio Math. **103** (1996), no. 3, 269–286.

[BM90] A. Brumer, O. McGuinness, *The behavior of the Mordell-Weil group of elliptic curves.* Bull. Amer. Math. Soc. (N.S.) **23** (1990), no. 2, 375–382, data available online at `http://modular.fas.harvard.edu/~oisin`

[CS03] F. Calegari, W. Stein, *Conjectures About Discriminants of Hecke Algebras of Prime Level*, submitted (2003).

[Cre] J. E. Cremona, *Elliptic Curves of conductor* $\leq$ 17000, electronic tables available online at `http://www.maths.nott.ac.uk/personal/jec/ftp/data`

[Del01] C. Delaunay, *Heuristics on Tate-Shafarevitch Groups of Elliptic Curves Defined over* **Q**. Experiment. Math. **10** (2001), no. 2, 191–196.

[Del83] P. Deligne, *Preuve des conjectures de Tate et de Shafarevitch (d'après G. Faltings).* (French). Seminaire Bourbaki, Vol. 1983/84. Astérisque No. 121-122, (1985), 25–41.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[Eme01] M. Emerton, *Optimal Quotients of Modular Jacobians,* preprint (2001).

[Frey87] G. Frey, *Links between solutions of* $A - B = C$ *and elliptic curves.* In *Number Theory (Ulm, 1987),* edited by H. P. Schlickewei and E. Wirsing, 31–62, Lecture Notes in Mathematics, 1380, Springer-Verlag, New York, 1989.

[Guy94] R. K. Guy, *Unsolved problems in number theory,* Springer-Verlag, 1994.

[HL22] G. H. Hardy, J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes.* Acta. Math. **44** (1922), 1–70.

[LO91] S. Ling, J. Oesterlé, The Shimura subgroup of $J_0(N)$. Astérisque **196–197** (1991), 171–203.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal.* Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33–186.

[Maz98] B. Mazur, *Three Lectures about the Arithmetic of Elliptic Curves.* Rough, unedited, and preliminary notes from lectures given at the 1998 Arizona Winter School. Available at `http://swc.math.arizona.edu/notes/files/98MazurLN.ps`

[Mer96] L. Merel, *L'accouplement de Weil entre le sous-groupe de Shimura et le sous-groupe cuspidal de $J_0(p)$*, J. Reine Angew. Math. **477** (1996), 71–115.

[MO89] J.-F. Mestre, J. Oesterlé, *Courbes de Weil semi-stables de discriminant une puissance m-ième* (French). J. Reine Angew. Math. **400** (1989), 173–184.

[Mum70] D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.

[Neu71] O. Neumann, *Elliptische Kurven mit vorgeschriebenem Reduktionsverhalten. I, II* (German). Math. Nachr. **49** (1971), 107–123, **56** (1973), 269–280.

[Ogg74] A. Ogg, *Hyperelliptic modular curves.* Bull. Soc. Math. France **102** (1974), 449–462.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 143–232.

[Set75] B. Setzer, *Elliptic Curves of prime conductor.* J. London Math. Soc. (2), **10** (1975), 367–378.

[Ste89] G. Stevens, *Stickelberger elements and modular parametrizations of elliptic curves.* Invent. Math. **98** (1989), no. 1, 75–106.

[SW02] W. A. Stein, M. Watkins, *A Database of Elliptic Curves—First Report.* In *Algorithmic number theory* (Sydney 2002), 267–275, edited by C. Fieker and D. Kohel, Lecture Notes in Comput. Sci., 2369, Springer, Berlin, 2002.

[Tan97] S.-L. Tang, *Congruences between modular forms, cyclic isogenies of modular elliptic curves, and integrality of p-adic L-function.* Trans. Amer. Math. Soc. **349** (1997), no. 2, 837–856.

[Vat03] V. Vatsal, *Multiplicative subgroups of $J_0(N)$ and applications to elliptic curves,* preprint (2003).

[Wat02] M. Watkins, *Computing the modular degree of an elliptic curve,* Experiment. Math. **11** (2002), no. 4, 487–502.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem.* Ann. of Math. (2) **141** (1995), no. 3, 443–551.

21 Conjectures About Discriminants of Hecke Algebras of Prime Level, with F. Calegari

Conjectures About Discriminants of Hecke Algebras of Prime Level

Frank Calegari⋆⋆⋆[1] and William A. Stein†[2]

[1] Harvard University
fcale@math.harvard.edu
http://www.math.harvard.edu/~fcale
[2] Harvard University,
was@math.harvard.edu
http://modular.fas.harvard.edu/

Abstract. In this paper, we study p-divisibility of discriminants of Hecke algebras associated to spaces of cusp forms of prime level. By considering cusp forms of weight bigger than 2, we are are led to make a precise conjecture about indexes of Hecke algebras in their normalisation which implies (if true) the surprising conjecture that there are no mod p congruences between non-conjugate newforms in $S_2(\Gamma_0(p))$, but there are almost always many such congruences when the weight is bigger than 2.

1 Basic Definitions

We first recall some commutative algebra related to discriminants, then introduce Hecke algebras of spaces of cusp forms.

1.1 Commutative Algebra

In this section we recall the definition of discriminant of a finite algebra and note that the discriminant is nonzero if and only if no base extension of the algebra contains nilpotents.

Let R be a ring and let A be an R-algebra that is free of finite rank as an R-module. The *trace* of $x \in A$ is the trace, in the sense of linear algebra, of left multiplication by x.

Definition 1 (Discriminant). *Let $\omega_1, \ldots, \omega_n$ be an R-basis for A. Then the* discriminant $\mathrm{disc}(A)$ *of A is the determinant of the $n \times n$ matrix $(\mathrm{tr}(\omega_i\omega_j))$.*

⋆⋆⋆ Supported in part by the American Institute of Mathematics

† Supported in part by a National Science Foundation Postdoctoral Fellowship

The discriminant is only well-defined modulo squares of units in R. When $R = \mathbf{Z}$ the discriminant is well defined, since the only units are ± 1.

We say that A is *separable over* R if for every extension R' of R, the ring $A \otimes R'$ contains no nilpotents.

Proposition 1. *Suppose R is a field. Then A has nonzero discriminant if and only if A is separable over R.*

Proof. For the convenience of the reader, we summarize the proof in [Mat86, §26]. If A contains a nilpotent then that nilpotent is in the kernel of the trace pairing, so the discriminant is 0. Conversely, if A is separable then we may assume that R is algebraically closed. Then A is an Artinian reduced ring, hence isomorphic as a ring to a finite product of copies of R, since R is algebraically closed. Thus the trace form on A is nondegenerate.

1.2 The Discriminant Valuation

We next introduce Hecke algebras attached to certain spaces of cusp forms of prime level p, define the discriminant valuation as the exponent of the largest power of p that divides the discriminant, and observe that there are eigenform congruences modulo p exactly when the discriminant valuation is positive. We then present an example to illustrate the definitions.

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$. In this paper, we will only consider $\Gamma = \Gamma_0(p)$ for p prime. For any positive integer k, let $S_k(\Gamma)$ denote the space of holomorphic weight k cusp forms for Γ. Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_k(\Gamma))$$

be the associated Hecke algebra, which is generated by Hecke operators T_n for all integers n, including $n = p$ (we will sometimes write U_p for T_p). Then $\mathbf{T}$ is a commutative ring that is free as a module over $\mathbf{Z}$ of rank equal to $\dim S_k(\Gamma)$. We will also sometimes consider the image $\mathbf{T}^{\mathrm{new}}$ of $\mathbf{T}$ in $\mathrm{End}(S_k(\Gamma)^{\mathrm{new}})$.

Definition 2 (Discriminant Valuation). *Let p be a prime, k a positive integer, and suppose that $\Gamma = \Gamma_0(p)$. Let $\mathbf{T}$ be the corresponding Hecke algebra. Then the* discriminant valuation *of Γ in weight k is*

$$d_k(\Gamma) = \mathrm{ord}_p(\mathrm{disc}(\mathbf{T})).$$

We expect that $d_k(\Gamma)$ is finite for the following reason. The Hecke operators T_n, with n not divisible by p, are diagonalizable since they are self adjoint with respect to the Petersson inner product. When $k = 2$ one knows that U_p is diagonalizable since the level is square free, and when $k > 2$ one expects this (see [CE98]). If $\mathbf{T}$ contains no nilpotents, Proposition 1 implies that the discriminant of $\mathbf{T}$ is nonzero. Thus $d_k(\Gamma)$ is finite when $k = 2$ and conjectured to be finite when $k > 2$.

Let p be a prime and suppose that $\Gamma = \Gamma_0(p)$. A *normalised eigenform* is an element $f = \sum a_n q^n \in S_k(\Gamma)$ that is an eigenvector for *all* Hecke operators T_ℓ, including those that divide p, normalised so that $a_1 = 1$. The quantity $d_k(\Gamma)$ is of interest because it measures mod p congruences between normalised eigenforms in $S_k(\Gamma)$.

Proposition 2. *Assume that $d_k(\Gamma)$ is finite. The discriminant valuation $d_k(\Gamma)$ is positive (i.e., the discriminant is divisible by p) if and only if there is a congruence in characteristic p between two normalized eigenforms in $S_k(\Gamma)$. (The two congruent eigenforms might be Galois conjugate.)*

Proof. It follows from Proposition 1 that $d_k(\Gamma) > 0$ if and only if $\mathbf{T} \otimes \overline{\mathbf{F}}_p$ is not separable. The Artinian ring $\mathbf{T} \otimes \overline{\mathbf{F}}_p$ is not separable if and only if the number of ring homomorphisms $\mathbf{T} \otimes \overline{\mathbf{F}}_p \to \overline{\mathbf{F}}_p$ is less than

$$\dim_{\overline{\mathbf{F}}_p} \mathbf{T} \otimes \overline{\mathbf{F}}_p = \dim_{\mathbf{C}} S_k(\Gamma).$$

Since $d_k(\Gamma)$ is finite, the number of ring homomorphisms $\mathbf{T} \otimes \overline{\mathbf{Q}}_p \to \overline{\mathbf{Q}}_p$ equals $\dim_{\mathbf{C}} S_k(\Gamma)$. The proposition follows from the fact that for any ring R, there is a bijection between ring homomorphisms $\mathbf{T} \to R$ and normalised eigenforms with q-expansion in R.

The same proof also shows that a prime ℓ divides the discriminant of $\mathbf{T}$ if and only if there is a congruence mod ℓ between two normalized eigenforms in $S_k(\Gamma)$

Example 1. If $\Gamma = \Gamma_0(389)$ and $k = 2$, then $\dim_{\mathbf{C}} S_2(\Gamma) = 32$. Let f be the characteristic polynomial of T_2. One can check that f is square free and 389 exactly divides the discriminant of f. This implies that $d_2(\Gamma) = 1$ and that T_2 generates $\mathbf{T} \otimes \mathbf{Z}_{389}$ as an algebra over $\mathbf{Z}_{389}$. (If T_2 only generated a subring of $\mathbf{T} \otimes \mathbf{Z}_{389}$ of finite index > 1, then the discriminant of f would be divisible by 389^2.)

Modulo 389 the characteristic polynomial f is congruent to

$$(x+2)(x+56)(x+135)(x+158)(x+175)^2(x+315)(x+342)(x^2+387)\\(x^2+97x+164)(x^2+231x+64)(x^2+286x+63)(x^5+88x^4+196x^3+\\113x^2+168x+349)(x^{11}+276x^{10}+182x^9+13x^8+298x^7+316x^6+\\213x^5+248x^4+108x^3+283x^2+x+101)$$

The factor $(x+175)^2$ indicates that $\mathbf{T}\otimes\mathbf{F}_{389}$ is not separable over $\mathbf{F}_{389}$ since the image of $(\overline{f}/(x+175))(T_2)$ in $\mathbf{T}\otimes\mathbf{F}_{389}$ is nilpotent (it is nonzero but its square is 0). There are 32 eigenforms over $\mathbf{Q}_2$ but only 31 mod 389 eigenforms, so there must be a congruence. There is a newform F in $S_2(\Gamma_0(389),\overline{\mathbf{Z}}_{389})$ whose a_2 term is a root of

$$x^2+(-39+190\cdot 389+96\cdot 389^2+\cdots)x+(-106+43\cdot 389+19\cdot 389^2+\cdots).$$

There is a congruence between F and its $\mathrm{Gal}(\overline{\mathbf{Q}}_{389}/\mathbf{Q}_{389})$-conjugate.

2 Computing Discriminants

In this section we sketch the algorithm that we use for computing the discriminants mentioned in this paper.

This algorithm was inspired by a discussion of the second author with Hendrik Lenstra. We leave the details of converting the description below into standard matrix operations to the reader. Also, the modular symbols algorithms needed to compute Hecke operators are quite involved.

Let $\Gamma=\Gamma_0(p)$, and let $k\geq 2$ be an integer. The following sketches an algorithm for computing the discriminant of the Hecke algebra $\mathbf{T}$ acting on $S_k(\Gamma)$.

1. For any given n, we can explicitly compute a matrix that represents the action of Hecke operators T_n on $S_k(\Gamma)$ using modular symbols. We use the second author's MAGMA [BCP97] packages for computing with modular symbols, which builds on work of many people (including [Cre97] and [Mer94]).
2. Using the Sturm bound, as described in the appendix to [LS02], find an integer b such that $T_1,\ldots,T_b$ generate $\mathbf{T}$ as a $\mathbf{Z}$-module. (The integer b is $\lceil (k/12)\cdot[\mathrm{SL}_2(\mathbf{Z}):\Gamma]\rceil$.)
3. Find a subset B of the T_i that form a $\mathbf{Q}$-basis for $\mathbf{T}\otimes_{\mathbf{Z}}\mathbf{Q}$. (This uses Gauss elimination.)
4. View $\mathbf{T}$ as a ring of matrices acting on $\mathbf{Q}^d$, where $d=\dim(S_k(\Gamma))$ and try random sparse vectors $v\in\mathbf{Q}^d$ until we find one such that the set of vectors $C=\{T(v):T\in B\}$ are linearly independent.

5. Write each of $T_1(v), \dots, T_b(v)$ as $\mathbf{Q}$-linear combinations of the elements of C.
6. Find a $\mathbf{Z}$-basis D for the $\mathbf{Z}$-span of these $\mathbf{Q}$-linear combinations of elements of C. (This basis D corresponds to a $\mathbf{Z}$-basis for $\mathbf{T}$, but is much easier to find that directly looking for a $\mathbf{Z}$-basis in the space of $d \times d$ matrices that $\mathbf{T}$ is naturally computed in.)
7. Unwinding what we have done in the previous steps, find the trace pairing on the elements of D, and deduce the discriminant of $\mathbf{T}$ by computing the determinant of the trace pairing matrix.

A very time-consuming step, at least in our implementation, is computing D from $T_1(v), \dots, T_b(v)$ expressed in terms of C, and this explains why we embed $\mathbf{T}$ in $\mathbf{Q}^d$ instead of viewing the elements of $\mathbf{T}$ as vectors in $\mathbf{Q}^{d \times d}$.

An implementation by the second author of the above algorithm is included with the MAGMA computer algebra system. The relevant source code is in the file `Geometry/ModSym/linalg.m` in the `package` directory (or ask the second author of the apper to send you a copy `linalg.m`). We illustrate the use of MAGMA to compute discriminants below, which were run under MAGMA V2.10-21 for Linux on a computer with an Athlon 2800MP processor (2.1Ghz).

```
> M := ModularSymbols(389,2, +1);
> S := CuspidalSubspace(M);
> time D  := DiscriminantOfHeckeAlgebra(S);
Time: 0.750
> D;
629670054720061882880174736321392595498204931550235108311\
04000000
> Factorisation(D);
[ <2, 53>, <3, 4>, <5, 6>, <31, 2>, <37, 1>, <389, 1>, ...]
> M := ModularSymbols(997,2, +1); S := CuspidalSubspace(M);
> time D  := DiscriminantOfHeckeAlgebra(S);
Time: 55.600
```

The reason for the $+1$ in the construction of modular symbols is so that we compute on a space that is isomorphic as a $\mathbf{T}$-module to one copy of $S_2(\Gamma_0(p))$, instead of two copies.

3 Data About Discriminant Valuations

In this section we report on our extensive computations of $d_k(\Gamma_0(p))$. We first note that there is only one $p < 50000$ such that $d_2(\Gamma_0(p)) > 0$.

Next we give a table of values of $d_4(\Gamma_0(p))$, which seems to exhibit a nice pattern.

3.1 Weight Two

Theorem 1. *The only prime* $p < 60000$ *such that* $d_2(\Gamma_0(p)) > 0$ *is* $p = 389$*, with the possible exception of* 50923 *and* 51437*.*

Computations in this direction by the second author have been cited in [Rib99], [MS01], [OW02], and [MO02]. For example, Theorem 1 is used for $p < 1000$ in [MS01] as a crucial step in proving that if E is an elliptic curve over $\mathbf{Q}(\mu_p)$, with $17 \leq p < 1000$, then not all elements of $E(\overline{\mathbf{Q}})[p]$ are rational over $\mathbf{Q}(\mu_p)$.

Proof. This is the result of a large computer computation. The rest of this proof describes how we did the computation, so the reader has some idea how to replicate or extend the computation. The computation described below took about one week using a cluster equipped with 10 Athlon 2000MP processors. The computations are nontrivial; we compute spaces of modular symbols, supersingular points, and Hecke operators on spaces of dimensions up to 5000.

The aim is to determine whether or not p divides the discriminant of the Hecke algebra of level p for each $p < 60000$. If T is an operator with integral characteristic polynomial, we write $\mathrm{disc}(T)$ for $\mathrm{disc}(\mathrm{charpoly}(T))$, which also equals $\mathrm{disc}(\mathbf{Z}[T])$. We will often use that

$$\mathrm{disc}(T) \bmod p = \mathrm{disc}(\mathrm{charpoly}(T) \bmod p).$$

We ruled out the possibility that $d_k(\Gamma_0(p)) > 0$ for most levels $p < 60000$ by computing characteristic polynomials of Hecke operators using an algorithm that the second author and D. Kohel implemented in MAGMA ([BCP97]), which is based on the Mestre-Oesterle method of graphs [Mes86] (or contact the second author for an English translation). Our implementation is available as the “Module of Supersingular Points” package that comes with MAGMA. We computed $\mathrm{disc}(T_q)$ modulo p for several small primes q, and in most cases found a prime q such that this discriminant is nonzero. The following table summarises how often we used each prime q (note that there are 6057 primes up to 60000):

q	number of $p < 60000$ where q smallest s.t. $\mathrm{disc}(T_q) \neq 0 \bmod p$
2	5809 times
3	161 (largest: 59471)
5	43 (largest: 57793)
7	15 (largest: 58699)
11	15 (the smallest is 307; the largest 50971)
13	2 (they are 577 and 5417)
17	3 (they are 17209, 24533, and 47387)
19	1 (it is 15661)

The numbers in the right column sum to 6049, so 8 levels are missing. These are

$$389, 487, 2341, 7057, 15641, 28279, 50923, \text{ and } 51437.$$

(The last two are still being processed. 51437 has the property that $\mathrm{disc}(T_q) = 0$ for $q = 2, 3, \ldots, 17$.) We determined the situation with the remaining 6 levels using Hecke operators T_n with n composite.

p	How we rule level p out, if possible
389	p does divide discriminant
487	using charpoly(T_{12})
2341	using charpoly(T_6)
7057	using charpoly(T_{18})
15641	using charpoly(T_6)
28279	using charpoly(T_{34})

Computing T_n with n composite is very time consuming when p is large, so it is important to choose the right T_n quickly. For $p = 28279$, here is a trick we used to quickly find an n such that $\mathrm{disc}(T_n)$ is not divisible by p. This trick might be used to speed up the computation for some other levels. The key idea is to efficiently discover which T_n to compute. Computing T_n on the full space of modular symbols is difficult, but using projections we can compute T_n on subspaces of modular symbols with small dimension more quickly (see, e.g., [Ste00, §3.5.2]). Let M be the space of mod p modular symbols of level $p = 28279$, and let $f = \gcd(\mathrm{charpoly}(T_2), \mathrm{deriv}(\mathrm{charpoly}(T_2)))$. Let V be the kernel of $f(T_2)$ (this takes 7 minutes to compute). If $V = 0$, we would be done, since then $\mathrm{disc}(T_2) \neq 0 \in \mathbf{F}_p$. In fact, V has dimension 7. We find the first few integers n so that the charpoly of T_n on V has distinct roots, and they are $n = 34$, 47, 53, and 89. We then computed charpoly(T_{34}) directly on the whole space and found that it has distinct roots modulo p.

3.2 Some Data About Weight 4

The following are the valuations $d = d_4(\Gamma_0(p))$ at p of the discriminant of the Hecke algebras associated to $S_4(\Gamma_0(p))$ for $p < 500$. This data suggests a pattern, which motivates Conjecture 1 below.

p	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59
d	0	0	0	0	0	2	2	2	2	4	4	6	6	6	6	8	8
p	61	67	71	73	79	83	89	97	101	103	107	109	113	127	131	137	139
d	10	10	10	12	12	12	14	16	16	16	16	18	18	20	20	22	24
p	149	151	157	163	167	173	179	181	191	193	197	199	211	223	227	229	233
d	24	24	26	26	26	28	28	30	30	32	32	32	34	36	36	38	38
p	239	241	251	257	263	269	271	277	281	283	293	307	311	313	317	331	337
d	38	40	40	42	42	44	44	46	46	46	48	50	50	52	52	54	56
p	347	349	353	359	367	373	379	383	389	397	401	409	419	421	431	433	439
d	56	58	58	58	60	62	62	62	65	66	66	68	68	70	70	72	72
p	443	449	457	461	463	467	479	487	491	499							
d	72	74	76	76	76	76	78	80	80	82							

4 Speculations

Motivated by the promise of a pattern suggested by the table in Section 3.2, we computed $d_k(\Gamma_0(p))$ for many values of k and p. Our observations led us to the following results and conjectures.

Theorem 2. *Suppose p is a prime and $k \geq 4$ is an even integer. Then $d_k(\Gamma_0(p)) > 0$ unless*

$$\begin{aligned}(p,k) \in \{&(2,4),(2,6),(2,8),(2,10),\\ &(3,4),(3,6),(3,8),\\ &(5,4),(5,6),(7,4),(11,4)\},\end{aligned}$$

in which case $d_k(\Gamma_0(p)) = 0$.

Proof. From [Rib91], mod p eigenforms on $\Gamma_0(p)$ of weight k arise exactly from mod p eigenforms on $\Gamma_0(1)$ of weight $(k/2)(p+1)$. Moreover, there is an equality of dimensions of vector spaces:

$$\dim S_{(k/2)(p+1)}(\Gamma_0(1)) + \dim S_{(k/2)(p+1)-(p-1)}(\Gamma_0(1)) = \dim S_k(\Gamma_0(p)).$$

Thus the dimension of $S_k(\Gamma_0(p))$ is bigger than the number of mod p eigenforms whenever $\dim S_{(k/2)(p+1)-(p-1)}(\Gamma_0(1))$ is non-zero. The cases of dimension zero correspond exactly to the finite list of exceptions above, for which one can explicitly calculate that $d_k(\Gamma_0(p)) = 0$.

Note that for $k = 2$, however, there is a canonical identification of spaces

$$S_{(p+1)}(\mathit{\Gamma}_0(1), \overline{\mathbf{F}}_p) \simeq S_2(\mathit{\Gamma}_0(p), \overline{\mathbf{F}}_p),$$

described geometrically in [Gro90]. For $k = 4$, the data suggests that the discriminants $d_4(\mathit{\Gamma}_0(p))$ are significantly larger than zero for large p, and the table above suggests a formula of the form $2 \cdot \lfloor p/12 \rfloor$ (Not entirely coincidentally, this is the difference in dimension of the spaces $S_4(\mathit{\Gamma}_0(p))$ and $S_{2(p+1)}(\mathit{\Gamma}_0(1))$). This exact formula is not correct, however, as evidenced by the case when $p = 139$. If we consider the Hecke algebra $\mathbf{T}_4$ for $p = 139$ in more detail, however, we observe that $\mathbf{T}_4 \otimes \mathbf{Q}_{139}$ is *ramified* at 139, and in particular contains two copies of the field $\mathbf{Q}_{139}(\sqrt{139})$. Just as in the case when $k = 2$ and $p = 389$, there is a "self congruence" between the associated ramified eigenforms and their Galois conjugates. For all other p in the range of the table, there is no ramification, and all congruences take place between distinct eigenforms. Such congruences are measured by the *index* of the Hecke algebra, which is defined to be the index of $\mathbf{T}$ in its normalisation $\widetilde{\mathbf{T}}$. If we are only interested in mod p congruences (rather than mod ℓ congruences for $\ell \neq p$), one can restrict to the index of $\mathbf{T} \otimes \mathbf{Z}_p$ inside its normalisation. There is a direct relation between the discriminant and the index. Suppose that $\mathbf{T} \otimes \mathbf{Q}_p = \prod K_i$ for certain fields $K_i/\mathbf{Q}_p$ (We may assume here that $\mathbf{T}$ is not nilpotent, for otherwise both the discriminant and index are infinite). Then if $i_p(\mathit{\Gamma}) = \mathrm{ord}_p([\mathbf{T}, \widetilde{\mathbf{T}}])$, then

$$d_p(\mathit{\Gamma}) = 2 i_p(\mathit{\Gamma}) + \sum \mathrm{ord}_p(\mathit{\Delta}(K_i/\mathbf{Q}_p)).$$

If we now return to the example $k = 4$ and $p = 139$, we see that the discrepancy from the discriminant $d_p(\mathit{\Gamma}_0(139)) = 24$ to the estimate $2\lfloor 139/12 \rfloor = 22$ is exactly accounted for by the two eigenforms with coefficients in $\mathbf{Q}_{139}(\sqrt{139})$, which contribute 2 to the above formula. This leads us to predict that the index is exactly given by the formula $\lfloor p/12 \rfloor$. Note that for primes p this is exactly the dimension of $S_{p+3}(\mathit{\Gamma}_0(1))$. Similar computations lead to the following more general conjecture.

Let $k = 2m$ be an even integer and p a prime. Let $\mathbf{T}$ be the Hecke algebra associated to $S_k(\mathit{\Gamma}_0(p))$ and let $\widetilde{\mathbf{T}}$ be the integral closure of $\mathbf{T}$ in $\mathbf{T} \otimes \mathbf{Q}$ (which is a product of number fields).

Conjecture 1. Suppose $p \geq k - 1$. Then

$$\mathrm{ord}_p([\widetilde{\mathbf{T}} : \mathbf{T}]) = \left\lfloor \frac{p}{12} \right\rfloor \cdot \binom{m}{2} + a(p, m),$$

where

$$a(p,m) = \begin{cases} 0 & \text{if } p \equiv 1 \pmod{12}, \\ 3 \cdot \binom{\lceil \frac{m}{3} \rceil}{2} & \text{if } p \equiv 5 \pmod{12}, \\ 2 \cdot \binom{\lceil \frac{m}{2} \rceil}{2} & \text{if } p \equiv 7 \pmod{12}, \\ a(5,m) + a(7,m) & \text{if } p \equiv 11 \pmod{12}. \end{cases}$$

Here $\binom{x}{y}$ is the binomial coefficient "x choose y", and floor and ceiling are as usual. The conjecture is very false if $k \gg p$.

When $k = 2$, the conjecture specializes to the assertion that $[\widetilde{\mathbf{T}} : \mathbf{T}]$ is not divisible by p. A possibly more familiar concrete consequence of the conjecture is the following conjecture about elliptic curves. The modular degree of an elliptic curve E is the smallest degree of a surjective morphism $X_0(N) \to E$, where N is the conductor of E.

Conjecture 2. Suppose E is an elliptic curve of prime conductor p. Then p does not divide the modular degree m_E of E.

Using the algorithm in [Wat02], M. Watkins has computed modular degrees of a huge number of elliptic curves of prime conductor $p < 10^7$, and not found a counterexample. Looking at smaller data, there is only one elliptic curve E of prime conductor $p < 20000$ such that the modular degree of E is even as big as the conductor of E, and that is a curve of conductor 13723. This curve has equation $[1, 1, 1, -10481, 408636]$, modular degree $m_E = 16176 = 2^4 \cdot 3 \cdot 337$. The modular degree can be divisible by large primes. For example, there is a Neumann-Setzer elliptic curve of prime conductor 90687593 whose modular degree is 1280092043, which is over 14 times as big as 90687593. In general, for an elliptic curve of conductor N, one has the estimate $m_E \gg N^{7/6-\epsilon}$ (see [Wat04]).

5 Conjectures Inspired by Conjecture 1

First, some notation. Let p be an odd prime. Let $\mathit{\Gamma} = \mathit{\Gamma}_0(p)$, and let

$$S_k(R) := S_k(\mathit{\Gamma})^{\text{new}} \otimes R.$$

The spaces S_k carry an action of the Hecke algebra $\mathbf{T}_k^{\text{new}}$, and a Fricke involution w_p. If $\frac{1}{2} \in R$, the space S_k can be decomposed into $+$ and $-$ eigenspaces for w_p. We call the resulting spaces S_k^+ and S_k^- respectively.

Similarly, let M_k^+ and M_k^- be the $+1$ and -1 eigenspaces for w_p on the full spaces of new modular forms of weight k for $\mathit{\Gamma}_0(p)$.

It follows from [AL70, Lem. 7] (which is an explicit formula for the trace to lower level) and the fact that U_p and w_p both preserve the new subspace, that the action of the Hecke operator U_p on S_k is given by the formula

$$U_p = -p^{(k-2)/2} w_p.$$

This gives rise to two quotients of the Hecke algebra:

$$\mathbf{T}^+ = \mathbf{T}^{\mathrm{new}}/(U_p + p^{(k-2)/2}) \quad \text{and} \quad \mathbf{T}^- = \mathbf{T}^{\mathrm{new}}/(U_p - p^{(k-2)/2}).$$

where $\mathbf{T}^+$ and $\mathbf{T}^-$ act on S^+ and S^-, respectively. Recall that $\widetilde{\mathbf{T}}$ is the normalization (integral closure) of $\mathbf{T}$ in $\mathbf{T} \otimes \mathbf{Q}$. Let $\widetilde{\mathbf{T}}^{\mathrm{new}}$ denote the integral closure of $\mathbf{T}^{\mathrm{new}}$ in $\mathbf{T}^{\mathrm{new}} \otimes \mathbf{Q}$.

Lemma 1. *There are injections*

$$\mathbf{T}^{\mathrm{new}} \hookrightarrow \mathbf{T}^+ \oplus \mathbf{T}^- \hookrightarrow \widetilde{\mathbf{T}}^{\mathrm{new}}.$$

We now begin stating some conjectures regarding the rings $\mathbf{T}^{\pm}$.

Conjecture 3. Let $k < p - 1$. Then $\mathbf{T}^+$ and $\mathbf{T}^-$ are integrally closed. Equivalently, all congruences between distinct eigenforms in $S_k(\overline{\mathbf{Z}}_p)$ take place between $+$ and $-$ eigenforms.

Note that for $k = 2$, there cannot be any congruences between $+$ and $-$ forms because this would force $1 \equiv -1 \mod p$, which is false, because p is odd. Thus we recover the conjecture that $p \nmid [\widetilde{\mathbf{T}} : \mathbf{T}]$ when $k = 2$. Our further conjectures go on to describe explicitly the congruences between forms in S_k^+ and S_k^-.

Let E_2 be the non-holomorphic Eisenstein series of weight 2. The q-expansion of E_2 is given explicitly by

$$E_2 = 1 - 24 \sum_{n=1}^{\infty} q^n \left(\sum_{d|n} d \right).$$

Moreover, the function $E_2^* = E_2(\tau) - pE_2(p\tau)$ is holomorphic of weight 2 and level $\mathit{\Gamma}_0(p)$, and moreover on q-expansions, $E_2^* \equiv E_2 \mod p$.

Lemma 2. *Let $p > 3$. Let $f \in M_k(\mathit{\Gamma}_0(p), \overline{\mathbf{F}}_p)$ be a Hecke eigenform. Then θf is an eigenform inside $S_{k+2}(\mathit{\Gamma}_0(p), \overline{\mathbf{F}}_p)$.*

Proof. One knows that $\partial f = \theta f - kE_2 f/12$ is of weight $k+2$. On q-expansions, $E_2 \equiv E_2^* \mod p$, and thus for $p > 3$,

$$\theta f \equiv \partial f + kE_2^* f/12 \pmod{p}$$

is the reduction of a weight $k+2$ form of level p. It is easy to see that θf is a cuspidal Hecke eigenform.

Let us now assume Conjecture 3 and consider the implications for $k = 4$ in more detail. The space of modular forms $M_2(\Gamma_0(p), \overline{\mathbf{F}}_p)$ consists precisely of S_2 and the Eisenstein series E_2^*. The map θ defined above induces maps:

$$\theta : S_2^+(\overline{\mathbf{F}}_p) \to S_4(\overline{\mathbf{F}}_p), \qquad \theta : M_2^-(\overline{\mathbf{F}}_p) \to S_4(\overline{\mathbf{F}}_p).$$

The images are distinct, since $\theta f = \theta g$ implies (with some care about a_p) that $f = g$.

Conjecture 4. Let $f \in S_2(\overline{\mathbf{Z}}_p)$ and $g \in S_4(\overline{\mathbf{Z}}_p)$ be two eigenforms such that $\theta f \equiv g \mod p$. Then the eigenvalue of w_p on f and g have opposite signs.

Assuming this, we get inclusions:

$$\theta S_2^+(\overline{\mathbf{F}}_p) \hookrightarrow S_4^-(\overline{\mathbf{F}}_p), \qquad \theta M_2^-(\overline{\mathbf{F}}_p) \hookrightarrow S_4^+(\overline{\mathbf{F}}_p).$$

Now we are ready to state our main conjecture:

Conjecture 5. There is an Hecke equivariant exact sequence:

$$\begin{tikzcd} 0 \arrow[r] & \theta S_2^+(\overline{\mathbf{F}}_p) \arrow[r] & S_4^-(\overline{\mathbf{F}}_p) \arrow[r] & S_4^+(\overline{\mathbf{F}}_p) \arrow[r] & \theta M_2^-(\overline{\mathbf{F}}_p) \arrow[r] & 0. \end{tikzcd}$$

Moreover, the map $S_4^-(\overline{\mathbf{F}}_p) \to S_4^+(\overline{\mathbf{F}}_p)$ here is the largest such equivariant map between these spaces. Equivalently, a residual eigenform of weight 4 and level p occurs in both the $+$ and $-$ spaces if and only if it is not in the image of θ.

Let us give some consequences of our conjectures for the index of $\mathbf{T}^{\mathrm{new}}$ inside its normalisation. Fix a residual representation $\overline{\rho} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F}_q)$ and consider the associated maximal ideal $\mathfrak{m}$ inside $\mathbf{T}_4$. If $\overline{\rho}$ lies in the image of θ then our conjecture implies that it is not congruent to any other eigenform. If $\overline{\rho}$ is not in the image of θ, then it should arise exactly from a pair of eigenforms, one inside $S_4^+(\overline{\mathbf{Q}}_p)$ and one inside $S_4^-(\overline{\mathbf{Q}}_p)$. Suppose that $q = p^r$. If there is no ramification in $\mathbf{T} \otimes \mathbf{Q}$ over

p (this is often true), then the $+$ and $-$ eigenforms will both be defined over the ring $W(\mathbf{F}_q)$ of Witt vectors of $\mathbf{F}_q$. Since $U_p = p$ on S_4^- and $-p$ on S_4^+, these forms can be at most congruent modulo p. Thus the completed Hecke algebra $(\mathbf{T}_4)_{\mathfrak{m}}$ is exactly

$$\{(a,b) \in W(\mathbf{F}_q) \oplus W(\mathbf{F}_q), |a \equiv b \mod p\}.$$

One sees that this has index $q = p^r$ inside its normalisation. Thus the (log of the) total index is equal to $\sum r_i$ over all eigenforms that occur inside S_4^+ and S_4^-, which from our exact sequence we see is equal to

$$\dim S_4^- - \dim S_2^+.$$

Conjecture 1 when $k = 4$, would then follow from the equality of dimensions:

$$\dim S_4^-(\overline{\mathbf{F}}_p) - \dim S_2^+(\overline{\mathbf{F}}_p) = \left\lfloor \frac{p}{12} \right\rfloor.$$

We expect that something similar, but a little more complicated, should happen in general. In weight $2k$, there are mod p^{k-r} congruences exactly between forms in the image of θ^{r-1} but not of θ^r.

5.1 Examples

We write small s's and m's for dimensions below.

Let $p = 101$. Then $s_2^+ = 1$, $m_2^- = 7 + 1 = 8$, $s_4^- = 9$, $s_4^+ = 16$. We predict the index should be $9 - 1 = 8 = \lfloor 101/12 \rfloor$. In the table below, we show the characteristic polynomials of T_2 on S_4^- and S_4^+, and for weight 2, we take the characteristic polynomial of θT_2 (or the same, taking $F(x/2)$ where $F(x)$ is the characteristic polynomial of T_2). Note that we have to add the Eisenstein series, which has characteristic polynomial $x - 1 - 2$, which becomes $x - 6 \equiv x + 95 \mod 101$ under θ.

Factors of the Characteristic Polynomial of T_2 for $p = 101$.

$\theta S_2^+(\overline{\mathbf{F}}_{101})$	$S_4^-(\overline{\mathbf{F}}_{101})$	$S_4^+(\overline{\mathbf{F}}_{101})$	$\theta M_2^-(\overline{\mathbf{F}}_{101})$
(x)	(x) $(x+46)$ $(x^2+58x+100)$ $(x^5+2x^4+27x^3$ $+49x^2+7x+65)$	$(x+46)$ $(x+95)$ $(x^2+58x+100)$ $(x^2+90x+78)$ $(x^2+96x+36)$ $(x^3+16x^2+35x+72)$ $(x^5+2x^4+27x^3$ $+49x^2+7x+65)$	$(x+95)$ $(x^2+90x+78)$ $(x^2+96x+36)$ $(x^3+16x^2$ $+35x+72)$

Here are some further conjectures when $k > 4$.

Conjecture 6. Let p and k be such that $4 < k < p-1$. There is an Hecke equivariant exact sequence:

$$0 \longrightarrow \theta S^+_{k-2}(\overline{\mathbf{F}}_p) \longrightarrow S^-_k(\overline{\mathbf{F}}_p) \longrightarrow S^+_k(\overline{\mathbf{F}}_p) \longrightarrow \theta S^-_{k-2}(\overline{\mathbf{F}}_p) \longrightarrow 0.$$

Moreover, all forms not in the image of θ contribute maximally to the index (a factor of $p^{(k-2)/2}$). Thus the total index should be equal to

$$\frac{(k-2)}{2}(\dim S^+_k - \dim S^-_{k-2}) \quad + \quad \text{the index at level } p \text{ and weight } k-2.$$

This is the sum

$$\sum_{n=2}^{k} \frac{(2n-2)}{2}(s^+_{2n} - s^-_{2n-2}).$$

When $k = 4$, we need to add the Eisenstein series to S^-_2 in our previous conjecture. Note that $s^+_k - s^-_{k-2} = s^-_k - s^+_{k-2}$ for $k > 4$ (and with s^-_2 replaced by m^-_2 when $k = 2$). This follows from our conjectures, but can easily be proved directly. As an example, when $p = 101$, we have $s^+_2 = 1$, $s^-_4 = 9$, $s^+_6 = 17$, $s^-_8 = 26$, $s^+_{10} = 34$, $s^-_{12} = 42$, $s^+_{14} = 51$, and so we would predict the indexes I_k to be as given in the following table:

k	I_k?
2	0
4	$8 = 8 + 0$
6	$24 = 24 + 0$
8	$51 = 48 + 3$
10	$83 = 80 + 3$
12	$123 = 120 + 3$
14	$177 = 168 + 9$

This agrees with our conjectural formula, which says that the index should be equal in this case to

$$8\binom{k/2}{2} + 3\binom{\lceil k/6 \rceil}{2}.$$

it also agrees with computation.

References

[AL70] A. O. L. Atkin and J. Lehner, *Hecke operators on $\Gamma_0(m)$*, Math. Ann. **185** (1970), 134–160.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[CE98] R. F. Coleman and B. Edixhoven, *On the semi-simplicity of the U_p-operator on modular forms*, Math. Ann. **310** (1998), no. 1, 119–127. MR 99b:11043

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[Gro90] B. H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. 61 (1990), no. 2, 445–517. MR 1 074 305

[LS02] J.-C. Lario and R. Schoof, *Some computations with Hecke rings and deformation rings*, Experiment. Math. **11** (2002), no. 2, 303–311, With an appendix by Amod Agashe and William Stein. MR 2004b:11072

[Mat86] H. Matsumura, *Commutative ring theory*, Cambridge University Press, Cambridge, 1986, Translated from the Japanese by M. Reid. MR 88h:13001 ed., Cambridge University Press, Cambridge, 1997.

[Mer94] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations, Springer, 1994, pp. 59–94.

[Mes86] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242.

[MO02] F. Momose and S. Ozawa, *Rational points of modular curves $X_{\mathrm{split}}(p)$*, Preprint (2002).

[MS01] L. Merel and W. A. Stein, *The field generated by the points of small prime order on an elliptic curve*, Internat. Math. Res. Notices (2001), no. 20, 1075–1082. MR 1 857 596

[OW02] K. Ono and W. McGraw, *Modular form congruences and selmer groups*, Preprint (2002).

[Rib91] K. A. Ribet, *Letter to Gerd Faltings*, Unpublished (1991).

[Rib99] K. A. Ribet, *Torsion points on $J_0(N)$ and Galois representations*, Arithmetic theory of elliptic curves (Cetraro, 1997), Springer, Berlin, 1999, pp. 145–166. MR 2001b:11054

[Ste00] W. A. Stein, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley (2000).

[Wat02] M. Watkins, *Computing the modular degree of an elliptic curve*, Experiment. Math. **11** (2002), no. 4, 487–502 (2003). MR 1 969 641

[Wat04] M. Watkins, *Explicit lower bounds on the modular degree of an elliptic curve*, Preprint (2004).
`www.mathpreprints.com/math/Preprint/Gerund2/20040320/1/dp.pdf`

22 Book – A Brief Introduction to Classical and Adelic Algebraic Number Theory

A Brief Introduction to Classical and Adelic Algebraic Number Theory

William Stein
(based heavily on works of Swinnerton-Dyer and Cassels)

May 2004

Contents

Chapter 1

Preface

This book is based on notes I created for a one-semester undergraduate course on Algebraic Number Theory, which I taught at Harvard during Spring 2004. The primary sources for the course were chapter 1 of Swinnerton-Dyer's book *A Brief Guide to Algebraic Number Theory* [SD01] and chapter 2 of Cassels's article *Global Fields* [Cas67]. I wrote these notes by following closely the above two chapters; in some cases I added substantial text and examples. For example, chapter 1 of [SD01] is 30 pages, whereas my rewrite of it occupies over 100 pages. In contrast, I follow [Cas67] more closely. *I have no intent whatever to plagiarize. I acknowledge as such those chapters in this book which are simply a close rewrite of [Cas67].* My goal is to take the useful classical article ([Cas67]) and make it more accessible to students by modernizing the notation, and adding additional explanations and examples.

I have no intent to publish this book with a traditional publisher, so it will remain freely available indefinitely. If you have comments, corrections, suggestions for additions, etc., please send them to me!

Copyright: William Stein, 2004.

License: FREE! More precisely, this book my be freely redistributed, copied, or even sold without requiring you to obtain written permission from me. You may even extend or change this book, but this preface page must remain in any derived work, and any derived work must also remain free, including the LaTeX source files. In particular, I have no interest in making any money from this book.

Please send me any typos or corrections: `was@math.harvard.edu`.

Acknowledgement: This book closely builds on Swinnerton-Dyer's book [SD01] and Cassels's article [Cas67]. Many of the students of Math 129 at Harvard during Spring 2004 made helpful comments: Jennifer Balakrishnan, Peter Behrooz, Jonathan Bloom, David Escott Jayce Getz, Michael Hamburg, Deniz Kural, Danielle li, Andrew Ostergaard Gregory Price, Grant Schoenebeck, Jennifer Sinnott, Stephen Walker, Daniel Weissman, and Inna Zakharevich. Also the course assistant Matt Bainbridge made many helpful comments.

Chapter 2

Introduction

2.1 Mathematical background I assume you have

In addition to general mathematical maturity, this book assumes you have the following background:

- Basics of finite group theory
- Commutative rings, ideals, quotient rings
- Some elementary number theory
- Basic Galois theory of fields
- Point set topology
- Basic of topological rings, groups, and measure theory

For example, if you have never worked with finite groups before, you should read another book first. If you haven't seen much elementary ring theory, there is still hope, but you will have to do some additional reading and exercises. I will briefly review the basics of the Galois theory of number fields.

Some of the homework problems involve using a computer, but I'll give you examples which you can build on. I will not assume that you have a programming background or know much about algorithms. If you don't have PARI [ABC$^+$] or Magma [BCP97], and don't want to install either one on your computer, you might want to try the following online interface to PARI and Magma:

`http://modular.fas.harvard.edu/calc/`

2.2 What is algebraic number theory?

A *number field* K is a finite algebraic extension of the rational numbers $\mathbf{Q}$. Every such extension can be represented as all polynomials in an algebraic number α:

$$K = \mathbf{Q}(\alpha) = \left\{ \sum_{n=0}^{m} a_n \alpha^n : a_n \in \mathbf{Q} \right\}.$$

Here α is a root of a polynomial with coefficients in $\mathbf{Q}$.

Algebraic number theory involves using techniques from (mostly commutative) algebra and finite group theory to gain a deeper understanding of number fields. The main objects that we study in algebraic number theory are number fields, rings of integers of number fields, unit groups, ideal class groups,norms, traces, discriminants, prime ideals, Hilbert and other class fields and associated reciprocity laws, zeta and L-functions, and algorithms for computing each of the above.

2.2.1 Topics in this book

These are some of the main topics that are discussed in this book:

- Rings of integers of number fields
- Unique factorization of ideals in Dedekind domains
- Structure of the group of units of the ring of integers
- Finiteness of the group of equivalence classes of ideals of the ring of integers (the "class group")
- Decomposition and inertia groups, Frobenius elements
- Ramification
- Discriminant and different
- Quadratic and biquadratic fields
- Cyclotomic fields (and applications)
- How to use a computer to compute with many of the above objects (both algorithms and actual use of PARI and MAGMA).
- Valuations on fields
- Completions (p-adic fields)
- Adeles and Ideles

Note that we will not do anything nontrivial with zeta functions or L-functions. This is to keep the prerequisites to algebra, and so we will have more time to discuss algorithmic questions. Depending on time and your inclination, I may also talk about integer factorization, primality testing, or complex multiplication elliptic curves (which are closely related to quadratic imaginary fields).

2.3 Some applications of algebraic number theory

The following examples are meant to convince you that learning algebraic number theory now will be an excellent investment of your time. If an example below seems vague to you, it is safe to ignore it.

1. **Integer factorization** using the number field sieve. The number field sieve is the asymptotically fastest known algorithm for factoring general large integers (that don't have too special of a form). Recently, in December 2003, the number field sieve was used to factor the RSA-576 \$10000 challenge:

$$\begin{array}{l}
188198812920607963838697239461650439807163563379417382700\ldots \\
\ldots 633564229888597152346654853190606065047430453173880113033\ldots \\
\ldots 67161996923212057340318795506569962213051687593076502570\\
= 398075086424064937397125500550386491199064362342526708\ldots \\
\quad \ldots 638518957594638895726176858331 \\
\qquad \times 4727721461074353025362230719730482246329146953020971\ldots \\
\qquad\quad \ldots 6459852171130520711256363590397527
\end{array}$$

   (The ... indicates that the newline should be removed, not that there are missing digits.) For more information on the NFS, see the paper by Lenstra et al. on the Math 129 web page.

2. **Primality test:** Agrawal and his students Saxena and Kayal from India recently (2002) found the first ever deterministic polynomial-time (in the number of digits) primality test. There methods involve arithmetic in quotients of $(\mathbf{Z}/n\mathbf{Z})[x]$, which are best understood in the context of algebraic number theory. For example, Lenstra, Bernstein, and others have done that and improved the algorithm significantly.

3. **Deeper point of view** on questions in number theory:

   (a) Pell's Equation ($x^2 - dy^2 = 1$) $\Longrightarrow$ Units in real quadratic fields $\Longrightarrow$ Unit groups in number fields

   (b) Diophantine Equations $\Longrightarrow$ For which n does $x^n + y^n = z^n$ have a nontrivial solution in $\mathbf{Q}(\sqrt{2})$?

   (c) Integer Factorization $\Longrightarrow$ Factorization of ideals

   (d) Riemann Hypothesis $\Longrightarrow$ Generalized Riemann Hypothesis

(e) Deeper proof of Gauss's quadratic reciprocity law in terms of arithmetic of cyclotomic fields $\mathbf{Q}(e^{2\pi i/n})$, which leads to class field theory.

4. Wiles's proof of **Fermat's Last Theorem**, i.e., $x^n+y^n = z^n$ has no nontrivial integer solutions, uses methods from algebraic number theory extensively (in addition to many other deep techniques). Attempts to prove Fermat's Last Theorem long ago were hugely influential in the development of algebraic number theory (by Dedekind, Kummer, Kronecker, et al.).

5. **Arithmetic geometry:** This is a huge field that studies solutions to polynomial equations that lie in arithmetically interesting rings, such as the integers or number fields. A famous major triumph of arithmetic geometry is Faltings's proof of Mordell's Conjecture.

   Theorem 2.3.1 (Faltings). *Let X be a plane algebraic curve over a number field K. Assume that the manifold $X(\mathbf{C})$ of complex solutions to X has genus at least 2 (i.e., $X(\mathbf{C})$ is topologically a donut with two holes). Then the set $X(K)$ of points on X with coordinates in K is finite.*

   For example, Theorem 2.3.1 implies that for any $n \geq 4$ and any number field K, there are only finitely many solutions in K to $x^n + y^n = 1$. A famous open problem in arithmetic geometry is the **Birch and Swinnerton-Dyer conjecture**, which gives a deep conjectural criterion for exactly when $X(K)$ should be infinite when $X(\mathbf{C})$ is a torus.

Part I

Classical Viewpoint

Chapter 3

Finitely generated abelian groups

We will now prove the structure theorem for finitely generated abelian groups, since it will be crucial for much of what we will do later.

Let $\mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$ denote the ring of integers, and for each positive integer n let $\mathbf{Z}/n\mathbf{Z}$ denote the ring of integers modulo n, which is a cyclic abelian group of order n under addition.

Definition 3.0.2 (Finitely Generated). A group G is *finitely generated* if there exists $g_1, \ldots, g_n \in G$ such that every element of G can be obtained from the g_i.

Theorem 3.0.3 (Structure Theorem for Abelian Groups). *Let G be a finitely generated abelian group. Then there is an isomorphism*

$$G \cong (\mathbf{Z}/n_1\mathbf{Z}) \oplus (\mathbf{Z}/n_2\mathbf{Z}) \oplus \cdots \oplus (\mathbf{Z}/n_s\mathbf{Z}) \oplus \mathbf{Z}^r,$$

where $n_1 > 1$ and $n_1 \mid n_2 \mid \cdots \mid n_s$. Furthermore, the n_i and r are uniquely determined by G.

We will prove the theorem as follows. We first remark that any subgroup of a finitely generated free abelian group is finitely generated. Then we see that finitely generated abelian groups can be presented as quotients of finite rank free abelian groups, and such a presentation can be reinterpreted in terms of matrices over the integers. Next we describe how to use row and column operations over the integers to show that every matrix over the integers is equivalent to one in a canonical diagonal form, called the Smith normal form. We obtain a proof of the theorem by reinterpreting Smith normal form in terms of groups.

Proposition 3.0.4. *Suppose G is a free abelian group of finite rank n, and H is a subgroup of G. Then H is a free abelian group generated by at most n elements.*

The key reason that this is true is that G is a finitely generated module over the principal ideal domain $\mathbf{Z}$. We will give a complete proof of a beautiful generalization of this result in the context of Noetherian rings next time, but will not prove this proposition here.

Corollary 3.0.5. *Suppose G is a finitely generated abelian group. Then there are finitely generated free abelian groups F_1 and F_2 such that $G \cong F_1/F_2$.*

Proof. Let $x_1, \ldots, x_m$ be generators for G. Let $F_1 = \mathbf{Z}^m$ and let $\varphi : F_1 \to G$ be the map that sends the ith generator $(0, 0, \ldots, 1, \ldots, 0)$ of $\mathbf{Z}^m$ to x_i. Then φ is a surjective homomorphism, and by Proposition 3.0.4 the kernel F_2 of φ is a finitely generated free abelian group. This proves the corollary. □

Suppose G is a nonzero finitely generated abelian group. By the corollary, there are free abelian groups F_1 and F_2 such that $G \cong F_1/F_2$. Choosing a basis for F_1, we obtain an isomorphism $F_1 \cong \mathbf{Z}^n$, for some positive integer n. By Proposition 3.0.4, $F_2 \cong \mathbf{Z}^m$, for some integer m with $0 \leq m \leq n$, and the inclusion map $F_2 \hookrightarrow F_1$ induces a map $\mathbf{Z}^m \to \mathbf{Z}^n$. This homomorphism is left multiplication by the $n \times m$ matrix A whose columns are the images of the generators of F_2 in $\mathbf{Z}^n$. The *cokernel* of this homomorphism is the quotient of $\mathbf{Z}^n$ by the image of A, and the cokernel is isomorphic to G. By augmenting A with zero columns on the right we obtain a square $n \times n$ matrix A with the same cokernel. The following proposition implies that we may choose bases such that the matrix A is diagonal, and then the structure of the cokernel of A will be easy to understand.

Proposition 3.0.6 (Smith normal form). *Suppose A is an $n \times n$ integer matrix. Then there exist invertible integer matrices P and Q such that $A' = PAQ$ is a diagonal matrix with entries $n_1, n_2, \ldots, n_s, 0, \ldots, 0$, where $n_1 > 1$ and $n_1 \mid n_2 \mid \ldots \mid n_s$. This is called the Smith normal form of A.*

We will see in the proof of Theorem 3.0.3 that A' is uniquely determined by A.

Proof. The matrix P will be a product of matrices that define elementary row operations and Q will be a product corresponding to elementary column operations. The elementary operations are:

1. Add an integer multiple of one row to another (or a multiple of one column to another).

2. Interchange two rows or two columns.

3. Multiply a row by -1.

Each of these operations is given by left or right multiplying by an invertible matrix E with integer entries, where E is the result of applying the given operation to the identity matrix, and E is invertible because each operation can be reversed using another row or column operation over the integers.

To see that the proposition must be true, assume $A \neq 0$ and perform the following steps (compare [Art91, pg. 459]):

1. By permuting rows and columns, move a nonzero entry of A with smallest absolute value to the upper left corner of A. Now attempt to make all other entries in the first row and column 0 by adding multiples of row or column 1 to other rows (see step 2 below). If an operation produces a nonzero entry in the matrix with absolute value smaller than $|a_{11}|$, start the process over by permuting rows and columns to move that entry to the upper left corner of A. Since the integers $|a_{11}|$ are a decreasing sequence of positive integers, we will not have to move an entry to the upper left corner infinitely often.

2. Suppose a_{i1} is a nonzero entry in the first column, with $i > 1$. Using the division algorithm, write $a_{i1} = a_{11}q + r$, with $0 \leq r < a_{11}$. Now add $-q$ times the first row to the ith row. If $r > 0$, then go to step 1 (so that an entry with absolute value at most r is the upper left corner). Since we will only perform step 1 finitely many times, we may assume $r = 0$. Repeating this procedure we set all entries in the first column (except a_{11}) to 0. A similar process using column operations sets each entry in the first row (except a_{11}) to 0.

3. We may now assume that a_{11} is the only nonzero entry in the first row and column. If some entry a_{ij} of A is not divisible by a_{11}, add the column of A containing a_{ij} to the first column, thus producing an entry in the first column that is nonzero. When we perform step 2, the remainder r will be greater than 0. Permuting rows and columns results in a smaller $|a_{11}|$. Since $|a_{11}|$ can only shrink finitely many times, eventually we will get to a point where every a_{ij} is divisible by a_{11}. If a_{11} is negative, multiple the first row by -1.

After performing the above operations, the first row and column of A are zero except for a_{11} which is positive and divides all other entries of A. We repeat the above steps for the matrix B obtained from A by deleting the first row and column. The upper left entry of the resulting matrix will be divisible by a_{11}, since every entry of B is. Repeating the argument inductively proves the proposition. □

Example 3.0.7. The matrix $\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}$ is equivalent to $\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}$ and the matrix $\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{pmatrix}$ is equivalent to $\begin{pmatrix}1 & 0 & 0\\0 & 3 & 0\\0 & 0 & 0\end{pmatrix}$. Note that the determinants match, up to sign.

Theorem 3.0.3. Suppose G is a finitely generated abelian group, which we may assume is nonzero. As in the paragraph before Proposition 3.0.6, we use Corollary 3.0.5 to write G as a the cokernel of an $n \times n$ integer matrix A. By Proposition 3.0.6 there are isomorphisms $Q : \mathbf{Z}^n \to \mathbf{Z}^n$ and $P : \mathbf{Z}^n \to \mathbf{Z}^n$ such that $A' = PAQ$ is a diagonal matrix with entries $n_1, n_2, \ldots, n_s, 0, \ldots, 0$, where $n_1 > 1$ and $n_1 \mid n_2 \mid \ldots \mid n_s$. Then G is isomorphic to the cokernel of the diagonal matrix A', so

$$G \cong (\mathbf{Z}/n_1\mathbf{Z}) \oplus (\mathbf{Z}/n_2\mathbf{Z}) \oplus \cdots \oplus (\mathbf{Z}/n_s\mathbf{Z}) \oplus \mathbf{Z}^r, \qquad (3.0.1)$$

as claimed. The n_i are determined by G, because n_i is the smallest positive integer n such that nG requires at most $s + r - i$ generators (we see from the representation (3.0.1) of G as a product that n_i has this property and that no smaller positive integer does).

□

Chapter 4

Commutative Algebra

We will do some serious commutative algebra in this chapter, which will provide a powerful algebraic foundation for understanding the more refined number-theoretic structures associated to number fields.

In the first section we establish the standard properties of Noetherian rings and modules, including the Hilbert basis theorem. We also observe that finitely generated abelian groups are Noetherian **Z**-modules, which fills the gap in our proof of the structure theorem for finitely generated abelian groups. After establishing properties of Noetherian rings, we consider the rings of algebraic integers and discuss some of their properties.

4.1 Noetherian Rings and Modules

Let R be a commutative ring with unit element. We will frequently work with R-modules, which are like vector spaces but over a ring. More precisely, recall that an *R-module* is an additive abelian group M equipped with a map $R \times M \to M$ such that for all $r, r' \in R$ and all $m, m' \in M$ we have $(rr')m = r(r'm)$, $(r + r')m = rm + r'm$, $r(m + m') = rm + rm'$, and $1m = m$. A *submodule* is a subgroup of M that is preserved by the action of R.

Example 4.1.1. The set of abelian groups are in natural bijection with **Z**-modules.

A *homomorphism* of R-modules $\varphi : M \to N$ is a abelian group homomorphism such that for any $r \in R$ and $m \in M$ we have $\varphi(rm) = r\varphi(m)$. A *short exact sequence* of R-modules

$$0 \to L \xrightarrow{f} M \xrightarrow{g} N \to 0$$

is a specific choice of injective homomorphism $f : L \to M$ and a surjective homomorphism $g : M \to N$ such that $\mathrm{im}(f) = \ker(g)$.

Definition 4.1.2 (Noetherian). An R-module M is *Noetherian* if every submodule of M is finitely generated. A ring R is *Noetherian* if R is Noetherian as a module over itself, i.e., if every ideal of R is finitely generated.

Notice that any submodule M' of M is Noetherian, because if every submodule of M is finitely generated then so is every submodule of M', since submodules of M' are also submodules of M.

Definition 4.1.3 (Ascending chain condition). An R-module M satisfies the *ascending chain condition* if every sequences $M_1 \subset M_2 \subset M_3 \subset \cdots$ of submodules of M eventually stabilizes, i.e., there is some n such that $M_n = M_{n+1} = M_{n+2} = \cdots$.

Proposition 4.1.4. *If M is an R-module, then the following are equivalent:*

1. *M is Noetherian,*
2. *M satisfies the ascending chain condition, and*
3. *Every nonempty set of submodules of M contains at least one maximal element.*

Proof. $1 \implies 2$**:** Suppose $M_1 \subset M_2 \subset \cdots$ is a sequence of submodules of M. Then $M_\infty = \cup_{n=1}^\infty M_n$ is a submodule of M. Since M is Noetherian, there is a finite set $a_1, \ldots, a_m$ of generators for M. Each a_i must be contained in some M_j, so there is an n such that $a_1, \ldots, a_m \in M_n$. But then $M_k = M_n$ for all $k \geq n$, which proves that the ascending chain condition holds for M.
$2 \implies 3$**:** Suppose 3 were false, so there exists a nonempty set S of submodules of M that does not contain a maximal element. We will use S to construct an infinite ascending chain of submodules of M that does not stabilize. Note that S is infinite, otherwise it would contain a maximal element. Let M_1 be any element of S. Then there is an M_2 in S that contains M_1, otherwise S would contain the maximal element M_1. Continuing inductively in this way we find an M_3 in S that properly contains M_2, etc., and we produce an infinite ascending chain of submodules of M, which contradicts the ascending chain condition.
$3 \implies 1$**:** Suppose 1 is false, so there is a submodule M' of M that is not finitely generated. We will show that the set S of all finitely generated submodules of M' does not have a maximal element, which will be a contradiction. Suppose S does have a maximal element L. Since L is finitely generated and $L \subset M'$, and M' is not finitely generated, there is an $a \in M'$ such that $a \notin L$. Then $L' = L + Ra$ is an element of S that strictly contains the presumed maximal element L, a contradiction. $\square$

Lemma 4.1.5. *If*

$$0 \to L \xrightarrow{f} M \xrightarrow{g} N \to 0$$

is a short exact sequence of R-modules, then M is Noetherian if and only if both L and N are Noetherian.

Proof. First suppose that M is Noetherian. Then L is a submodule of M, so L is Noetherian. If N' is a submodule of N, then the inverse image of N' in M is a submodule of M, so it is finitely generated, hence its image N' is finitely generated. Thus N is Noetherian as well.

Next assume nothing about M, but suppose that both L and N are Noetherian. If M' is a submodule of M, then $M_0 = \varphi(L) \cap M'$ is isomorphic to a submodule of the Noetherian module L, so M_0 is generated by finitely many elements $a_1, \ldots, a_n$. The quotient M'/M_0 is isomorphic (via g) to a submodule of the Noetherian module N, so M'/M_0 is generated by finitely many elements $b_1, \ldots, b_m$. For each $i \leq m$, let c_i be a lift of b_i to M', modulo M_0. Then the elements $a_1, \ldots, a_n, c_1, \ldots, c_m$ generate M', for if $x \in M'$, then there is some element $y \in M_0$ such that $x - y$ is an R-linear combination of the c_i, and y is an R-linear combination of the a_i. □

Proposition 4.1.6. *Suppose R is a Noetherian ring. Then an R-module M is Noetherian if and only if it is finitely generated.*

Proof. If M is Noetherian then every submodule of M is finitely generated so M is finitely generated. Conversely, suppose M is finitely generated, say by elements $a_1, \ldots, a_n$. Then there is a surjective homomorphism from $R^n = R \oplus \cdots \oplus R$ to M that sends $(0, \ldots, 0, 1, 0, \ldots, 0)$ (1 in ith factor) to a_i. Using Lemma 4.1.5 and exact sequences of R-modules such as $0 \to R \to R \oplus R \to R \to 0$, we see inductively that R^n is Noetherian. Again by Lemma 4.1.5, homomorphic images of Noetherian modules are Noetherian, so M is Noetherian. □

Lemma 4.1.7. *Suppose $\varphi : R \to S$ is a surjective homomorphism of rings and R is Noetherian. Then S is Noetherian.*

Proof. The kernel of φ is an ideal I in R, and we have an exact sequence

$$0 \to I \to R \to S \to 0$$

with R Noetherian. By Lemma 4.1.5, it follows that S is a Noetherian R-modules. Suppose J is an ideal of S. Since J is an R-submodule of S, if we view J as an R-module, then J is finitely generated. Since R acts on J through S, the R-generators of J are also S-generators of J, so J is finitely generated as an ideal. Thus S is Noetherian. □

Theorem 4.1.8 (Hilbert Basis Theorem). *If R is a Noetherian ring and S is finitely generated as a ring over R, then S is Noetherian. In particular, for any n the polynomial ring $R[x_1, \ldots, x_n]$ and any of its quotients are Noetherian.*

Proof. Assume first that we have already shown that for any n the polynomial ring $R[x_1, \ldots, x_n]$ is Noetherian. Suppose S is finitely generated as a ring over R, so there are generators $s_1, \ldots, s_n$ for S. Then the map $x_i \mapsto s_i$ extends uniquely to a surjective homomorphism $\pi : R[x_1, \ldots, x_n] \to S$, and Lemma 4.1.7 implies that S is Noetherian.

The rings $R[x_1, \ldots, x_n]$ and $(R[x_1, \ldots, x_{n-1}])[x_n]$ are isomorphic, so it suffices to prove that if R is Noetherian then $R[x]$ is also Noetherian. (Our proof follows

[Art91, §12.5].) Thus suppose I is an ideal of $R[x]$ and that R is Noetherian. We will show that I is finitely generated.

Let A be the set of leading coefficients of polynomials in I along with 0. If $a, b \in A$ are nonzero with $a + b \neq 0$, then there are polynomials f and g in I with leading coefficients a and b. If $\deg(f) \leq \deg(g)$, then $a + b$ is the leading coefficient of $x^{\deg(g)-\deg(f)}f + g$, so $a + b \in A$. If $r \in R$ and $a \in A$ with $ra \neq 0$, then ra is the leading coefficient of rf, so $ra \in A$. Thus A is an ideal in R, so since R is Noetherian there exists $a_1, \ldots, a_n$ that generate A as an ideal. Since A is the set of leading coefficients of elements of I, and the a_j are in I, we can choose for each $j \leq n$ an element $f_j \in I$ with leading coefficient a_j. By multipying the f_j by some power of x, we may assume that the f_j all have the same degree d.

Let $S_{<d}$ be the set of elements of I that have degree strictly less than d. This set is closed under addition and under multiplication by elements of R, so $S_{<d}$ is a module over R. The module $S_{<d}$ is submodule of the R-module of polynomials of degree less than n, which is Noetherian because it is generated by $1, x, \ldots, x^{n-1}$. Thus $S_{<d}$ is finitely generated, and we may choose generators $h_1, \ldots, h_m$ for $S_{<d}$.

Suppose $g \in I$ is an arbitrary element. We will show by induction on the degree of g that g is an $R[x]$-linear combination of $f_1, \ldots, f_n, h_1, \ldots h_m$. Thus suppose this statement is true for all elements of I of degree less than the degree of g. If the degree of g is less than d, then $g \in S_{<d}$, so g is in the $R[x]$-ideal generated by $h_1, \ldots, h_m$. Next suppose that g has degree $e \geq d$. Then the leading coefficient b of g lies in the ideal A of leading coefficients of g, so there exist $r_i \in R$ such that $b = r_1 a_1 + \cdots + r_n a_n$. Since f_i has leading coefficient a_i, the difference $g - x^{e-d} r_i f_i$ has degree less than the degree e of g. By induction $g - x^{e-d} r_i f_i$ is an $R[x]$ linear combination of $f_1, \ldots, f_n, h_1, \ldots h_m$, so g is also an $R[x]$ linear combination of $f_1, \ldots, f_n, h_1, \ldots h_m$. Since each f_i and h_j lies in I, it follows that I is generated by $f_1, \ldots, f_n, h_1, \ldots h_m$, so I is finitely generated, as required. □

Properties of Noetherian rings and modules will be crucial in the rest of this course. We have proved above that Noetherian rings have many desirable properties.

4.1.1 Z is Noetherian

The ring $\mathbf{Z}$ of integers is Noetherian because every ideal of $\mathbf{Z}$ is generated by one element.

Proposition 4.1.9. *Every ideal of the ring* $\mathbf{Z}$ *of integers is principal.*

Proof. Suppose I is a nonzero ideal in $\mathbf{Z}$. Let d the least positive element of I. Suppose that $a \in I$ is any nonzero element of I. Using the division algorithm, write $a = dq + r$, where q is an integer and $0 \leq r < d$. We have $r = a - dq \in I$ and $r < d$, so our assumption that d is minimal implies that $r = 0$, so $a = dq$ is in the ideal generated by d. Thus I is the principal ideal generated by d. □

Proposition 4.1.6 and 4.1.9 together imply that any finitely generated abelian group is Noetherian. This means that subgroups of finitely generated abelian groups are finitely generated, which provides the missing step in our proof of the structure theorem for finitely generated abelian groups.

Chapter 5

Rings of Algebraic Integers

In this chapter we will learn about rings of algebraic integers and discuss some of their properties. We will prove that the ring of integers $\mathcal{O}_K$ of a number field is Noetherian. We will also develop some basic properties of norms, traces, and discriminants, and give more properties of rings of integers in the general context of Dedekind domains.

5.1 Rings of Algebraic Integers

Fix an algebraic closure $\overline{\mathbf{Q}}$ of $\mathbf{Q}$. For example, $\overline{\mathbf{Q}}$ could be the subfield of the complex numbers $\mathbf{C}$ generated by all roots in $\mathbf{C}$ of all polynomials with coefficients in $\mathbf{Q}$.

Much of this course is about algebraic integers.

Definition 5.1.1 (Algebraic Integer). An element $\alpha \in \overline{\mathbf{Q}}$ is an *algebraic integer* if it is a root of some monic polynomial with coefficients in $\mathbf{Z}$.

Definition 5.1.2 (Minimal Polynomial). The *minimal polynomial* of $\alpha \in \mathbf{Q}$ is the monic polynomial $f \in \mathbf{Q}[x]$ of least positive degree such that $f(\alpha) = 0$.

The minimal polynomial of α divides any polynomial h such that $h(\alpha) = 0$, for the following reason. If $h(\alpha) = 0$, use the division algorithm to write $h = qf + r$, where $0 \leq \deg(r) < \deg(f)$. We have $r(\alpha) = h(\alpha) - q(\alpha)f(\alpha) = 0$, so α is a root of r. However, f is the polynomial of least positive degree with root α, so $r = 0$.

Lemma 5.1.3. *If α is an algebraic integer, then the minimal polynomial of α has coefficients in $\mathbf{Z}$.*

Proof. Suppose $f \in \mathbf{Q}[x]$ is the minimal polynomial of α and $g \in \mathbf{Z}[x]$ is a monic integral polynomial such that $g(\alpha) = 0$. As mentioned after the definition of minimal polynomial, we have $g = fh$, for some $h \in \mathbf{Q}[x]$. If $f \not\in \mathbf{Z}[x]$, then some prime p divides the denominator of some coefficient of f. Let p^i be the largest power of p that divides some denominator of some coefficient f, and likewise let p^j be the largest power of p that divides some denominator of a coefficient of g. Then $p^{i+j}g = (p^i f)(p^j g)$, and if we reduce both sides modulo p, then the left hand side is 0 but the right hand side is a product of two nonzero polynomials in $\mathbf{F}_p[x]$, hence nonzero, a contradiction. □

Proposition 5.1.4. *An element $\alpha \in \overline{\mathbf{Q}}$ is integral if and only if $\mathbf{Z}[\alpha]$ is finitely generated as a $\mathbf{Z}$-module.*

Proof. Suppose α is integral and let $f \in \mathbf{Z}[x]$ be the monic minimal polynomial of α (that $f \in \mathbf{Z}[x]$ is Lemma 5.1.3). Then $\mathbf{Z}[\alpha]$ is generated by $1, \alpha, \alpha^2, \ldots, \alpha^{d-1}$, where d is the degree of f. Conversely, suppose $\alpha \in \overline{\mathbf{Q}}$ is such that $\mathbf{Z}[\alpha]$ is finitely generated, say by elements $f_1(\alpha), \ldots, f_n(\alpha)$. Let d be any integer bigger than the degree of any f_i. Then there exist integers a_i such that $\alpha^d = \sum a_i f_i(\alpha)$, hence α satisfies the monic polynomial $x^d - \sum a_i f_i(x) \in \mathbf{Z}[x]$, so α is integral. □

The rational number $\alpha = 1/2$ is not integral. Note that $G = \mathbf{Z}[1/2]$ is not a finitely generated $\mathbf{Z}$-module, since G is infinite and $G/2G = 0$.

Proposition 5.1.5. *The set $\overline{\mathbf{Z}}$ of all algebraic integers is a ring, i.e., the sum and product of two algebraic integers is again an algebraic integer.*

Proof. Suppose $\alpha, \beta \in \mathbf{Z}$, and let m, n be the degrees of the minimal polynomials of α, β, respectively. Then $1, \alpha, \ldots, \alpha^{m-1}$ span $\mathbf{Z}[\alpha]$ and $1, \beta, \ldots, \beta^{n-1}$ span $\mathbf{Z}[\beta]$ as $\mathbf{Z}$-module. Thus the elements $\alpha^i \beta^j$ for $i \leq m, j \leq n$ span $\mathbf{Z}[\alpha, \beta]$. Since $\mathbf{Z}[\alpha + \beta]$ is a submodule of the finitely-generated module $\mathbf{Z}[\alpha, \beta]$, it is finitely generated, so $\alpha + \beta$ is integral. Likewise, $\mathbf{Z}[\alpha\beta]$ is a submodule of $\mathbf{Z}[\alpha, \beta]$, so it is also finitely generated and $\alpha\beta$ is integral. □

Recall that a *number field* is a subfield K of $\overline{\mathbf{Q}}$ such that the degree $[K : \mathbf{Q}] := \dim_{\mathbf{Q}}(K)$ is finite.

Definition 5.1.6 (Ring of Integers). The *ring of integers* of a number field K is the ring

$$\mathcal{O}_K = K \cap \overline{\mathbf{Z}} = \{x \in K : \ x \text{ is an algebraic integer}\}.$$

The field $\mathbf{Q}$ of rational numbers is a number field of degree 1, and the ring of integers of $\mathbf{Q}$ is $\mathbf{Z}$. The field $K = \mathbf{Q}(i)$ of Gaussian integers has degree 2 and $\mathcal{O}_K = \mathbf{Z}[i]$. The field $K = \mathbf{Q}(\sqrt{5})$ has ring of integers $\mathcal{O}_K = \mathbf{Z}[(1+\sqrt{5})/2]$. Note that the Golden ratio $(1+\sqrt{5})/2$ satisfies $x^2 - x - 1$. According to Magma, the ring of integers of $K = \mathbf{Q}(\sqrt[3]{9})$ is $\mathbf{Z}[\sqrt[3]{3}]$, where $\sqrt[3]{3} = \frac{1}{3}(\sqrt[3]{9})^2$.

Definition 5.1.7 (Order). An *order* in $\mathcal{O}_K$ is any subring R of $\mathcal{O}_K$ such that the quotient $\mathcal{O}_K/R$ of abelian groups is finite. (Note that R must contain 1 because it is a ring, and for us every ring has a 1.)

As noted above, $\mathbf{Z}[i]$ is the ring of integers of $\mathbf{Q}(i)$. For every nonzero integer n, the subring $\mathbf{Z} + ni\mathbf{Z}$ of $\mathbf{Z}[i]$ is an order. The subring $\mathbf{Z}$ of $\mathbf{Z}[i]$ is not an order, because $\mathbf{Z}$ does not have finite index in $\mathbf{Z}[i]$. Also the subgroup $2\mathbf{Z}+i\mathbf{Z}$ of $\mathbf{Z}[i]$ is not an order because it is not a ring.

We will frequently consider orders in practice because they are often much easier to write down explicitly than $\mathcal{O}_K$. For example, if $K=\mathbf{Q}(\alpha)$ and α is an algebraic integer, then $\mathbf{Z}[\alpha]$ is an order in $\mathcal{O}_K$, but frequently $\mathbf{Z}[\alpha]\neq\mathcal{O}_K$.

Lemma 5.1.8. *Let $\mathcal{O}_K$ be the ring of integers of a number field. Then $\mathcal{O}_K\cap\mathbf{Q}=\mathbf{Z}$ and $\mathbf{Q}\mathcal{O}_K=K$.*

Proof. Suppose $\alpha\in\mathcal{O}_K\cap\mathbf{Q}$ with $\alpha=a/b$ in lowest terms and $b>0$. The monic minimal polynomial of α is $bx-a\in\mathbf{Z}[x]$, so if $b\neq 1$ then Lemma 5.1.3 implies that α is not an algebraic integer, a contradiction.

To prove that $\mathbf{Q}\mathcal{O}_K=K$, suppose $\alpha\in K$, and let $f(x)\in\mathbf{Q}[x]$ be the minimal monic polynomial of α. For any positive integer d, the minimal monic polynomial of $d\alpha$ is $d^{\deg(f)}f(x/d)$, i.e., the polynomial obtained from $f(x)$ by multiplying the coefficient of $x^{\deg(f)}$ by 1, multiplying the coefficient of $x^{\deg(f)-1}$ by d, multiplying the coefficient of $x^{\deg(f)-2}$ by d^2, etc. If d is the least common multiple of the denominators of the coefficients of f, then the minimal monic polynomial of $d\alpha$ has integer coefficients, so $d\alpha$ is integral and $d\alpha\in\mathcal{O}_K$. This proves that $\mathbf{Q}\mathcal{O}_K=K$. □

In the next two sections we will develop some basic properties of norms and traces, and deduce further properties of rings of integers.

5.2 Norms and Traces

Before discussing norms and traces we introduce some notation for field extensions. If $K\subset L$ are number fields, we let $[L:K]$ denote the dimension of L viewed as a K-vector space. If K is a number field and $a\in\overline{\mathbf{Q}}$, let $K(a)$ be the number field generated by a, which is the smallest number field that contains a. If $a\in\overline{\mathbf{Q}}$ then a has a minimal polynomial $f(x)\in\mathbf{Q}[x]$, and the *Galois conjugates* of a are the roots of f. For example the element $\sqrt{2}$ has minimal polynomial x^2-2 and the Galois conjugates of $\sqrt{2}$ are $\pm\sqrt{2}$.

Suppose $K\subset L$ is an inclusion of number fields and let $a\in L$. Then left multiplication by a defines a K-linear transformation $\ell_a:L\to L$. (The transformation ℓ_a is K-linear because L is commutative.)

Definition 5.2.1 (Norm and Trace). The *norm* and *trace* of a from L to K are

$$\mathrm{Norm}_{L/K}(a)=\mathrm{Det}(\ell_a)\quad\text{ and }\quad\mathrm{tr}_{L/K}(a)=\mathrm{tr}(\ell_a).$$

It is standard from linear algebra that determinants are multiplicative and traces are additive, so for $a,b\in L$ we have

$$\mathrm{Norm}_{L/K}(ab)=\mathrm{Norm}_{L/K}(a)\cdot\mathrm{Norm}_{L/K}(b)$$

and

$$\mathrm{tr}_{L/K}(a+b)=\mathrm{tr}_{L/K}(a)+\mathrm{tr}_{L/K}(b).$$

Note that if $f \in \mathbf{Q}[x]$ is the characteristic polynomial of ℓ_a, then the constant term of f is $(-1)^{\deg(f)}\operatorname{Det}(\ell_a)$, and the coefficient of $x^{\deg(f)-1}$ is $-\operatorname{tr}(\ell_a)$.

Proposition 5.2.2. *Let $a \in L$ and let $\sigma_1, \ldots, \sigma_d$, where $d = [L : K]$, be the distinct field embeddings $L \hookrightarrow \overline{\mathbf{Q}}$ that fix every element of K. Then*

$$\operatorname{Norm}_{L/K}(a) = \prod_{i=1}^{d} \sigma_i(a) \qquad \textit{and} \qquad \operatorname{tr}_{L/K}(a) = \sum_{i=1}^{d} \sigma_i(a).$$

Proof. We prove the proposition by computing the characteristic polynomial F of a. Let $f \in K[x]$ be the minimal polynomial of a over K, and note that f has distinct roots (since it is the polynomial in $K[x]$ of least degree that is satisfied by a). Since f is irreducible, $[K(a) : K] = \deg(f)$, and a satisfies a polynomial if and only if ℓ_a does, the characteristic polynomial of ℓ_a acting on $K(a)$ is f. Let $b_1, \ldots, b_n$ be a basis for L over $K(a)$ and note that $1, \ldots, a^m$ is a basis for $K(a)/K$, where $m = \deg(f) - 1$. Then $a^i b_j$ is a basis for L over K, and left multiplication by a acts the same way on the span of $b_j, ab_j, \ldots, a^m b_j$ as on the span of $b_k, ab_k, \ldots, a^m b_k$, for any pair $j, k \leq n$. Thus the matrix of ℓ_a on L is a block direct sum of copies of the matrix of ℓ_a acting on $K(a)$, so the characteristic polynomial of ℓ_a on L is $f^{[L:K(a)]}$. The proposition follows because the roots of $f^{[L:K(a)]}$ are exactly the images $\sigma_i(a)$, with multiplicity $[L : K(a)]$ (since each embedding of $K(a)$ into $\overline{\mathbf{Q}}$ extends in exactly $[L : K(a)]$ ways to L by Exercise 9). □

The following corollary asserts that the norm and trace behave well in towers.

Corollary 5.2.3. *Suppose $K \subset L \subset M$ is a tower of number fields, and let $a \in M$. Then*

$$\operatorname{Norm}_{M/K}(a) = \operatorname{Norm}_{L/K}(\operatorname{Norm}_{M/L}(a)) \qquad \textit{and} \qquad \operatorname{tr}_{M/K}(a) = \operatorname{tr}_{L/K}(\operatorname{tr}_{M/L}(a)).$$

Proof. For the first equation, both sides are the product of $\sigma_i(a)$, where σ_i runs through the embeddings of M into K. To see this, suppose $\sigma : L \to \overline{\mathbf{Q}}$ fixes K. If σ' is an extension of σ to M, and $\tau_1, \ldots, \tau_d$ are the embeddings of M into $\overline{\mathbf{Q}}$ that fix L, then $\tau_1\sigma', \ldots, \tau_d\sigma'$ are exactly the extensions of σ to M. For the second statement, both sides are the sum of the $\sigma_i(a)$. □

The norm and trace down to $\mathbf{Q}$ of an algebraic integer a is an element of $\mathbf{Z}$, because the minimal polynomial of a has integer coefficients, and the characteristic polynomial of a is a power of the minimal polynomial, as we saw in the proof of Proposition 5.2.2.

Proposition 5.2.4. *Let K be a number field. The ring of integers $\mathcal{O}_K$ is a lattice in K, i.e., $\mathbf{Q}\mathcal{O}_K = K$ and $\mathcal{O}_K$ is an abelian group of rank $[K : \mathbf{Q}]$.*

Proof. We saw in Lemma 5.1.8 that $\mathbf{Q}\mathcal{O}_K = K$. Thus there exists a basis $a_1, \ldots, a_n$ for K, where each a_i is in $\mathcal{O}_K$. Suppose that as $x = \sum c_i a_i \in \mathcal{O}_K$ varies over all elements of $\mathcal{O}_K$ the denominators of the coefficients c_i are arbitrarily large. Then subtracting off integer multiples of the a_i, we see that as $x = \sum c_i a_i \in \mathcal{O}_K$ varies over elements of $\mathcal{O}_K$ with c_i between 0 and 1, the denominators of the c_i are also arbitrarily large. This implies that there are infinitely many elements of $\mathcal{O}_K$ in the bounded subset

$$S = \{c_1a_1 + \cdots + c_na_n : c_i \in \mathbf{Q},\ 0 \leq c_i \leq 1\} \subset K.$$

Thus for any $\varepsilon > 0$, there are elements $a, b \in \mathcal{O}_K$ such that the coefficients of $a - b$ are all less than ε (otherwise the elements of $\mathcal{O}_K$ would all be a "distance" of least ε from each other, so only finitely many of them would fit in S).

As mentioned above, the norms of elements of $\mathcal{O}_K$ are integers. Since the norm of an element is the determinant of left multiplication by that element, the norm is a homogenous polynomial of degree n in the indeterminate coefficients c_i. If the c_i get arbitrarily small for elements of $\mathcal{O}_K$, then the values of the norm polynomial get arbitrarily small, which would imply that there are elements of $\mathcal{O}_K$ with positive norm too small to be in $\mathbf{Z}$, a contradiction. So the set S contains only finitely many elements of $\mathcal{O}_K$. Thus the denominators of the c_i are bounded, so for some d, we have that $\mathcal{O}_K$ has finite index in $A = \frac{1}{d}\mathbf{Z}a_1 + \cdots + \frac{1}{d}\mathbf{Z}a_n$. Since A is isomorphic to $\mathbf{Z}^n$, it follows from the structure theorem for finitely generated abelian groups that $\mathcal{O}_K$ is isomorphic as a $\mathbf{Z}$-module to $\mathbf{Z}^n$, as claimed. □

Corollary 5.2.5. *The ring of integers $\mathcal{O}_K$ of a number field is Noetherian.*

Proof. By Proposition 5.2.4, the ring $\mathcal{O}_K$ is finitely generated as a module over $\mathbf{Z}$, so it is certainly finitely generated as a ring over $\mathbf{Z}$. By the Hilbert Basis Theorem, $\mathcal{O}_K$ is Noetherian. □

Chapter 6

Unique Factorization of Ideals

In this chapter we will deduce, with complete proofs, the most important basic property of the ring of integers $\mathcal{O}_K$ of an algebraic number, namely that every nonzero ideals can be written uniquely as products of prime ideals. After proving this fundamental theorem, we will compute some examples using Magma. The next chapter will consist mostly of examples illustrating the substantial theory we will have already developed, so hang in there!

6.1 Dedekind Domains

Recall (Corollary 5.2.5) that we proved that the ring of integers $\mathcal{O}_K$ of a number field is Noetherian. As we saw before using norms, the ring $\mathcal{O}_K$ is finitely generated as a module over $\mathbf{Z}$, so it is certainly finitely generated as a ring over $\mathbf{Z}$. By the Hilbert Basis Theorem, $\mathcal{O}_K$ is Noetherian.

If R is an integral domain, the *field of fractions* of R is the field of all elements a/b, where $a, b \in R$. The field of fractions of R is the smallest field that contains R. For example, the field of fractions of $\mathbf{Z}$ is $\mathbf{Q}$ and of $\mathbf{Z}[(1+\sqrt{5})/2]$ is $\mathbf{Q}(\sqrt{5})$.

Definition 6.1.1 (Integrally Closed). An integral domain R is *integrally closed in its field of fractions* if whenever α is in the field of fractions of R and α satisfies a monic polynomial $f \in R[x]$, then $\alpha \in R$.

Proposition 6.1.2. *If K is any number field, then $\mathcal{O}_K$ is integrally closed. Also, the ring $\overline{\mathbf{Z}}$ of all algebraic integers is integrally closed.*

Proof. We first prove that $\overline{\mathbf{Z}}$ is integrally closed. Suppose $c \in \overline{\mathbf{Q}}$ is integral over $\overline{\mathbf{Z}}$, so there is a monic polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ with $a_i \in \overline{\mathbf{Z}}$ and $f(c) = 0$. The a_i all lie in the ring of integers $\mathcal{O}_K$ of the number field $K = \mathbf{Q}(a_0, a_1, \ldots a_{n-1})$, and $\mathcal{O}_K$ is finitely generated as a $\mathbf{Z}$-module, so $\mathbf{Z}[a_0, \ldots, a_{n-1}]$ is finitely generated as a $\mathbf{Z}$-module. Since $f(c) = 0$, we can write c^n as a $\mathbf{Z}[a_0, \ldots, a_{n-1}]$-linear combination of c^i for $i < n$, so the ring $\mathbf{Z}[a_0, \ldots, a_{n-1}, c]$ is also finitely generated as a $\mathbf{Z}$-module. Thus $\mathbf{Z}[c]$ is finitely generated as $\mathbf{Z}$-module because it is a submodule of a finitely generated $\mathbf{Z}$-module, which implies that c is integral over $\mathbf{Z}$.

Suppose $c \in K$ is integral over $\mathcal{O}_K$. Then since $\overline{\mathbf{Z}}$ is integrally closed, c is an element of $\overline{\mathbf{Z}}$, so $c \in K \cap \overline{\mathbf{Z}} = \mathcal{O}_K$, as required. □

Definition 6.1.3 (Dedekind Domain). An integral domain R is a *Dedekind domain* if it is Noetherian, integrally closed in its field of fractions, and every nonzero prime ideal of R is maximal.

The ring $\mathbf{Q} \oplus \mathbf{Q}$ is Noetherian, integrally closed in its field of fractions, and the two prime ideals are maximal. However, it is not a Dedekind domain because it is not an integral domain. The ring $\mathbf{Z}[\sqrt{5}]$ is not a Dedekind domain because it is not integrally closed in its field of fractions, as $(1+\sqrt{5})/2$ is integrally over $\mathbf{Z}$ and lies in $\mathbf{Q}(\sqrt{5})$, but not in $\mathbf{Z}[\sqrt{5}]$. The ring $\mathbf{Z}$ is a Dedekind domain, as is any ring of integers $\mathcal{O}_K$ of a number field, as we will see below. Also, any field K is a Dedekind domain, since it is a domain, it is trivially integrally closed in itself, and there are no nonzero prime ideals so that condition that they be maximal is empty.

Proposition 6.1.4. *The ring of integers $\mathcal{O}_K$ of a number field is a Dedekind domain.*

Proof. By Proposition 6.1.2, the ring $\mathcal{O}_K$ is integrally closed, and by Proposition 5.2.5 it is Noetherian. Suppose that $\mathfrak{p}$ is a nonzero prime ideal of $\mathcal{O}_K$. Let $\alpha \in \mathfrak{p}$ be a nonzero element, and let $f(x) \in \mathbf{Z}[x]$ be the minimal polynomial of α. Then

$$f(\alpha) = \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0,$$

so $a_0 = -(\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha) \in \mathfrak{p}$. Since f is irreducible, a_0 is a nonzero element of $\mathbf{Z}$ that lies in $\mathfrak{p}$. Every element of the finitely generated abelian group $\mathcal{O}_K/\mathfrak{p}$ is killed by a_0, so $\mathcal{O}_K/\mathfrak{p}$ is a finite set. Since $\mathfrak{p}$ is prime, $\mathcal{O}_K/\mathfrak{p}$ is an integral domain. Every finite integral domain is a field, so $\mathfrak{p}$ is maximal, which completes the proof. □

If I and J are ideals in a ring R, the product IJ is the ideal generated by all products of elements in I with elements in J:

$$IJ = (ab : a \in I, b \in J) \subset R.$$

Note that the set of all products ab, with $a \in I$ and $b \in J$, need not be an ideal, so it is important to take the ideal generated by that set. (See the homework problems for examples.)

Definition 6.1.5 (Fractional Ideal). A *fractional ideal* is an $\mathcal{O}_K$-submodule of $I \subset K$ that is finitely generated as an $\mathcal{O}_K$-module.

To avoid confusion, we will sometimes call a genuine ideal $I \subset \mathcal{O}_K$ an *integral ideal*. Also, since fractional ideals are finitely generated, we can clear denominators of a generating set to see that every fractional ideal is of the form $aI = \{ab : b \in I\}$ for some $a \in K$ and ideal $I \subset \mathcal{O}_K$.

For example, the collection $\frac{1}{2}\mathbf{Z}$ of rational numbers with denominator 1 or 2 is a fractional ideal of $\mathbf{Z}$.

Theorem 6.1.6. *The set of nonzero fractional ideals of a Dedekind domain R is an abelian group under ideal multiplication.*

Before proving Theorem 6.1.6 we prove a lemma. For the rest of this section $\mathcal{O}_K$ is the ring of integers of a number field K.

Definition 6.1.7 (Divides for Ideals). Suppose that I, J are ideals of $\mathcal{O}_K$. Then I *divides* J if $I \supset J$.

To see that this notion of divides is sensible, suppose $K = \mathbf{Q}$, so $\mathcal{O}_K = \mathbf{Z}$. Then $I = (n)$ and $J = (m)$ for some integer n and m, and I divides J means that $(n) \supset (m)$, i.e., that there exists an integer c such that $m = cn$, which exactly means that n divides m, as expected.

Lemma 6.1.8. *Suppose I is an ideal of $\mathcal{O}_K$. Then there exist prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ such that $\mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_n \subset I$. In other words, I divides a product of prime ideals. (By convention the empty product is the unit ideal. Also, if $I = 0$, then we take $\mathfrak{p}_1 = (0)$, which is a prime ideal.)*

Proof. The key idea is to use that $\mathcal{O}_K$ is Noetherian to deduce that the set S of ideals that do not satisfy the lemma is empty. If S is nonempty, then because $\mathcal{O}_K$ is Noetherian, there is an ideal $I \in S$ that is maximal as an element of S. If I were prime, then I would trivially contain a product of primes, so I is not prime. By definition of prime ideal, there exists $a, b \in \mathcal{O}_K$ such that $ab \in I$ but $a \notin I$ and $b \notin I$. Let $J_1 = I + (a)$ and $J_2 = I + (b)$. Then neither J_1 nor J_2 is in S, since I is maximal, so both J_1 and J_2 contain a product of prime ideals. Thus so does I, since

$$J_1 J_2 = I^2 + I(b) + (a)I + (ab) \subset I,$$

which is a contradiction. Thus S is empty, which completes the proof. □

We are now ready to prove the theorem.

Proof of Theorem 6.1.6. The product of two fractional ideals is again finitely generated, so it is a fractional ideal, and $I\mathcal{O}_K = \mathcal{O}_K$ for any nonzero ideal I, so to prove that the set of fractional ideals under multiplication is a group it suffices to show the existence of inverses. We will first prove that if $\mathfrak{p}$ is a prime ideal, then $\mathfrak{p}$ has an inverse, then we will prove that nonzero integral ideals have inverses, and finally observe that every fractional ideal has an inverse.

Suppose $\mathfrak{p}$ is a nonzero prime ideal of $\mathcal{O}_K$. We will show that the $\mathcal{O}_K$-module

$$I = \{a \in K : a\mathfrak{p} \subset \mathcal{O}_K\}$$

is a fractional ideal of $\mathcal{O}_K$ such that $I\mathfrak{p} = \mathcal{O}_K$, so that I is an inverse of $\mathfrak{p}$.

For the rest of the proof, fix a nonzero element $b \in \mathfrak{p}$. Since I is an $\mathcal{O}_K$-module, $bI \subset \mathcal{O}_K$ is an $\mathcal{O}_K$ ideal, hence I is a fractional ideal. Since $\mathcal{O}_K \subset I$ we have $\mathfrak{p} \subset I\mathfrak{p} \subset \mathcal{O}_K$, hence either $\mathfrak{p} = I\mathfrak{p}$ or $I\mathfrak{p} = \mathcal{O}_K$. If $I\mathfrak{p} = \mathcal{O}_K$, we are done since then I is an inverse of $\mathfrak{p}$. Thus suppose that $I\mathfrak{p} = \mathfrak{p}$. Our strategy is to show that there is some $d \in I$ not in $\mathcal{O}_K$; such a d would leave $\mathfrak{p}$ invariant (i.e., $d\mathfrak{p} \subset \mathfrak{p}$), so since $\mathfrak{p}$ is an $\mathcal{O}_K$-module it will follow that $d \in \mathcal{O}_K$, a contradiction.

By Lemma 6.1.8, we can choose a product $\mathfrak{p}_1, \ldots, \mathfrak{p}_m$, with m minimal, such that

$$\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_m \subset (b) \subset \mathfrak{p}.$$

If no $\mathfrak{p}_i$ is contained in $\mathfrak{p}$, then we can choose for each i an $a_i \in \mathfrak{p}_i$ with $a_i \notin \mathfrak{p}$; but then $\prod a_i \in \mathfrak{p}$, which contradicts that $\mathfrak{p}$ is a prime ideal. Thus some $\mathfrak{p}_i$, say $\mathfrak{p}_1$, is contained in $\mathfrak{p}$, which implies that $\mathfrak{p}_1 = \mathfrak{p}$ since every nonzero prime ideal is maximal. Because m is minimal, $\mathfrak{p}_2 \cdots \mathfrak{p}_m$ is not a subset of (b), so there exists $c \in \mathfrak{p}_2 \cdots \mathfrak{p}_m$ that does not lie in (b). Then $\mathfrak{p}(c) \subset (b)$, so by definition of I we have $d = c/b \in I$. However, $d \notin \mathcal{O}_K$, since if it were then c would be in (b). We have thus found our element $d \in I$ that does not lie in $\mathcal{O}_K$. To finish the proof that $\mathfrak{p}$ has an inverse, we observe that d preserves the $\mathcal{O}_K$-module $\mathfrak{p}$, and is hence in $\mathcal{O}_K$, a contradiction. More precisely, if $b_1, \ldots, b_n$ is a basis for $\mathfrak{p}$ as a $\mathbf{Z}$-module, then the action of d on $\mathfrak{p}$ is given by a matrix with entries in $\mathbf{Z}$, so the minimal polynomial of d has coefficients in $\mathbf{Z}$. This implies that d is integral over $\mathbf{Z}$, so $d \in \mathcal{O}_K$, since $\mathcal{O}_K$ is integrally closed by Proposition 6.1.2. (Note how this argument depends strongly on the fact that $\mathcal{O}_K$ is integrally closed!)

So far we have proved that if $\mathfrak{p}$ is a prime ideal of $\mathcal{O}_K$, then $\mathfrak{p}^{-1} = \{a \in \mathbf{K} : a\mathfrak{p} \subset \mathcal{O}_K\}$ is the inverse of $\mathfrak{p}$ in the monoid of nonzero fractional ideals of $\mathcal{O}_K$. As mentioned after Definition 6.1.5 [on Tuesday], every nonzero fractional ideal is of the form aI for $a \in K$ and I an integral ideal, so since (a) has inverse $(1/a)$, it suffices to show that every integral ideal I has an inverse. If not, then there is a nonzero integral ideal I that is maximal among all nonzero integral ideals that do not have an inverse. Every ideal is contained in a maximal ideal, so there is a nonzero prime ideal $\mathfrak{p}$ such that $I \subset \mathfrak{p}$. Multiplying both sides of this inclusion by $\mathfrak{p}^{-1}$ and using that $\mathcal{O}_K \subset \mathfrak{p}^{-1}$, we see that $I \subset \mathfrak{p}^{-1}I \subset \mathcal{O}_K$. If $I = \mathfrak{p}^{-1}I$, then arguing as in the proof that $\mathfrak{p}^{-1}$ is the inverse of $\mathfrak{p}$, we see that each element of $\mathfrak{p}^{-1}$ preserves the finitely generated $\mathbf{Z}$-module I and is hence integral. But then $\mathfrak{p}^{-1} \subset \mathcal{O}_K$, which implies that $\mathcal{O}_K = \mathfrak{p}\mathfrak{p}^{-1} \subset \mathfrak{p}$, a contradiction. Thus $I \neq \mathfrak{p}^{-1}I$. Because I is maximal among ideals that do not have an inverse, the ideal $\mathfrak{p}^{-1}I$ does have an inverse, call it J. Then $\mathfrak{p}J$ is the inverse of I, since $\mathcal{O}_K = (\mathfrak{p}J)(\mathfrak{p}^{-1}I) = JI$. □

We can finally deduce the crucial Theorem 6.1.10, which will allow us to show that any nonzero ideal of a Dedekind domain can be expressed uniquely as a product of primes (up to order). Thus unique factorization holds for ideals in a Dedekind domain, and it is this unique factorization that initially motivated the introduction of rings of integers of number fields over a century ago.

Theorem 6.1.9. *Suppose I is an integral ideal of $\mathcal{O}_K$. Then I can be written as a product*

$$I = \mathfrak{p}_1 \cdots \mathfrak{p}_n$$

of prime ideals of $\mathcal{O}_K$, and this representation is unique up to order. (Exception: If $I = 0$, then the representation is not unique.)

Proof. Suppose I is an ideal that is maximal among the set of all ideals in $\mathcal{O}_K$ that can not be written as a product of primes. Every ideal is contained in a maximal ideal, so I is contained in a nonzero prime ideal $\mathfrak{p}$. If $I\mathfrak{p}^{-1} = I$, then by Theorem 6.1.6 we can cancel I from both sides of this equation to see that $\mathfrak{p}^{-1} = \mathcal{O}_K$, a contradiction. Thus I is strictly contained in $I\mathfrak{p}^{-1}$, so by our maximality assumption on I there are maximal ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ such that $I\mathfrak{p}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_n$. Then $I = \mathfrak{p} \cdot \mathfrak{p}_1 \cdots \mathfrak{p}_n$, a contradiction. Thus every ideal can be written as a product of primes.

Suppose $\mathfrak{p}_1 \cdots \mathfrak{p}_n = \mathfrak{q}_1 \cdots \mathfrak{q}_m$. If no $\mathfrak{q}_i$ is contained in $\mathfrak{p}_1$, then for each i there is an $a_i \in \mathfrak{q}_i$ such that $a_i \notin \mathfrak{p}_1$. But the product of the a_i is in the $\mathfrak{p}_1 \cdots \mathfrak{p}_n$, which is a subset of $\mathfrak{p}_1$, which contradicts the fact that $\mathfrak{p}_1$ is a prime ideal. Thus $\mathfrak{q}_i = \mathfrak{p}_1$ for some i. We can thus cancel $\mathfrak{q}_i$ and $\mathfrak{p}_1$ from both sides of the equation. Repeating this argument finishes the proof of uniqueness. □

Corollary 6.1.10. *If I is a fractional ideal of $\mathcal{O}_K$ then there exists prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ and $\mathfrak{q}_1, \ldots, \mathfrak{q}_m$, unique up to order, such that*

$$I = (\mathfrak{p}_1 \cdots \mathfrak{p}_n)(\mathfrak{q}_1 \cdots \mathfrak{q}_m)^{-1}.$$

Proof. We have $I = (a/b)J$ for some $a, b \in \mathcal{O}_K$ and integral ideal J. Applying Theorem 6.1.10 to (a), (b), and J gives an expression as claimed. For uniqueness, if one has two such product expressions, multiply through by the denominators and use the uniqueness part of Theorem 6.1.10 □

Example 6.1.11. The ring of integers of $K = \mathbf{Q}(\sqrt{-6})$ is $\mathcal{O}_K = \mathbf{Z}[\sqrt{-6}]$. In $\mathcal{O}_K$, we have

$$6 = -\sqrt{-6}\sqrt{-6} = 2 \cdot 3.$$

If $ab = \sqrt{-6}$, with $a, b \in \mathcal{O}_K$ and neither a unit, then $\mathrm{Norm}(a)\,\mathrm{Norm}(b) = 6$, so without loss $\mathrm{Norm}(a) = 2$ and $\mathrm{Norm}(b) = 3$. If $a = c + d\sqrt{-6}$, then $\mathrm{Norm}(a) = c^2 + 6d^2$; since the equation $c^2 + 6d^2 = 2$ has no solution with $c, d \in \mathbf{Z}$, there is no element in $\mathcal{O}_K$ with norm 2, so $\sqrt{-6}$ is irreducible. Also, $\sqrt{-6}$ is not a unit times 2 or times 3, since again the norms would not match up. Thus 6 can not be written uniquely as a product of irreducibles in $\mathcal{O}_K$. Theorem 6.1.9, however, implies that the principal ideal (6) can, however, be written uniquely as a product of prime ideals. Using MAGMA we find such a decomposition:

```
> R<x> := PolynomialRing(RationalField());
> K := NumberField(x^2+6);
> OK := MaximalOrder(K);
         > [K!b : b in Basis(OK)];
         [
             1,
             K.1    // this is sqrt(-6)
         ]
         > Factorization(6*OK);
         [
             <Prime Ideal of OK
             Two element generators:
                 [2, 0]
                 [2, 1], 2>,
             <Prime Ideal of OK
             Two element generators:
                 [3, 0]
                 [3, 1], 2>
         ]
```

The output means that

$$(6) = (2, 2 + \sqrt{-6})^2 \cdot (3, 3 + \sqrt{-6})^2,$$

where each of the ideals $(2, 2 + \sqrt{-6})$ and $(3, 3 + \sqrt{-6})$ is prime. I will discuss algorithms for computing such a decomposition in detail, probably next week. The first idea is to write $(6) = (2)(3)$, and hence reduce to the case of writing the (p), for $p \in \mathbf{Z}$ prime, as a product of primes. Next one decomposes the Artinian ring $\mathcal{O}_K \otimes \mathbf{F}_p$ as a product of local Artinian rings.

Chapter 7

Computing

7.1 Algorithms for Algebraic Number Theory

I think the best overall reference for algorithms for doing basic algebraic number theory computations is [Coh93].

Our main long-term algorithmic goals for this book (which we won't succeed at reaching) are to understand good algorithms for solving the following problems in particular cases:

- **Ring of integers:** Given a number field K (by giving a polynomial), compute the full ring $\mathcal{O}_K$ of integers.

- **Decomposition of primes:** Given a prime number $p \in \mathbf{Z}$, find the decomposition of the ideal $p\mathcal{O}_K$ as a product of prime ideals of $\mathcal{O}_K$.

- **Class group:** Compute the group of equivalence classes of nonzero ideals of $\mathcal{O}_K$, where I and J are equivalent if there exists $\alpha \in \mathcal{O}_K$ such that $IJ^{-1} = (\alpha)$.

- **Units:** Compute generators for the group of units of $\mathcal{O}_K$.

As we will see, somewhat surprisingly it turns out that algorithmically by far the most time-consuming step in computing the ring of integers $\mathcal{O}_K$ seems to be to factor the discriminant of a polynomial whose root generates the field K. The algorithm(s) for computing $\mathcal{O}_K$ are quite complicated to describe, but the first step is to factor this discriminant, and it takes much longer in practice than all the other complicated steps.

7.2 Using MAGMA

This section is a first introduction to MAGMA for algebraic number theory. MAGMA is a good general purpose package for doing algebraic number theory computations. You can use it via the web page `http://modular.fas.harvard.edu/calc`. MAGMA is not free, but student discounts are available.

The following examples illustrate what we've done so far in the course using Magma, and a little of where we are going. Feel free to ask questions as we go.

7.2.1 Smith Normal Form

On the first day of class we learned about Smith normal forms of matrices.

```
> A := Matrix(2,2,[1,2,3,4]);
> A;
[1 2]
[3 4]
> SmithForm(A);
[1 0]
[0 2]

[ 1  0]
[-1  1]

[-1  2]
[ 1 -1]
```

As you can see, Magma computed the Smith form, which is $\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$. What are the other two matrices it output? To see what any Magma command does, type the command by itself with no arguments followed by a semicolon.

```
> SmithForm;
Intrinsic 'SmithForm'

Signatures:

    (<Mtrx> X) -> Mtrx, AlgMatElt, AlgMatElt
    [
        k: RngIntElt,
        NormType: MonStgElt,
        Partial: BoolElt,
        RightInverse: BoolElt
    ]

        The smith form S of X, together with unimodular matrices
        P and Q such that P * X * Q = S.
```

As you can see, `SmithForm` returns three arguments, a matrix and matrices P and Q that transform the input matrix to Smith normal form. The syntax to "receive" three return arguments is natural, but uncommon in other programming languages:

```
> S, P, Q := SmithForm(A);
> S;
[1 0]
[0 2]
> P;
[ 1  0]
[-1  1]
> Q;
[-1  2]
[ 1 -1]
> P*A*Q;
[1 0]
[0 2]
```

Next, let's test the limits. We make a 10 × 10 integer matrix with entries between 0 and 99, and compute its Smith normal form.

```
> A := Matrix(10,10,[Random(100) : i in [1..100]]);
> time B := SmithForm(A);
Time: 0.000
```

Let's print the first row of A, the first and last row of B, and the diagonal of B:

```
> A[1];
( 4 48 84  3 58 61 53 26  9  5)
> B[1];
(1 0 0 0 0 0 0 0 0 0)
> B[10];
(0 0 0 0 0 0 0 0 0 51805501538039733)
> [B[i,i] : i in [1..10]];
[ 1, 1, 1, 1, 1, 1, 1, 1, 1, 51805501538039733 ]
```

Let's see how big we have to make A in order to slow down MAGMA. These timings below are on a 1.6Ghz Pentium 4-M laptop running Magma V2.11 under VMware Linux. I tried exactly the same computation running Magma V2.17 natively under Windows XP on the same machine, and it takes *twice* as long to do each computation, which is strange.

```
> n := 50; A := Matrix(n,n,[Random(100) : i in [1..n^2]]);
> time B := SmithForm(A);
Time: 0.050
> n := 100; A := Matrix(n,n,[Random(100) : i in [1..n^2]]);
> time B := SmithForm(A);
Time: 0.800
> n := 150; A := Matrix(n,n,[Random(100) : i in [1..n^2]]);
> time B := SmithForm(A);
```

```
Time: 4.900
> n := 200; A := Matrix(n,n,[Random(100) : i in [1..n^2]]);
> time B := SmithForm(A);
Time: 19.160
```

Magma can also work with finitely generated abelian groups.

```
> G := AbelianGroup([3,5,18]);
> G;
Abelian Group isomorphic to Z/3 + Z/90
Defined on 3 generators
Relations:
    3*G.1 = 0
    5*G.2 = 0
    18*G.3 = 0
> #G;
270
> H := sub<G | [G.1+G.2]>;
> #H;
15
> G/H;
Abelian Group isomorphic to Z/18
```

7.2.2 $\overline{\mathbf{Q}}$ and Number Fields

Magma has many commands for doing basic arithmetic with $\overline{\mathbf{Q}}$.

```
> Qbar := AlgebraicClosure(RationalField());
> Qbar;
> S<x> := PolynomialRing(Qbar);
> r := Roots(x^3-2);
> r;
[
    <r1, 1>,
    <r2, 1>,
    <r3, 1>
]
> a := r[1][1];
> MinimalPolynomial(a);
x^3 - 2
> s := Roots(x^2-7);
> b := s[1][1];
> MinimalPolynomial(b);
x^2 - 7
> a+b;
```

```
r4 + r1
> MinimalPolynomial(a+b);
x^6 - 21*x^4 - 4*x^3 + 147*x^2 - 84*x - 339
> Trace(a+b);
0
> Norm(a+b);
-339
```

There are few commands for general algebraic number fields, so usually we work in specific finitely generated subfields:

```
> MinimalPolynomial(a+b);
x^6 - 21*x^4 - 4*x^3 + 147*x^2 - 84*x - 339
> K := NumberField($1) ;  // $1 = result of previous computation.
> K;
Number Field with defining polynomial x^6 - 21*x^4 - 4*x^3 +
    147*x^2 - 84*x - 339 over the Rational Field
```

We can also define relative extensions of number fields and pass to the corresponding absolute extension.

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^3-2);   // a is the image of x in Q[x]/(x^3-2)
> a;
a
> a^3;
2
> S<y> := PolynomialRing(K);
> L<b> := NumberField(y^2-a);
> L;
Number Field with defining polynomial y^2 - a over K
> b^2;
a
> b^6;
2
> AbsoluteField(L);
Number Field with defining polynomial x^6 - 2 over the Rational
Field
```

7.2.3 Rings of integers

Magma computes rings of integers of number fields.

```
> RingOfIntegers(K);
Maximal Equation Order with defining polynomial x^3 - 2 over ZZ
> RingOfIntegers(L);
```

```
Maximal Equation Order with defining polynomial x^2 + [0, -1, 0]
over its ground order
```

Sometimes the ring of integers of $\mathbf{Q}(a)$ isn't just $\mathbf{Z}[a]$. First a simple example, then a more complicated one:

```
> K<a> := NumberField(2*x^2-3);   // doesn't have to be monic
> 2*a^2 - 3;
0
> K;
Number Field with defining polynomial x^2 - 3/2 over the Rational
Field
> O := RingOfIntegers(K);
> O;
Maximal Order of Equation Order with defining polynomial 2*x^2 -
    3 over ZZ
> Basis(O);
[
    O.1,
    O.2
]
> [K!x : x in Basis(O)];
[
    1,
    2*a       // this is Sqrt(3)
]
```

Here's are some more examples:

```
> procedure ints(f)   // (procedures don't return anything; functions do)
      K<a> := NumberField(f);
      O := MaximalOrder(K);
      print [K!z : z in Basis(O)];
  end procedure;
> ints(x^2-5);
[
    1,
    1/2*(a + 1)
]
> ints(x^2+5);
[
    1,
    a
]
> ints(x^3-17);
```

```
[
    1,
    a,
    1/3*(a^2 + 2*a + 1)
]
> ints(CyclotomicPolynomial(7));
[
    1,
    a,
    a^2,
    a^3,
    a^4,
    a^5
]
> ints(x^5+&+[Random(10)*x^i : i in [0..4]]);  // RANDOM
[
    1,
    a,
    a^2,
    a^3,
    a^4
]
> ints(x^5+&+[Random(10)*x^i : i in [0..4]]);  // RANDOM
[
    1,
    a,
    a^2,
    1/2*(a^3 + a),
    1/16*(a^4 + 7*a^3 + 11*a^2 + 7*a + 14)
]
```

Lets find out how high of a degree Magma can easily deal with.

```
> d := 10; time ints(x^10+&+[Random(10)*x^i : i in [0..d-1]]);
[
    1, a, a^2, a^3, a^4, a^5, a^6, a^7, a^8, a^9
]
Time: 0.030
> d := 15; time ints(x^10+&+[Random(10)*x^i : i in [0..d-1]]);
[
    1,
    7*a,
    7*a^2 + 4*a,
    7*a^3 + 4*a^2 + 4*a,
    7*a^4 + 4*a^3 + 4*a^2 + a,
    7*a^5 + 4*a^4 + 4*a^3 + a^2 + a,
    7*a^6 + 4*a^5 + 4*a^4 + a^3 + a^2 + 4*a,
    7*a^7 + 4*a^6 + 4*a^5 + a^4 + a^3 + 4*a^2,
    7*a^8 + 4*a^7 + 4*a^6 + a^5 + a^4 + 4*a^3 + 4*a,
    7*a^9 + 4*a^8 + 4*a^7 + a^6 + a^5 + 4*a^4 + 4*a^2 + 5*a,
    7*a^10 + 4*a^9 + 4*a^8 + a^7 + a^6 + 4*a^5 + 4*a^3 + 5*a^2 +  4*a,
  ...
]
Time: 0.480
> d := 20; time ints(x^10+&+[Random(10)*x^i : i in [0..d-1]]);
[
    1,
    2*a,
    4*a^2,
    8*a^3,
    8*a^4 + 2*a^2 + a,
    8*a^5 + 2*a^3 + 3*a^2,
 ...]
Time: 3.940
> d := 25; time ints(x^10+&+[Random(10)*x^i : i in [0..d-1]]);
... I stopped it after a few minutes...
```

We can also define orders in rings of integers.

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^3-2);
> O := Order([2*a]);
> O;
Transformation of Order over
Equation Order with defining polynomial x^3 - 2 over ZZ
Transformation Matrix:
[1 0 0]
[0 2 0]
[0 0 4]
> OK := MaximalOrder(K);
> Index(OK,O);
8
> Discriminant(O);
-6912
> Discriminant(OK);
-108
> 6912/108;
64    // perfect square...
```

7.2.4 Ideals

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^3-2);
> O := Order([2*a]);
> O;
Transformation of Order over
Equation Order with defining polynomial x^3 - 2 over ZZ
Transformation Matrix:
[1 0 0]
[0 2 0]
[0 0 4]
> OK := MaximalOrder(K);
> Index(OK,O);
8
> Discriminant(O);
-6912
> Discriminant(OK);
-108
> 6912/108;
64    // perfect square...
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^2-7);
> K<a> := NumberField(x^2-5);
> Discriminant(K);
20   // ????????? Yuck!
> OK := MaximalOrder(K);
> Discriminant(OK);
5    // better
> Discriminant(NumberField(x^2-20));
80
> I := 7*OK;
> I;
Principal Ideal of OK
Generator:
    [7, 0]
> J := (OK!a)*OK;    // the ! computes the natural image of a in OK
> J;
Principal Ideal of OK
Generator:
    [-1, 2]
> I*J;
Principal Ideal of OK
Generator:
    [-7, 14]
> J*I;
Principal Ideal of OK
Generator:
    [-7, 14]
> I+J;
Principal Ideal of OK
Generator:
    [1, 0]
>
> Factorization(I);
[
    <Principal Prime Ideal of OK
    Generator:
        [7, 0], 1>
]
> Factorization(3*OK);
[
    <Principal Prime Ideal of OK
    Generator:
        [3, 0], 1>
]
> Factorization(5*OK);
[
    <Prime Ideal of OK
    Two element generators:
        [5, 0]
        [4, 2], 2>
]
> Factorization(11*OK);
[
    <Prime Ideal of OK
    Two element generators:
        [11, 0]
        [14, 2], 1>,
    <Prime Ideal of OK
    Two element generators:
        [11, 0]
        [17, 2], 1>
]
```

We can even work with fractional ideals in MAGMA.

```
> K<a> := NumberField(x^2-5);
> OK := MaximalOrder(K);
> I := 7*OK;
> J := (OK!a)*OK;
> M := I/J;
> M;
Fractional Principal Ideal of OK
Generator:
    -7/5*OK.1 + 14/5*OK.2
> Factorization(M);
[
    <Prime Ideal of OK
    Two element generators:
        [5, 0]
        [4, 2], -1>,
    <Principal Prime Ideal of OK
    Generator:
        [7, 0], 1>
]
```

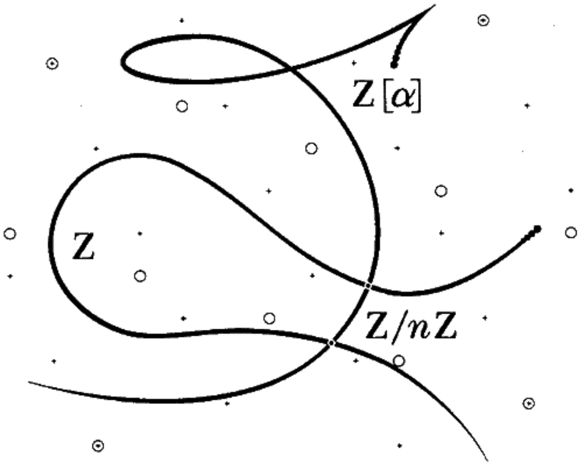
In the next chapter, we will learn about discriminants and an algorithm for "factoring primes", that is writing an ideal $p\mathcal{O}_K$ as a product of prime ideals of $\mathcal{O}_K$.

Chapter 8

Factoring Primes

First we will learn how, if $p \in \mathbf{Z}$ is a prime and $\mathcal{O}_K$ is the ring of integers of a number field, to write $p\mathcal{O}_K$ as a product of primes of $\mathcal{O}_K$. Then I will sketch the main results and definitions that we will study in detail during the next few chapters. We will cover discriminants and norms of ideals, define the class group of $\mathcal{O}_K$ and prove that it is finite and computable, and define the group of units of $\mathcal{O}_K$, determine its structure, and prove that it is also computable.

8.1 Factoring Primes



A diagram from [LL93].

"The obvious mathematical breakthrough would be development of an easy way to factor large prime numbers." –Bill Gates, *The Road Ahead*, pg. 265

Let $K = \mathbf{Q}(\alpha)$ be a number field, and let $\mathcal{O}_K$ be the ring of integers of K. To employ our geometric intuition, as the Lenstras did on the cover of [LL93], it is helpful to view $\mathcal{O}_K$ as a one-dimensional scheme

$$X = \operatorname{Spec}(\mathcal{O}_K) = \{ \text{ all prime ideals of } \mathcal{O}_K \ \}$$

over

$$Y = \operatorname{Spec}(\mathbf{Z}) = \{(0)\} \cup \{p\mathbf{Z} : p \in \mathbf{Z} \text{ is prime } \}.$$

There is a natural map $\pi : X \to Y$ that sends a prime ideal $\mathfrak{p} \in X$ to $\mathfrak{p} \cap \mathbf{Z} \in Y$. For much more on this point of view, see [EH00, Ch. 2].

Ideals were originally introduced by Kummer because, as we proved last Tuesday, in rings of integers of number fields ideals factor uniquely as products of primes ideals, which is something that is not true for general algebraic integers. (The failure of unique factorization for algebraic integers was used by Liouville to destroy Lamé's purported 1847 "proof" of Fermat's Last Theorem.)

If $p \in \mathbf{Z}$ is a prime number, then the ideal $p\mathcal{O}_K$ of $\mathcal{O}_K$ factors uniquely as a product $\prod \mathfrak{p}_i^{e_i}$, where the $\mathfrak{p}_i$ are maximal ideals of $\mathcal{O}_K$. We may imagine the decomposition of $p\mathcal{O}_K$ into prime ideals geometrically as the fiber $\pi^{-1}(p\mathbf{Z})$ (with multiplicities).

How can we compute $\pi^{-1}(p\mathbf{Z})$ in practice?

Example 8.1.1. The following MAGMA session shows the commands needed to compute the factorization of $p\mathcal{O}_K$ in MAGMA for K the number field defined by a root of $x^5 + 7x^4 + 3x^2 - x + 1$.

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^5 + 7*x^4 + 3*x^2 - x + 1);
> OK := MaximalOrder(K);
> I := 2*OK;
> Factorization(I);
[
<Principal Prime Ideal of OK
Generator:
[2, 0, 0, 0, 0], 1>
]
> J := 5*OK;
> Factorization(J);
[
<Prime Ideal of OK
Two element generators:
[5, 0, 0, 0, 0]
[2, 1, 0, 0, 0], 1>,
<Prime Ideal of OK
Two element generators:
[5, 0, 0, 0, 0]
[3, 1, 0, 0, 0], 2>,
<Prime Ideal of OK
Two element generators:
[5, 0, 0, 0, 0]
[2, 4, 1, 0, 0], 1>
]
> [K!OK.i : i in [1..5]];
[ 1, a, a^2, a^3, a^4 ]
```

Thus $2\mathcal{O}_K$ is already a prime ideal, and

$$5\mathcal{O}_K = (5, 2+a) \cdot (5, 3+a)^2 \cdot (5, 2+4a+a^2).$$

Notice that in this example $\mathcal{O}_K = \mathbf{Z}[a]$. (Warning: There are examples of $\mathcal{O}_K$ such that $\mathcal{O}_K \neq \mathbf{Z}[a]$ for any $a \in \mathcal{O}_K$, as Example 8.1.6 below illustrates.) When $\mathcal{O}_K = \mathbf{Z}[a]$ it is very easy to factor $p\mathcal{O}_K$, as we will see below. The following factorization gives a hint as to why:

$$x^5 + 7x^4 + 3x^2 - x + 1 \equiv (x+2) \cdot (x+3)^2 \cdot (x^2 + 4x + 2) \pmod 5.$$

The exponent 2 of $(5, 3+a)^2$ in the factorization of $5\mathcal{O}_K$ above suggests "ramification", in the sense that the cover $X \to Y$ has less points (counting their "size", i.e., their residue class degree) in its fiber over 5 than it has generically. Here's a suggestive picture:

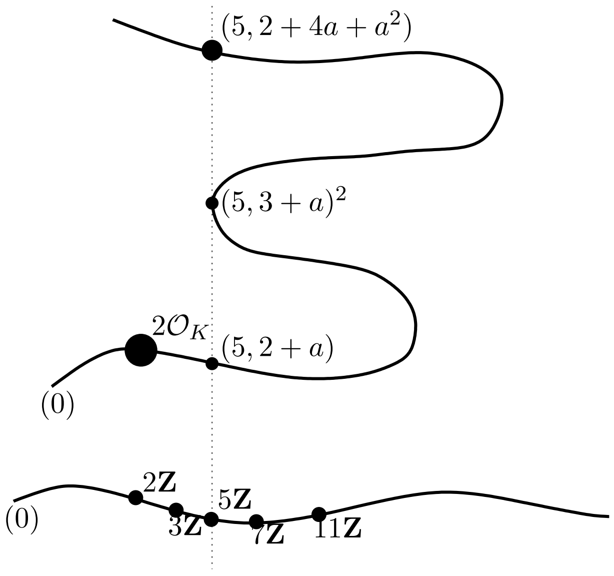


Diagram of $\mathrm{Spec}(\mathcal{O}_K) \to \mathrm{Spec}(\mathbf{Z})$

8.1.1 A Method for Factoring that Often Works

Suppose $a \in \mathcal{O}_K$ is such that $K = \mathbf{Q}(a)$, and let $g(x)$ be the minimal polynomial of a. Then $\mathbf{Z}[a] \subset \mathcal{O}_K$, and we have a diagram of schemes

$$\begin{tikzcd}
(??) \arrow[r, hook] \arrow[d] & \mathrm{Spec}(\mathcal{O}_K) \arrow[d] \\
\bigcup \mathrm{Spec}(\mathbf{F}_p[x]/(\overline{g}_i^{e_i})) \arrow[r, hook] \arrow[d] & \mathrm{Spec}(\mathbf{Z}[a]) \arrow[d] \\
\mathrm{Spec}(\mathbf{F}_p) \arrow[r, hook] & \mathrm{Spec}(\mathbf{Z})
\end{tikzcd}$$

where $\overline{g} = \prod_i \overline{g}_i^{e_i}$ is the factorization of the image of g in $\mathbf{F}_p[x]$.

The cover $\pi : \mathrm{Spec}(\mathbf{Z}[a]) \to \mathrm{Spec}(\mathbf{Z})$ is easy to understand because it is defined by the single equation $g(x)$. To give a maximal ideal $\mathfrak{p}$ of $\mathbf{Z}[a]$ such that $\pi(\mathfrak{p}) = p\mathbf{Z}$ is the same as giving a homomorphism $\varphi : \mathbf{Z}[x]/(g) \to \overline{\mathbf{F}}_p$ (up to automorphisms of the image), which is in turn the same as giving a root of g in $\overline{\mathbf{F}}_p$ (up to automorphism), which is the same as giving an irreducible factor of the reduction of g modulo p.

Lemma 8.1.2. *Suppose the index of $\mathbf{Z}[a]$ in $\mathcal{O}_K$ is coprime to p. Then the primes $\mathfrak{p}_i$ in the factorization of $p\mathbf{Z}[a]$ do not decompose further going from $\mathbf{Z}[a]$ to $\mathcal{O}_K$, so finding the prime ideals of $\mathbf{Z}[a]$ that contain p yields the factorization of $p\mathcal{O}_K$.*

Proof. Hi-brow argument: By hypothesis we have an exact sequence of abelian groups

$$0 \to \mathbf{Z}[a] \to \mathcal{O}_K \to H \to 0,$$

where H is a finite abelian group of order coprime to p. Tensor product is right exact, and there is an exact sequence

$$\mathrm{Tor}_1(H, \mathbf{F}_p) \to \mathbf{Z}[a] \otimes \mathbf{F}_p \to \mathcal{O}_K \otimes \mathbf{F}_p \to H \otimes \mathbf{F}_p \to 0,$$

and $\mathrm{Tor}_1(H, \mathbf{F}_p) = H \otimes \mathbf{F}_p = 0$, so $\mathbf{Z}[a] \otimes \mathbf{F}_p \cong \mathcal{O}_K \otimes \mathbf{F}_p$.

Low-brow argument: The inclusion map $\mathbf{Z}[a] \hookrightarrow \mathcal{O}_K$ is defined by a matrix over $\mathbf{Z}$ that has determinant $\pm[\mathcal{O}_K : \mathbf{Z}[a]]$, which is coprime to p. The reduction of this matrix modulo p is invertible, so it defines an isomorphism $\mathbf{Z}[a] \otimes \mathbf{F}_p \to \mathcal{O}_K \otimes \mathbf{F}_p$. Any homomorphism $\mathcal{O}_K \to \overline{\mathbf{F}}_p$ is the composition of a homomorphism $\mathcal{O}_K \to \mathcal{O}_K \otimes \mathbf{F}_p$ with a homomorphism $\mathcal{O}_K \otimes \mathbf{F}_p \to \overline{\mathbf{F}}_p$. Since $\mathcal{O}_K \otimes \mathbf{F}_p \cong \mathbf{Z}[a] \otimes \mathbf{F}_p$, the homomorphisms $\mathcal{O}_K \to \overline{\mathbf{F}}_p$ are in bijection with the homomorphisms $\mathbf{Z}[a] \to \overline{\mathbf{F}}_p$, which proves the lemma. □

As suggested in the proof of the lemma, we find all homomorphisms $\mathcal{O}_K \to \overline{\mathbf{F}}_p$ by finding all homomorphism $\mathbf{Z}[a] \to \overline{\mathbf{F}}_p$. In terms of ideals, if $\mathfrak{p} = (g(a), p)\mathbf{Z}[a]$ is a maximal ideal of $\mathbf{Z}[a]$, then the ideal $\mathfrak{p}' = (g(a), p)\mathcal{O}_K$ of $\mathcal{O}_K$ is also maximal, since

$$\mathcal{O}_K/\mathfrak{p}' \cong (\mathcal{O}_K \otimes \mathbf{F}_p)/(g(\tilde{a})) \cong (\mathbf{Z}[a] \otimes \mathbf{F}_p)/(g(\tilde{a})) \subset \overline{\mathbf{F}}_p.$$

We formalize the above discussion in the following theorem:

Theorem 8.1.3. *Let $f(x)$ denote the minimal polynomial of a over $\mathbf{Q}$. Suppose that $p \nmid [\mathcal{O}_K : \mathbf{Z}[a]]$ is a prime. Let*

$$\overline{f} = \prod_{i=1}^{t} \overline{f}_i^{e_i} \in \mathbf{F}_p[x]$$

where the $\overline{f}_i$ are distinct monic irreducible polynomials. Let $\mathfrak{p}_i = (p, f_i(a))$ where $f_i \in \mathbf{Z}[x]$ is a lift of $\overline{f}_i$ in $\mathbf{F}_p[X]$. Then

$$p\mathcal{O}_K = \prod_{i=1}^{t} \mathfrak{p}_i^{e_i}.$$

We return to the example from above, in which $K = \mathbf{Q}(a)$, where a is a root of $x^5+7x^4+3x^2-x+1$. According to Magma, the maximal order $\mathcal{O}_K$ has discriminant 2945785:

```
> Discriminant(MaximalOrder(K));
2945785
```

The order $\mathbf{Z}[a]$ has the same discriminant as $\mathcal{O}_K$, so $\mathbf{Z}[a] = \mathcal{O}_K$ and we can apply the above theorem.

```
> Discriminant(x^5 + 7*x^4 + 3*x^2 - x + 1);
2945785
```

We have

$$x^5 + 7x^4 + 3x^2 - x + 1 \equiv (x+2) \cdot (x+3)^2 \cdot (x^2 + 4x + 2) \pmod{5},$$

which yields the factorization of $5\mathcal{O}_K$ given before the theorem.

If we replace a by $b = 7a$, then the index of $\mathbf{Z}[b]$ in $\mathcal{O}_K$ will be a power of 7, which is coprime to 5, so the above method will still work.

```
> f:=MinimalPolynomial(7*a);
> f;
x^5 + 49*x^4 + 1029*x^2 - 2401*x + 16807
> Discriminant(f);
235050861175510968365785
> Discriminant(f)/Discriminant(MaximalOrder(K));
79792266297612001    // coprime to 5
> S<t> := PolynomialRing(GF(5));
> Factorization(S!f);
[
    <t + 1, 2>,
    <t + 4, 1>,
    <t^2 + 3*t + 3, 1>
]
```

Thus 5 factors in $\mathcal{O}_K$ as

$$5\mathcal{O}_K = (5, 7a+1)^2 \cdot (5, 7a+4) \cdot (5, (7a)^2 + 3(7a) + 3).$$

If we replace a by $b = 5a$ and try the above algorithm with $\mathbf{Z}[b]$, then the method fails because the index of $\mathbf{Z}[b]$ in $\mathcal{O}_K$ is divisible by 5.

```
> f:=MinimalPolynomial(5*a);
> f;
x^5 + 35*x^4 + 375*x^2 - 625*x + 3125
> Discriminant(f) / Discriminant(MaximalOrder(K));
95367431640625    // divisible by 5
> Factorization(S!f);
[
    <t, 5>
]
```

8.1.2 A Method for Factoring that Always Works

There are numbers fields K such that $\mathcal{O}_K$ is not of the form $\mathbf{Z}[a]$ for any $a \in K$. Even worse, Dedekind found a field K such that $2 \mid [\mathcal{O}_K : \mathbf{Z}[a]]$ for *all* $a \in \mathcal{O}_K$, so there is no choice of a such that Theorem 8.1.3 can be used to factor 2 for K (see Example 8.1.6 below).

Most algebraic number theory books do not describe an algorithm for decomposing primes in the general case. Fortunately, Cohen's book [Coh93, §6.2]) describes how to solve the general problem. The solutions are somewhat surprising, since the algorithms are much more sophisticated than the one suggested by Theorem 8.1.3. However, these complicated algorithms all run very quickly in practice, even without assuming the maximal order is already known.

For simplicity we consider the following slightly easier problem whose solution contains the key ideas: *Let $\mathcal{O}$ be any order in $\mathcal{O}_K$ and let p be a prime of $\mathbf{Z}$. Find the prime ideals of $\mathcal{O}$ that contain p.*

To go from this special case to the general case, given a prime p that we wish to factor in $\mathcal{O}_K$, we find a p-maximal order $\mathcal{O}$, i.e., an order $\mathcal{O}$ such that $[\mathcal{O}_K : \mathcal{O}]$ is coprime to p. A p-maximal order can be found very quickly in practice using the "round 2" or "round 4" algorithms. (Remark: Later we will see that to compute $\mathcal{O}_K$, we take the sum of p-maximal orders, one for every p such that p^2 divides $\mathrm{Disc}(\mathcal{O}_K)$. The time-consuming part of this computation of $\mathcal{O}_K$ is finding the primes p such that $p^2 \mid \mathrm{Disc}(\mathcal{O}_K)$, not finding the p-maximal orders. Thus a fast algorithm for factoring integers would not only break many cryptosystems, but would massively speed up computation of the ring of integers of a number field.)

Algorithm 8.1.4. Suppose $\mathcal{O}$ is an order in the ring $\mathcal{O}_K$ of integers of a number field K. For any prime $p \in \mathbf{Z}$, the following (sketch of an) algorithm computes the set of maximal ideals of $\mathcal{O}$ that contain p.

Sketch of algorithm. Let $K = \mathbf{Q}(a)$ be a number field given by an algebraic integer a as a root of its minimal monic polynomial f of degree n. We assume that an order $\mathcal{O}$ has been given by a basis $w_1, \ldots, w_n$ and that $\mathcal{O}$ that contains $\mathbf{Z}[a]$. Each of the following steps can be carried out efficiently using little more than linear algebra over $\mathbf{F}_p$. The details are in [Coh93, §6.2.5].

1. [Check if easy] If $p \nmid \mathrm{disc}(\mathbf{Z}[a])/\mathrm{disc}(\mathcal{O})$ (so $p \nmid [\mathcal{O} : \mathbf{Z}[a]]$), then by a slight modification of Theorem 8.1.3, we easily factor $p\mathcal{O}$.

2. [Compute radical] Let I be the *radical* of $p\mathcal{O}$, which is the ideal of elements $x \in \mathcal{O}$ such that $x^m \in p\mathcal{O}$ for some positive integer m. Using linear algebra over the finite field $\mathbf{F}_p$, we can quickly compute a basis for $I/p\mathcal{O}$. (We never compute $I \subset \mathcal{O}$.)

3. [Compute quotient by radical] Compute an $\mathbf{F}_p$ basis for
$$A = \mathcal{O}/I = (\mathcal{O}/p\mathcal{O})/(I/p\mathcal{O}).$$
The second equality comes from the fact that $p\mathcal{O} \subset I$, which is clear by definition. Note that $\mathcal{O}/p\mathcal{O} \cong \mathcal{O} \otimes \mathbf{F}_p$ is obtained by simply reducing the basis $w_1, \ldots, w_n$ modulo p.

4. [Decompose quotient] The ring A is a finite Artin ring with no nilpotents, so it decomposes as a product $A \cong \prod \mathbf{F}_p[x]/g_i(x)$ of fields. We can quickly find such a decomposition explicitly, as described in [Coh93, §6.2.5].

5. [Compute the maximal ideals over p] Each maximal ideal $\mathfrak{p}_i$ lying over p is the kernel of $\mathcal{O} \to A \to \mathbf{F}_p[x]/g_i(x)$.

The algorithm finds all primes of $\mathcal{O}$ that contain the radical I of $p\mathcal{O}$. Every such prime clearly contains p, so to see that the algorithm is correct, we must prove that the primes $\mathfrak{p}$ of $\mathcal{O}$ that contain p also contain I. If $\mathfrak{p}$ is a prime of $\mathcal{O}$ that contains p, then $p\mathcal{O} \subset \mathfrak{p}$. If $x \in I$ then $x^m \in p\mathcal{O}$ for some m, so $x^m \in \mathfrak{p}$ which implies that $x \in \mathfrak{p}$ by primality of $\mathfrak{p}$. Thus $\mathfrak{p}$ contains I, as required.

8.1.3 Essential Discriminant Divisors

Definition 8.1.5. A prime p is an *essential discriminant divisor* if $p \mid [\mathcal{O}_K : \mathbf{Z}[a]]$ for *every* $a \in \mathcal{O}_K$.

Since $[\mathcal{O}_K : \mathbf{Z}[a]]$ is the absolute value of $\mathrm{Disc}(f(x))/\mathrm{Disc}(\mathcal{O}_K)$, where $f(x)$ is the characteristic polynomial of $f(x)$, an essential discriminant divisor divides the discriminant of the characteristic polynomial of any element of $\mathcal{O}_K$.

Example 8.1.6 *(Dedekind).* Let $K = \mathbf{Q}(a)$ be the cubic field defined by a root a of the polynomial $f = x^3 + x^2 - 2x + 8$. We will use MAGMA, which implements the algorithm described in the previous section, to show that 2 is an essential discriminant divisor for K.

```
> K<a> := NumberField(x^3 + x^2 - 2*x + 8);
> OK := MaximalOrder(K);
> Factorization(2*OK);
[
<Prime Ideal of OK
Basis:
[2 0 0]
[0 1 0]
[0 0 1], 1>,
<Prime Ideal of OK
Basis:
[1 0 1]
[0 1 0]
[0 0 2], 1>,
<Prime Ideal of OK
Basis:
[1 0 1]
[0 1 1]
[0 0 2], 1>
]
```

Thus $2\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3$, with the $\mathfrak{p}_i$ distinct. Moreover, one can check that $\mathcal{O}_K/\mathfrak{p}_i \cong \mathbf{F}_2$. If $\mathcal{O}_K = \mathbf{Z}[a]$ for some $a \in \mathcal{O}_K$ with minimal polynomial g, then $\overline{g}(x) \in \mathbf{F}_2[x]$ must be a product of three *distinct* linear factors, which is impossible.

Chapter 9

Chinese Remainder Theorem

In this section we will prove the Chinese Remainder Theorem for rings of integers, deduce several surprising and useful consequences, then learn about discriminants, and finally norms of ideals. We will also define the class group of $\mathcal{O}_K$ and state the main theorem about it. The tools we develop here illustrate the power of what we have already proved about rings of integers, and will be used over and over again to prove other deeper results in algebraic number theory. It is essentially to understand everything we discuss in this chapter very well.

9.1 The Chinese Remainder Theorem

Recall that the Chinese Remainder Theorem from elementary number theory asserts that if $n_1, \ldots, n_r$ are integers that are coprime in pairs, and $a_1, \ldots, a_r$ are integers, then there exists an integer a such that $a \equiv a_i \pmod{n_i}$ for each $i = 1, \ldots, r$. In terms of rings, the Chinese Remainder Theorem asserts that the natural map

$$\mathbf{Z}/(n_1 \cdots n_r)\mathbf{Z} \to (\mathbf{Z}/n_1\mathbf{Z}) \oplus \cdots \oplus (\mathbf{Z}/n_r\mathbf{Z})$$

is an isomorphism. This result generalizes to rings of integers of number fields.

Lemma 9.1.1. *If I and J are coprime ideals in $\mathcal{O}_K$, then $I \cap J = IJ$.*

Proof. The ideal $I \cap J$ is the largest ideal of $\mathcal{O}_K$ that is divisible by (contained in) both I and J. Since I and J are coprime, $I \cap J$ is divisible by IJ, i.e., $I \cap J \subset IJ$. By definition of ideal $IJ \subset I \cap J$, which completes the proof. □

Remark 9.1.2. This lemma is true for any ring R and ideals $I, J \subset R$ such that $I + J = R$. For the general proof, choose $x \in I$ and $y \in J$ such that $x + y = 1$. If $c \in I \cap J$ then

$$c = c \cdot 1 = c \cdot (x + y) = cx + cy \in IJ + IJ = IJ,$$

so $I \cap J \subset IJ$, and the other inclusion is obvious by definition.

Theorem 9.1.3 (Chinese Remainder Theorem). *Suppose $I_1, \ldots, I_r$ are ideals of $\mathcal{O}_K$ such that $I_m + I_n = \mathcal{O}_K$ for any $m \neq n$. Then the natural homomorphism $\mathcal{O}_K \to \bigoplus_{n=1}^r (\mathcal{O}_K/I_n)$ induces an isomorphism*

$$\mathcal{O}_K/\left(\prod_{n=1}^r I_n\right) \to \bigoplus_{n=1}^r (\mathcal{O}_K/I_n).$$

Thus given any $a_n \in I_n$ then there exists $a \in \mathcal{O}_K$ such that $a \equiv a_n \pmod{I_n}$ for $n = 1, \ldots, r$.

Proof. First assume that we know the theorem in the case when the I_n are powers of prime ideals. Then we can deduce the general case by noting that each $\mathcal{O}_K/I_n$ is isomorphic to a product $\prod \mathcal{O}_K/\mathfrak{p}_m^{e_m}$, where $I_n = \prod \mathfrak{p}_m^{e_m}$, and $\mathcal{O}_K/(\prod_n I_n)$ is isomorphic to the product of the $\mathcal{O}_K/\mathfrak{p}^e$, where the $\mathfrak{p}$ and e run through the same prime powers as appear on the right hand side.

It thus suffices to prove that if $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ are distinct prime ideals of $\mathcal{O}_K$ and $e_1, \ldots, e_r$ are positive integers, then

$$\psi : \mathcal{O}_K/\left(\prod_{n=1}^r \mathfrak{p}_n^{e_n}\right) \to \bigoplus_{n=1}^r (\mathcal{O}_K/\mathfrak{p}_n^{e_n})$$

is an isomorphism. Let $\varphi : \mathcal{O}_K \to \oplus_{n=1}^r (\mathcal{O}_K/\mathfrak{p}_n^{e_n})$ be the natural map induced by reduction mod $\mathfrak{p}_n^{e_n}$. Then kernel of φ is $\cap_{n=1}^r \mathfrak{p}_n^{e_n}$, which by Lemma 9.1.1 is equal to $\prod_{n=1}^r \mathfrak{p}_n^{e_n}$, so ψ is injective. Note that the projection $\mathcal{O}_K \to \mathcal{O}_K/\mathfrak{p}_n^{e_n}$ of φ onto each factor is obviously surjective, so it suffices to show that the element $(1, 0, \ldots, 0)$ is in the image of φ (and the similar elements for the other factors). Since $J = \prod_{n=2}^r \mathfrak{p}_n^{e_n}$ is not divisible by $\mathfrak{p}_1$, hence not contained in $\mathfrak{p}_1$, there is an element $a \in J$ with $a \notin \mathfrak{p}_1$. Since $\mathfrak{p}_1$ is maximal, $\mathcal{O}_K/\mathfrak{p}_1$ is a field, so there exists $b \in \mathcal{O}_K$ such that $ab = 1 - c$, for some $c \in \mathfrak{p}_1$. Then

$$1 - c^{n_1} = (1 - c)(1 + c + c^2 + \cdots + c^{n_1 - 1}) = ab(1 + c + c^2 + \cdots + c^{n_1 - 1})$$

is congruent to 0 mod $\mathfrak{p}_n^{e_n}$ for each $n \geq 2$ since it is in $\prod_{n=2}^r \mathfrak{p}_n^{e_n}$, and it is congruent to 1 modulo $\mathfrak{p}_1^{n_1}$. □

Remark 9.1.4. In fact, the surjectivity part of the above proof is easy to prove for any commutative ring; indeed, the above proof illustrates how trying to prove something in a special case can result in a more complicated proof!! Suppose R is a ring and I, J are ideals in R such that $I + J = R$. Choose $x \in I$ and $y \in J$ such that $x + y = 1$. Then $x = 1 - y$ maps to $(0, 1)$ in $R/I \oplus R/J$ and $y = 1 - x$ maps to $(1, 0)$ in $R/I \oplus R/J$. Thus the map $R/(I \cap J) \to R/I \oplus R/J$ is surjective. Also, as mentioned above, $R/(I \cap J) = R/(IJ)$.

Example 9.1.5. The MAGMA command `ChineseRemainderTheorem` implements the algorithm suggested by the above theorem. In the following example, we compute a prime over (3) and a prime over (5) of the ring of integers of $\mathbf{Q}(\sqrt[3]{2})$, and find an element of $\mathcal{O}_K$ that is congruent to $\sqrt[3]{2}$ modulo one prime and 1 modulo the other.

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^3-2);
> OK := MaximalOrder(K);
> I := Factorization(3*OK)[1][1];
> J := Factorization(5*OK)[1][1];
> I;
Prime Ideal of OK
Two element generators:
    [3, 0, 0]
    [4, 1, 0]
> J;
Prime Ideal of OK
Two element generators:
    [5, 0, 0]
    [7, 1, 0]
> b := ChineseRemainderTheorem(I, J, OK!a, OK!1);
> b - a in I;
true
> b - 1 in J;
true
> K!b;
-4
```

The element found by the Chinese Remainder Theorem algorithm in this case is -4.

The following lemma is a nice application of the Chinese Remainder Theorem. We will use it to prove that every ideal of $\mathcal{O}_K$ can be generated by two elements. Suppose I is a nonzero integral ideals of $\mathcal{O}_K$. If $a \in I$, then $(a) \subset I$, so I divides (a) and the quotient $(a)/I$ is an integral ideal. The following lemma asserts that (a) can be chosen so the quotient $(a)/I$ is coprime to any given ideal.

Lemma 9.1.6. *If I, J are nonzero integral ideals in $\mathcal{O}_K$, then there exists an $a \in I$ such that $(a)/I$ is coprime to J.*

Proof. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be the prime divisors of J. For each n, let v_n be the largest power of $\mathfrak{p}_n$ that divides I. Choose an element $a_n \in \mathfrak{p}_n^{v_n}$ that is not in $\mathfrak{p}_n^{v_n+1}$ (there is such an element since $\mathfrak{p}_n^{v_n} \neq \mathfrak{p}_n^{v_n+1}$, by unique factorization). By Theorem 9.1.3, there exists $a \in \mathcal{O}_K$ such that

$$a \equiv a_n \pmod{\mathfrak{p}_n^{v_n+1}}$$

for all $n = 1, \ldots, r$ and also

$$a \equiv 0 \pmod{I/\prod \mathfrak{p}_n^{v_n}}.$$

(We are applying the theorem with the coprime integral ideals $\mathfrak{p}_n^{v_n+1}$, for $n = 1, \ldots, r$ and the integral ideal $I/\prod \mathfrak{p}_n^{v_n}$.)

To complete the proof we must show that $(a)/I$ is not divisible by any $\mathfrak{p}_n$, or equivalently, that the $\mathfrak{p}_n^{v_n}$ exactly divides (a). Because $a \equiv a_n \pmod{\mathfrak{p}_n^{v_n+1}}$, there is $b \in \mathfrak{p}_n^{v_n+1}$ such that $a = a_n + b$. Since $a_n \in \mathfrak{p}_n^{v_n}$, it follows that $a \in \mathfrak{p}_n^{v_n}$, so $\mathfrak{p}_n^{v_n}$ divides (a). If $a \in \mathfrak{p}_n^{v_n+1}$, then $a_n = a - b \in \mathfrak{p}_n^{v_n+1}$, a contradiction, so $\mathfrak{p}_n^{v_n+1}$ does not divide (a), which completes the proof. □

Suppose I is a nonzero ideal of $\mathcal{O}_K$. As an abelian group $\mathcal{O}_K$ is free of rank equal to the degree $[K : \mathbf{Q}]$ of K, and I is of finite index in $\mathcal{O}_K$, so I can be generated as an abelian group, hence as an ideal, by $[K : \mathbf{Q}]$ generators. The following proposition asserts something much better, namely that I can be generated *as an ideal* in $\mathcal{O}_K$ by at most two elements.

Proposition 9.1.7. *Suppose I is a fractional ideal in the ring $\mathcal{O}_K$ of integers of a number field. Then there exist $a, b \in K$ such that $I = (a, b)$.*

Proof. If $I = (0)$, then I is generated by 1 element and we are done. If I is not an integral ideal, then there is $x \in K$ such that xI is an integral ideal, and the number of generators of xI is the same as the number of generators of I, so we may assume that I is an integral ideal.

Let a be any nonzero element of the integral ideal I. We will show that there is some $b \in I$ such that $I = (a, b)$. Let $J = (b)$. By Lemma 9.1.6, there exists $a \in I$ such that $(a)/I$ is coprime to (b). The ideal $(a, b) = (a) + (b)$ is the greatest common divisor of (a) and (b), so I divides (a, b), since I divides both (a) and (b). Suppose $\mathfrak{p}^n$ is a prime power that divides (a, b), so $\mathfrak{p}^n$ divides both (a) and (b). Because $(a)/I$ and (b) are coprime and $\mathfrak{p}^n$ divides (b), we see that $\mathfrak{p}^n$ does not divide $(a)/I$, so $\mathfrak{p}^n$ must divide I. Thus (a, b) divides I, so $(a, b) = I$ as claimed. □

We can also use Theorem 9.1.3 to determine the $\mathcal{O}_K$-module structure of the successive quotients $\mathfrak{p}^n/\mathfrak{p}^{n+1}$.

Proposition 9.1.8. *Let $\mathfrak{p}$ be a nonzero prime ideal of $\mathcal{O}_K$, and let $n \geq 0$ be an integer. Then $\mathfrak{p}^n/\mathfrak{p}^{n+1} \cong \mathcal{O}_K/\mathfrak{p}$ as $\mathcal{O}_K$-modules.*

Proof. (Compare page 13 of Swinnerton-Dyer.) Since $\mathfrak{p}^n \neq \mathfrak{p}^{n+1}$ (by unique factorization), we can fix an element $b \in \mathfrak{p}^n$ such that $b \notin \mathfrak{p}^{n+1}$. Let $\varphi : \mathcal{O}_K \to \mathfrak{p}^n/\mathfrak{p}^{n+1}$ be the $\mathcal{O}_K$-module morphism defined by $\varphi(a) = ab$. The kernel of φ is $\mathfrak{p}$ since clearly $\varphi(\mathfrak{p}) = 0$ and if $\varphi(a) = 0$ then $ab \in \mathfrak{p}^{n+1}$, so $\mathfrak{p}^{n+1} \mid (a)(b)$, so $\mathfrak{p} \mid (a)$, since $\mathfrak{p}^{n+1}$ does not divide (b). Thus φ induces an injective $\mathcal{O}_K$-module homomorphism $\mathcal{O}_K/\mathfrak{p} \hookrightarrow \mathfrak{p}^n/\mathfrak{p}^{n+1}$.

It remains to show that φ is surjective, and this is where we will use Theorem 9.1.3. Suppose $c \in \mathfrak{p}^n$. By Theorem 9.1.3 there exists $d \in \mathcal{O}_K$ such that

$$d \equiv c \pmod{\mathfrak{p}^{n+1}} \qquad \text{and} \qquad d \equiv 0 \pmod{(b)/\mathfrak{p}^n}.$$

We have $\mathfrak{p}^n \mid (c)$ since $c \in \mathfrak{p}^n$ and $(b)/\mathfrak{p}^n \mid (d)$ by the second displayed condition, so $(b) = \mathfrak{p}^n \cdot (b)/\mathfrak{p}^n \mid (d)$, hence $d/b \in \mathcal{O}_K$. Finally

$$\varphi\left(\frac{d}{b}\right) \quad = \quad \frac{d}{b} \cdot b \pmod{\mathfrak{p}^{n+1}} \quad = \quad b \pmod{p^{n+1}} \quad = \quad c \pmod{p^{n+1}},$$

so φ is surjective. □

Chapter 10

Discrimannts, Norms, and Finiteness of the Class Group

10.1 Preliminary Remarks

Let K be a number field of degree n. Then there are n embeddings

$$\sigma_1, \ldots, \sigma_n : K \hookrightarrow \mathbf{C}.$$

Let $\sigma : K \to \mathbf{C}^n$ be the product map $a \mapsto (\sigma_1(a), \ldots, \sigma_n(a))$. Let $V = \mathbf{R}\sigma(K)$ be the $\mathbf{R}$-span of $\sigma(K)$ inside $\mathbf{C}^n$.

Proposition 10.1.1. *The $\mathbf{R}$-vector space $V = \mathbf{R}\sigma(K)$ spanned by the image $\sigma(K)$ has dimension n.*

Proof. We prove this by showing that the image $\sigma(\mathcal{O}_K)$ is discrete. If $\sigma(\mathcal{O}_K)$ were not discrete it would contain elements all of whose coordinates are simultaneously arbitrarily small. The norm of an element $a \in \mathcal{O}_K$ is the product of the entries of $\sigma(a)$, so the norms of nonzero elements of $\mathcal{O}_K$ would go to 0. This is a contradiction, since the norms of elements of $\mathcal{O}_K$ are integers.

The fact that $\sigma(\mathcal{O}_K)$ is discrete in $\mathbf{C}^n$ implies that $\mathbf{R}\sigma(\mathcal{O}_K)$ has dimension equal to the rank n of $\sigma(\mathcal{O}_K)$, as claimed. This last assertion is not obvious, and requires observing that if L if a free abelian group that is discrete in a real vector space W and $\mathbf{R}L = W$, then the rank of L equals the dimension of W. Here's why this is true. If $x_1, \ldots, x_m \in L$ are a basis for $\mathbf{R}L$, then $\mathbf{Z}x_1 + \cdots + \mathbf{Z}x_m$ has finite index in L, since otherwise there would be infinitely many elements of L in a fundamental domain for $\mathbf{Z}x_1 + \cdots + \mathbf{Z}x_m$, which would contradict discreteness of L. Thus the rank of L is $m = \dim(\mathbf{R}L)$, as claimed. □

Since $\sigma(\mathcal{O}_K)$ is a lattice in V, the volume of $V/\sigma(\mathcal{O}_K)$ is finite. Suppose $w_1, \ldots, w_n$ is a basis for $\mathcal{O}_K$. Then if A is the matrix whose ith row is $\sigma(w_i)$, then $|\operatorname{Det}(A)|$ is the volume of $V/\sigma(\mathcal{O}_K)$. (Take this determinant as the definition of the volume—we won't be using "volume" here except in a formal motivating way.)

Example 10.1.2. Let $\mathcal{O}_K = \mathbf{Z}[i]$ be the ring of integers of $K = \mathbf{Q}(i)$. Then $w_1 = 1$, $w_2 = i$ is a basis for $\mathcal{O}_K$. The map $\sigma : K \to \mathbf{C}^2$ is given by

$$\sigma(a + bi) = (a + bi, a - bi) \in \mathbf{C}^2.$$

The image $\sigma(\mathcal{O}_K)$ is spanned by $(1, 1)$ and $(i, -i)$. The volume determinant is

$$\left| \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \right| = |-2i| = 2.$$

Let $\mathcal{O}_K = \mathbf{Z}[\sqrt{2}]$ be the ring of integers of $K = \mathbf{Q}(\sqrt{2})$. The map σ is

$$\sigma(a + b\sqrt{2}) = (a + b\sqrt{2}, a - b\sqrt{2}) \in \mathbf{R}^2,$$

and

$$A = \begin{pmatrix} 1 & 1 \\ \sqrt{2} & -\sqrt{2} \end{pmatrix},$$

which has determinant $-2\sqrt{2}$, so the volume of the ring of integers is $2\sqrt{2}$.

As the above example illustrates, the volume of the ring of integers is not a great invariant of $\mathcal{O}_K$. For example, it need not be an integer. If we consider $\mathrm{Det}(A)^2$ instead, we obtain a number that is a well-defined integer which can be either positive or negative. In the next section we will do just this.

10.2 Discriminants

Suppose $w_1, \ldots, w_n$ are a basis for a number field K, which we view as a $\mathbf{Q}$-vector space. Let $\sigma : K \hookrightarrow \mathbf{C}^n$ be the embedding $\sigma(a) = (\sigma_1(a), \ldots, \sigma_n(a))$, where $\sigma_1, \ldots, \sigma_n$ are the distinct embeddings of K into $\mathbf{C}$. Let A be the matrix whose rows are $\sigma(w_1), \ldots, \sigma(w_n)$. The quantity $\mathrm{Det}(A)$ depends on the ordering of the w_i, and need not be an integer.

If we consider $\mathrm{Det}(A)^2$ instead, we obtain a number that is a well-defined integer which can be either positive or negative. Note that

$$\begin{aligned} \mathrm{Det}(A)^2 &= \mathrm{Det}(AA) = \mathrm{Det}(AA^t) \\ &= \mathrm{Det}\left(\sum_{k=1,\ldots,n} \sigma_k(w_i)\sigma_k(w_j) \right) \\ &= \mathrm{Det}(\mathrm{Tr}(w_i w_j)_{1 \le i,j \le n}), \end{aligned}$$

so $\mathrm{Det}(A)^2$ can be defined purely in terms of the trace without mentioning the embeddings σ_i. Also, changing the basis for $\mathcal{O}_K$ is the same as left multiplying A by an integer matrix U of determinant ± 1, which does not change the squared determinant, since $\mathrm{Det}(UA)^2 = \mathrm{Det}(U)^2 \, \mathrm{Det}(A)^2 = \mathrm{Det}(A)^2$. Thus $\mathrm{Det}(A)^2$ is well defined, and does not depend on the choice of basis.

If we view K as a $\mathbf{Q}$-vector space, then $(x, y) \mapsto \mathrm{Tr}(xy)$ defines a bilinear pairing $K \times K \to \mathbf{Q}$ on K, which we call the *trace pairing.* The following lemma asserts that this pairing is nondegenerate, so $\mathrm{Det}(\mathrm{Tr}(w_i w_j)) \neq 0$ hence $\mathrm{Det}(A) \neq 0$.

Lemma 10.2.1. *The trace pairing is nondegenerate.*

Proof. If the trace pairing is degenerate, then there exists $a \in K$ such that for every $b \in K$ we have $\mathrm{Tr}(ab) = 0$. In particularly, taking $b = a^{-1}$ we see that $0 = \mathrm{Tr}(aa^{-1}) = \mathrm{Tr}(1) = [K : \mathbf{Q}] > 0$, which is absurd. □

Definition 10.2.2 (Discriminant). Suppose $a_1, \ldots, a_n$ is any $\mathbf{Q}$-basis of K. The *discriminant* of $a_1, \ldots, a_n$ is

$$\mathrm{Disc}(a_1, \ldots, a_n) = \mathrm{Det}(\mathrm{Tr}(a_i a_j)_{1 \leq i,j \leq n}) \in \mathbf{Q}.$$

The *discriminant* $\mathrm{Disc}(\mathcal{O})$ of an order $\mathcal{O}$ in $\mathcal{O}_K$ is the discriminant of any basis for $\mathcal{O}$. The *discriminant* $d_K = \mathrm{Disc}(K)$ of the number field K is the discrimimant of $\mathcal{O}_K$.

Note that the discriminants defined above are all nonzero by Lemma 10.2.1.

Warning: In Magma $\mathrm{Disc}(K)$ is defined to be the discriminant of the polynomial you happened to use to define K, which is (in my opinion) a poor choice and goes against most of the literature.

The following proposition asserts that the discriminant of an order $\mathcal{O}$ in $\mathcal{O}_K$ is bigger than $\mathrm{disc}(\mathcal{O}_K)$ by a factor of the square of the index.

Proposition 10.2.3. *Suppose $\mathcal{O}$ is an order in $\mathcal{O}_K$. Then*

$$\mathrm{Disc}(\mathcal{O}) = \mathrm{Disc}(\mathcal{O}_K) \cdot [\mathcal{O}_K : \mathcal{O}]^2.$$

Proof. Let A be a matrix whose rows are the images via σ of a basis for $\mathcal{O}_K$, and let B be a matrix whose rows are the images via σ of a basis for $\mathcal{O}$. Since $\mathcal{O} \subset \mathcal{O}_K$ has finite index, there is an integer matrix C such that $CA = B$, and $|\mathrm{Det}(C)| = [\mathcal{O}_K : \mathcal{O}]$. Then

$$\mathrm{Disc}(\mathcal{O}) = \mathrm{Det}(B)^2 = \mathrm{Det}(CA)^2 = \mathrm{Det}(C)^2 \,\mathrm{Det}(A)^2 = [\mathcal{O}_K : \mathcal{O}]^2 \cdot \mathrm{Disc}(\mathcal{O}_K).$$

□

This result is enough to give an algorithm for computing $\mathcal{O}_K$, albeit a potentially slow one. Given K, find some order $\mathcal{O} \subset K$, and compute $d = \mathrm{Disc}(\mathcal{O})$. Factor d, and use the factorization to write $d = s \cdot f^2$, where f^2 is the largest square that divides d. Then the index of $\mathcal{O}$ in $\mathcal{O}_K$ is a divisor of f, and we (tediously) can enumerate all rings R with $\mathcal{O} \subset R \subset K$ and $[R : \mathcal{O}] \mid f$, until we find the largest one all of whose elements are integral.

Example 10.2.4. Consider the ring $\mathcal{O}_K = \mathbf{Z}[(1+\sqrt{5})/2]$ of integers of $K = \mathbf{Q}(\sqrt{5})$. The discriminant of the basis $1, a = (1+\sqrt{5})/2$ is

$$\mathrm{Disc}(\mathcal{O}_K) = \left|\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}\right| = 5.$$

Let $\mathcal{O} = \mathbf{Z}[\sqrt{5}]$ be the order generated by $\sqrt{5}$. Then $\mathcal{O}$ has basis $1, \sqrt{5}$, so

$$\mathrm{Disc}(\mathcal{O}) = \left|\begin{pmatrix} 2 & 0 \\ 0 & 10 \end{pmatrix}\right| = 20 = [\mathcal{O}_K : \mathcal{O}]^2 \cdot 5.$$

10.3 Norms of Ideals

In this section we extend the notion of norm to ideals. This will be helpful in proving of class groups in the next section. For example, we will prove that the group of fractional ideals modulo principal fractional ideals of a number field is finite by showing that every ideal is equivalent to an ideal with norm at most some a priori bound.

Definition 10.3.1 (Lattice Index). If L and M are two lattices in vector space V, then the *lattice index* $[L : M]$ is by definition the absolute value of the determinant of any linear automorphism A of V such that $A(L) = M$.

The lattice index has the following properties:

- If $M \subset L$, then $[L : M] = \#(L/M)$.
- If M, L, N are lattices then $[L : N] = [L : M] \cdot [M : N]$.

Definition 10.3.2 (Norm of Fractional Ideal). Suppose I is a fractional ideal of $\mathcal{O}_K$. The *norm* of I is the lattice index

$$\mathrm{Norm}(I) = [\mathcal{O}_K : I] \in \mathbf{Q}_{\geq 0},$$

or 0 if $I = 0$.

Note that if I is an integral ideal, then $\mathrm{Norm}(I) = \#(\mathcal{O}_K/I)$.

Lemma 10.3.3. *Suppose $a \in K$ and I is an integral ideal. Then*

$$\mathrm{Norm}(aI) = |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)|\,\mathrm{Norm}(I).$$

Proof. By properties of the lattice index mentioned above we have

$$[\mathcal{O}_K : aI] = [\mathcal{O}_K : I] \cdot [I : aI] = \mathrm{Norm}(I) \cdot |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)|.$$

Here we have used that $[I : aI] = |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)|$, which is because left multiplication ℓ_a is an automorphism of K that sends I onto aI, so $[I : aI] = |\,\mathrm{Det}(\ell_a)| = |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)|$. □

Proposition 10.3.4. *If I and J are fractional ideals, then*

$$\mathrm{Norm}(IJ) = \mathrm{Norm}(I) \cdot \mathrm{Norm}(J).$$

Proof. By Lemma 10.3.3, it suffices to prove this when I and J are integral ideals. If I and J are coprime, then Theorem 9.1.3 (Chinese Remainder Theorem) implies that $\mathrm{Norm}(IJ) = \mathrm{Norm}(I)\cdot\mathrm{Norm}(J)$. Thus we reduce to the case when $I = \mathfrak{p}^m$ and $J = \mathfrak{p}^k$ for some prime ideal $\mathfrak{p}$ and integers m, k. By Proposition 9.1.8 (consequence of CRT that $\mathcal{O}_K/\mathfrak{p} \cong \mathfrak{p}^n/\mathfrak{p}^{n+1}$), the filtration of $\mathcal{O}_K/\mathfrak{p}^n$ given by powers of $\mathfrak{p}$ has successive quotients isomorphic to $\mathcal{O}_K/\mathfrak{p}$, so we see that $\#(\mathcal{O}_K/\mathfrak{p}^n) = \#(\mathcal{O}_K/\mathfrak{p})^n$, which proves that $\mathrm{Norm}(\mathfrak{p}^n) = \mathrm{Norm}(\mathfrak{p})^n$. □

Lemma 10.3.5. *Fix a number field K. Let B be a positive integer. There are only finitely many integral ideals I of $\mathcal{O}_K$ with norm at most B.*

Proof. An integral ideal I is a subgroup of $\mathcal{O}_K$ of index equal to the norm of I. If G is any finitely generated abelian group, then there are only finitely many subgroups of G of index at most B, since the subgroups of index dividing an integer n are all subgroups of G that contain nG, and the group G/nG is finite. This proves the lemma. □

10.4 Finiteness of the Class Group via Geometry of Numbers

We have seen examples in which $\mathcal{O}_K$ is not a unique factorization domain. If $\mathcal{O}_K$ is a principal ideal domain, then it is a unique factorization domain, so it is of interest to understand how badly $\mathcal{O}_K$ fails to be a principal ideal domain. The class group of $\mathcal{O}_K$ measures this failure. As one sees in a course on Class Field Theory, the class group and its generalizations also yield deep insight into the possible abelian Galois extensions of K.

Definition 10.4.1 (Class Group). Let $\mathcal{O}_K$ be the ring of integers of a number field K. The *class group* C_K of K is the group of nonzero fractional ideals modulo the sugroup of principal fractional ideals (a), for $a \in K$.

Note that if we let $\mathrm{Div}(K)$ denote the group of nonzero fractional ideals, then there is an exact sequence

$$0 \to \mathcal{O}_K^* \to K^* \to \mathrm{Div}(K) \to C_K \to 0.$$

A basic theorem in algebraic number theory is that the class group C_K is finite, which follows from the first part of the following theorem and the fact that there are only finitely many ideals of norm less than a given integer.

Theorem 10.4.2 (Finiteness of the Class Group). *Let K be a number field. There is a constant $C_{r,s}$ that depends only on the number r, s of real and pairs of complex conjugate embeddings of K such that every ideal class of $\mathcal{O}_K$ contains an integral ideal of norm at most $C_{r,s}\sqrt{|d_K|}$, where $d_K = \mathrm{Disc}(\mathcal{O}_K)$. Thus by Lemma 10.3.5 the class group C_K of K is finite. One can choose $C_{r,s}$ such that every ideal class in C_K contains an integral ideal of norm at most*

$$\sqrt{|d_K|} \cdot \left(\frac{4}{\pi}\right)^s \frac{n!}{n^n}.$$

The explicit bound in the theorem is called the Minkowski bound, and I think it is the best known unconditional general bound (though there are better bounds in certain special cases).

Before proving Theorem 10.4.2, we prove a few lemmas. The strategy of the proof will be to start with any nonzero ideal I, and prove that there is some nonzero $a \in K$, with very small norm, such that aI is an integral ideal. Then $\mathrm{Norm}(aI) = \mathrm{Norm}_{K/\mathbf{Q}}(a)\,\mathrm{Norm}(I)$ will be small, since $\mathrm{Norm}_{K/\mathbf{Q}}(a)$ is small. The trick is to determine precisely how small an a we can choose subject to the condition that aI be an integral ideal, i.e., that $a \in I^{-1}$.

Let S be a subset of $V = \mathbf{R}^n$. Then S is *convex* if whenever $x, y \in S$ then the line connecting x and y lies entirely in S. We say that S is *symmetric about the origin* if whenever $x \in S$ then $-x \in S$ also. If L is a lattice in V, then the *volume* of V/L is the volume of the compact real manifold V/L, which is the same thing as the absolute value of the determinant of any matrix whose rows form a basis for L.

Lemma 10.4.3 (Blichfeld). *Let L be a lattice in $V = \mathbf{R}^n$, and let S be a bounded closed convex subset of V that is symmetric about the origin. Assume that* $\mathrm{Vol}(S) \geq 2^n\,\mathrm{Vol}(V/L)$*. Then S contains a nonzero element of L.*

Proof. First assume that $\mathrm{Vol}(S) > 2^n \cdot \mathrm{Vol}(V/L)$. If the map $\pi : \frac{1}{2}S \to V/L$ is injective, then

$$\frac{1}{2^n}\,\mathrm{Vol}(S) = \mathrm{Vol}\left(\frac{1}{2}S\right) \leq \mathrm{Vol}(V/L),$$

a contradiction. Thus π is not injective, so there exist $P_1 \neq P_2 \in \frac{1}{2}S$ such that $P_1 - P_2 \in L$. By symmetry $-P_2 \in \frac{1}{2}S$. By convexity, the average $\frac{1}{2}(P_1 - P_2)$ of P_1 and $-P_2$ is also in $\frac{1}{2}S$. Thus $0 \neq P_1 - P_2 \in S \cap L$, as claimed.

Next assume that $\mathrm{Vol}(S) = 2^n \cdot \mathrm{Vol}(V/L)$. Then for all $\varepsilon > 0$ there is $0 \neq Q_\varepsilon \in L \cap (1+\varepsilon)S$, since $\mathrm{Vol}((1+\varepsilon)S) > \mathrm{Vol}(S) = 2^n \cdot \mathrm{Vol}(V/L)$. If $\varepsilon < 1$ then the Q_ε are all in $L \cap 2S$, which is finite since $2S$ is bounded and L is discrete. Hence there exists $Q = Q_\varepsilon \in L \cap (1+\varepsilon)S$ for arbitrarily small ε. Since S is closed, $Q \in L \cap S$. □

Lemma 10.4.4. *If L_1 and L_2 are lattices in V, then*

$$\mathrm{Vol}(V/L_2) = \mathrm{Vol}(V/L_1) \cdot [L_1 : L_2].$$

Proof. Let A be an automorphism of V such that $A(L_1) = L_2$. Then A defines an isomorphism of real manifolds $V/L_1 \to V/L_2$ that changes volume by a factor of $|\mathrm{Det}(A)| = [L_1 : L_2]$. The claimed formula then follows. □

Fix a number field K with ring of integers $\mathcal{O}_K$. Let $\sigma : K \to V = \mathbf{R}^n$ be the embedding

$$\begin{aligned}\sigma(x) = \big(&\sigma_1(x), \sigma_2(x), \ldots, \sigma_r(x),\\ &\mathrm{Re}(\sigma_{r+1}(x)), \ldots, \mathrm{Re}(\sigma_{r+s}(x)), \mathrm{Im}(\sigma_{r+1}(x)), \ldots, \mathrm{Im}(\sigma_{r+s}(x))\big),\end{aligned}$$

where $\sigma_1, \ldots, \sigma_r$ are the real embeddings of K and $\sigma_{r+1}, \ldots, \sigma_{r+s}$ are half the complex embeddings of K, with one representative of each pair of complex conjugate embeddings. Note that this σ is *not* exactly the same as the one at the beginning of Section 10.2.

Lemma 10.4.5.
$$\mathrm{Vol}(V/\sigma(\mathcal{O}_K)) = 2^{-s}\sqrt{|d_K|}.$$

Proof. Let $L = \sigma(\mathcal{O}_K)$. From a basis $w_1, \ldots, w_n$ for $\mathcal{O}_K$ we obtain a matrix A whose ith row is

$$(\sigma_1(w_i), \cdots, \sigma_r(w_i), \mathrm{Re}(\sigma_{r+1}(w_i)), \ldots, \mathrm{Re}(\sigma_{r+s}(w_1)), \mathrm{Im}(\sigma_{r+1}(w_i)), \ldots, \mathrm{Im}(\sigma_{r+s}(w_1)))$$

and whose determinant has absolute value equal to the volume of V/L. By doing the following three column operations, we obtain a matrix whose rows are exactly the images of the w_i under *all* embeddings of K into $\mathbf{C}$, which is the matrix that came up when we defined d_K.

1. Add $i = \sqrt{-1}$ times each column with entries $\mathrm{Im}(\sigma_{r+j}(w_i))$ to the column with entries $\mathrm{Re}(\sigma_{r+j}(w_i))$.

2. Multiply all columns $\mathrm{Im}(\sigma_{r+j}(w_i))$ by $-2i$, thus changing the determinant by $(-2i)^s$.

3. Add each columns with entries $\mathrm{Re}(\sigma_{r+j}(w_i))$ to the the column with entries $-2i\mathrm{Im}(\sigma_{r+j}(w_i))$.

Recalling the definition of discriminant, we see that if B is the matrix constructed by the above three operations, then $\mathrm{Det}(B)^2 = d_K$. Thus

$$\mathrm{Vol}(V/L) = |\,\mathrm{Det}(A)| = |(-2i)^{-s} \cdot \mathrm{Det}(B)| = 2^{-s}\sqrt{|d_K|}.$$

□

Lemma 10.4.6. *If I is a nonzero fractional ideal for $\mathcal{O}_K$, then $\sigma(I)$ is a lattice in V, and*
$$\mathrm{Vol}(V/\sigma(I)) = 2^{-s}\sqrt{|d_K|} \cdot \mathrm{Norm}(I).$$

Proof. We know that $[\mathcal{O}_K : I] = \mathrm{Norm}(I)$ is a nonzero rational number. Lemma 10.4.5 implies that $\sigma(\mathcal{O}_K)$ is a lattice in V, since $\sigma(\mathcal{O}_K)$ has rank n as abelian group and spans V, so $\sigma(I)$ is also a lattice in V. For the volume formula, combine Lemmas 10.4.4–10.4.5 to get

$$\mathrm{Vol}(V/\sigma(I)) = \mathrm{Vol}(V/\sigma(\mathcal{O}_K)) \cdot [\mathcal{O}_K : I] = 2^{-s}\sqrt{|d_K|}\,\mathrm{Norm}(I).$$

□

Proof of Theorem 10.4.2. Let K be a number field with ring of integers $\mathcal{O}_K$, let $\sigma : K \hookrightarrow V \cong \mathbf{R}^n$ be as above, and let $f : V \to \mathbf{R}$ be the function defined by

$$f(x_1, \ldots, x_n) = |x_1 \cdots x_r \cdot (x_{r+1}^2 + x_{(r+1)+s}^2) \cdots (x_{r+s}^2 + x_n^2).$$

Notice that if $x \in K$ then $f(\sigma(x)) = |\,\mathrm{Norm}_{K/\mathbf{Q}}(x)|$.

Let $S \subset V$ be any closed, bounded, convex, subset that is symmetric with respect to the origin and has positive volume. Since S is closed and bounded,

$$M = \max\{f(x) : x \in S\}$$

exists.

Suppose I is any nonzero fractional ideal of $\mathcal{O}_K$. Our goal is to prove there is an integral ideal aI with small norm. We will do this by finding an appropriate $a \in I^{-1}$. By Lemma 10.4.6,

$$c = \mathrm{Vol}(V/I^{-1}) = \frac{2^{-s}\sqrt{|d_K|}}{\mathrm{Norm}(I)}.$$

Let $\lambda = 2 \cdot \left(\frac{c}{v}\right)^{1/n}$, where $v = \mathrm{Vol}(S)$. Then

$$\mathrm{Vol}(\lambda S) = \lambda^n \,\mathrm{Vol}(S) = 2^n \frac{c}{v} \cdot v = 2^n \cdot c = 2^n \,\mathrm{Vol}(V/I^{-1}),$$

so by Lemma 10.4.3 there exists $0 \neq a \in I^{-1} \cap \lambda S$. Since M is the largest norm of an element of S, the largest norm of an element of $I^{-1} \cap \lambda S$ is at most $\lambda^n M$, so

$$|\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| \leq \lambda^n M.$$

Since $a \in I^{-1}$, we have $aI \subset \mathcal{O}_K$, so aI is an integral ideal of $\mathcal{O}_K$ that is equivalent to I, and

$$\begin{aligned}
\mathrm{Norm}(aI) &= |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| \cdot \mathrm{Norm}(I)\\
&\leq \lambda^n M \cdot \mathrm{Norm}(I)\\
&\leq 2^n \frac{c}{v} M \cdot \mathrm{Norm}(I)\\
&\leq 2^n \cdot 2^{-s}\sqrt{|d_K|} \cdot M \cdot v^{-1}\\
&= 2^{r+s}\sqrt{|d_K|} \cdot M \cdot v^{-1}.
\end{aligned}$$

Notice that the right hand side is independent of I. It depends only on r, s, $|d_K|$, and our choice of S. This completes the proof of the theorem, except for the assertion that S can be chosen to give the claim at the end of the theorem, which we leave as an exercise. □

Corollary 10.4.7. *Suppose that $K \neq \mathbf{Q}$ is a number field. Then $|d_K| > 1$.*

Proof. Applying Theorem 10.4.2 to the unit ideal, we get the bound

$$1 \leq \sqrt{|d_K|} \cdot \left(\frac{4}{\pi}\right)^s \frac{n!}{n^n}.$$

Thus

$$\sqrt{|d_K|} \geq \left(\frac{\pi}{4}\right)^s \frac{n^n}{n!},$$

and the right hand quantity is strictly bigger than 1 for any $s \leq n/2$ and any $n > 1$ (exercise). □

10.4.1 An Open Problem

Conjecture 10.4.8. *There are infinitely many number fields K such that the class group of K has order* 1.

For example, if we consider real quadratic fields $K = \mathbf{Q}(\sqrt{d})$, with d positive and square free, many class numbers are probably 1, as suggested by the MAGMA output below. It looks like 1's will keep appearing infinitely often, and indeed Cohen and Lenstra conjecture that they do. Nobody has found a way to prove this yet.

```
> for d in [2..1000] do
     if d eq SquareFree(d) then
        h := ClassNumber(NumberField(x^2-d));
        if h eq 1 then
           printf "%o, ", d;
        end if;
     end if;
   end for;

2, 3, 5, 6, 7, 11, 13, 14, 17, 19, 21, 22, 23, 29, 31, 33, 37,
38, 41, 43, 46, 47, 53, 57, 59, 61, 62, 67, 69, 71, 73, 77, 83,
86, 89, 93, 94, 97, 101, 103, 107, 109, 113, 118, 127, 129, 131,
133, 134, 137, 139, 141, 149, 151, 157, 158, 161, 163, 166, 167,
173, 177, 179, 181, 191, 193, 197, 199, 201, 206, 209, 211, 213,
214, 217, 227, 233, 237, 239, 241, 249, 251, 253, 262, 263, 269,
271, 277, 278, 281, 283, 293, 301, 302, 307, 309, 311, 313, 317,
329, 331, 334, 337, 341, 347, 349, 353, 358, 367, 373, 379, 381,
382, 383, 389, 393, 397, 398, 409, 413, 417, 419, 421, 422, 431,
433, 437, 446, 449, 453, 454, 457, 461, 463, 467, 478, 479, 487,
489, 491, 497, 501, 502, 503, 509, 517, 521, 523, 526, 537, 541,
542, 547, 553, 557, 563, 566, 569, 571, 573, 581, 587, 589, 593,
597, 599, 601, 607, 613, 614, 617, 619, 622, 631, 633, 641, 643,
647, 649, 653, 661, 662, 669, 673, 677, 681, 683, 691, 694, 701,
709, 713, 717, 718, 719, 721, 734, 737, 739, 743, 749, 751, 753,
757, 758, 766, 769, 773, 781, 787, 789, 797, 809, 811, 813, 821,
823, 827, 829, 838, 849, 853, 857, 859, 862, 863, 869, 877, 878,
881, 883, 886, 887, 889, 893, 907, 911, 913, 917, 919, 921, 926,
929, 933, 937, 941, 947, 953, 958, 967, 971, 973, 974, 977, 983,
989, 991, 997, 998,
```

In contrast, if we look at class numbers of quadratic imaginary fields, only a few at the beginning have class number 1.

```
> for d in [1..1000] do
     if d eq SquareFree(d) then
        h := ClassNumber(NumberField(x^2+d));
```

```
           if h eq 1 then
              printf "%o, ", d;
           end if;
        end if;
     end for;
  1, 2, 3, 7, 11, 19, 43, 67, 163
```

It is a theorem that the above list of 9 fields is the complete list with class number 1. More generally, it is possible (in theory), using deep work of Gross, Zagier, and Goldfeld involving zeta functions and elliptic curves, to enumerate all quadratic number fields with a given class number.

Chapter 11

Computing Class Groups

In this chapter we discuss how to compute class groups in some examples, then introduce the group of units. We will prove the main structure theorem for the group of units in the next chapter.

11.1 Remarks on Computing the Class Group

If $\mathfrak{p}$ is a prime of $\mathcal{O}_K$, then the intersection $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ is a prime ideal of $\mathbf{Z}$. We say that $\mathfrak{p}$ *lies over* $p \in \mathbf{Z}$. Note $\mathfrak{p}$ lies over $p \in \mathbf{Z}$ if and only if $\mathfrak{p}$ is one of the prime factors in the factorization of the ideal $p\mathcal{O}_K$. Geometrically, $\mathfrak{p}$ is a point of $\mathrm{Spec}(\mathcal{O}_K)$ that lies over the point $p\mathbf{Z}$ of $\mathrm{Spec}(\mathbf{Z})$ under the map induced by the inclusion $\mathbf{Z} \hookrightarrow \mathcal{O}_K$.

Lemma 11.1.1. *Let K be a number field with ring of integers $\mathcal{O}_K$. Then the class group $\mathrm{Cl}(K)$ is generated by the prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$ lying over primes $p \in \mathbf{Z}$ with $p \leq B_K = \sqrt{|d_K|} \cdot \left(\frac{4}{\pi}\right)^s \cdot \frac{n!}{n^n}$, where s is the number of complex conjugate pairs of embeddings $K \hookrightarrow \mathbf{C}$.*

Proof. We proved before that every ideal class in $\mathrm{Cl}(K)$ is represented by an ideal I with $\mathrm{Norm}(I) \leq B_K$. Write $I = \prod_{i=1}^{m} \mathfrak{p}_i^{e_i}$, with each $e_i \geq 1$. Then by multiplicativity of the norm, each $\mathfrak{p}_i$ also satisfies $\mathrm{Norm}(\mathfrak{p}_i) \leq B_K$. If $\mathfrak{p}_i \cap \mathbf{Z} = p\mathbf{Z}$, then $p \mid \mathrm{Norm}(\mathfrak{p}_i)$, since p is the residue characteristic of $\mathcal{O}_K/\mathfrak{p}$, so $p \leq B_K$. Thus I is a product of primes $\mathfrak{p}$ that satisfies the norm bound of the lemma, whcih proves the lemma. □

This is a sketch of how to compute $\mathrm{Cl}(K)$:

1. Use the "factoring primes" algorithm to list all prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$ that appear in the factorization of a prime $p \in \mathbf{Z}$ with $p \leq B_K$.

2. Find the group generated by the ideal classes $[\mathfrak{p}]$, where the $\mathfrak{p}$ are the prime ideals found in step 1. (In general, one must think more carefully about how to do this step.)

The following three examples illustrate computation of $\mathrm{Cl}(K)$ for $K = \mathbf{Q}(i), \mathbf{Q}(\sqrt{5})$ and $\mathbf{Q}(\sqrt{-6})$.

Example 11.1.2. We compute the class group of $K = \mathbf{Q}(i)$. We have

$$n = 2, \quad r = 0, \quad s = 1, \quad d_K = -4,$$

so

$$B_K = \sqrt{4} \cdot \left(\frac{4}{\pi}\right)^1 \cdot \left(\frac{2!}{2^2}\right) = \frac{8}{\pi} < 3.$$

Thus $\mathrm{Cl}(K)$ is generated by the prime divisors of 2. We have

$$2\mathcal{O}_K = (1+i)^2,$$

so $\mathrm{Cl}(K)$ is generated by the principal prime ideal $\mathfrak{p} = (1+i)$. Thus $\mathrm{Cl}(K) = 0$ is trivial.

Example 11.1.3. We compute the class group of $K = \mathbf{Q}(\sqrt{5})$. We have

$$n = 2, \quad r = 2, \quad s = 0, \quad d_K = 5,$$

so

$$B = \sqrt{5} \cdot \left(\frac{4}{\pi}\right)^0 \cdot \left(\frac{2!}{2^2}\right) < 3.$$

Thus $\mathrm{Cl}(K)$ is generated by the primes that divide 2. We have $\mathcal{O}_K = \mathbf{Z}[\gamma]$, where $\gamma = \frac{1+\sqrt{5}}{2}$ satisfies $x^2 - x - 1$. The polynomial $x^2 - x - 1$ is irreducible mod 2, so $2\mathcal{O}_K$ is prime. Since it is principal, we see that $\mathrm{Cl}(K) = 1$ is trivial.

Example 11.1.4. In this example, we compute the class group of $K = \mathbf{Q}(\sqrt{-6})$. We have

$$n = 2, \quad r = 0, \quad s = 1, \quad d_K = -24,$$

so

$$B = \sqrt{24} \cdot \frac{4}{\pi} \cdot \left(\frac{2!}{2^2}\right) \sim 3.1.$$

Thus $\mathrm{Cl}(K)$ is generated by the prime ideals lying over 2 and 3. We have $\mathcal{O}_K = \mathbf{Z}[\sqrt{-6}]$, and $\sqrt{-6}$ satisfies $x^2 + 6 = 0$. Factoring $x^2 + 6$ modulo 2 and 3 we see that the class group is generated by the prime ideals

$$\mathfrak{p}_2 = (2, \sqrt{-6}) \qquad \text{and} \qquad \mathfrak{p}_3 = (3, \sqrt{-6}).$$

Also, $\mathfrak{p}_2^2 = 2\mathcal{O}_K$ and $\mathfrak{p}_3^2 = 3\mathcal{O}_K$, so $\mathfrak{p}_2$ and $\mathfrak{p}_3$ define elements of order dividing 2 in $\mathrm{Cl}(K)$.

Is either $\mathfrak{p}_2$ or $\mathfrak{p}_3$ principal? Fortunately, there is an easier norm trick that allows us to decide. Suppose $\mathfrak{p}_2 = (\alpha)$, where $\alpha = a + b\sqrt{-6}$. Then

$$2 = \mathrm{Norm}(\mathfrak{p}_2) = |\,\mathrm{Norm}(\alpha)| = (a + b\sqrt{-6})(a - b\sqrt{-6}) = a^2 + 6b^2.$$

Trying the first few values of $a, b \in \mathbf{Z}$, we see that this equation has no solutions, so $\mathfrak{p}_2$ can not be principal. By a similar argument, we see that $\mathfrak{p}_3$ is not principal either. Thus $\mathfrak{p}_2$ and $\mathfrak{p}_3$ define elements of order 2 in $\mathrm{Cl}(K)$.

Does the class of $\mathfrak{p}_2$ equal the class of $\mathfrak{p}_3$? Since $\mathfrak{p}_2$ and $\mathfrak{p}_3$ define classes of order 2, we can decide this by finding the class of $\mathfrak{p}_2 \cdot \mathfrak{p}_3$. We have

$$\mathfrak{p}_2 \cdot \mathfrak{p}_3 = (2, \sqrt{-6}) \cdot (3, \sqrt{-6}) = (6, 2\sqrt{-6}, 3\sqrt{-6}) \subset (\sqrt{-6}).$$

The ideals on both sides of the inclusion have norm 6, so by multiplicativity of the norm, they must be the same ideal. Thus $\mathfrak{p}_2 \cdot \mathfrak{p}_3 = (\sqrt{-6})$ is principal, so $\mathfrak{p}_2$ and $\mathfrak{p}_3$ represent the same element of $\mathrm{Cl}(K)$. We conclude that

$$\mathrm{Cl}(K) = \langle \mathfrak{p}_2 \rangle = \mathbf{Z}/2\mathbf{Z}.$$

Chapter 12

Dirichlet's Unit Theorem

In this chapter we will prove the main structure theorem for the group of units of the ring of integers of a number field. The answer is remarkably simple: if K has r real and s complex embeddings, then

$$\mathcal{O}_K^* \approx \mathbf{Z}^{r+s-1} \oplus W,$$

where W is the finite cyclic group of roots of unity in K. Examples will follow on Thursday (application: the solutions to Pell's equation $x^2 - dy^2 = 1$, for $d > 1$ squarefree, form a free abelian group of rank 1).

12.1 The Group of Units

Definition 12.1.1 (Unit Group). The *group of units* U_K associated to a number field K is the group of elements of $\mathcal{O}_K$ that have an inverse in $\mathcal{O}_K$.

Theorem 12.1.2 (Dirichlet). *The group U_K is the product of a finite cyclic group of roots of unity with a free abelian group of rank $r + s - 1$, where r is the number of real embeddings of K and s is the number of complex conjugate pairs of embeddings.*

We prove the theorem by defining a map $\varphi : U_K \to \mathbf{R}^{r+s}$, and showing that the kernel of φ is finite and the image of φ is a lattice in a hyperplane in $\mathbf{R}^{r+s}$. The trickiest part of the proof is showing that the image of φ spans a hyperplane, and we do this by a clever application of Blichfeldt's lemma (that if S is closed, bounded, symmetric, etc., and has volume at least $2^n \cdot \mathrm{Vol}(V/L)$, then $S \cap L$ contains a nonzero element).

Remark 12.1.3. Theorem 12.1.2 is due to Dirichlet who lived 1805–1859. Thomas Hirst described Dirichlet as follows:

> He is a rather tall, lanky-looking man, with moustache and beard about to turn grey with a somewhat harsh voice and rather deaf. He was unwashed, with his cup of coffee and cigar. One of his failings is forgetting time, he pulls his watch out, finds it past three, and runs out without even finishing the sentence.

Koch wrote that:

> ... important parts of mathematics were influenced by Dirichlet. His proofs characteristically started with surprisingly simple observations, followed by extremely sharp analysis of the remaining problem.

I think Koch's observation nicely describes the proof we will give of Theorem 12.1.2.

The following proposition explains how to think about units in terms of the norm.

Proposition 12.1.4. *An element $a \in \mathcal{O}_K$ is a unit if and only if* $\mathrm{Norm}_{K/\mathbf{Q}}(a) = \pm 1$.

Proof. Write $\mathrm{Norm} = \mathrm{Norm}_{K/\mathbf{Q}}$. If a is a unit, then a^{-1} is also a unit, and $1 = \mathrm{Norm}(a)\,\mathrm{Norm}(a^{-1})$. Since both $\mathrm{Norm}(a)$ and $\mathrm{Norm}(a^{-1})$ are integers, it follows that $\mathrm{Norm}(a) = \pm 1$. Conversely, if $a \in \mathcal{O}_K$ and $\mathrm{Norm}(a) = \pm 1$, then the equation $aa^{-1} = 1 = \pm\,\mathrm{Norm}(a)$ implies that $a^{-1} = \pm\,\mathrm{Norm}(a)/a$. But $\mathrm{Norm}(a)$ is the product of the images of a in $\mathbf{C}$ by all embeddings of K into $\mathbf{C}$, so $\mathrm{Norm}(a)/a$ is also a product of images of a in $\mathbf{C}$, hence a product of algebraic integers, hence an algebraic integer. Thus $a^{-1} \in \mathcal{O}_K$, which proves that a is a unit. □

Let r be the number of real and s the number of complex conjugate embeddings of K into $\mathbf{C}$, so $n = [K : \mathbf{Q}] = r + 2s$. Define a map

$$\varphi : U_K \to \mathbf{R}^{r+s}$$

by

$$\varphi(a) = (\log|\sigma_1(a)|, \ldots, \log|\sigma_{r+s}(a)|).$$

Lemma 12.1.5. *The image of φ lies in the hyperplane*

$$H = \{(x_1, \ldots, x_{r+s}) \in \mathbf{R}^{r+s} : x_1 + \cdots + x_r + 2x_{r+1} + \cdots + 2x_{r+s} = 0\}. \quad (12.1.1)$$

Proof. If $a \in U_K$, then by Proposition 12.1.4,

$$\left(\prod_{i=1}^{r} |\sigma_i(a)|\right) \cdot \left(\prod_{i=r+1}^{s} |\sigma_i(a)|^2\right) = 1.$$

Taking logs of both sides proves the lemma. □

Lemma 12.1.6. *The kernel of φ is finite.*

Proof. We have

$$\begin{aligned} \mathrm{Ker}(\varphi) &\subset \{a \in \mathcal{O}_K : |\sigma_i(a)| = 1 \text{ for all } i = 1, \ldots, r+2s\} \\ &\subset \sigma(\mathcal{O}_K) \cap X, \end{aligned}$$

where X is the bounded subset of $\mathbf{R}^{r+2s}$ of elements all of whose coordinates have absolute value at most 1. Since $\sigma(\mathcal{O}_K)$ is a lattice (see Proposition 5.2.4), the intersection $\sigma(\mathcal{O}_K) \cap X$ is finite, so $\mathrm{Ker}(\varphi)$ is finite. □

Lemma 12.1.7. *The kernel of φ is a finite cyclic group.*

Proof. It is a general fact that any finite subgroup of the multiplicative group of a field is cyclic. [Homework.] □

To prove Theorem 12.1.2, it suffices to proove that $\mathrm{Im}(\varphi)$ is a lattice in the hyperplane H from (12.1.1), which we view as a vector space of dimension $r+s-1$.

Define an embedding

$$\sigma : K \hookrightarrow \mathbf{R}^n \tag{12.1.2}$$

given by $\sigma(x) = (\sigma_1(x), \ldots, \sigma_{r+s}(x))$, where we view $\mathbf{C} \cong \mathbf{R} \times \mathbf{R}$ via $a + bi \mapsto (a, b)$. Note that this is exactly the same as the embedding

$$x \mapsto \big(\sigma_1(x), \sigma_2(x), \ldots, \sigma_r(x),\\ \mathrm{Re}(\sigma_{r+1}(x)), \ldots, \mathrm{Re}(\sigma_{r+s}(x)), \mathrm{Im}(\sigma_{r+1}(x)), \ldots, \mathrm{Im}(\sigma_{r+s}(x))\big),$$

from before, except that we have re-ordered the last s imaginary components to be next to their corresponding real parts.

Lemma 12.1.8. *The image of φ is discrete in $\mathbf{R}^{r+s}$.*

Proof. Suppose X is any bounded subset of $\mathbf{R}^{r+s}$. Then for any $u \in Y = \varphi^{-1}(X)$ the coordinates of $\sigma(u)$ are bounded in terms of X (since log is an increasing function). Thus $\sigma(Y)$ is a bounded subset of $\mathbf{R}^n$. Since $\sigma(Y) \subset \sigma(\mathcal{O}_K)$, and $\sigma(\mathcal{O}_K)$ is a lattice in $\mathbf{R}^n$, it follows that $\sigma(Y)$ is finite. Since σ is injective, Y is finite, and φ has finite kernel, so $\varphi(U_K) \cap X$ is finite, which implies that $\varphi(U_K)$ is discrete. □

To finish the proof of Theorem 12.1.2, we will show that the image of φ spans H. Let W be the $\mathbf{R}$-span of the image $\varphi(U_K)$, and note that W is a subspace of H. We will show that $W = H$ indirectly by showing that if $v \notin H^\perp$, where $\perp$ is with respect to the dot product on $\mathbf{R}^{r+s}$, then $v \notin W^\perp$. This will show that $W^\perp \subset H^\perp$, hence that $H \subset W$, as required.

Thus suppose $z = (z_1, \ldots, z_{r+s}) \notin H^\perp$. Define a function $f : K^* \to \mathbf{R}$ by

$$f(x) = z_1 \log|\sigma_1(x)| + \cdots z_{r+s} \log|\sigma_{r+s}(x)|. \tag{12.1.3}$$

To show that $z \notin W^\perp$ we show that there exists some $u \in U_K$ with $f(u) \neq 0$.

Let

$$A = \sqrt{|d_K|} \cdot \left(\frac{2}{\pi}\right)^s \in \mathbf{R}_{>0}.$$

Choose any positive real numbers $c_1, \ldots, c_{r+s} \in \mathbf{R}_{>0}$ such that

$$c_1 \cdots c_r \cdot (c_{r+1} \cdots c_{r+s})^2 = A.$$

Let

$$S = \{(x_1, \ldots, x_n) \in \mathbf{R}^n :\\ |x_i| \leq c_i \text{ for } 1 \leq i \leq r,\\ |x_i^2 + x_{i+s}^2| \leq c_i^2 \text{ for } r < i \leq r+s\} \subset \mathbf{R}^n.$$

Then S is closed, bounded, convex, symmetric with respect to the origin, and of dimension $r+2s$, since S is a product of r intervals and s discs, each of which has these properties. Viewing S as a product of intervals and discs, we see that the volume of S is

$$\mathrm{Vol}(S) = \prod_{i=1}^{r}(2c_i) \cdot \prod_{i=1}^{s}(\pi c_i^2) = 2^r \cdot \pi^s \cdot A.$$

Recall *Blichfeldt's lemma* that if L is a lattice and S is closed, bounded, etc., and has volume at least $2^n \cdot \mathrm{Vol}(V/L)$, then $S \cap L$ contains a nonzero element. To apply this lemma, we take $L = \sigma(\mathcal{O}_K) \subset \mathbf{R}^n$, where σ is as in (12.1.2). We showed, when proving finiteness of the class group, that $\mathrm{Vol}(\mathbf{R}^n/L) = 2^{-s}\sqrt{|d_K|}$. To check the hypothesis to Blichfeld's lemma, note that

$$\mathrm{Vol}(S) = 2^{r+s}\sqrt{|d_K|} = 2^n 2^{-s}\sqrt{|d_K|} = 2^n \,\mathrm{Vol}(\mathbf{R}^n/L).$$

Thus there exists a nonzero element $a \in S \cap \sigma(\mathcal{O}_K)$, i.e., a nonzero $a \in \mathcal{O}_K$ such that $|\sigma_i(a)| \leq c_i$ for $1 \leq i \leq r+s$. We then have

$$\begin{aligned} |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| &= \left|\prod_{i=1}^{r+2s} \sigma_i(a)\right| \\ &= \prod_{i=1}^{r} |\sigma_i(a)| \cdot \prod_{i=r+1}^{s} |\sigma_i(a)|^2 \\ &\leq c_1 \cdots c_r \cdot (c_{r+1} \cdots c_{r+s})^2 = A. \end{aligned}$$

Since $a \in \mathcal{O}_K$ is nonzero, we also have

$$|\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| \geq 1.$$

Moreover, if for any $i \leq r$, we have $|\sigma_i(a)| < \frac{c_i}{A}$, then

$$1 \leq |\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| < c_1 \cdots \frac{c_i}{A} \cdots c_r \cdot (c_{r+1} \cdots c_{r+s})^2 = \frac{A}{A} = 1,$$

a contradiction, so $|\sigma_i(a)| \geq \frac{c_i}{A}$ for $i = 1, \ldots, r$. Likewise, $|\sigma_i(a)|^2 \geq \frac{c_i^2}{A}$, for $i = r+1, \ldots, r+s$. Rewriting this we have

$$\frac{c_i}{|\sigma_i(a)|} \leq A \quad \text{for } i \leq r \quad \text{and} \quad \left(\frac{c_i}{|\sigma_i(a)|}\right)^2 \leq A \quad \text{for } i = r+1, \ldots, r+s.$$

Our strategy is to use an appropriately chosen a to construct a unit $u \in U_K$ such $f(u) \neq 0$. First, let $b_1, \ldots, b_m$ be representative generators for the finitely many nonzero principal ideals of $\mathcal{O}_K$ of norm at most A. Since $|\,\mathrm{Norm}_{K/\mathbf{Q}}(a)| \leq A$, we have $(a) = (b_j)$, for some j, so there is a unit $u \in \mathcal{O}_K$ such that $a = ub_j$.

Let

$$s = s(c_1, \ldots, c_{r+s}) = z_1 \log(c_1) + \cdots + z_{r+s} \log(c_{r+s}),$$

and recall $f : K^* \to \mathbf{R}$ defined in (12.1.3) above. We first show that

$$|f(u) - s| \leq B = |f(b_j)| + \log(A) \cdot \left(\sum_{i=1}^{r} |z_i| + \frac{1}{2} \cdot \sum_{i=r+1}^{s} |z_i| \right). \qquad (12.1.4)$$

We have

$$\begin{aligned}
|f(u) - s| &= |f(a) - f(b_j) - s| \\
&\leq |f(b_j)| + |s - f(a)| \\
&= |f(b_j)| + |z_1(\log(c_1) - \log(|\sigma_1(a)|)) + \cdots + z_{r+s}(\log(c_{r+s}) - \log(|\sigma_{r+s}(a)|))| \\
&= |f(b_j)| + |z_1 \cdot \log(c_1/|\sigma_1(a)|) + \cdots + \frac{z_{r+s}}{2} \cdot \log((c_{r+s}/|\sigma_{r+s}(a)|)^2)| \\
&\leq |f(b_j)| + \log(A) \cdot \left(\sum_{i=1}^{r} |z_i| + \frac{1}{2} \cdot \sum_{i=r+1}^{s} |z_i| \right).
\end{aligned}$$

The amazing thing about (12.1.4) is that the bound B on the right hand side does not depend on the c_i. Suppose we can choose positive real numbers c_i such that

$$c_1 \cdots c_r \cdot (c_{r+1} \cdots c_{r+s})^2 = A$$

and $s = s(c_1, \ldots, c_{r+s})$ is such that $|s| > B$. Then $|f(u) - s| \leq B$ would imply that $|f(u)| > 0$, which is exactly what we aimed to prove. It is possible to choose such c_i, by proceeding as follows. If $r + s = 1$, then we are trying to prove that $\varphi(U_K)$ is a lattice in $\mathbf{R}^0 = \mathbf{R}^{r+s-1}$, which is automatically true, so assume $r + s > 1$. Then there are at least two distinct c_i. Let j be such that $z_j \neq 0$ (which exists since $z \neq 0$). Then $|z_j \log(c_j)| \to \infty$ as $c_j \to \infty$, so we choose c_j very large and the other c_i, for $i \neq j$, in any way we want subject to the condition

$$\prod_{i=1, i\neq j}^{r} c_i \cdot \prod_{i=r+1}^{s} c_i^2 = \frac{A}{c_j}.$$

Since it is possible to choose the c_i as needed, it is possible to find a unit u such that $f(u) > 0$. We conclude that $z \notin W^\perp$, so $W^\perp \subset Z^\perp$, whence $Z \subset W$, which finishes the proof Theorem 12.1.2.

12.2 Finishing the proof of Dirichlet's Unit Theorem

We begin by finishing Dirichlet's proof that the group of units U_K of $\mathcal{O}_K$ is isomorphic to $\mathbf{Z}^{r+s-1} \oplus \mathbf{Z}/m\mathbf{Z}$, where r is the number of real embeddings, s is half the number of complex embeddings, and m is the number of roots of unity in K. Recall that we defined a map $\varphi : U_K \to \mathbf{R}^{r+s}$ by

$$\varphi(x) = (\log |\sigma_1(x)|, \ldots, \log |\sigma_{r+s}(x)|).$$

Without much trouble, we proved that the kernel of φ if finite and the image φ is discrete, and in the last section we were finishing the proof that the image of φ spans the subspace H of elements of $\mathbf{R}^{r+s}$ that are orthogonal to $v = (1, \ldots, 1, 2, \ldots, 2)$, where r of the entries are 1's and s of them are 2's. The somewhat indirect route we followed was to suppose

$$z \notin H^{\perp} = \mathrm{Span}(v),$$

i.e., that z is not a multiple of v, and prove that z is not orthogonal to some element of $\varphi(U_K)$. Writing $W = \mathrm{Span}(\varphi(U_K))$, this would show that $W^{\perp} \subset H^{\perp}$, so $H \subset W$. We ran into two problems: (1) we ran out of time, and (2) the notes contained an incomplete argument that a quantity $s = s(c_1, \ldots, c_{r+s})$ can be chosen to be arbitrarily large. We will finish going through a complete proof, then compute many examples of unit groups using Magma.

Recall that $f : K^* \to \mathbf{R}$ was defined by

$$f(x) = z_1 \log |\sigma_1(x)| + \cdots + z_{r+s} \log |\sigma_{r+s}(x)| = z \bullet \varphi(x) \qquad \text{(dot product)},$$

and our goal is to show that there is a $u \in U_K$ such that $f(u) \neq 0$.

Our strategy is to use an appropriately chosen a to construct a unit $u \in U_K$ such $f(u) \neq 0$. Recall that we used Blichfeld's lemma to find an $a \in \mathcal{O}_K$ such that $1 \leq |\operatorname{Norm}_{K/\mathbf{Q}}(a)| \leq A$, and

$$\frac{c_i}{|\sigma_i(a)|} \leq A \quad \text{for } i \leq r \quad \text{and} \quad \left(\frac{c_i}{|\sigma_i(a)|}\right)^2 \leq A \quad \text{for } i = r+1, \ldots, r+s. \tag{12.2.1}$$

Let $b_1, \ldots, b_m$ be representative generators for the finitely many nonzero principal ideals of $\mathcal{O}_K$ of norm at most $A = A_K = \sqrt{|d_K|} \cdot \left(\frac{2}{\pi}\right)^s$. Modify the b_i to have the property that $|f(b_i)|$ is minimal among generators of (b_i) (this is possible because ideals are discrete). Note that the set $\{|f(b_i)| : i = 1, \ldots, m\}$ depends only on A. Since $|\operatorname{Norm}_{K/\mathbf{Q}}(a)| \leq A$, we have $(a) = (b_j)$, for some j, so there is a unit $u \in \mathcal{O}_K$ such that $a = ub_j$.

Let

$$s = s(c_1, \ldots, c_{r+s}) = z_1 \log(c_1) + \cdots + z_{r+s} \log(c_{r+s}) \in \mathbf{R}.$$

Lemma 12.2.1. *We have*

$$|f(u) - s| \leq B = \max_i(|f(b_i)|) + \log(A) \cdot \left(\sum_{i=1}^{r} |z_i| + \frac{1}{2} \cdot \sum_{i=r+1}^{s} |z_i|\right),$$

and B depends only on K and our fixed choice of $z \in H^{\perp}$.

Proof. By properties of logarithms, $f(u) = f(a/b_j) = f(a) - f(b_j)$. We next use the triangle inequality $|a + b| \leq |a| + |b|$ in various ways, properties of logarithms, and the bounds (12.2.1) in the following computation:

$$\begin{aligned}|f(u)-s| &= |f(a)-f(b_j)-s|\\ &\le |f(b_j)|+|s-f(a)|\\ &= |f(b_j)|+|z_1(\log(c_1)-\log(|\sigma_1(a)|))+\cdots+z_{r+s}(\log(c_{r+s})-\log(|\sigma_{r+s}(a)|))|\\ &= |f(b_j)|+|z_1\cdot\log(c_1/|\sigma_1(a)|)+\cdots+\frac{1}{2}\cdot z_{r+s}\log((c_{r+s}/|\sigma_{r+s}(a)|)^2)|\\ &\le |f(b_j)|+\log(A)\cdot\left(\sum_{i=1}^{r}|z_i|+\frac{1}{2}\cdot\sum_{i=r+1}^{s}|z_i|\right).\end{aligned}$$

The inequality of the lemma now follows. That B only depends on K and our choice of z follows from the formula for A and how we chose the b_i. □

The amazing thing about Lemma 12.2.1 is that the bound B on the right hand side does not depend on the c_i. Suppose we could somehow cleverly choose the positive real numbers c_i in such a way that

$$c_1\cdots c_r\cdot(c_{r+1}\cdots c_{r+s})^2=A \qquad\text{and}\qquad |s(c_1,\ldots,c_{r+s})|>B.$$

Then the facts that $|f(u)-s|\le B$ and $|s|>B$ would together imply that $|f(u)|>0$ (since $f(u)$ is closer to s than s is to 0), which is exactly what we aimed to prove. We finish the proof by showing that it is possible to choose such c_i. Note that if we change the c_i, then a could change, hence the j such that a/b_j is a unit could change, but the b_j don't change, just the subscript j. Also note that if $r+s=1$, then we are trying to prove that $\varphi(U_K)$ is a lattice in $\mathbf{R}^0=\mathbf{R}^{r+s-1}$, which is automatically true, so we may assume that $r+s>1$.

Lemma 12.2.2. *Assume $r+s>1$. Then there is a choice of $c_1,\ldots,c_{r+s}\in\mathbf{R}_{>0}$ such that*

$$|z_1\log(c_1)+\cdots+z_{r+s}\log(c_{r+s})|>B.$$

Proof. It is easier if we write

$$\begin{aligned}&z_1\log(c_1)+\cdots+z_{r+s}\log(c_{r+s})=\\ &\qquad z_1\log(c_1)+\cdots+z_r\log(c_r)+\frac{1}{2}\cdot z_{r+1}\log(c_{r+1}^2)+\cdots+\frac{1}{2}\cdot z_{r+s}\log(c_{r+s}^2)\\ &\qquad = w_1\log(d_1)+\cdots+w_r\log(d_r)+w_{r+1}\log(d_{r+1})+\cdots+\cdot w_{r+s}\log(d_{r+s}),\end{aligned}$$

where $w_i=z_i$ and $d_i=c_i$ for $i\le r$, and $w_i=\frac{1}{2}z_i$ and $d_i=c_i^2$ for $r<i\le s$,

The condition that $z\notin H^\perp$ is that the w_i are not all the same, and in our new coordinates the lemma is equivalent to showing that $|\sum_{i=1}^{r+s}w_i\log(d_i)|>B$, subject to the condition that $\prod_{i=1}^{r+s}d_i=A$. Order the w_i so that $w_1\ne0$. By hypothesis there exists a w_j such that $w_j\ne w_1$, and again re-ordering we may assume that

$j = 2$. Set $d_3 = \cdots = d_{r+s} = 1$. Then $d_1 d_2 = A$ and $\log(1) = 0$, so

$$\begin{aligned}\left|\sum_{i=1}^{r+s} w_i \log(d_i)\right| &= |w_1 \log(d_1) + w_2 \log(d_2)| \\ &= |w_1 \log(d_1) + w_2 \log(A/d_1)| \\ &= |(w_1 - w_2)\log(d_1) + w_2 \log(A)|\end{aligned}$$

Since $w_1 \neq w_2$, we have $|(w_1 - w_2)\log(d_1) + w_2 \log(A)| \to \infty$ as $d_1 \to \infty$. □

12.3 Some Examples of Units in Number Fields

The classical Pell's equation is, given square-free $d > 0$, to find all positive integer solutions (x, y) to the equation $x^2 - dy^2 = 1$. Note that if $x + y\sqrt{d} \in \mathbf{Q}(\sqrt{d})$, then

$$\mathrm{Norm}(x + y\sqrt{d}) = (x + y\sqrt{d})(x - y\sqrt{d}) = x^2 - dy^2.$$

The solutions to Pell's equation thus form a finite-index subgroup of the group of units in the ring of integers of $\mathbf{Q}(\sqrt{d})$. Dirichlet's unit theorem implies that for any d the solutions to Pell's equation form an infinite cyclic group, a fact that takes substantial work to prove using only elementary number theory (for example, using continued fractions).

We first solve the Pell equation $x^2 - 5y^2 = 1$ by finding the units of a field using Magma (we will likely discuss algorithms for computing unit groups later in the course...).

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^2-5);
> G, phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/2 + Z
Defined on 2 generators
Relations:
    2*G.1 = 0
> K!phi(G.1);
-1
> u := K!phi(G.2); u;
1/2*(a + 1)
> u^2;
1/2*(a + 3)
> u^3;
a + 2
> Norm(u);
-1
> Norm(u^3);
```

```
-1
> Norm(u^6);
1
> fund := u^6;
> fund;
4*a + 9
> 9^2 - 5*4^2;
1
> fund^2;
72*a + 161
> fund^3;
1292*a + 2889
> fund^4;
23184*a + 51841
> fund^5;
416020*a + 930249
```

I think in practice for solving Pell's equation it's best to use the ideas in the paper [Len02]. A review of this paper says: "This wonderful article begins with history and some elementary facts and proceeds to greater and greater depth about the existence of solutions to Pell equations and then later the algorithmic issues of finding those solutions. The cattle problem is discussed, as are modern smooth number methods for solving Pell equations and the algorithmic issues of representing very large solutions in a reasonable way." You can get the paper freely online from the Notices web page.

The simplest solutions to Pell's equation can be huge, even when d is quite small. Read Lenstra's paper for some awesome examples from antiquity.

```
K<a> := NumberField(x^2-NextPrime(10^7));
> G, phi := UnitGroup(K);
> K!phi(G.2);
    163580259880346328225592238121094625499142677693142915506747253O\
    0034006410036576787289043881624927126642399817503030943657 56\
    1063163927237760168060379588379147781761197418407544570282 37\
    89975945910042889569323816504809803 9*a +
    5172866928858149674701706723683467983036290343735752029750 75\
    6050587149580808939912744279034480986438365128783512278562 69\
    0868566790783049793210477650310733452599026227120591649690 08\
    633603603640331175663456220418293622 2240930
```

The Magma `Signature` command returns the number of real and complex conjugate embeddings of K into $\mathbf{C}$. The command `UnitGroup`, which we used above, returns the unit group U_K as an abstract abelian group and a homomorphism $U_K \to \mathcal{O}_K$. Note that we have to bang (!) into K to get the units as elements of K.

First we consider $K = \mathbf{Q}(i)$.

```
> R<x> := PolynomialRing(RationalField());
> K<a> := NumberField(x^2+1);
> Signature(K);
0 1    // r=0, s=1
> G,phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/4
Defined on 1 generator
Relations:
    4*G.1 = 0
> K!phi(G.1);
-a
```

Next we consider $K = \mathbf{Q}(\sqrt[3]{2})$.

```
> K<a> := NumberField(x^3-2);
> Signature(K);
1 1
> G,phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/2 + Z
Defined on 2 generators
Relations:
    2*G.1 = 0
> K!phi(G.2);
-a + 1
```

The Conjugates command returns the sequence $(\sigma_1(x), \ldots, \sigma_{r+2s}(x))$ of all embeddings of $x \in K$ into $\mathbf{C}$. The Logs command returns the sequence

$$(\log(|\sigma_1(x)|), \ldots, \log(|\sigma_{r+s}(x)|)).$$

Continuing the above example, we have

```
> Conjugates(K!phi(G.2));
[ -0.2599210498948731647672106072782283505702514647009999999999995,
1.6299605249474365823836053036391141752851257323513843923104 -
1.09112363597172140356007261418980888132587333874018547370560*i,
1.6299605249474365823836053036391141752851257323513843923104 +
1.09112363597172140356007261418980888132587333874018547370560*i ]
> Logs(K!phi(G.2));   // image of infinite order unit -- generates a lattice
[ -1.34737734832938410091818789144565304628306227332099999999989\
, 0.673688674164692050459093945722826523141531136660328899999 ]
> Logs(K!phi(G.1));   // image of -1
[ 0.E-57, 0.E-57 ]
```

Let's try a field such that $r + s - 1 = 2$. First, one with $r = 0$ and $s = 3$:

```
> K<a> := NumberField(x^6+x+1);
> Signature(K);
0 3
> G, phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/2 + Z + Z
Defined on 3 generators
Relations:
    2*G.1 = 0
> u1 := K!phi(G.2); u1;
a
> u2 := K!phi(G.3); u2;
-2*a^5 - a^3 + a^2 + a
> Logs(u1);
[ 0.11877157353322375762475480482285510811783185904379239999998,
0.04864390975267339963515094053332998614834212839311989999997,
-0.16741548328589715725990574535618509426617398743691229999999 ]
> Logs(u2);
[ 1.6502294567845884711894772749682228152154948421589999999997,
-2.0963853913452777953249166008337095194338210890229999999997,
0.44615593456068932413543932586548670421832624686433469999994 ]
```

Notice that the log image of u_1 is clearly not a real multiple of the log image of u_2 (e.g., the scalar would have to be positive because of the first coefficient, but negative because of the second). This illustrates the fact that the log images of u_1 and u_2 span a two-dimensional space.

Next we compute a field with $r = 3$ and $s = 0$. (A field with $s = 0$ is called "totally real".)

```
> K<a> := NumberField(x^3 + x^2 - 5*x - 1);
> Signature(K);
3 0
> G, phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/2 + Z + Z
Defined on 3 generators
Relations:
    2*G.1 = 0
> u1 := K!phi(G.2); u1;
1/2*(a^2 + 2*a - 1)
> u2 := K!phi(G.3); u2;
a
> Logs(u1);
```

```
[ 1.16761574692758757159598251863681302946987760474899999999995,
-0.392848724581398261291798625834359518758414226430443699999996,
-0.774767022346189310304183892802453510711463378318176699999998 ]
> Logs(u2);
[ 0.643542946228861877385181722768646725775795402446308199999,
-1.64022415032231714691015055517008505755834642266699999999999,
0.996681204093455269524968832401438331782551020220549899999998 ]
```

A family of fields with $r = 0$ (totally complex) is the *cyclotomic fields* $\mathbf{Q}(\zeta_n)$. The degree of $\mathbf{Q}(\zeta_n)$ over $\mathbf{Q}$ is $\varphi(n)$ and $r = 0$, so $s = \varphi(n)/2$ (assuming $n > 2$).

```
> K := CyclotomicField(11); K;
Cyclotomic Field of order 11 and degree 10
> G, phi := UnitGroup(K);
> G;
Abelian Group isomorphic to Z/22 + Z + Z + Z + Z
Defined on 5 generators
Relations:
    22*G.1 = 0
> u := K!phi(G.2); u;
zeta_11^9 + zeta_11^8 + zeta_11^7 + zeta_11^6 + zeta_11^5 +
    zeta_11^3 + zeta_11^2 + zeta_11 + 1
> Logs(u);
[ -1.25656632417872848745322215929976803991663080388899999999969,
0.651796894033140007971792388468509918282328440230327399999,
-0.185330046559862140949221639201972215564315421718192699999999,
0.520284982030074939330698573411850755138895506527223699999998,
0.269814494675375681099952836621379582059722278850091599999993 ]
> K!phi(G.3);
zeta_11^9 + zeta_11^7 + zeta_11^6 + zeta_11^5 + zeta_11^4 +
    zeta_11^3 + zeta_11^2 + zeta_11 + 1
> K!phi(G.4);
zeta_11^9 + zeta_11^8 + zeta_11^7 + zeta_11^6 + zeta_11^5 +
    zeta_11^4 + zeta_11^3 + zeta_11^2 + zeta_11
> K!phi(G.5);
zeta_11^9 + zeta_11^8 + zeta_11^7 + zeta_11^6 + zeta_11^5 +
    zeta_11^4 + zeta_11^2 + zeta_11 + 1
```

How far can we go computing unit groups of cyclotomic fields directly with Magma?

```
> time G,phi := UnitGroup(CyclotomicField(13));
Time: 2.210
> time G,phi := UnitGroup(CyclotomicField(17));
Time: 8.600
```

```
> time G,phi := UnitGroup(CyclotomicField(23));
.... I waited over 10 minutes (usage of 300MB RAM) and gave up.
```

12.4 Preview

In the next chapter we will study extra structure in the case when K is Galois over $\mathbf{Q}$; the results are nicely algebraic, beautiful, and have interesting ramifications. We'll learn about Frobenius elements, the Artin symbol, decomposition groups, and how the Galois group of K is related to Galois groups of residue class fields. These are the basic structures needed to make any sense of representations of Galois groups, which is at the heart of much of number theory.

Chapter 13

Decomposition and Inertia Groups

13.1 Galois Extensions

Suppose $K \subset \mathbf{C}$ is a number field. Then K is *Galois* if every field homomorphism $K \to \mathbf{C}$ has image K, or equivalently, $\#\operatorname{Aut}(K) = [K : \mathbf{Q}]$. More generally, we have the following definition.

Definition 13.1.1 (Galois). An extension K/L of number fields is *Galois* if $\#\operatorname{Aut}(K/L) = [K : L]$, where $\operatorname{Aut}(K/L)$ is the group of automorphisms of K that fix L. We write $\operatorname{Gal}(K/L) = \operatorname{Aut}(K/L)$.

For example, $\mathbf{Q}$ is Galois (over itself), any quadratic extension K/L is Galois, since it is of the form $L(\sqrt{a})$, for some $a \in L$, and the nontrivial embedding is induced by $\sqrt{a} \mapsto -\sqrt{a}$, so there is always one nontrivial automorphism. If $f \in L[x]$ is an irreducible cubic polynomial, and a is a root of f, then one proves in a course in Galois theory that $L(a)$ is Galois over L if and only if the discriminant of f is a perfect square in L. Random number fields of degree bigger than 2 are rarely Galois (I will not justify this claim further in this course).

If $K/\mathbf{Q}$ is a number field, then the Galois closure K^{gc} of K is the field generated by all images of K under all embeddings in $\mathbf{C}$ (more generally, if K/L is an extension, the Galois closure of K over L is the field generated by images of embeddings $K \to \mathbf{C}$ that are the identity map on L). If $K = \mathbf{Q}(a)$, then K^{gc} is generated by each of the conjugates of a, and is hence Galois over $\mathbf{Q}$, since the image under an embedding of any polynomial in the conjugates of a is again a polynomial in conjugates of a.

How much bigger can the degree of K^{gc} be as compared to the degree of $K = \mathbf{Q}(a)$? There is a natural embedding of $\operatorname{Gal}(K^{\mathrm{gc}}/\mathbf{Q})$ into the group of permutations of the conjugates of a. If there are n conjugates of a, then this is an embedding $\operatorname{Gal}(K^{\mathrm{gc}}/\mathbf{Q}) \hookrightarrow S_n$, where S_n is the symmetric group on n symbols, which has order $n!$. Thus the degree of the K^{gc} over $\mathbf{Q}$ is a divisor of $n!$. Also the Galois group is a transitive subgroup of S_n, which constrains the possibilities further. When

$n = 2$, we recover the fact that quadratic extensions are Galois. When $n = 3$, we see that the Galois closure of a cubic extension is either the cubic extension or a quadratic extension of the cubic extension. It turns out that that Galois closure of a cubic extension is obtained by adjoining the square root of the discriminant. For an extension K of degree 5, it is "frequently" the case that the Galois closure has degree 120, and in fact it is a difficult and interesting problem to find examples of degree 5 extension in which the Galois closure has degree smaller than 120 (according to Magma: the only possibilities for the order of a transitive proper subgroup of S_5 are 5, 10, 20, and 60; there are five transitive subgroups of S_5 out of the total of 19 subgroups of S_5).

Let n be a positive integer. Consider the field $K = \mathbf{Q}(\zeta_n)$, where $\zeta_n = e^{2\pi i/n}$ is a primitive nth root of unity. If $\sigma : K \to \mathbf{C}$ is an embedding, then $\sigma(\zeta_n)$ is also an nth root of unity, and the group of nth roots of unity is cyclic, so $\sigma(\zeta_n) = \zeta_n^m$ for some m which is invertible modulo n. Thus K is Galois and $\text{Gal}(K/\mathbf{Q}) \hookrightarrow (\mathbf{Z}/n\mathbf{Z})^*$. However, $[K : \mathbf{Q}] = n$, so this map is an isomorphism. (Side note: Taking a p-adic limit and using the maps $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \text{Gal}(\mathbf{Q}(\zeta_{p^r})/\mathbf{Q})$, we obtain a homomorphism $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_p^*$, which is called the p-adic cyclotomic character.)

Compositums of Galois extensions are Galois. For example, the biquadratic field $K = \mathbf{Q}(\sqrt{5}, \sqrt{-1})$ is a Galois extension of $\mathbf{Q}$ of degree 4.

Fix a number field K that is Galois over a subfield L. Then the Galois group $G = \text{Gal}(K/L)$ acts on many of the object that we have associated to K, including:

- the integers $\mathcal{O}_K$,
- the units U_K,
- the group of nonzero fractional ideals of $\mathcal{O}_K$,
- the class group $\text{Cl}(K)$, and
- the set $S_\mathfrak{p}$ of prime ideals $\mathfrak{P}$ lying over a given prime $\mathfrak{p}$ of $\mathcal{O}_L$.

In the next section we will be concerned with the action of $\text{Gal}(K/L)$ on $S_\mathfrak{p}$, though actions on each of the other objects, especially $\text{Cl}(K)$, will be of further interest.

13.2 Decomposition of Primes

Fix a prime $\mathfrak{p} \subset \mathcal{O}_K$ and write $\mathfrak{p}\mathcal{O}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g}$, so $S_\mathfrak{p} = \{\mathfrak{P}_1, \dots, \mathfrak{P}_g\}$.

Definition 13.2.1 (Residue class degree). Suppose $\mathfrak{P}$ is a prime of $\mathcal{O}_K$ lying over $\mathfrak{p}$. Then the *residue class degree* of $\mathfrak{P}$ is

$$f_{\mathfrak{P}/\mathfrak{p}} = [\mathcal{O}_K/\mathfrak{P} : \mathcal{O}_L/\mathfrak{p}],$$

i.e., the degree of the extension of residue class fields.

If $M/K/L$ is a tower of field extensions and $\mathfrak{q}$ is a prime of M over $\mathfrak{P}$, then

$$f_{\mathfrak{q}/\mathfrak{p}} = [\mathcal{O}_M/\mathfrak{q} : \mathcal{O}_L/\mathfrak{p}] = [\mathcal{O}_M/\mathfrak{q} : \mathcal{O}_K/\mathfrak{P}] \cdot [\mathcal{O}_K/\mathfrak{P} : \mathcal{O}_L/\mathfrak{p}] = f_{\mathfrak{q}/\mathfrak{P}} \cdot f_{\mathfrak{P}/\mathfrak{p}},$$

so the residue class degree is multiplicative in towers.

Note that if $\sigma \in \mathrm{Gal}(K/L)$ and $\mathfrak{P} \in S_p$, then σ induces an isomorphism of finite fields $\mathcal{O}_K/\mathfrak{P} \to \mathcal{O}_K/\sigma(\mathfrak{P})$ that fixes the common subfield $\mathcal{O}_L/\mathfrak{p}$. Thus the residue class degrees of $\mathfrak{P}$ and $\sigma(\mathfrak{P})$ are the same. In fact, much more is true.

Theorem 13.2.2. *Suppose K/L is a Galois extension of number fields, and let $\mathfrak{p}$ be a prime of $\mathcal{O}_L$. Write $\mathfrak{p}\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{P}_i^{e_i}$, and let $f_i = f_{\mathfrak{P}_i/\mathfrak{p}}$. Then $G = \mathrm{Gal}(K/L)$ acts transitively on the set $S_{\mathfrak{p}}$ of primes $\mathfrak{P}_i$,*

$$e_1 = \cdots = e_g, \qquad f_1 = \cdots = f_g,$$

and $efg = [K : L]$, where e is the common value of the e_i and f is the common value of the f_i.

Proof. For simplicity, we will give the proof only in the case $L = \mathbf{Q}$, but the proof works in general. Suppose $p \in \mathbf{Z}$ and $p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_g^{e_g}$, and $S = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_g\}$. We will first prove that G acts transitively on S. Let $\mathfrak{p} = \mathfrak{p}_i$ for some i. Recall that we proved long ago, using the Chinese Remainder Theorem (Theorem 9.1.3) that there exists $a \in \mathfrak{p}$ such that $(a)/\mathfrak{p}$ is an integral ideal that is coprime to $p\mathcal{O}_K$. The product

$$I = \prod_{\sigma \in G} \sigma((a)/\mathfrak{p}) = \prod_{\sigma \in G} \frac{(\sigma(a))\mathcal{O}_K}{\sigma(\mathfrak{p})} = \frac{(\mathrm{Norm}_{K/\mathbf{Q}}(a))\mathcal{O}_K}{\displaystyle\prod_{\sigma \in G} \sigma(\mathfrak{p})} \qquad (13.2.1)$$

is a nonzero integral $\mathcal{O}_K$ ideal since it is a product of nonzero integral $\mathcal{O}_K$ ideals. Since $a \in \mathfrak{p}$ we have that $\mathrm{Norm}_{K/\mathbf{Q}}(a) \in \mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$. Thus the numerator of the rightmost expression in (13.2.1) is divisible by $p\mathcal{O}_K$. Also, because $(a)/\mathfrak{p}$ is coprime to $p\mathcal{O}_K$, each $\sigma((a)/\mathfrak{p})$ is coprime to $p\mathcal{O}_K$ as well. Thus I is coprime to $p\mathcal{O}_K$. Thus the denominator of the rightmost expression in (13.2.1) must also be divisibly by $p\mathcal{O}_K$ in order to cancel the $p\mathcal{O}_K$ in the numerator. Thus for any i we have

$$\prod_{j=1}^{g} \mathfrak{p}_j^{e_j} = p\mathcal{O}_K \;\Big|\; \prod_{\sigma \in G} \sigma(\mathfrak{p}_i),$$

which in particular implies that G acts transitively on the $\mathfrak{p}_i$.

Choose some j and suppose that $k \neq j$ is another index. Because G acts transitively, there exists $\sigma \in G$ such that $\sigma(\mathfrak{p}_k) = \mathfrak{p}_j$. Applying σ to the factorization $p\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$, we see that

$$\prod_{i=1}^{g} \mathfrak{p}_i^{e_i} = \prod_{i=1}^{g} \sigma(\mathfrak{p}_i)^{e_i}.$$

Taking $\mathrm{ord}_{\mathfrak{p}_j}$ on both sides we get $e_j = e_k$. Thus $e_1 = e_2 = \cdots = e_g$.

As was mentioned right before the statement of the theorem, for any $\sigma \in G$ we have $\mathcal{O}_K/\mathfrak{p}_i \cong \mathcal{O}_K/\sigma(\mathfrak{p}_i)$, so by transitivity $f_1 = f_2 = \cdots = f_g$. Since $\mathcal{O}_K$ is a lattice in K, we have

$$\begin{aligned}[K:\mathbf{Q}] &= \dim_{\mathbf{Z}} \mathcal{O}_K = \dim_{\mathbf{F}_p} \mathcal{O}_K/p\mathcal{O}_K \\ &= \dim_{\mathbf{F}_p}\left(\bigoplus_{i=1}^{g} \mathcal{O}_K/\mathfrak{p}_i^{e_i}\right) = \sum_{i=1}^{g} e_i f_i = efg,\end{aligned}$$

which completes the proof. □

The rest of this section illustrates the theorem for quadratic fields and a cubic field and its Galois closure.

13.2.1 Quadratic Extensions

Suppose $K/\mathbf{Q}$ is a quadratic field. Then K is Galois, so for each prime $p \in \mathbf{Z}$ we have $2 = efg$. There are exactly three possibilties:

- **Ramified:** $e = 2$, $f = g = 1$: The prime p ramifies in $\mathcal{O}_K$, so $p\mathcal{O}_K = \mathfrak{p}^2$. There are only finitely many such primes, since if $f(x)$ is the minimal polynomial of a generator for $\mathcal{O}_K$, then p ramifies if and only if $f(x)$ has a multiple root modulo p. However, $f(x)$ has a multiple root modulo p if and only if p divides the discriminant of $f(x)$, which is nonzero because $f(x)$ is irreducible over $\mathbf{Z}$. (This argument shows there are only finitely many ramified primes in any number field. In fact, we will later show that the ramified primes are exactly the ones that divide the discriminant.)

- **Inert:** $e = 1$, $f = 2$, $g = 1$: The prime p is inert in $\mathcal{O}_K$, so $p\mathcal{O}_K = \mathfrak{p}$ is prime. This happens 50% of the time, which is suggested by quadratic reciprocity (but not proved this way), as we will see illustrated below for a particular example.

- **Split:** $e = f = 1$, $g = 2$: The prime p splits in $\mathcal{O}_K$, in the sense that $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$ with $\mathfrak{p}_1 \neq \mathfrak{p}_2$. This happens the other 50% of the time.

Suppose, in particular, that $K = \mathbf{Q}(\sqrt{5})$, so $\mathcal{O}_K = \mathbf{Z}[\gamma]$, where $\gamma = (1+\sqrt{5})/2$. Then $p = 5$ is ramified, since $p\mathcal{O}_K = (\sqrt{5})^2$. More generally, the order $\mathbf{Z}[\sqrt{5}]$ has index 2 in $\mathcal{O}_K$, so for any prime $p \neq 2$ we can determine the factorization of p in $\mathcal{O}_K$ by finding the factorization of the polynomial $x^2 - 5 \in \mathbf{F}_p[x]$. The polynomial $x^2 - 5$ splits as a product of two distinct factors in $\mathbf{F}_p[x]$ if and only if $e = f = 1$ and $g = 2$. For $p \neq 2, 5$ this is the case if and only if 5 is a square in $\mathbf{F}_p$, i.e., if $\left(\frac{5}{p}\right) = 1$, where $\left(\frac{5}{p}\right)$ is $+1$ if 5 is a square mod p and -1 if 5 is not. By quadratic reciprocity,

$$\left(\frac{5}{p}\right) = (-1)^{\frac{5-1}{2}\cdot\frac{p-1}{2}} \cdot \left(\frac{p}{5}\right) = \left(\frac{p}{5}\right) = \begin{cases} +1 & \text{if } p \equiv \pm 1 \pmod{5} \\ -1 & \text{if } p \equiv \pm 2 \pmod{5}. \end{cases}$$

Thus whether p splits or is inert in $\mathcal{O}_K$ is determined by the residue class of p modulo 5.

13.2.2 The Cube Roots of Two

Suppose $K/\mathbf{Q}$ is not Galois. Then e_i, f_i, and g are defined for each prime $p \in \mathbf{Z}$, but we need not have $e_1 = \cdots = e_g$ or $f_1 = \cdots = f_g$. We do still have that $\sum_{i=1}^{g} e_i f_i = n$, by the Chinese Remainder Theorem.

For example, let $K = \mathbf{Q}(\sqrt[3]{2})$. We know that $\mathcal{O}_K = \mathbf{Z}[\sqrt[3]{2}]$. Thus $2\mathcal{O}_K = (\sqrt[3]{2})^3$, so for 2 we have $e = 3$ and $f = g = 1$. To factor $3\mathcal{O}_K$, we note that working modulo 3 we have

$$x^3 - 2 = (x-2)(x^2+2x+1) = (x-2)(x+1)^2 \in \mathbf{F}_3[x],$$

so

$$3\mathcal{O}_K = (3, \sqrt[3]{2} - 2) \cdot (3, \sqrt[3]{2} + 1)^2.$$

Thus $e_1 = 1$, $e_2 = 2$, $f_1 = f_2 = 1$, and $g = 2$. Next, working modulo 5 we have

$$x^3 - 2 = (x+2)(x^2+3x+4) \in \mathbf{F}_5[x],$$

and the quadratic factor is irreducible. Thus

$$5\mathcal{O}_K = (5, \sqrt[3]{2} + 2) \cdot (5, \sqrt[3]{2}^2 + 3\sqrt[3]{2} + 4).$$

Thus here $e_1 = e_2 = 1$, $f_1 = 1$, $f_2 = 2$, and $g = 2$.

Next we consider what happens in the Galois closure of K. Since the three embeddings of $\sqrt[3]{2}$ in $\mathbf{C}$ are $\sqrt[3]{2}$, $\zeta_3\sqrt[3]{2}$, and $\zeta_3^2\sqrt[3]{2}$, we have

$$M = K^{\mathrm{gc}} = \mathbf{Q}(\sqrt[3]{2}, \zeta_3) = K.L,$$

where $L = \mathbf{Q}(\zeta_3) = \mathbf{Q}(\sqrt{-3})$, since $\zeta_3 = (-1+\sqrt{-3})/2$ is a primitive cube root of unity. The notation $K.L$ means the "compositum of K and L", which is the smallest field generated by K and L.

Let's figure out e, f, and g for the prime $p = 3$ relative to the degree six Galois field $M/\mathbf{Q}$ by using Theorem 13.2.2 and what we can easily determine about K and L. First, we know that $efg = 6$. We have $3\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2^2$, so $3\mathcal{O}_M = \mathfrak{p}_1\mathcal{O}_M \cdot (\mathfrak{p}_2\mathcal{O}_M)^2$, and the prime factors of $\mathfrak{p}_1\mathcal{O}_M$ are disjoint from the prime factors of $\mathfrak{p}_2\mathcal{O}_M$. Thus $e > 1$ is even and also $g > 1$. The only possibility for e, f, g satisfying these two conditions is $e = 2$, $f = 1$, $g = 3$, so we conclude that $3\mathcal{O}_M = \mathfrak{q}_1^2\mathfrak{q}_2^2\mathfrak{q}_3^2$ without doing any further work, and without actually knowing the $\mathfrak{q}_i$ explicitly.

Here's another interesting deduction that we can make "by hand". Suppose for the moment that $\mathcal{O}_M = \mathbf{Z}[\sqrt[3]{2}, \zeta_3]$ (this will turn out to be false). Then the factorization of $(\sqrt{-3}) \subset \mathcal{O}_L$ in $\mathcal{O}_M$ would be exactly reflected by the factorization of $x^3 - 2$ in $\mathbf{F}_3 = \mathcal{O}_L/(\sqrt{-3})$. Modulo 3 we have $x^3 - 2 = x^3 + 1 = (x+1)^3$, which would imply that $(\sqrt{-3}) = \mathfrak{q}^3$ for some prime $\mathfrak{q}$ of $\mathcal{O}_M$, i.e., that $e = 6$ and $f = g = 1$, which is incorrect. Thus $\mathcal{O}_M \neq \mathbf{Z}[\sqrt[3]{2}, \zeta_3]$. Indeed, this conclusion agrees with the following Magma computation, which asserts that $[\mathcal{O}_M : \mathbf{Z}[\sqrt[3]{2}, \zeta_3]] = 24$:

```
> R<x> := PolynomialRing(RationalField());
> K := NumberField(x^3-2);
> L := NumberField(x^2+3);
> M := CompositeFields(K,L)[1];
> O_M := MaximalOrder(M);
> a := M!K.1;
> b := M!L.1;
> O := Order([a,b]);
> Index(O_M,O);
24
```

Chapter 14

Decomposition Groups and Galois Representations

14.1 The Decomposition Group

Suppose K is a number field that is Galois over $\mathbf{Q}$ with group $G = \mathrm{Gal}(K/\mathbf{Q})$. Fix a prime $\mathfrak{p} \subset \mathcal{O}_K$ lying over $p \in \mathbf{Z}$.

Definition 14.1.1 (Decomposition group). The *decomposition group* of $\mathfrak{p}$ is the subgroup

$$D_\mathfrak{p} = \{\sigma \in G : \sigma(\mathfrak{p}) = \mathfrak{p}\} \leq G.$$

(Note: The decomposition group is called the "splitting group" in Swinnerton-Dyer. Everybody I know calls it the decomposition group, so we will too.)

Let $\mathbf{F}_\mathfrak{p} = \mathcal{O}_K/\mathfrak{p}$ denote the residue class field of $\mathfrak{p}$. In this section we will prove that there is a natural exact sequence

$$1 \to I_\mathfrak{p} \to D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p) \to 1,$$

where $I_\mathfrak{p}$ is the *inertia subgroup* of $D_\mathfrak{p}$, and $\#I_\mathfrak{p} = e$. The most interesting part of the proof is showing that the natural map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbf{F}_\mathfrak{p}/\mathbf{F}_p)$ is surjective.

We will also discuss the structure of $D_\mathfrak{p}$ and introduce Frobenius elements, which play a crucial roll in understanding Galois representations.

Recall that G acts on the set of primes $\mathfrak{p}$ lying over p. Thus the decomposition group is the stabilizer in G of $\mathfrak{p}$. The orbit-stabilizer theorem implies that $[G : D_\mathfrak{p}]$ equals the orbit of $\mathfrak{p}$, which by Theorem 13.2.2 equals the number g of primes lying over p, so $[G : D_\mathfrak{p}] = g$.

Lemma 14.1.2. *The decomposition subgroups $D_\mathfrak{p}$ corresponding to primes $\mathfrak{p}$ lying over a given p are all conjugate in G.*

Proof. We have $\tau(\sigma(\tau^{-1}(\mathfrak{p}))) = \mathfrak{p}$ if and only if $\sigma(\tau^{-1}(\mathfrak{p})) = \tau^{-1}\mathfrak{p}$. Thus $\tau\sigma\tau^{-1} \in D_p$ if and only if $\sigma \in D_{\tau^{-1}\mathfrak{p}}$, so $\tau^{-1}D_p\tau = D_{\tau^{-1}\mathfrak{p}}$. The lemma now follows because, by Theorem 13.2.2, G acts transitively on the set of $\mathfrak{p}$ lying over p. □

The decomposition group is extremely useful because it allows us to see the extension $K/\mathbf{Q}$ as a tower of extensions, such that at each step in the tower we understand well the splitting behavior of the primes lying over p. Now might be a good time to glance ahead at Figure 14.1.2 on page 101.

We characterize the fixed field of $D = D_{\mathfrak{p}}$ as follows.

Proposition 14.1.3. *The fixed field K^D of D*

$$K^D = \{a \in K : \sigma(a) = a \textit{ for all } \sigma \in D\}$$

is the smallest subfield $L \subset K$ such that $\mathfrak{p} \cap L$ does not split in K (i.e., $g(K/L) = 1$).

Proof. First suppose $L = K^D$, and note that by Galois theory $\mathrm{Gal}(K/L) \cong D$, and by Theorem 13.2.2, the group D acts transitively on the primes of K lying over $\mathfrak{p} \cap L$. One of these primes is $\mathfrak{p}$, and D fixes $\mathfrak{p}$ by definition, so there is only one prime of K lying over $\mathfrak{p} \cap L$, i.e., $\mathfrak{p} \cap L$ does not split in K. Conversely, if $L \subset K$ is such that $\mathfrak{p} \cap L$ does not split in K, then $\mathrm{Gal}(K/L)$ fixes $\mathfrak{p}$ (since it is the only prime over $\mathfrak{p} \cap L$), so $\mathrm{Gal}(K/L) \subset D$, hence $K^D \subset L$. □

Thus p does not split in going from K^D to K—it does some combination of ramifying and staying inert. To fill in more of the picture, the following proposition asserts that p splits completely and does not ramify in $K^D/\mathbf{Q}$.

Proposition 14.1.4. *Let $L = K^D$ for our fixed prime p and Galois extension $K/\mathbf{Q}$. Let $e = e(L/\mathbf{Q}), f = f(L/\mathbf{Q}), g = g(L/\mathbf{Q})$ be for $L/\mathbf{Q}$ and p. Then $e = f = 1$ and $g = [L : \mathbf{Q}]$, i.e., p does not ramify and splits completely in L. Also $f(K/\mathbf{Q}) = f(K/L)$ and $e(K/\mathbf{Q}) = e(K/L)$.*

Proof. As mentioned right after Definition 14.1.1, the orbit-stabilizer theorem implies that $g(K/\mathbf{Q}) = [G : D]$, and by Galois theory $[G : D] = [L : \mathbf{Q}]$. Thus

$$\begin{aligned} e(K/L) \cdot f(K/L) &= [K : L] = [K : \mathbf{Q}]/[L : \mathbf{Q}] \\ &= \frac{e(K/\mathbf{Q}) \cdot f(K/\mathbf{Q}) \cdot g(K/\mathbf{Q})}{[L : \mathbf{Q}]} = e(K/\mathbf{Q}) \cdot f(K/\mathbf{Q}). \end{aligned}$$

Now $e(K/L) \leq e(K/\mathbf{Q})$ and $f(K/L) \leq f(K/\mathbf{Q})$, so we must have $e(K/L) = e(K/\mathbf{Q})$ and $f(K/L) = f(K/\mathbf{Q})$. Since $e(K/\mathbf{Q}) = e(K/L) \cdot e(L/\mathbf{Q})$ and $f(K/\mathbf{Q}) = f(K/L) \cdot f(L/\mathbf{Q})$, the proposition follows. □

14.1.1 Galois groups of finite fields

Each $\sigma \in D = D_{\mathfrak{p}}$ acts in a well-defined way on the finite field $\mathbf{F}_{\mathfrak{p}} = \mathcal{O}_K/\mathfrak{p}$, so we obtain a homomorphism

$$\varphi : D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p).$$

We pause for a moment and derive a few basic properties of $\mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$, which are in fact general properties of Galois groups for finite fields. Let $f = [\mathbf{F}_{\mathfrak{p}} : \mathbf{F}_p]$.

The group $\mathrm{Aut}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ contains the element Frob_p defined by

$$\mathrm{Frob}_p(x) = x^p,$$

because $(xy)^p = x^p y^p$ and

$$(x+y)^p = x^p + px^{p-1}y + \cdots + y^p \equiv x^p + y^p \pmod{p}.$$

By Exercise 29 (see Chapter 22), the group $\mathbf{F}_{\mathfrak{p}}^*$ is cyclic, so there is an element $a \in \mathbf{F}_{\mathfrak{p}}^*$ of order $p^f - 1$, and $\mathbf{F}_{\mathfrak{p}} = \mathbf{F}_p(a)$. Then $\mathrm{Frob}_p^n(a) = a^{p^n} = a$ if and only if $(p^f - 1) \mid p^n - 1$ which is the case preciselywhen $f \mid n$, so the order of Frob_p is f. Since the order of the automorphism group of a field extension is at most the degree of the extension, we conclude that $\mathrm{Aut}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ is generated by Frob_p. Also, since $\mathrm{Aut}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ has order equal to the degree, we conclude that $\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p$ is Galois, with group $\mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ cyclic of order f generated by Frob_p. (Anther general fact: Up to isomorphism there is exactly one finite field of each degree. Indeed, if there were two of degree f, then both could be characterized as the set of roots in the compositum of $x^{p^f} - 1$, hence they would be equal.)

14.1.2 The Exact Sequence

There is a natural reduction homomorphism

$$\varphi : D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p).$$

Theorem 14.1.5. *The homomorphism φ is surjective.*

Proof. Let $\tilde{a} \in \mathbf{F}_{\mathfrak{p}}$ be an element such that $\mathbf{F}_{\mathfrak{p}} = \mathbf{F}_p(a)$. Lift $\tilde{a}$ to an algebraic integer $a \in \mathcal{O}_K$, and let $f = \prod_{\sigma \in D_p}(x - \sigma(a)) \in K^D[x]$ be the characteristic polynomial of a over K^D. Using Proposition 14.1.4 we see that f reduces to the minimal polynomial $\tilde{f} = \prod(x - \widetilde{\sigma(a)}) \in \mathbf{F}_p[x]$ of $\tilde{a}$ (by the Proposition the coefficients of $\tilde{f}$ are in $\mathbf{F}_p$, and $\tilde{a}$ satisfies $\tilde{f}$, and the degree of $\tilde{f}$ equals the degree of the minimal polynomial of $\tilde{a}$). The roots of $\tilde{f}$ are of the form $\tilde{\sigma}(a)$, and the element $\mathrm{Frob}_p(a)$ is also a root of $\tilde{f}$, so it is of the form $\widetilde{\sigma(a)}$. We conclude that the generator Frob_p of $\mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ is in the image of φ, which proves the theorem. □

Definition 14.1.6 (Inertia Group). The *inertia group* is the kernel $I_{\mathfrak{p}}$ of $D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$.

Combining everything so far, we find an exact sequence of groups

$$1 \to I_{\mathfrak{p}} \to D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p) \to 1. \tag{14.1.1}$$

The inertia group is a measure of how p ramifies in K.

Corollary 14.1.7. *We have $\#I_{\mathfrak{p}} = e(\mathfrak{p}/p)$, where $\mathfrak{p}$ is a prime of K over p.*

Proof. The sequence (14.1.1) implies that $\#I_{\mathfrak{p}} = \#D_{\mathfrak{p}}/f(K/\mathbf{Q})$. Applying Propositions 14.1.3–14.1.4, we have

$$\#D_{\mathfrak{p}} = [K : L] = \frac{[K : \mathbf{Q}]}{g} = \frac{efg}{g} = ef.$$

Dividing both sides by $f = f(K/\mathbf{Q})$ proves the corollary. □

We have the following characterization of $I_{\mathfrak{p}}$.

Proposition 14.1.8. *Let $K/\mathbf{Q}$ be a Galois extension with group G, let $\mathfrak{p}$ be a prime lying over a prime p. Then*

$$I_{\mathfrak{p}} = \{\sigma \in G : \sigma(a) = a \pmod{\mathfrak{p}} \textit{ for all } a \in \mathcal{O}_K\}.$$

Proof. By definition $I_{\mathfrak{p}} = \{\sigma \in D_{\mathfrak{p}} : \sigma(a) = a \pmod{\mathfrak{p}}$ for all $a \in \mathcal{O}_K\}$, so it suffices to show that if $\sigma \notin D_{\mathfrak{p}}$, then there exists $a \in \mathcal{O}_K$ such that $\sigma(a) = a \pmod{\mathfrak{p}}$. If $\sigma \notin D_{\mathfrak{p}}$, we have $\sigma^{-1}(\mathfrak{p}) \neq \mathfrak{p}$, so since both are maximal ideals, there exists $a \in \mathfrak{p}$ with $a \notin \sigma^{-1}(\mathfrak{p})$, i.e., $\sigma(a) \notin \mathfrak{p}$. Thus $\sigma(a) \not\equiv a \pmod{\mathfrak{p}}$. □

Figure 14.1.2 is a picture of the splitting behavior of a prime $p \in \mathbf{Z}$.

14.2 Frobenius Elements

Suppose that $K/\mathbf{Q}$ is a finite Galois extension with group G and p is a prime such that $e = 1$ (i.e., an unramified prime). Then $I = I_{\mathfrak{p}} = 1$ for any $\mathfrak{p} \mid p$, so the map φ of Theorem 14.1.5 is a canonical isomorphism $D_{\mathfrak{p}} \cong \mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$. By Section 14.1.1, the group $\mathrm{Gal}(\mathbf{F}_{\mathfrak{p}}/\mathbf{F}_p)$ is cyclic with canonical generator Frob_p. The *Frobenius element* corresponding to $\mathfrak{p}$ is $\mathrm{Frob}_{\mathfrak{p}} \in D_{\mathfrak{p}}$. It is the unique element of G such that for all $a \in \mathcal{O}_K$ we have

$$\mathrm{Frob}_{\mathfrak{p}}(a) \equiv a^p \pmod{\mathfrak{p}}.$$

(To see this argue as in the proof of Proposition 14.1.8.) Just as the primes $\mathfrak{p}$ and decomposition groups D are all conjugate, the Frobenius elements over a given prime are conjugate.

Proposition 14.2.1. *For each $\sigma \in G$, we have*

$$\mathrm{Frob}_{\sigma\mathfrak{p}} = \sigma \, \mathrm{Frob}_{\mathfrak{p}} \, \sigma^{-1}.$$

In particular, the Frobenius elements lying over a given prime are all conjugate.

Proof. Fix $\sigma \in G$. For any $a \in \mathcal{O}_K$ we have $\mathrm{Frob}_{\mathfrak{p}}(\sigma^{-1}(a)) - \sigma^{-1}(a) \in \mathfrak{p}$. Multiply by σ we see that $\sigma \, \mathrm{Frob}_{\mathfrak{p}}(\sigma^{-1}(a)) - a \in \sigma\mathfrak{p}$, which proves the proposition. □

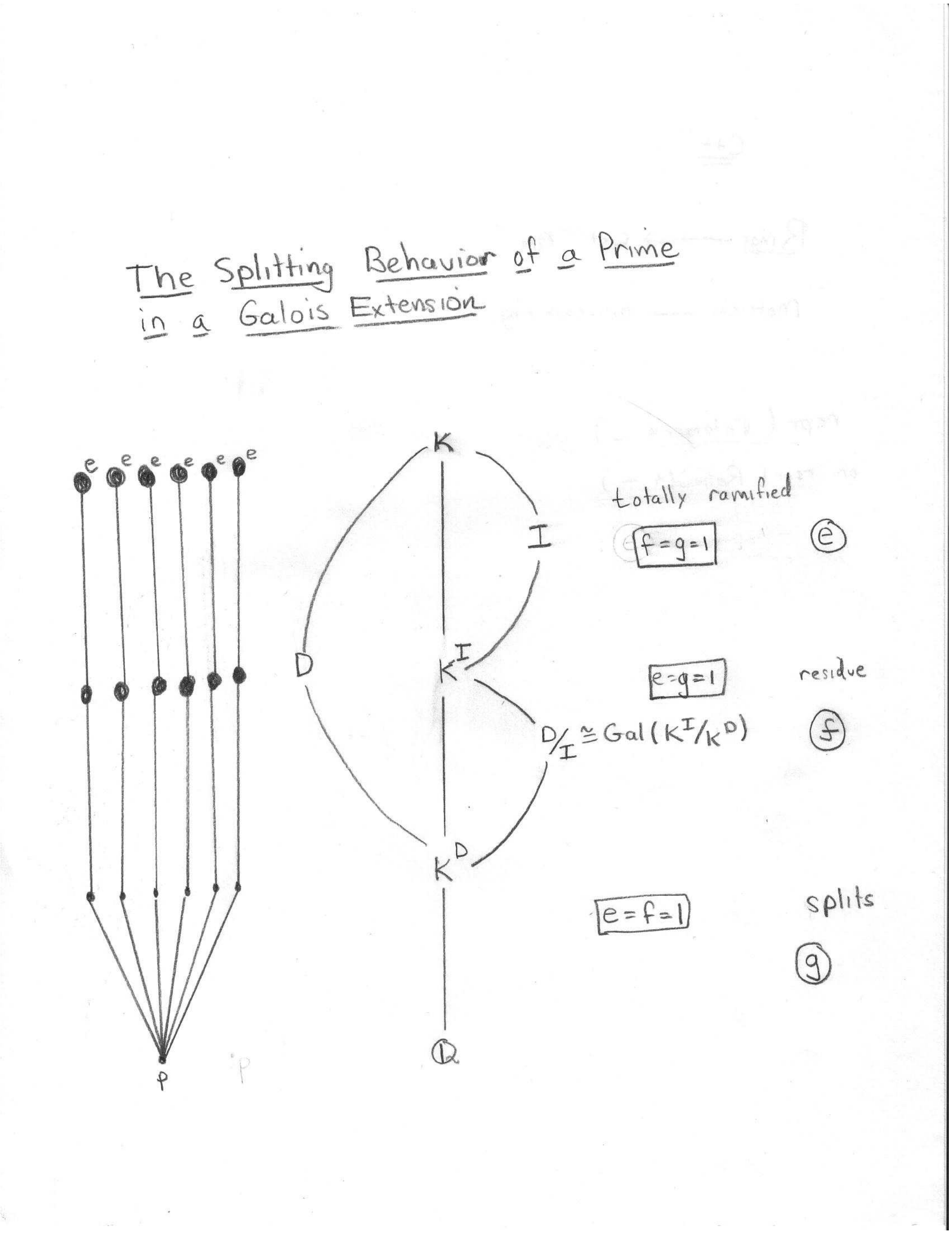


Figure 14.1.1: The Splitting of Behavior of a Prime in a Galois Extension

Thus the conjugacy class of $\mathrm{Frob}_{\mathfrak{p}}$ in G is a well defined function of p. For example, if G is abelian, then $\mathrm{Frob}_{\mathfrak{p}}$ does not depend on the choice of $\mathfrak{p}$ lying over p and we obtain a well defined symbol $\left(\frac{K/\mathbf{Q}}{p}\right) = \mathrm{Frob}_{\mathfrak{p}} \in G$ called the *Artin symbol.* It extends to a map from the free abelian group on unramified primes to the group G (the fractional ideals of $\mathbf{Z}$). Class field theory (for $\mathbf{Q}$) sets up a natural bijection between abelian Galois extensions of $\mathbf{Q}$ and certain maps from certain subgroups of the group of fractional ideals for $\mathbf{Z}$. We have just described one direction of this bijection, which associates to an abelian extension the Artin symbol (which induces a homomorphism). The Kronecker-Weber theorem asserts that the abelian extensions of $\mathbf{Q}$ are exactly the subfields of the fields $\mathbf{Q}(\zeta_n)$, as n varies over all positive integers. By Galois theory there is a correspondence between the subfields of $\mathbf{Q}(\zeta_n)$ (which has Galois group $(\mathbf{Z}/n\mathbf{Z})^*$) and the subgroups of $(\mathbf{Z}/n\mathbf{Z})^*$. Giving an abelian extension of $\mathbf{Q}$ is *exactly the same* as giving an integer n and a subgroup of $(\mathbf{Z}/n\mathbf{Z})^*$. Even more importantly, the reciprocity map $p \mapsto \left(\frac{Q(\zeta_n)/\mathbf{Q}}{p}\right)$ is simply $p \mapsto p \in (\mathbf{Z}/n\mathbf{Z})^*$. This is a nice generalization of quadratic reciprocity: for $\mathbf{Q}(\zeta_n)$, the efg for a prime p depends in a simple way on nothing but $p \mod n$.

14.3 Galois Representations and a Conjecture of Artin

The Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is an object of central importance in number theory, and I've often heard that in some sense number theory is the study of this group. A good way to study a group is to study how it acts on various objects, that is, to study its representations.

Endow $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ with the topology which has as a basis of open neighborhoods of the origin the subgroups $\mathrm{Gal}(\overline{\mathbf{Q}}/K)$, where K varies over finite Galois extensions of $\mathbf{Q}$. (Note: This is **not** the topology got by taking as a basis of open neighborhoods the collection of finite-index normal subgroups of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.) Fix a positive integer n and let $\mathrm{GL}_n(\mathbf{C})$ be the group of $n \times n$ invertible matrices over $\mathbf{C}$ with the discrete topology.

Definition 14.3.1. A *complex n-dimensional representation* of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a continuous homomorphism

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C}).$$

For ρ to be continuous means that there is a finite Galois extension $K/\mathbf{Q}$ such that ρ factors through $\mathrm{Gal}(K/\mathbf{Q})$:

$$\begin{tikzcd}
\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \arrow[rr, "\rho"] \arrow[dr] & & \mathrm{GL}_n(\mathbf{C}) \\
& \mathrm{Gal}(K/\mathbf{Q}) \arrow[ur, "\rho'"'] &
\end{tikzcd}$$

For example, one could take K to be the fixed field of $\ker(\rho)$. (Note that continous implies that the image of ρ is finite, but using Zorn's lemma one can show that there are homomorphisms $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \{\pm 1\}$ with finite image that are not continuous, since they do not factor through the Galois group of any finite Galois extension.)

Fix a Galois representation ρ and a finite Galois extension K such that ρ factors through $\mathrm{Gal}(K/\mathbf{Q})$. For each prime $p \in \mathbf{Z}$ that is not ramified in K, there is an element $\mathrm{Frob}_{\mathfrak{p}} \in \mathrm{Gal}(K/\mathbf{Q})$ that is well-defined up to conjugation by elements of $\mathrm{Gal}(K/\mathbf{Q})$. This means that $\rho'(\mathrm{Frob}_p) \in \mathrm{GL}_n(\mathbf{C})$ is well-defined up to conjugation. Thus the characteristic polynomial $F_p \in \mathbf{C}[x]$ is a well-defined invariant of p and ρ. Let

$$R_p(x) = x^{\deg(F_p)} \cdot F_p(1/x) = 1 + \cdots + \mathrm{Det}(\mathrm{Frob}_p) \cdot x^{\deg(F_p)}$$

be the polynomial obtain by reversing the order of the coefficients of F_p. Following E. Artin, set

$$L(\rho, s) = \prod_{p \text{ unramified}} \frac{1}{R_p(p^{-s})}. \tag{14.3.1}$$

We view. $L(\rho, s)$ as a function of a single complex variable s. One can prove that $L(\rho, s)$ is holomorphic on some right half plane, and extends to a meromorphic function on all $\mathbf{C}$.

Conjecture 14.3.2 (Artin). *The L-series of any continuous representation*

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C})$$

is an entire function on all $\mathbf{C}$, except possibly at 1.

This conjecture asserts that there is some way to analytically continue $L(\rho, s)$ to the whole complex plane, except possibly at 1. (A standard fact from complex analysis is that this analytic continuation must be unique.) The simple pole at $s = 1$ corresponds to the trivial representation (the Riemann zeta function), and if $n \geq 2$ and ρ is irreducible, then the conjecture is that ρ extends to a holomorphic function on all $\mathbf{C}$.

The conjecture follows from class field theory for $\mathbf{Q}$ when $n = 1$. When $n = 2$ and the image of ρ in $\mathrm{PGL}_2(\mathbf{C})$ is a solvable group, the conjecture is known, and is a deep theorem of Langlands and others (see [Lan80]), which played a crucial roll in Wiles's proof of Fermat's Last Theorem. When $n = 2$ and the projective image is not solvable, the only possibility is that the projective image is isomorphic to the alternating group A_5. Because A_5 is the symmetric group of the icosahedron, these representations are called *icosahedral*. In this case, Joe Buhler's Harvard Ph.D. thesis gave the first example, there is a whole book [Fre94], which proves Artin's conjecture for 7 icosahedral representation (none of which are twists of each other). Kevin Buzzard and I (Stein) proved the conjecture for 8 more examples. Subsequently, Richard Taylor, Kevin Buzzard, and Mark Dickinson proved the conjecture for an infinite class of icosahedral Galois representations (disjoint from the examples). The general problem for $n = 2$ is still open, but perhaps Taylor and others are still making progress toward it.

Part II

Adelic Viewpoint

Chapter 15

Valuations

The rest of this book is a partial rewrite of [Cas67] meant to make it more accessible. I have attempted to add examples and details of the implicit exercises and remarks that are left to the reader.

15.1 Valuations

Definition 15.1.1 (Valuation). A *valuation* $|\cdot|$ on a field K is a function defined on K with values in $\mathbf{R}_{\geq 0}$ satisfying the following axioms:

(1) $|a| = 0$ if and only if $a = 0$,

(2) $|ab| = |a|\,|b|$, and

(3) there is a constant $C \geq 1$ such that $|1 + a| \leq C$ whenever $|a| \leq 1$.

The *trivial valuation* is the valuation for which $|a| = 1$ for all $a \neq 0$. We will often tacitly exclude the trivial valuation from consideration.

From (2) we have

$$|1| = |1| \cdot |1|,$$

so $|1| = 1$ by (1). If $w \in K$ and $w^n = 1$, then $|w| = 1$ by (2). In particular, the only valuation of a finite field is the trivial one. The same argument shows that $|-1| = |1|$, so

$$|-a| = |a| \qquad \text{all } a \in K.$$

Definition 15.1.2 (Equivalent). Two valuations $|\cdot|_1$ and $|\cdot|_2$ on the same field are *equivalent* if there exists $c > 0$ such that

$$|a|_2 = |a|_1^c$$

for all $a \in K$.

Note that if $|\cdot|_1$ is a valuation, then $|\cdot|_2 = |\cdot|_1^c$ is also a valuation. Also, equivalence of valuations is an equivalence relation.

If $|\cdot|$ is a valuation and C is the constant from Axiom (3), then there is a $c > 0$ such that $C^c = 2$ (i.e., $c = \log(C)/\log(2)$). Then we can take 2 as constant for the equivalent valuation $|\cdot|^c$. Thus every valuation is equivalent to a valuation with $C = 2$. Note that if $C = 1$, e.g., if $|\cdot|$ is the trivial valuation, then we could simply take $C = 2$ in Axiom (3).

Proposition 15.1.3. *Suppose $|\cdot|$ is a valuation with $C = 2$. Then for all $a, b \in K$ we have*

$$|a + b| \leq |a| + |b| \qquad \textit{(triangle inequality).} \tag{15.1.1}$$

Proof. Suppose $a_1, a_2 \in K$ with $|a_1| \geq |a_2|$. Then $a = a_2/a_1$ satisfies $|a| \leq 1$. By Axiom (3) we have $|1 + a| \leq 2$, so multiplying by a_1 we see that

$$|a_1 + a_2| \leq 2|a_1| = 2 \cdot \max\{|a_1|, |a_2|\}.$$

Also we have

$$|a_1 + a_2 + a_3 + a_4| \leq 2 \cdot \max\{|a_1 + a_2|, |a_3 + a_4|\} \leq 4 \cdot \max\{|a_1|, |a_2|, |a_3|, |a_4|\},$$

and inductively we have for any $r > 0$ that

$$|a_1 + a_2 + \cdots + a_{2^r}| \leq 2^r \cdot \max |a_j|.$$

If n is any positive integer, let r be such that $2^{r-1} \leq n \leq 2^r$. Thenn

$$|a_1 + a_2 + \cdots + a_n| \leq 2^r \cdot \max\{|a_j|\} \leq 2n \cdot \max\{|a_j|\},$$

since $2^r \leq 2n$. In particular,

$$|n| \leq 2n \cdot |1| = 2n \qquad \text{(for } n > 0). \tag{15.1.2}$$

Applying (15.1.2) to $\left|\binom{n}{j}\right|$ and using the binomial expansion, we have for any $a, b \in K$ that

$$\begin{aligned}
|a + b|^n &= \left|\sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}\right| \\
&\leq 2(n+1) \max_j \left\{ \left|\binom{n}{j}\right| |a|^j\, |b|^{n-j} \right\} \\
&\leq 2(n+1) \max_j \left\{ 2\binom{n}{j} |a|^j\, |b|^{n-j} \right\} \\
&\leq 4(n+1) \max_j \left\{ \binom{n}{j} |a|^j\, |b|^{n-j} \right\} \\
&\leq 4(n+1)(|a| + |b|)^n.
\end{aligned}$$

Now take nth roots of both sides to obtain

$$|a+b| \leq \sqrt[n]{4(n+1)} \cdot (|a|+|b|).$$

We have by elementary calculus that

$$\lim_{n\to\infty} \sqrt[n]{4(n+1)} = 1,$$

so $|a+b| \leq |a|+|b|$. (The "elementary calculus": We instead prove that $\sqrt[n]{n} \to 1$, since the argument is the same and the notation is simpler. First, for any $n \geq 1$ we have $\sqrt[n]{n} \geq 1$, since upon taking nth powers this is equivalent to $n \geq 1^n$, which is true by hypothesis. Second, suppose there is an $\varepsilon > 0$ such that $\sqrt[n]{n} \geq 1+\varepsilon$ for all $n \geq 1$. Then taking logs of boths sides we see that $\frac{1}{n}\log(n) \geq \log(1+\varepsilon) > 0$. But $\log(n)/n \to 0$, so there is no such ε. Thus $\sqrt[n]{n} \to 1$ as $n \to \infty$.) □

Note that Axioms (1), (2) and Equation (15.1.1) imply Axiom (3) with $C = 2$. We take Axiom (3) instead of Equation (15.1.1) for the technical reason that we will want to call the square of the absolute value of the complex numbers a valuation.

Lemma 15.1.4. *Suppose $a, b \in K$, and $|\cdot|$ is a valuation on K with $C \leq 2$. Then*

$$\Big| |a| - |b| \Big| \leq |a-b|\,.$$

(Here the big absolute value on the outside of the left-hand side of the inequality is the usual absolute value on real numbers, but the other absolute values are a valuation on an arbitrary field K.)

Proof. We have

$$|a| = |b + (a-b)| \leq |b| + |a-b|,$$

so $|a| - |b| \leq |a-b|$. The same argument with a and b swapped implies that $|b| - |a| \leq |a-b|$, which proves the lemma. □

15.2 Types of Valuations

We define two important properties of valuations, both of which apply to equivalence classes of valuations (i.e., the property holds for $|\cdot|$ if and only if it holds for a valuation equivalent to $|\cdot|$).

Definition 15.2.1 (Discrete). A valuation $|\cdot|$ is *discrete* if there is a $\delta > 0$ such that for any $a \in K$

$$1 - \delta < |a| < 1 + \delta \implies |a| = 1.$$

Thus the absolute values are bounded away from 1.

To say that $|\cdot|$ is discrete is the same as saying that the set

$$G = \bigl\{\log|a| : a \in K, a \neq 0\bigr\} \subset \mathbf{R}$$

forms a discrete subgroup of the reals under addition (because the elements of the group G are bounded away from 0).

Proposition 15.2.2. *A nonzero discrete subgroup G of $\mathbf{R}$ is free on one generator.*

Proof. Since G is discrete there is a positive $m \in G$ such that for any positive $x \in G$ we have $m \leq x$. Suppose $x \in G$ is an arbitrary positive element. By subtracting off integer multiples of m, we find that there is a unique n such that

$$0 \leq x - nm < m.$$

Since $x - nm \in G$ and $0 < x - nm < m$, it follows that $x - nm = 0$, so x is a multiple of m. □

By Proposition 15.2.2, the set of $\log|a|$ for nonzero $a \in K$ is free on one generator, so there is a $c < 1$ such that $|a|$, for $a \neq 0$, runs precisely through the set

$$c^{\mathbf{Z}} = \{c^m : m \in \mathbf{Z}\}$$

(Note: we can replace c by c^{-1} to see that we can assume that $c < 1$).

Definition 15.2.3 (Order). If $|a| = c^m$, we call $m = \mathrm{ord}(a)$ the *order* of a.

Axiom (2) of valuations translates into

$$\mathrm{ord}(ab) = \mathrm{ord}(a) + \mathrm{ord}(b).$$

Definition 15.2.4 (Non-archimedean). A valuation $|\cdot|$ is *non-archimedean* if we can take $C = 1$ in Axiom (3), i.e., if

$$|a + b| \leq \max\bigl\{|a|, |b|\bigr\}. \tag{15.2.1}$$

If $|\cdot|$ is not non-archimedean then it is *archimedean*.

Note that if we can take $C = 1$ for $|\cdot|$ then we can take $C = 1$ for any valuation equivalent to $|\cdot|$. To see that (15.2.1) is equivalent to Axiom (3) with $C = 1$, suppose $|b| \leq |a|$. Then $|b/a| \leq 1$, so Axiom (3) asserts that $|1 + b/a| \leq 1$, which implies that $|a + b| \leq |a| = \max\{|a|, |b|\}$, and conversely.

We note at once the following consequence:

Lemma 15.2.5. *Suppose $|\cdot|$ is a non-archimedean valuation. If $a, b \in K$ with $|b| < |a|$, then $|a + b| = |a|$.*

Proof. Note that $|a+b| \leq \max\{|a|,|b|\} = |a|$, which is true even if $|b| = |a|$. Also,

$$|a| = |(a+b) - b| \leq \max\{|a+b|, |b|\} = |a+b|,$$

where for the last equality we have used that $|b| < |a|$ (if $\max\{|a+b|,|b|\} = |b|$, then $|a| \leq |b|$, a contradiction).

□

Definition 15.2.6 (Ring of Integers). Suppose $|\cdot|$ is a non-archimedean absolute value on a field K. Then

$$\mathcal{O} = \{a \in K : |a| \leq 1\}$$

is a ring called the *ring of integers* of K with respect to $|\cdot|$.

Lemma 15.2.7. *Two non-archimedean valuations $|\cdot|_1$ and $|\cdot|_2$ are equivalent if and only if they give the same $\mathcal{O}$.*

We will prove this modulo the claim (to be proved later in Section 16.1) that valuations are equivalent if (and only if) they induce the same topology.

Proof. Suppose suppose $|\cdot|_1$ is equivalent to $|\cdot|_2$, so $|\cdot|_1 = |\cdot|_2^c$, for some $c > 0$. Then $|c|_1 \leq 1$ if and only if $|c|_2^c \leq 1$, i.e., if $|c|_2 \leq 1^{1/c} = 1$. Thus $\mathcal{O}_1 = \mathcal{O}_2$.

Conversely, suppose $\mathcal{O}_1 = \mathcal{O}_2$. Then $|a|_1 < |b|_1$ if and only if $a/b \in \mathcal{O}_1$ and $b/a \notin \mathcal{O}_1$, so

$$|a|_1 < |b|_1 \iff |a|_2 < |b|_2. \tag{15.2.2}$$

The topology induced by $|\;|_1$ has as basis of open neighborhoods the set of open balls

$$B_1(z,r) = \{x \in K : |x - z|_1 < r\},$$

for $r > 0$, and likewise for $|\;|_2$. Since the absolute values $|b|_1$ get arbitrarily close to 0, the set $\mathcal{U}$ of open balls $B_1(z, |b|_1)$ also forms a basis of the topology induced by $|\;|_1$ (and similarly for $|\;|_2$). By (15.2.2) we have

$$B_1(z, |b|_1) = B_2(z, |b|_2),$$

so the two topologies both have $\mathcal{U}$ as a basis, hence are equal. That equal topologies imply equivalence of the corresponding valuations will be proved in Section 16.1. □

The set of $a \in \mathcal{O}$ with $|a| < 1$ forms an ideal $\mathfrak{p}$ in $\mathcal{O}$. The ideal $\mathfrak{p}$ is maximal, since if $a \in \mathcal{O}$ and $a \notin \mathfrak{p}$ then $|a| = 1$, so $|1/a| = 1/|a| = 1$, hence $1/a \in \mathcal{O}$, so a is a unit.

Lemma 15.2.8. *A non-archimedean valuation $|\cdot|$ is discrete if and only if $\mathfrak{p}$ is a principal ideal.*

Proof. First suppose that $|\cdot|$ is discrete. Choose $\pi \in \mathfrak{p}$ with $|\pi|$ maximal, which we can do since

$$S = \{\log|a| : a \in \mathfrak{p}\} \subset (-\infty, 1],$$

so the discrete set S is bounded above. Suppose $a \in \mathfrak{p}$. Then

$$\left|\frac{a}{\pi}\right| = \frac{|a|}{|\pi|} \leq 1,$$

so $a/\pi \in \mathcal{O}$. Thus

$$a = \pi \cdot \frac{a}{\pi} \in \pi\mathcal{O}.$$

Conversely, suppose $\mathfrak{p} = (\pi)$ is principal. For any $a \in \mathfrak{p}$ we have $a = \pi b$ with $b \in \mathcal{O}$. Thus

$$|a| = |\pi| \cdot |b| \leq |\pi| < 1.$$

Thus $\{|a| : |a| < 1\}$ is bounded away from 1, which is exactly the definition of discrete. □

Example 15.2.9. For any prime p, define the p-adic valuation $|\cdot|_p : \mathbf{Q} \to \mathbf{R}$ as follows. Write a nonzero $\alpha \in K$ as $p^n \cdot \frac{a}{b}$, where $\gcd(a,p) = \gcd(b,p) = 1$. Then

$$\left|p^n \cdot \frac{a}{b}\right|_p := p^{-n} = \left(\frac{1}{p}\right)^n.$$

This valuation is both discrete and non-archimedean. The ring $\mathcal{O}$ is the local ring

$$\mathbf{Z}_{(p)} = \left\{\frac{a}{b} \in \mathbf{Q} : p \nmid b\right\},$$

which has maximal ideal generated by p. Note that $\mathrm{ord}(p^n \cdot \frac{a}{b}) = p^n$.

We will using the following lemma later (e.g., in the proof of Corollary 16.2.4 and Theorem 15.3.2).

Lemma 15.2.10. *A valuation $|\cdot|$ is non-archimedean if and only if $|n| \leq 1$ for all n in the ring generated by 1 in K.*

Note that we cannot identify the ring generated by 1 with $\mathbf{Z}$ in general, because K might have characteristic $p > 0$.

Proof. If $|\cdot|$ is non-archimedean, then $|1| \leq 1$, so by Axiom (3) with $a = 1$, we have $|1 + 1| \leq 1$. By induction it follows that $|n| \leq 1$.

Conversely, suppose $|n| \leq 1$ for all integer multiples n of 1. This condition is also true if we replace $|\cdot|$ by any equivalent valuation, so replace $|\cdot|$ by one with $C \leq 2$, so that the triangle inequality holds. Suppose $a \in K$ with $|a| \leq 1$. Then by the triangle inequality,

$$\begin{aligned}|1+a|^n &= |(1+a)^n|\\ &\leq \sum_{j=0}^{n} \left|\binom{n}{j}\right| |a|\\ &\leq 1 + 1 + \cdots + 1 = n.\end{aligned}$$

Now take nth roots of both sides to get

$$|1+a| \leq \sqrt[n]{n},$$

and take the limit as $n \to \infty$ to see that $|1+a| \leq 1$. This proves that one can take $C = 1$ in Axiom (3), hence that $|\cdot|$ is non-archimedean. □

15.3 Examples of Valuations

The archetypal example of an archimedean valuation is the absolute value on the complex numbers. It is essentially the only one:

Theorem 15.3.1 (Gelfand-Tornheim). *Any field K with an archimedean valuation is isomorphic to a subfield of $\mathbf{C}$, the valuation being equivalent to that induced by the usual absolute value on $\mathbf{C}$.*

We do not prove this here as we do not need it. For a proof, see [Art59, pg. 45, 67].

There are many non-archimedean valuations. On the rationals $\mathbf{Q}$ there is one for every prime $p > 0$, the p-adic valuation, as in Example 15.2.9.

Theorem 15.3.2 (Ostrowski). *The nontrivial valuations on $\mathbf{Q}$ are those equivalent to $|\cdot|_p$, for some prime p, and the usual absolute value $|\cdot|_\infty$.*

Remark 15.3.3. Before giving the proof, we pause with a brief remark about Ostrowski. According to

`http://www-gap.dcs.st-and.ac.uk/~history/Mathematicians/Ostrowski.html`

Ostrowski was a Ukrainian mathematician who lived 1893–1986. Gautschi writes about Ostrowski as follows: "... you are able, on the one hand, to emphasise the abstract and axiomatic side of mathematics, as for example in your theory of general norms, or, on the other hand, to concentrate on the concrete and constructive aspects of mathematics, as in your study of numerical methods, and to do both with equal ease. *You delight in finding short and succinct proofs, of which you have given many examples* ..." [italics mine]

We will now give an example of one of these short and succinct proofs.

Proof. Suppose $|\cdot|$ is a nontrivial valuation on $\mathbf{Q}$.

Nonarchimedean case: Suppose $|c| \leq 1$ for all $c \in \mathbf{Z}$, so by Lemma 15.2.10, $|\cdot|$ is nonarchimedean. Since $|\cdot|$ is nontrivial, the set

$$\mathfrak{p} = \{a \in \mathbf{Z} : |a| < 1\}$$

is nonzero. Also $\mathfrak{p}$ is an ideal and if $|ab| < 1$, then $|a|\,|b| = |ab| < 1$, so $|a| < 1$ or $|b| < 1$, so $\mathfrak{p}$ is a prime ideal of $\mathbf{Z}$. Thus $\mathfrak{p} = p\mathbf{Z}$, for some prime number p. Since every element of $\mathbf{Z}$ has valuation at most 1, if $u \in \mathbf{Z}$ with $\gcd(u,p) = 1$, then $u \notin \mathfrak{p}$, so $|u| = 1$. Let $\alpha = \log_{|p|} \frac{1}{p}$, so $|p|^{\alpha} = \frac{1}{p}$. Then for any r and any $u \in \mathbf{Z}$ with $\gcd(u, p) = 1$, we have

$$|up^r|^{\alpha} = |u|^{\alpha} |p|^{\alpha r} = |p|^{\alpha r} = p^{-r} = |up^r|_p .$$

Thus $|\cdot|^{\alpha} = |\cdot|_p$ on $\mathbf{Z}$, hence on $\mathbf{Q}$ by multiplicativity, so $|\cdot|$ is equivalent to $|\cdot|_p$, as claimed.

Archimedean case: By replacing $|\cdot|$ by a power of $|\cdot|$, we may assume without loss that $|\cdot|$ satisfies the triangle inequality. We first make some general remarks about any valuation that satisfies the triangle inequality. Suppose $a \in \mathbf{Z}$ is greater than 1. Consider, for any $b \in \mathbf{Z}$ the base-a expansion of b:

$$b = b_m a^m + b_{m-1} a^{m-1} + \cdots + b_0,$$

where

$$0 \le b_j < a \qquad (0 \le j \le m),$$

and $b_m \neq 0$. Since $a^m \le b$, taking logs we see that $m \log(a) \le \log(b)$, so

$$m \le \frac{\log(b)}{\log(a)}.$$

Let $M = \max_{1 \le d < a} |d|$. Then by the triangle inequality for $|\cdot|$, we have

$$\begin{aligned}
|b| &\le |b_m|\, a^m + \cdots + |b_1|\, |a| + |b_0| \\
&\le M \cdot (|a|^m + \cdots + |a| + 1) \\
&\le M \cdot (m+1) \cdot \max(1, |a|^m) \\
&\le M \cdot \left(\frac{\log(b)}{\log(a)} + 1\right) \cdot \max\left(1, |a|^{\log(b)/\log(a)}\right),
\end{aligned}$$

where in the last step we use that $m \le \frac{\log(b)}{\log(a)}$. Setting $b = c^n$, for $c \in \mathbf{Z}$, in the above inequality and taking nth roots, we have

$$\begin{aligned}
|c| &\le \left(M \cdot \left(\frac{\log(c^n)}{\log(a)} + 1\right) \cdot \max(1, |a|^{log(c^n)/\log(a)})\right)^{1/n} \\
&= M^{1/n} \cdot \left(\frac{\log(c^n)}{\log(a)} + 1\right)^{1/n} \cdot \max\left(1, |a|^{\log(c^n)/\log(a)}\right)^{1/n}.
\end{aligned}$$

The first factor $M^{1/n}$ converges to 1 as $n \to \infty$, since $M \ge 1$ (because $|1| = 1$). The second factor is

$$\left(\frac{\log(c^n)}{\log(a)} + 1\right)^{1/n} = \left(n \cdot \frac{\log(c)}{\log(a)} + 1\right)^{1/n}$$

which also converges to 1, for the same reason that $n^{1/n} \to 1$ (because $\log(n^{1/n}) = \frac{1}{n}\log(n) \to 0$ as $n \to \infty$). The third factor is

$$\max\left(1, |a|^{\log(c^n)/\log(a)}\right)^{1/n} = \begin{cases} 1 & \text{if } |a| < 1, \\ |a|^{\log(c)/\log(a)} & \text{if } |a| \ge 1. \end{cases}$$

Putting this all together, we see that

$$|c| \leq \max\left(1, |a|^{\frac{\log(c)}{\log(a)}}\right).$$

Our assumption that $|\cdot|$ is nonarchimedean implies that there is $c \in \mathbf{Z}$ with $c > 1$ and $|c| > 1$. Then for all $a \in \mathbf{Z}$ with $a > 1$ we have

$$1 < |c| \leq \max\left(1, |a|^{\frac{\log(c)}{\log(a)}}\right), \tag{15.3.1}$$

so $1 < |a|^{\log(c)/\log(a)}$, so $1 < |a|$ as well (i.e., any $a \in \mathbf{Z}$ with $a > 1$ automatically satisfies $|a| > 1$). Also, taking the $1/\log(c)$ power on both sides of (15.3.1) we see that

$$|c|^{\frac{1}{\log(c)}} \leq |a|^{\frac{1}{\log(a)}}. \tag{15.3.2}$$

Because, as mentioned above, $|a| > 1$, we can interchange the roll of a and c to obtain the reverse inequality of (15.3.2). We thus have

$$|c| = |a|^{\frac{\log(c)}{\log(a)}}.$$

Letting $\alpha = \log(2) \cdot \log_{|2|}(e)$ and setting $a = 2$, we have

$$|c|^{\alpha} = |2|^{\frac{\alpha}{\log(2)} \cdot \log(c)} = \left(|2|^{\log_{|2|}(e)}\right)^{\log(c)} = e^{\log(c)} = c = |c|_{\infty}.$$

Thus for all integers $c \in \mathbf{Z}$ with $c > 1$ we have $|c|^{\alpha} = |c|_{\infty}$, which implies that $|\cdot|$ is equivalent to $|\cdot|_{\infty}$. □

Let k be any field and let $K = k(t)$, where t is transcendental. Fix a real number $c > 1$. If $p = p(t)$ is an irreducible polynomial in the ring $k[t]$, we define a valuation by

$$\left|p^a \cdot \frac{u}{v}\right|_p = c^{-\deg(p)\cdot a}, \tag{15.3.3}$$

where $a \in \mathbf{Z}$ and $u, v \in k[t]$ with $p \nmid u$ and $p \nmid v$.

Remark 15.3.4. This definition differs from the one page 46 of [Cassels-Frohlich, Ch. 2] in two ways. First, we assume that $c > 1$ instead of $c < 1$, since otherwise $|\cdot|_p$ does not satisfy Axiom 3 of a valuation. Also, we write $c^{-\deg(p)\cdot a}$ instead of c^{-a}, so that the product formula will hold. (For more about the product formula, see Section 20.1.)

In addition there is a a non-archimedean valuation $|\cdot|_{\infty}$ defined by

$$\left|\frac{u}{v}\right|_{\infty} = c^{\deg(u)-\deg(v)}. \tag{15.3.4}$$

This definition differs from the one in [Cas67, pg. 46] in two ways. First, we assume that $c > 1$ instead of $c < 1$, since otherwise $|\cdot|_p$ does not satisfy Axiom 3 of a valuation. Here's why: Recall that Axiom 3 for a non-archimedean valuation on K asserts that whenever $a \in K$ and $|a| \leq 1$, then $|a+1| \leq 1$. Set $a = p - 1$, where $p = p(t) \in K[t]$ is an irreducible polynomial. Then $|a| = c^0 = 1$, since $\mathrm{ord}_p(p-1) = 0$. However, $|a+1| = |p-1+1| = |p| = c^1 < 1$, since $\mathrm{ord}_p(p) = 1$. If we take $c > 1$ instead of $c < 1$, as I propose, then $|p| = c^1 > 1$, as required.

Note the (albeit imperfect) analogy between $K = k(t)$ and $\mathbf{Q}$. If $s = t^{-1}$, so $k(t) = k(s)$, the valuation $|\cdot|_\infty$ is of the type (15.3.3) belonging to the irreducible polynomial $p(s) = s$.

The reader is urged to prove the following lemma as a homework problem.

Lemma 15.3.5. *The only nontrivial valuations on $k(t)$ which are trivial on k are equivalent to the valuation (15.3.3) or (15.3.4).*

For example, if k is a finite field, there are no nontrivial valuations on k, so the only nontrivial valuations on $k(t)$ are equivalent to (15.3.3) or (15.3.4).

Chapter 16

Topology and Completeness

16.1 Topology

A valuation $|\cdot|$ on a field K induces a topology in which a basis for the neighborhoods of a are the *open balls*

$$B(a,d) = \{x \in K : |x-a| < d\}$$

for $d > 0$.

Lemma 16.1.1. *Equivalent valuations induce the same topology.*

Proof. If $|\cdot|_1 = |\cdot|_2^r$, then $|x-a|_1 < d$ if and only if $|x-a|_2^r < d$ if and only if $|x-a|_2 < d^{1/r}$ so $B_1(a,d) = B_2(a,d^{1/r})$. Thus the basis of open neighborhoods of a for $|\cdot|_1$ and $|\cdot|_2$ are identical. □

A valuation satisfying the triangle inequality gives a metric for the topology on defining the distance from a to b to be $|a-b|$. Assume for the rest of this section that we only consider valuations that satisfy the triangle inequality.

Lemma 16.1.2. *A field with the topology induced by a valuation is a* topological field, *i.e., the operations sum, product, and reciprocal are continuous.*

Proof. For example (product) the triangle inequality implies that

$$|(a+\varepsilon)(b+\delta) - ab| \leq |\varepsilon|\,|\delta| + |a|\,|\delta| + |b|\,|\varepsilon|$$

is small when $|\varepsilon|$ and $|\delta|$ are small (for fixed a,b). □

Lemma 16.1.3. *Suppose two valuations $|\cdot|_1$ and $|\cdot|_2$ on the same field K induce the same topology. Then for any sequence $\{x_n\}$ in K we have*

$$|x_n|_1 \to 0 \iff |x_n|_2 \to 0.$$

Proof. It suffices to prove that if $|x_n|_1 \to 0$ then $|x_n|_2 \to 0$, since the proof of the other implication is the same. Let $\varepsilon > 0$. The topologies induced by the two absolute values are the same, so $B_2(0, \varepsilon)$ can be covered by open balls $B_1(a_i, r_i)$. One of these open balls $B_1(a, r)$ contains 0. There is $\varepsilon' > 0$ such that

$$B_1(0, \varepsilon') \subset B_1(a, r) \subset B_2(0, \varepsilon).$$

Since $|x_n|_1 \to 0$, there exists N such that for $n \geq N$ we have $|x_n|_1 < \varepsilon'$. For such n, we have $x_n \in B_1(0, \varepsilon')$, so $x_n \in B_2(0, \varepsilon)$, so $|x_n|_2 < \varepsilon$. Thus $|x_n|_2 \to 0$. □

Proposition 16.1.4. *If two valuations $|\cdot|_1$ and $|\cdot|_2$ on the same field induce the same topology, then they are equivalent in the sense that there is a positive real α such that $|\cdot|_1 = |\cdot|_2^\alpha$.*

Proof. If $x \in K$ and $i = 1, 2$, then $|x^n|_i \to 0$ if and only if $|x|_i^n \to 0$, which is the case if and only if $|x|_i < 1$. Thus Lemma 16.1.3 implies that $|x|_1 < 1$ if and only if $|x|_2 < 1$. On taking reciprocals we see that $|x|_1 > 1$ if and only if $|x|_2 > 1$, so finally $|x|_1 = 1$ if and only if $|x|_2 = 1$.

Let now $w, z \in K$ be nonzero elements with $|w|_i \neq 1$ and $|z|_i \neq 1$. On applying the foregoing to

$$x = w^m z^n \qquad (m, n \in \mathbf{Z})$$

we see that

$$m \log |w|_1 + n \log |z|_1 \geq 0$$

if and only if

$$m \log |w|_2 + n \log |z|_2 \geq 0.$$

Dividing through by $\log |z|_i$, and rearranging, we see that for every rational number $\alpha = -n/m$,

$$\frac{\log |w|_1}{\log |z|_1} \geq \alpha \iff \frac{\log |w|_2}{\log |z|_2} \geq \alpha.$$

Thus

$$\frac{\log |w|_1}{\log |z|_1} = \frac{\log |w|_2}{\log |z|_2},$$

so

$$\frac{\log |w|_1}{\log |w|_2} = \frac{\log |z|_1}{\log |z|_2}.$$

Since this equality does not depend on the choice of z, we see that there is a constant c ($= \log |z|_1 / \log |z|_2$) such that $\log |w|_1 / \log |w|_2 = c$ for all w. Thus $\log |w|_1 = c \cdot \log |w|_2$, so $|w|_1 = |w|_2^c$, which implies that $|\cdot|_1$ is equivalent to $|\cdot|_2$. □

16.2 Completeness

We recall the definition of metric on a set X.

Definition 16.2.1 (Metric). A *metric* on a set X is a map

$$d : X \times X \to \mathbf{R}$$

such that for all $x, y, z \in X$,

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$,

2. $d(x, y) = d(y, x)$, and

3. $d(x, z) \leq d(x, y) + d(y, z)$.

A *Cauchy sequence* is a sequence (x_n) in X such that for all $\varepsilon > 0$ there exists M such that for all $n, m > M$ we have $d(x_n, x_m) < \varepsilon$. The *completion* of X is the set of Cauchy sequences (x_n) in X modulo the equivalence relation in which two Cauchy sequences (x_n) and (y_n) are equivalent if $\lim_{n\to\infty} d(x_n, y_n) = 0$. A metric space is *complete* if every Cauchy sequence converges, and one can show that the completion of X with respect to a metric is complete.

For example, $d(x, y) = |x - y|$ (usual archimedean absolute value) defines a metric on $\mathbf{Q}$. The completion of $\mathbf{Q}$ with respect to this metric is the field $\mathbf{R}$ of real numbers. More generally, whenever $|\cdot|$ is a valuation on a field K that satisfies the triangle inequality, then $d(x, y) = |x - y|$ defines a metric on K. Consider for the rest of this section only valuations that satisfy the triangle inequality.

Definition 16.2.2 (Complete). A field K is *complete* with respect to a valuation $|\cdot|$ if given any Cauchy sequence a_n, $(n = 1, 2, \ldots)$, i.e., one for which

$$|a_m - a_n| \to 0 \qquad (m, n \to \infty, \infty),$$

there is an $a^* \in K$ such that

$$a_n \to a^* \qquad \text{w.r.t. } |\cdot|$$

(i.e., $|a_n - a^*| \to 0$).

Theorem 16.2.3. *Every field K with valuation $v = |\cdot|$ can be embedded in a complete field K_v with a valuation $|\cdot|$ extending the original one in such a way that K_v is the closure of K with respect to $|\cdot|$. Further K_v is unique up to a unique isomorphism fixing K.*

Proof. Define K_v to be the completion of K with respect to the metric defined by $|\cdot|$. Thus K_v is the set of equivalence classes of Cauchy sequences, and there is a natural injective map from K to K_v sending an element $a \in K$ to the constant Cauchy sequence (a). Because the field operations on K are continuous, they induce well-defined field operations on equivalence classes of Cauchy sequences componentwise. Also, define a valuation on K_v by

$$|(a_n)_{n=1}^{\infty}| = \lim_{n\to\infty} |a_n|,$$

and note that this is well defined and extends the valuation on K.

To see that K_v is unique up to a unique isomorphism fixing K, we observe that there are no nontrivial continuous automorphisms $K_v \to K_v$ that fix K. This is because, by denseness, a continuous automorphism $\sigma : K_v \to K_v$ is determined by what it does to K, and by assumption σ is the identity map on K. More precisely, suppose $a \in K_v$ and n is a positive integer. Then by continuity there is $\delta > 0$ (with $\delta < 1/n$) such that if $a_n \in K_v$ and $|a - a_n| < \delta$ then $|\sigma(a) - \sigma(a_n)| < 1/n$. Since K is dense in K_v, we can choose the a_n above to be an element of K. Then by hypothesis $\sigma(a_n) = a_n$, so $|\sigma(a) - a_n| < 1/n$. Thus $\sigma(a) = \lim_{n\to\infty} a_n = a$. □

Corollary 16.2.4. *The valuation $|\cdot|$ is non-archimedean on K_v if and only if it is so on K. If $|\cdot|$ is non-archimedean, then the set of values taken by $|\cdot|$ on K and K_v are the same.*

Proof. The first part follows from Lemma 15.2.10 which asserts that a valuation is non-archimedean if and only if $|n| < 1$ for all integers n. Since the valuation on K_v extends the valuation on K, and all n are in K, the first statement follows.

For the second, suppose that $|\cdot|$ is non-archimedean (but not necessarily discrete). Suppose $b \in K_v$ with $b \neq 0$. First I claim that there is $c \in K$ such that $|b - c| < |b|$. To see this, let $c' = b - \frac{b}{a}$, where a is some element of K_v with $|a| > 1$, note that $|b - c'| = \left|\frac{b}{a}\right| < |b|$, and choose $c \in K$ such that $|c - c'| < |b - c'|$, so

$$|b - c| = \left|b - c' - (c - c')\right| \leq \max\left(\left|b - c'\right|, \left|c - c'\right|\right) = \left|b - c'\right| < |b|.$$

Since $|\cdot|$ is non-archimedean, we have

$$|b| = |(b - c) + c| \leq \max\left(|b - c|, |c|\right) = |c|,$$

where in the last equality we use that $|b - c| < |b|$. Also,

$$|c| = |b + (c - b)| \leq \max\left(|b|, |c - b|\right) = |b|,$$

so $|b| = |c|$, which is in the set of values of $|\cdot|$ on K. □

16.2.1 p-adic Numbers

This section is about the p-adic numbers $\mathbf{Q}_p$, which are the completion of $\mathbf{Q}$ with respect to the p-adic valuation. Alternatively, to give a p-adic *integer* in $\mathbf{Z}_p$ is the same as giving for every prime power p^r an element $a_r \in \mathbf{Z}/p^r\mathbf{Z}$ such that if $s \leq r$ then a_s is the reduction of a_r modulo p^s. The field $\mathbf{Q}_p$ is then the field of fractions of $\mathbf{Z}_p$.

We begin with the definition of the N-adic numbers for any positive integer N. Section 16.2.1 is about the N-adics in the special case $N = 10$; these are fun because they can be represented as decimal expansions that go off infinitely far to the left. Section 16.2.3 is about how the topology of $\mathbf{Q}_N$ is nothing like the topology of $\mathbf{R}$. Finally, in Section 16.2.4 we state the Hasse-Minkowski theorem, which shows how to use p-adic numbers to decide whether or not a quadratic equation in n variables has a rational zero.

The N-adic Numbers

Lemma 16.2.5. *Let N be a positive integer. Then for any nonzero rational number α there exists a unique $e \in \mathbf{Z}$ and integers a, b, with b positive, such that $\alpha = N^e \cdot \frac{a}{b}$ with $N \nmid a$, $\gcd(a, b) = 1$, and $\gcd(N, b) = 1$.*

Proof. Write $\alpha = c/d$ with $c, d \in \mathbf{Z}$ and $d > 0$. First suppose d is exactly divisible by a power of N, so for some r we have $N^r \mid d$ but $\gcd(N, d/N^r) = 1$. Then

$$\frac{c}{d} = N^{-r}\frac{c}{d/N^r}.$$

If N^s is the largest power of N that divides c, then $e = s - r$, $a = c/N^s$, $b = d/N^r$ satisfy the conclusion of the lemma.

By unique factorization of integers, there is a smallest multiple f of d such that fd is exactly divisible by N. Now apply the above argument with c and d replaced by cf and df. □

Definition 16.2.6 (N-adic valuation). Let N be a positive integer. For any positive $\alpha \in \mathbf{Q}$, the *N-adic valuation* of α is e, where e is as in Lemma 16.2.5. The N-adic valuation of 0 is ∞.

We denote the N-adic valuation of α by $\mathrm{ord}_N(\alpha)$. (Note: Here we are using "valuation" in a different way than in the rest of the text. This valuation is not an absolute value, but the logarithm of one.)

Definition 16.2.7 (N-adic metric). For $x, y \in \mathbf{Q}$ the *N-adic distance* between x and y is

$$d_N(x, y) = N^{-\mathrm{ord}_N(x-y)}.$$

We let $d_N(x, x) = 0$, since $\mathrm{ord}_N(x - x) = \mathrm{ord}_N(0) = \infty$.

For example, $x, y \in \mathbf{Z}$ are close in the N-adic metric if their difference is divisible by a large power of N. E.g., if $N = 10$ then 93427 and 13427 are close because their difference is 80000, which is divisible by a large power of 10.

Proposition 16.2.8. *The distance d_N on $\mathbf{Q}$ defined above is a metric. Moreover, for all $x, y, z \in \mathbf{Q}$ we have*

$$d(x, z) \leq \max(d(x, y), d(y, z)).$$

(This is the "nonarchimedean" triangle inequality.)

Proof. The first two properties of Definition 16.2.1 are immediate. For the third, we first prove that if $\alpha, \beta \in \mathbf{Q}$ then

$$\mathrm{ord}_N(\alpha+\beta) \geq \min(\mathrm{ord}_N(\alpha), \mathrm{ord}_N(\beta)).$$

Assume, without loss, that $\mathrm{ord}_N(\alpha) \leq \mathrm{ord}_N(\beta)$ and that both α and β are nonzero. Using Lemma 16.2.5 write $\alpha = N^e(a/b)$ and $\beta = N^f(c/d)$ with a or c possibly negative. Then

$$\alpha+\beta = N^e\left(\frac{a}{b} + N^{f-e}\frac{c}{d}\right) = N^e\left(\frac{ad + bcN^{f-e}}{bd}\right).$$

Since $\gcd(N, bd) = 1$ it follows that $\mathrm{ord}_N(\alpha+\beta) \geq e$. Now suppose $x, y, z \in \mathbf{Q}$. Then

$$x - z = (x-y) + (y-z),$$

so

$$\mathrm{ord}_N(x-z) \geq \min(\mathrm{ord}_N(x-y), \mathrm{ord}_N(y-z)),$$

hence $d_N(x,z) \leq \max(d_N(x,y), d_N(y,z))$. □

We can finally define the N-adic numbers.

Definition 16.2.9 (The N-adic Numbers). The set of *N-adic numbers*, denoted $\mathbf{Q}_N$, is the completion of $\mathbf{Q}$ with respect to the metric d_N.

The set $\mathbf{Q}_N$ is a ring, but it need not be a field as you will show in Exercises 57 and 58. It is a field if and only if N is prime. Also, $\mathbf{Q}_N$ has a "bizarre" topology, as we will see in Section 16.2.3.

The 10-adic Numbers

It's a familiar fact that every real number can be written in the form

$$d_n \dots d_1 d_0 . d_{-1} d_{-2} \dots = d_n 10^n + \cdots + d_1 10 + d_0 + d_{-1}10^{-1} + d_{-2}10^{-2} + \cdots$$

where each digit d_i is between 0 and 9, and the sequence can continue indefinitely to the right.

The 10-adic numbers also have decimal expansions, but everything is backward! To get a feeling for why this might be the case, we consider Euler's nonsensical series

$$\sum_{n=1}^{\infty}(-1)^{n+1}n! = 1! - 2! + 3! - 4! + 5! - 6! + \cdots.$$

One can prove (see Exercise 55) that this series converges in $\mathbf{Q}_{10}$ to some element $\alpha \in \mathbf{Q}_{10}$.

What is α? How can we write it down? First note that for all $M \geq 5$, the terms of the sum are divisible by 10, so the difference between α and $1! - 2! + 3! - 4!$ is divisible by 10. Thus we can compute α modulo 10 by computing $1! - 2! + 3! - 4!$ modulo 10. Likewise, we can compute α modulo 100 by compute $1! - 2! + \cdots + 9! - 10!$, etc. We obtain the following table:

α	mod 10^r
1	mod 10
81	mod 10^2
981	mod 10^3
2981	mod 10^4
22981	mod 10^5
422981	mod 10^6

Continuing we see that

$$1! - 2! + 3! - 4! + \cdots = \ldots 637838364422981 \qquad \text{in } \mathbf{Q}_{10}\ !$$

Here's another example. Reducing 1/7 modulo larger and larger powers of 10 we see that

$$\frac{1}{7} = \ldots 857142857143 \qquad \text{in } \mathbf{Q}_{10}.$$

Here's another example, but with a decimal point.

$$\frac{1}{70} = \frac{1}{10} \cdot \frac{1}{7} = \ldots 85714285714.3$$

We have

$$\frac{1}{3} + \frac{1}{7} = \ldots 66667 + \ldots 57143 = \frac{10}{21} = \ldots 23810,$$

which illustrates that addition with carrying works as usual.

Fermat's Last Theorem in $\mathbf{Z}_{10}$

An amusing observation, which people often argued about on USENET news back in the 1990s, is that Fermat's last theorem is false in $\mathbf{Z}_{10}$. For example, $x^3 + y^3 = z^3$ has a nontrivial solution, namely $x = 1$, $y = 2$, and $z = \ldots 60569$. Here z is a cube root of 9 in $\mathbf{Z}_{10}$. Note that it takes some work to prove that there is a cube root of 9 in $\mathbf{Z}_{10}$ (see Exercise 56).

16.2.2 The Field of p-adic Numbers

The ring $\mathbf{Q}_{10}$ of 10-adic numbers is isomorphic to $\mathbf{Q}_2 \times \mathbf{Q}_5$ (see Exercise 58), so it is not a field. For example, the element $\ldots 8212890625$ corresponding to $(1, 0)$ under this isomorphism has no inverse. (To compute n digits of $(1, 0)$ use the Chinese remainder theorem to find a number that is 1 modulo 2^n and 0 modulo 5^n.)

If p is prime then $\mathbf{Q}_p$ is a field (see Exercise 57). Since $p \neq 10$ it is a little more complicated to write p-adic numbers down. People typically write p-adic numbers in the form

$$\frac{a_{-d}}{p^d} + \cdots + \frac{a_{-1}}{p} + a_0 + a_1 p + a_2 p^2 + a_3 p^3 + \cdots$$

where $0 \le a_i < p$ for each i.

16.2.3 The Topology of $\mathbf{Q}_N$ (is Weird)

Definition 16.2.10 (Connected). Let X be a topological space. A subset S of X is *disconnected* if there exist open subsets $U_1, U_2 \subset X$ with $U_1 \cap U_2 \cap S = \emptyset$ and $S = (S \cap U_1) \cup (S \cap U_2)$ with $S \cap U_1$ and $S \cap U_2$ nonempty. If S is not disconnected it is *connected*.

The topology on $\mathbf{Q}_N$ is induced by d_N, so every open set is a union of open balls

$$B(x, r) = \{y \in \mathbf{Q}_N : d_N(x, y) < r\}.$$

Recall Proposition 16.2.8, which asserts that for all x, y, z,

$$d(x, z) \leq \max(d(x, y), d(y, z)).$$

This translates into the following shocking and bizarre lemma:

Lemma 16.2.11. *Suppose $x \in \mathbf{Q}_N$ and $r > 0$. If $y \in \mathbf{Q}_N$ and $d_N(x, y) \geq r$, then $B(x, r) \cap B(y, r) = \emptyset$.*

Proof. Suppose $z \in B(x, r)$ and $z \in B(y, r)$. Then

$$r \leq d_N(x, y) \leq \max(d_N(x, z), d_N(z, y)) < r,$$

a contradiction. □

You should draw a picture to illustrates Lemma 16.2.11.

Lemma 16.2.12. *The open ball $B(x, r)$ is also closed.*

Proof. Suppose $y \notin B(x, r)$. Then $r \leq d(x, y)$ so

$$B(y, d(x, y)) \cap B(x, r) \subset B(y, d(x, y)) \cap B(x, d(x, y)) = \emptyset.$$

Thus the complement of $B(x, r)$ is a union of open balls. □

The lemmas imply that $\mathbf{Q}_N$ is *totally disconnected*, in the following sense.

Proposition 16.2.13. *The only connected subsets of $\mathbf{Q}_N$ are the singleton sets $\{x\}$ for $x \in \mathbf{Q}_N$ and the empty set.*

Proof. Suppose $S \subset \mathbf{Q}_N$ is a nonempty connected set and x, y are distinct elements of S. Let $r = d_N(x, y) > 0$. Let $U_1 = B(x, r)$ and U_2 be the complement of U_1, which is open by Lemma 16.2.12. Then U_1 and U_2 satisfies the conditions of Definition 16.2.10, so S is not connected, a contradiction. □

16.2.4 The Local-to-Global Principle of Hasse and Minkowski

Section 16.2.3 might have convinced you that $\mathbf{Q}_N$ is a bizarre pathology. In fact, $\mathbf{Q}_N$ is omnipresent in number theory, as the following two fundamental examples illustrate.

In the statement of the following theorem, a *nontrivial solution* to a homogeneous polynomial equation is a solution where not all indeterminates are 0.

Theorem 16.2.14 (Hasse-Minkowski). *The quadratic equation*

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2 = 0, \tag{16.2.1}$$

with $a_i \in \mathbf{Q}^\times$, has a nontrivial solution with $x_1, \ldots, x_n$ in $\mathbf{Q}$ if and only if (16.2.1) has a solution in $\mathbf{R}$ and in $\mathbf{Q}_p$ for all primes p.

This theorem is very useful in practice because the p-adic condition turns out to be easy to check. For more details, including a complete proof, see [Ser73, IV.3.2].

The analogue of Theorem 16.2.14 for cubic equations is false. For example, Selmer proved that the cubic

$$3x^3 + 4y^3 + 5z^3 = 0$$

has a solution other than $(0,0,0)$ in $\mathbf{R}$ and in $\mathbf{Q}_p$ for all primes p but has no solution other than $(0,0,0)$ in $\mathbf{Q}$ (for a proof see [Cas91, §18]).

Open Problem. Give an algorithm that decides whether or not a cubic

$$ax^3 + by^3 + cz^3 = 0$$

has a nontrivial solution in $\mathbf{Q}$.

This open problem is closely related to the Birch and Swinnerton-Dyer Conjecture for elliptic curves. The truth of the conjecture would follow if we knew that "Shafarevich-Tate Groups" of elliptic curves were finite.

16.3 Weak Approximation

The following theorem asserts that inequivalent valuations are in fact almost totally indepedent. For our purposes it will be superseded by the strong approximation theorem (Theorem 20.4.4).

Theorem 16.3.1 (Weak Approximation). *Let $|\cdot|_n$, for $1 \leq n \leq N$, be inequivalent nontrivial valuations of a field K. For each n, let K_n be the topological space consisting of the set of elements of K with the topology induced by $|\cdot|_n$. Let Δ be the image of K in the topological product*

$$A = \prod_{1 \leq n \leq N} K_n$$

equipped with the product topology. Then Δ is dense in A.

The conclusion of the theorem may be expressed in a less topological manner as follows: given any $a_n \in K$, for $1 \leq n \leq N$, and real $\varepsilon > 0$, there is an $b \in K$ such that simultaneously

$$|a_n - b|_n < \varepsilon \qquad (1 \leq n \leq N).$$

If $K = \mathbf{Q}$ and the $|\cdot|$ are p-adic valuations, Theorem 16.3.1 is related to the Chinese Remainder Theorem (Theorem 9.1.3), but the strong approximation theorem (Theorem 20.4.4) is the real generalization.

Proof. We note first that it will be enough to find, for each n, an element $c_n \in K$ such that

$$|c_n|_n > 1 \quad \text{and} \quad |c_n|_m < 1 \quad \text{for } n \neq m,$$

where $1 \leq n, m \leq N$. For then as $r \to +\infty$, we have

$$\frac{c_n^r}{1 + c_n^r} = \frac{1}{1 + \left(\frac{1}{c_n}\right)^r} \to \begin{cases} 1 & \text{with respect to } |\cdot|_n \text{ and} \\ 0 & \text{with respect to } |\cdot|_m\text{, for } m \neq n. \end{cases}$$

It is then enough to take

$$b = \sum_{n=1}^{N} \frac{c_n^r}{1 + c_n^r} \cdot a_n$$

By symmetry it is enough to show the existence of $c = c_1$ with

$$|c|_1 > 1 \qquad \text{and} \qquad |c|_n < 1 \quad \text{for} \quad 2 \leq n \leq N.$$

We will do this by induction on N.

First suppose $N = 2$. Since $|\cdot|_1$ and $|\cdot|_2$ are inequivalent (and all absolute values are assumed nontrivial) there is an $a \in K$ such that

$$|a|_1 < 1 \qquad \text{and} \qquad |a|_2 \geq 1 \tag{16.3.1}$$

and similarly a b such that

$$|b|_1 \geq 1 \qquad \text{and} \qquad |b|_2 < 1.$$

Then $c = \dfrac{b}{a}$ will do.

Remark 16.3.2. It is not completely clear that one can choose an a such that (16.3.1) is satisfied. Suppose it were impossible. Then because the valuations are nontrivial, we would have that for any $a \in K$ if $|a|_1 < 1$ then $|a|_2 < 1$. This implies the converse statement: if $a \in K$ and $|a|_2 < 1$ then $|a|_1 < 1$. To see this, suppose there is an $a \in K$ such that $|a|_2 < 1$ and $|a|_1 \geq 1$. Choose $y \in K$ such that $|y|_1 < 1$. Then for any integer $n > 0$ we have $|y/a^n|_1 < 1$, so by hypothesis $|y/a^n|_2 < 1$. Thus $|y|_2 < |a|_2^n < 1$ for all n. Since $|a|_2 < 1$ we have $|a|_2^n \to 0$ as $n \to \infty$, so $|y|_2 = 0$, a contradiction since $y \neq 0$. Thus $|a|_1 < 1$ if and only if $|a|_2 < 1$, and we have proved before that this implies that $|\cdot|_1$ is equivalent to $|\cdot|_2$.

Next suppose $N \geq 3$. By the case $N-1$, there is an $a \in K$ such that

$$|a|_1 > 1 \qquad \text{and} \qquad |a|_n < 1 \quad \text{for} \quad 2 \leq n \leq N-1.$$

By the case for $N = 2$ there is a $b \in K$ such that

$$|b|_1 > 1 \qquad \text{and} \qquad |b|_N < 1.$$

Then put

$$c = \begin{cases} a & \text{if } |a|_N < 1 \\ a^r \cdot b & \text{if } |a|_N = 1 \\ \dfrac{a^r}{1+a^r} \cdot b & \text{if } |a|_N > 1 \end{cases}$$

where $r \in \mathbf{Z}$ is sufficiently large so that $|c|_1 > 1$ and $|c|_n < 1$ for $2 \leq n \leq N$. □

Example 16.3.3. Suppose $K = \mathbf{Q}$, let $|\cdot|_1$ be the archimedean absolute value and let $|\cdot|_2$ be the 2-adic absolute value. Let $a_1 = -1$, $a_2 = 8$, and $\varepsilon = 1/10$, as in the remark right after Theorem 16.3.1. Then the theorem implies that there is an element $b \in \mathbf{Q}$ such that

$$|-1-b|_1 < \frac{1}{10} \qquad \text{and} \qquad |8-b|_2 < \frac{1}{10}.$$

As in the proof of the theorem, we can find such a b by finding a $c_1, c_2 \in \mathbf{Q}$ such that $|c_1|_1 > 1$ and $|c_1|_2 < 1$, and a $|c_2|_1 < 1$ and $|c_2|_2 > 1$. For example, $c_1 = 2$ and $c_2 = 1/2$ works, since $|2|_1 = 2$ and $|2|_2 = 1/2$ and $|1/2|_1 = 1/2$ and $|1/2|_2 = 2$. Again following the proof, we see that for sufficiently large r we can take

$$\begin{aligned} b_r &= \frac{c_1^r}{1+c_1^r} \cdot a_1 + \frac{c_2^r}{1+c_2^r} \cdot a_2 \\ &= \frac{2^r}{1+2^r} \cdot (-1) + \frac{(1/2)^r}{1+(1/2)^r} \cdot 8. \end{aligned}$$

We have $b_1 = 2$, $b_2 = 4/5$, $b_3 = 0$, $b_4 = -8/17$, $b_5 = -8/11$, $b_6 = -56/55$. None of the b_i work for $i < 6$, but b_6 works.

Chapter 17

Adic Numbers: The Finite Residue Field Case

17.1 Finite Residue Field Case

Let K be a field with a non-archimedean valuation $v = |\cdot|$. Recall that the set of $a \in K$ with $|a| \leq 1$ forms a ring $\mathcal{O}$, the ring of integers for v. The set of $u \in K$ with $|u| = 1$ are a group U under multiplication, the group of units for v. Finally, the set of $a \in K$ with $|a| < 1$ is a maximal ideal $\mathfrak{p}$, so the quotient ring $\mathcal{O}/\mathfrak{p}$ is a field. In this section we consider the case when $\mathcal{O}/\mathfrak{p}$ is a finite field of order a prime power q. For example, K could be $\mathbf{Q}$ and $|\cdot|$ could be a p-adic valuation, or K could be a number field and $|\cdot|$ could be the valuation corresponding to a maximal ideal of the ring of integers. Among other things, we will discuss in more depth the topological and measure-theoretic nature of the completion of K at v.

Suppose further for the rest of this section that $|\cdot|$ is discrete. Then by Lemma 15.2.8, the ideal $\mathfrak{p}$ is a principal ideal (π), say, and every $a \in K$ is of the form $a = \pi^n \varepsilon$, where $n \in \mathbf{Z}$ and $\varepsilon \in U$ is a unit. We call

$$n = \operatorname{ord}(a) = \operatorname{ord}_\pi(a) = \operatorname{ord}_\mathfrak{p}(a) = \operatorname{ord}_v(a)$$

the ord of a at v. (Some authors, including me (!) also call this integer the *valuation* of a with respect to v.) If $\mathfrak{p} = (\pi')$, then π/π' is a unit, and conversely, so $\operatorname{ord}(a)$ is independent of the choice of π.

Let $\mathcal{O}_v$ and $\mathfrak{p}_v$ be defined with respect to the completion K_v of K at v.

Lemma 17.1.1. *There is a natural isomorphism*

$$\varphi : \mathcal{O}_v/\mathfrak{p}_v \to \mathcal{O}/\mathfrak{p},$$

and $\mathfrak{p}_v = (\pi)$ *as an* $\mathcal{O}_v$*-ideal.*

Proof. We may view $\mathcal{O}_v$ as the set of equivalence classes of Cauchy sequences (a_n) in K such that $a_n \in \mathcal{O}$ for n sufficiently large. For any ε, given such a sequence

(a_n), there is N such that for $n, m \geq N$, we have $|a_n - a_m| < \varepsilon$. In particular, we can choose N such that $n, m \geq N$ implies that $a_n \equiv a_m \pmod{\mathfrak{p}}$. Let $\varphi((a_n)) = a_N \pmod{\mathfrak{p}}$, which is well-defined. The map φ is surjective because the constant sequences are in $\mathcal{O}_v$. Its kernel is the set of Cauchy sequences whose elements are eventually all in $\mathfrak{p}$, which is exactly $\mathfrak{p}_v$. This proves the first part of the lemma. The second part is true because any element of $\mathfrak{p}_v$ is a sequence all of whose terms are eventually in $\mathfrak{p}$, hence all a multiple of π (we can set to 0 a finite number of terms of the sequence without changing the equivalence class of the sequence). □

Assume for the rest of this section that K is complete with respect to $|\cdot|$.

Lemma 17.1.2. *Then ring $\mathcal{O}$ is precisely the set of infinite sums*

$$a = \sum_{j=0}^{\infty} a_j \cdot \pi^j \tag{17.1.1}$$

where the a_j run independently through some set $\mathcal{R}$ of representatives of $\mathcal{O}$ in $\mathcal{O}/\mathfrak{p}$.

By (17.1.1) is meant the limit of the Cauchy sequence $\sum_{j=0}^{n} a_j \cdot \pi^j$ as $j \to \infty$.

Proof. There is a uniquely defined $a_0 \in \mathcal{R}$ such that $|a - a_0| < 1$. Then $a' = \pi^{-1} \cdot (a - a_0) \in \mathcal{O}$. Now define $a_1 \in \mathcal{R}$ by $|a' - a_1| < 1$. And so on. □

Example 17.1.3. Suppose $K = \mathbf{Q}$ and $|\cdot| = |\cdot|_p$ is the p-adic valuation, for some prime p. We can take $\mathcal{R} = \{0, 1, \ldots, p-1\}$. The lemma asserts that

$$\mathcal{O} = \mathbf{Z}_p = \left\{ \sum_{j=0}^{\infty} a_n p^n : 0 \leq a_n \leq p - 1 \right\}.$$

Notice that $\mathcal{O}$ is uncountable since there are p choices for each p-adic "digit". We can do arithmetic with elements of $\mathbf{Z}_p$, which can be thought of "backwards" as numbers in base p. For example, with $p = 3$ we have

$$\begin{aligned}
&(1 + 2\cdot 3 + 3^2 + \cdots) + (2 + 2\cdot 3 + 3^2 + \cdots)\\
&= 3 + 4\cdot 3 + 2\cdot 3^2 + \cdots \qquad \text{not in canonical form}\\
&= 0 + 2\cdot 3 + 3\cdot 3 + 2\cdot 3^2 + \cdots \qquad \text{still not canonical}\\
&= 0 + 2\cdot 3 + 0\cdot 3^2 + \cdots
\end{aligned}$$

Basic arithmetic with the p-adics in MAGMA is really weird (even weirder than it was a year ago... There are presumably efficiency advantages to using the MAGMA formalization, and it's supposed to be better for working with extension fields. But I can't get it to do even the calculation below in a way that is clear.) In PARI (gp) the p-adics work as expected:

```
? a = 1 + 2*3 + 3^2 + O(3^3);
? b = 2 + 2*3 + 3^2 + O(3^3);
? a+b
%3 = 2*3 + O(3^3)
? sqrt(1+2*3+O(3^20))
%5 = 1 + 3 + 3^2 + 2*3^4 + 2*3^7 + 3^8 + 3^9 + 2*3^10 + 2*3^12
       + 2*3^13 + 2*3^14 + 3^15 + 2*3^17 + 3^18 + 2*3^19 + O(3^20)
? 1/sqrt(1+2*3+O(3^20))
%6 = 1 + 2*3 + 2*3^2 + 2*3^7 + 2*3^10 + 2*3^11 + 2*3^12 + 2*3^13
       + 2*3^14 + 3^15 + 2*3^16 + 2*3^17 + 3^18 + 3^19 + O(3^20)
```

Theorem 17.1.4. *Under the conditions of the preceding lemma, $\mathcal{O}$ is compact with respect to the $|\cdot|$-topology.*

Proof. Let V_λ, for λ running through some index set Λ, be some family of open sets that cover $\mathcal{O}$. We must show that there is a finite subcover. We suppose not.

Let $\mathcal{R}$ be a set of representatives for $\mathcal{O}/\mathfrak{p}$. Then $\mathcal{O}$ is the union of the finite number of cosets $a + \pi\mathcal{O}$, for $a \in \mathcal{R}$. Hence for at lest one $a_0 \in \mathcal{R}$ the set $a_0 + \pi\mathcal{O}$ is not covered by finitely many of the V_λ. Then similarly there is an $a_1 \in \mathcal{R}$ such that $a_0 + a_1\pi + \pi^2\mathcal{O}$ is not finitely covered. And so on. Let

$$a = a_0 + a_1\pi + a_2\pi^2 + \cdots \in \mathcal{O}.$$

Then $a \in V_{\lambda_0}$ for some $\lambda_0 \in \Lambda$. Since V_{λ_0} is an open set, $a + \pi^J \cdot \mathcal{O} \subset V_{\lambda_0}$ for some J (since those are exactly the open balls that form a basis for the topology). This is a contradiction because we constructed a so that none of the sets $a + \pi^n \cdot \mathcal{O}$, for each n, are not covered by any finite subset of the V_λ. $\square$

Definition 17.1.5 (Locally compact). A topological space X is *locally compact* at a point x if there is some compact subset C of X that contains a neighborhood of x. The space X is locally compact if it is locally compact at each point in X.

Corollary 17.1.6. *The complete local field K is locally compact.*

Proof. If $x \in K$, then $x \in C = x + \mathcal{O}$, and C is a compact subset of K by Theorem 17.1.4. Also C contains the neighborhood $x + \pi\mathcal{O} = B(x, 1)$ of x. Thus K is locally compact at x. $\square$

Remark 17.1.7. The converse is also true. If K is locally compact with respect to a non-archimedean valuation $|\cdot|$, then

1. K is complete,
2. the residue field is finite, and
3. the valuation is discrete.

For there is a compact neighbourhood C of 0. Let π be any nonzero with $|\pi| < 1$. Then $\pi^n \cdot \mathcal{O} \subset C$ for sufficiently large n, so $\pi^n \cdot \mathcal{O}$ is compact, being closed. Hence $\mathcal{O}$ is compact. Since $|\cdot|$ is a metric, $\mathcal{O}$ is sequentially compact, i.e., every fundamental sequence in $\mathcal{O}$ has a limit, which implies (1). Let a_λ (for $\lambda \in \Lambda$) be a set of representatives in $\mathcal{O}$ of $\mathcal{O}/\mathfrak{p}$. Then $\mathcal{O}_\lambda = \{z : |z - a_\lambda| < 1\}$ is an open covering of $\mathcal{O}$. Thus (2) holds since $\mathcal{O}$ is compact. Finally, $\mathfrak{p}$ is compact, being a closed subset of $\mathcal{O}$. Let S_n be the set of $a \in K$ with $|a| < 1 - 1/n$. Then S_n (for $1 \leq n < \infty$) is an open covering of $\mathfrak{p}$, so $\mathfrak{p} = S_n$ for some n, i.e., (3) is true.

If we allow $|\cdot|$ to be archimedean the only further possibilities are $k = \mathbf{R}$ and $k = \mathbf{C}$ with $|\cdot|$ equivalent to the usual absolute value.

We denote by K^+ the commutative topological group whose points are the elements of K, whose group law is addition and whose topology is that induced by $|\cdot|$. General theory tells us that there is an invariant Haar measure defined on K^+ and that this measure is unique up to a multiplicative constant.

Definition 17.1.8 (Haar Measure). A *Haar measure* on a locally compact topological group G is a translation invariant measure such that every open set can be covered by open sets with finite measure.

Lemma 17.1.9. *Haar measure of any compact subset C of G is finite.*

Proof. The whole group G is open, so there is a covering U_α of G by open sets each of which has finite measure. Since C is compact, there is a finite subset of the U_α that covers C. The measure of C is at most the sum of the measures of these finitely many U_α, hence finite. □

Remark 17.1.10. Usually one defined Haar measure to be a translation invariant measure such that the measure of compact sets is finite. Because of local compactness, this definition is equivalent to Definition 17.1.8. We take this alternative viewpoint because Haar measure is constructed naturally on the topological groups we will consider by defining the measure on each member of a basis of open sets for the topology.

We now deduce what any such measure μ on $G = K^+$ must be. Since $\mathcal{O}$ is compact (Theorem 17.1.4), the measure of $\mathcal{O}$ is finite. Since μ is translation invariant,

$$\mu_n = \mu(a + \pi^n \mathcal{O})$$

is independent of a. Further,

$$a + \pi^n \mathcal{O} = \bigcup_{1 \leq j \leq q} a + \pi^n a_j + \pi^{n+1}\mathcal{O}, \qquad \text{(disjoint union)}$$

where a_j (for $1 \leq j \leq q$) is a set of representatives of $\mathcal{O}/\mathfrak{p}$. Hence

$$\mu_n = q \cdot \mu_{n+1}.$$

If we normalize μ by putting

$$\mu(\mathcal{O}) = 1$$

we have $\mu_0 = 1$, hence $\mu_1 = q$, and in general

$$\mu_n = q^{-n}.$$

Conversely, without the theory of Haar measure, we could *define* μ to be the necessarily unique measure on K^+ such that $\mu(\mathcal{O}) = 1$ that is translation invariant. This would have to be the μ we just found above.

Everything so far in this section has depended not on the valuation $|\cdot|$ but only on its equivalence class. The above considerations now single out one valuation in the equivalence class as particularly important.

Definition 17.1.11 (Normalized valuation). Let K be a field equipped with a discrete valuation $|\cdot|$ and residue class field with $q < \infty$ elements. We say that $|\cdot|$ is *normalized* if

$$|\pi| = \frac{1}{q},$$

where $\mathfrak{p} = (\pi)$ is the maximal ideal of $\mathcal{O}$.

Example 17.1.12. The normalized valuation on the p-adic numbers $\mathbf{Q}_p$ is $|u \cdot p^n| = p^{-n}$, where u is a rational number whose numerator and denominator are coprime to p.

Next suppose $K = \mathbf{Q}_p(\sqrt{p})$. Then the p-adic valuation on $\mathbf{Q}_p$ extends uniquely to one on K such that $\left|\sqrt{p}\right|^2 = |p| = 1/p$. Since $\pi = \sqrt{p}$ for K, this valuation is not normalized. (Note that the ord of $\pi = \sqrt{p}$ is $1/2$.) The normalized valuation is $v = |\cdot|' = |\cdot|^2$. Note that $|\cdot|' \, p = 1/p^2$, or $\mathrm{ord}_v(p) = 2$ instead of 1.

Finally suppose that $K = \mathbf{Q}_p(\sqrt{q})$ where $x^2 - q$ has not root mod p. Then the residue class field degree is 2, and the normalized valuation must satisfy $\left|\sqrt{q}\right| = 1/p^2$.

The following proposition makes clear why this is the best choice of normalization.

Theorem 17.1.13. *Suppose further that K is complete with respect to the normalized valuation $|\cdot|$. Then*

$$\mu(a + b\mathcal{O}) = |b|\,,$$

where μ is the Haar measure on K^+ normalized so that $\mu(\mathcal{O}) = 1$.

Proof. Since μ is translation invariant, $\mu(a + b\mathcal{O}) = \mu(b\mathcal{O})$. Write $b = u \cdot \pi^n$, where u is a unit. Then since $u \cdot \mathcal{O} = \mathcal{O}$, we have

$$\mu(b\mathcal{O}) = \mu(u \cdot \pi^n \cdot \mathcal{O}) = \mu(\pi^n \cdot u \cdot \mathcal{O}) = \mu(\pi^n \cdot \mathcal{O}) = q^{-n} = |\pi^n| = |b|\,.$$

Here we have $\mu(\pi^n \cdot \mathcal{O}) = q^{-n}$ by the discussion before Definition 17.1.11. □

We can express the result of the theorem in a more suggestive way. Let $b \in K$ with $b \neq 0$, and let μ be a Haar measure on K^+ (not necessarily normalized as in the theorem). Then we can define a new Haar measure μ_b on K^+ by putting $\mu_b(E) = \mu(bE)$ for $E \subset K^+$. But Haar measure is unique up to a multiplicative constant and so $\mu_b(E) = \mu(bE) = c \cdot \mu(E)$ for all measurable sets E, where the factor c depends only on b. Putting $E = \mathcal{O}$, shows that the theorem implies that c is just $|b|$, when $|\cdot|$ is the normalized valuation.

Remark 17.1.14. The theory of locally compact topological groups leads to the consideration of the dual (character) group of K^+. It turns out that it is isomorphic to K^+. We do not need this fact for class field theory, so do not prove it here. For a proof and applications see Tate's thesis or Lang's *Algebraic Numbers*, and for generalizations see Weil's *Adeles and Algebraic Groups* and Godement's Bourbaki seminars 171 and 176. The determination of the character group of K^* is local class field theory.

The set of nonzero elements of K is a group K^* under multiplication. Multiplication and inverses are continuous with respect to the topology induced on K^* as a subset of K, so K^* is a topological group with this topology. We have

$$U_1 \subset U \subset K^*$$

where U is the group of units of $\mathcal{O} \subset K$ and U_1 is the group of 1-units, i.e., those units $\varepsilon \in U$ with $|\varepsilon - 1| < 1$, so

$$U_1 = 1 + \pi\mathcal{O}.$$

The set U is the open ball about 0 of radius 1, so is open, and because the metric is nonarchimedean U is also closed. Likewise, U_1 is both open and closed.

The quotient $K^*/U = \{\pi^n \cdot U : n \in \mathbf{Z}\}$ is isomorphic to the additive group $\mathbf{Z}^+$ of integers with the discrete topology, where the map is

$$\pi^n \cdot U \mapsto n \qquad \text{for } n \in \mathbf{Z}.$$

The quotient U/U_1 is isomorphic to the multiplicative group $\mathbf{F}^*$ of the nonzero elements of the residue class field, where the finite gorup $\mathbf{F}^*$ has the discrete topology. Note that $\mathbf{F}^*$ is cyclic of order $q-1$, and Hensel's lemma implies that K^* contains a primitive $(q-1)$th root of unity ζ. Thus K^* has the following structure:

$$K^* = \{\pi^n\zeta^m\varepsilon : n \in \mathbf{Z}, m \in \mathbf{Z}/(q-1)\mathbf{Z}, \varepsilon \in U_1\} \cong \mathbf{Z} \ \times \ \mathbf{Z}/(q-1)\mathbf{Z} \ \times \ U_1.$$

(How to apply Hensel's lemma: Let $f(x) = x^{q-1} - 1$ and let $a \in \mathcal{O}$ be such that a mod $\mathfrak{p}$ generates K^*. Then $|f(a)| < 1$ and $|f'(a)| = 1$. By Hensel's lemma there is a $\zeta \in K$ such that $f(\zeta) = 0$ and $\zeta \equiv a \pmod{\mathfrak{p}}$.)

Since U is compact and the cosets of U cover K, we see that K^* is locally compact.

Lemma 17.1.15. *The additive Haar measure μ on K^+, when restricted to U_1 gives a measure on U_1 that is also invariant under multiplication, so gives a Haar measure on U_1.*

Proof. It suffices to show that

$$\mu(1+\pi^n\mathcal{O}) = \mu(u\cdot(1+\pi^n\mathcal{O})),$$

for any $u \in U_1$ and $n > 0$. Write $u = 1 + a_1\pi + a_2\pi^2 + \cdots$. We have

$$\begin{aligned} u\cdot(1+\pi^n\mathcal{O}) &= (1 + a_1\pi + a_2\pi^2 + \cdots)\cdot(1+\pi^n\mathcal{O}) \\ &= 1 + a_1\pi + a_2\pi^2 + \cdots + \pi^n\mathcal{O} \\ &= a_1\pi + a_2\pi^2 + \cdots + (1+\pi^n\mathcal{O}), \end{aligned}$$

which is an additive translate of $1+\pi^n\mathcal{O}$, hence has the same measure. □

Thus μ gives a Haar measure on K^* by translating U_1 around to cover K^*.

Lemma 17.1.16. *The topological spaces K^+ and K^* are totally disconnected (the only connected sets are points).*

Proof. The proof is the same as that of Proposition 16.2.13. The point is that the non-archimedean triangle inequality forces the complement an open disc to be open, hence any set with at least two distinct elements "falls apart" into a disjoint union of two disjoint open subsets. □

Remark 17.1.17. Note that K^* and K^+ are locally isomorphic if K has characteristic 0. We have the exponential map

$$a \mapsto \exp(a) = \sum_{n=0}^{\infty} \frac{a^n}{n!}$$

defined for all sufficiently small a with its inverse

$$\log(a) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(a-1)^n}{n},$$

which is defined for all a sufficiently close to 1.

Chapter 18

Normed Spaces and Tensor Products

Much of this chapter is preparation for what we will do later when we will prove that if K is complete with respect to a valuation (and locally compact) and L is a finite extension of K, then there is a *unique* valuation on L that extends the valuation on K. Also, if K is a number field, $v = |\cdot|$ is a valuation on K, K_v is the completion of K with respect to v, and L is a finite extension of K, we'll prove that

$$K_v \otimes_K L = \bigoplus_{j=1}^{J} L_j,$$

where the L_j are the completions of L with respect to the equivalence classes of extensions of v to L. In particular, if L is a number field defined by a root of $f(x) \in \mathbf{Q}[x]$, then

$$\mathbf{Q}_p \otimes_{\mathbf{Q}} L = \bigoplus_{j=1}^{J} L_j,$$

where the L_j correspond to the irreducible factors of the polynomial $f(x) \in \mathbf{Q}_p[x]$ (hence the extensions of $|\cdot|_p$ correspond to irreducible factors of $f(x)$ over $\mathbf{Q}_p[x]$).

In preparation for this clean view of the local nature of number fields, we will prove that the norms on a finite-dimensional vector space over a complete field are all equivalent. We will also explicitly construct tensor products of fields and deduce some of their properties.

18.1 Normed Spaces

Definition 18.1.1 (Norm). Let K be a field with valuation $|\cdot|$ and let V be a vector space over K. A real-valued function $\|\cdot\|$ on V is called a *norm* if

1. $\|v\| > 0$ for all nonzero $v \in V$ (positivity).

2. $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$ (triangle inequality).

3. $\|av\| = |a|\,\|v\|$ for all $a \in K$ and $v \in V$ (homogeneity).

Note that setting $\|v\| = 1$ for all $v \neq 0$ does *not* define a norm unless the absolute value on K is trivial, as $1 = \|av\| = |a|\,\|v\| = |a|$. We assume for the rest of this section that $|\cdot|$ is not trivial.

Definition 18.1.2 (Equivalent). Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the same vector space V are *equivalent* if there exists positive real numbers c_1 and c_2 such that for all $v \in V$

$$\|v\|_1 \leq c_1 \|v\|_2 \qquad \text{and} \qquad \|v\|_2 \leq c_2 \|v\|_1 .$$

Lemma 18.1.3. *Suppose that K is a field that is complete with respect to a valuation $|\cdot|$ and that V is a finite dimensional K vector space. Continue to assume, as mentioned above, that K is complete with respect to $|\cdot|$. Then any two norms on V are equivalent.*

Remark 18.1.4. As we shall see soon (see Theorem 19.1.8), the lemma is usually false if we do not assume that K is complete. For example, when $K = \mathbf{Q}$ and $|\cdot|_p$ is the p-adic valuation, and V is a number field, then there may be several extensions of $|\cdot|_p$ to inequivalent norms on V.

If two norms are equivalent then the corresponding topologies on V are equal, since very open ball for $\|\cdot\|_1$ is contained in an open ball for $\|\cdot\|_2$, and conversely. (The converse is also true, since, as we will show, all norms on V are equivalent.)

Proof. Let $v_1, \ldots, v_N$ be a basis for V. Define the max norm $\|\cdot\|_0$ by

$$\left\| \sum_{n=1}^{N} a_n v_n \right\|_0 = \max \{|a_n| : n = 1, \ldots, N\} .$$

It is enough to show that any norm $\|\cdot\|$ is equivalent to $\|\cdot\|_0$. We have

$$\begin{aligned} \left\| \sum_{n=1}^{N} a_n v_n \right\| &\leq \sum_{n=1}^{N} |a_n|\,\|v_n\| \\ &\leq \sum_{n=1}^{N} \max |a_n|\,\|v_n\| \\ &= c_1 \cdot \left\| \sum_{n=1}^{N} a_n v_n \right\|_0 , \end{aligned}$$

where $c_1 = \sum_{n=1}^{N} \|v_n\|$.

To finish the proof, we show that there is a $c_2 \in \mathbf{R}$ such that for all $v \in V$,

$$\|v\|_0 \leq c_2 \cdot \|v\| .$$

We will only prove this in the case when K is not just merely complete with respect to $|\cdot|$ but also locally compact. This will be the case of primary interest to us. For a proof in the general case, see the original article by Cassels (page 53).

By what we have already shown, the function $\|v\|$ is continuous in the $\|\cdot\|_0$-topology, so by local compactness it attains its lower bound δ on the unit circle $\{v \in V : \|v\|_0 = 1\}$. (Why is the unit circle compact? With respect to $\|\cdot\|_0$, the topology on V is the same as that of a product of copies of K. If the valuation is archimedean then $K \cong \mathbf{R}$ or $\mathbf{C}$ with the standard topology and the unit circle is compact. If the valuation is non-archimedean, then we saw (see Remark 17.1.7) that if K is locally compact, then the valuation is discrete, in which case we showed that the unit disc is compact, hence the unit circle is also compact since it is closed.) Note that $\delta > 0$ by part 1 of Definition 18.1.1. Also, by definition of $\|\cdot\|_0$, for any $v \in V$ there exists $a \in K$ such that $\|v\|_0 = |a|$ (just take the max coefficient in our basis). Thus we can write any $v \in V$ as $a \cdot w$ where $a \in K$ and $w \in V$ with $\|w\|_0 = 1$. We then have

$$\frac{\|v\|_0}{\|v\|} = \frac{\|aw\|_0}{\|aw\|} = \frac{|a|\,\|w\|_0}{|a|\,\|w\|} = \frac{1}{\|w\|} \leq \frac{1}{\delta}.$$

Thus for all v we have

$$\|v\|_0 \leq c_2 \cdot \|v\|,$$

where $c_2 = 1/\delta$, which proves the theorem. □

18.2 Tensor Products

We need only a special case of the tensor product construction. Let A and B be commutative rings containing a field K and suppose that B is of finite dimension N over K, say, with basis

$$1 = w_1, w_2, \ldots, w_N.$$

Then B is determined up to isomorphism as a ring over K by the multiplication table $(c_{i,j,n})$ defined by

$$w_i \cdot w_j = \sum_{n=1}^{N} c_{i,j,n} \cdot w_n.$$

We define a new ring C containing K whose elements are the set of all expressions

$$\sum_{n=1}^{N} a_n \underline{w}_n$$

where the $\underline{w}_n$ have the same multiplication rule

$$\underline{w}_i \cdot \underline{w}_j = \sum_{n=1}^{N} c_{i,j,n} \cdot \underline{w}_n$$

as the w_n.

There are injective ring homomorphisms

$$i : A \hookrightarrow C, \qquad i(a) = a\underline{w}_1 \qquad \text{(note that } \underline{w}_1 = 1\text{)}$$

and

$$j : B \hookrightarrow C, \qquad j\left(\sum_{n=1}^{N} c_n w_n\right) = \sum_{n=1}^{N} c_n \underline{w}_n.$$

Moreover C is defined, up to isomorphism, by A and B and is independent of the particular choice of basis w_n of B (i.e., a change of basis of B induces a canonical isomorphism of the C defined by the first basis to the C defined by the second basis). We write

$$C = A \otimes_K B$$

since C is, in fact, a special case of the ring tensor product.

Let us now suppose, further, that A is a topological ring, i.e., has a topology with respect to which addition and multiplication are continuous. Then the map

$$C \to A \oplus \cdots \oplus A, \qquad \sum_{m=1}^{N} a_m \underline{w}_m \mapsto (a_1, \ldots, a_N)$$

defines a bijection between C and the product of N copies of A (considered as sets). We give C the product topology. It is readily verified that this topology is independent of the choice of basis $w_1, \ldots, w_N$ and that multiplication and addition on C are continuous, so C is a topological ring. We call this topology on C the *tensor product topology.*

Now drop our assumption that A and B have a topology, but suppose that A and B are not merely rings but fields. Recall that a finite extension L/K of fields is *separable* if the number of embeddings $L \hookrightarrow \overline{K}$ that fix K equals the degree of L over K, where $\overline{K}$ is an algebraic closure of K. The primitive element theorem from Galois theory asserts that any such extension is generated by a single element, i.e., $L = K(a)$ for some $a \in L$.

Lemma 18.2.1. *Let A and B be fields containing the field K and suppose that B is a separable extension of finite degree $N = [B : K]$. Then $C = A \otimes_K B$ is the direct sum of a finite number of fields K_j, each containing an isomorphic image of A and an isomorphic image of B.*

Proof. By the primitive element theorem, we have $B = K(b)$, where b is a root of some separable irreducible polynomial $f(x) \in K[x]$ of degree N. Then $1, b, \ldots, b^{N-1}$ is a basis for B over K, so

$$A \otimes_K B = A[\underline{b}] \cong A[x]/(f(x))$$

where $1, \underline{b}, \underline{b}^2, \ldots, \underline{b}^{N-1}$ are linearly independent over A and $\underline{b}$ satisfies $f(\underline{b}) = 0$.

Although the polynomial $f(x)$ is irreducible as an element of $K[x]$, it need not be irreducible in $A[x]$. Since A is a field, we have a factorization

$$f(x) = \prod_{j=1}^{J} g_j(x)$$

where $g_j(x) \in A[x]$ is irreducible. The $g_j(x)$ are distinct because $f(x)$ is separable (i.e., has distinct roots in any algebraic closure).

For each j, let $\underline{b}_j \in \overline{A}$ be a root of $g_j(x)$, where $\overline{A}$ is a fixed algebraic closure of the field A. Let $K_j = A(\underline{b}_j)$. Then the map

$$\varphi_j : A \otimes_K B \to K_j \tag{18.2.1}$$

given by sending any polynomial $h(\underline{b})$ in $\underline{b}$ (where $h \in A[x]$) to $h(\underline{b}_j)$ is a ring homomorphism, because the image of $\underline{b}$ satisfies the polynomial $f(x)$, and $A \otimes_K B \cong A[x]/(f(x))$.

By the Chinese Remainder Theorem, the maps from (18.2.1) combine to define a ring isomorphism

$$A \otimes_K B \cong A[x]/(f(x)) \cong \bigoplus_{j=1}^{J} A[x]/(g_j(x)) \cong \bigoplus_{j=1}^{J} K_j.$$

Each K_j is of the form $A[x]/(g_j(x))$, so contains an isomorphic image of A. It thus remains to show that the ring homomorphisms

$$\lambda_j : B \xrightarrow{b \mapsto 1 \otimes b} A \otimes_K B \xrightarrow{\varphi_j} K_j$$

are injections. Since B and K_j are both fields, λ_j is either the 0 map or injective. However, λ_j is not the 0 map since $\lambda_j(1) = 1 \in K_j$. □

Example 18.2.2. If A and B are finite extensions of $\mathbf{Q}$, then $A \otimes_{\mathbf{Q}} B$ is an algebra of degree $[A : \mathbf{Q}] \cdot [B : \mathbf{Q}]$. For example, suppose A is generated by a root of $x^2 + 1$ and B is generated by a root of $x^3 - 2$. We can view $A \otimes_{\mathbf{Q}} B$ as either $A[x]/(x^3 - 2)$ or $B[x]/(x^2 + 1)$. The polynomial $x^2 + 1$ is irreducible over $\mathbf{Q}$, and if it factored over the cubic field B, then there would be a root of $x^2 + 1$ in B, i.e., the quadratic field $A = \mathbf{Q}(i)$ would be a subfield of the cubic field $B = \mathbf{Q}(\sqrt[3]{2})$, which is impossible. Thus $x^2 + 1$ is irreducible over B, so $A \otimes_{\mathbf{Q}} B = A.B = \mathbf{Q}(i, \sqrt[3]{2})$ is a degree 6 extension of $\mathbf{Q}$. Notice that $A.B$ contains a copy A and a copy of B. By the primitive element theorem the composite field $A.B$ can be generated by the root of a single polynomial. For example, the minimal polynomial of $i + \sqrt[3]{2}$ is $x^6 + 3x^4 - 4x^3 + 3x^2 + 12x + 5$, hence $\mathbf{Q}(i + \sqrt[3]{2}) = A.B$.

Example 18.2.3. The case $A \cong B$ is even more exciting. For example, suppose $A = B = \mathbf{Q}(i)$. Using the Chinese Remainder Theorem we have that

$$\mathbf{Q}(i) \otimes_{\mathbf{Q}} \mathbf{Q}(i) \cong \mathbf{Q}(i)[x]/(x^2 + 1) \cong \mathbf{Q}(i)[x]/((x - i)(x + i)) \cong \mathbf{Q}(i) \oplus \mathbf{Q}(i),$$

since $(x-i)$ and $(x+i)$ are coprime. The last isomorphism sends $a+bx$, with $a,b\in\mathbf{Q}(i)$, to $(a+bi,a-bi)$. Since $\mathbf{Q}(i)\oplus\mathbf{Q}(i)$ has zero divisors, the tensor product $\mathbf{Q}(i)\otimes_{\mathbf{Q}}\mathbf{Q}(i)$ must also have zero divisors. For example, $(1,0)$ and $(0,1)$ is a zero divisor pair on the right hand side, and we can trace back to the elements of the tensor product that they define. First, by solving the system

$$a+bi=1 \qquad \text{and} \qquad a-bi=0$$

we see that $(1,0)$ corresponds to $a=1/2$ and $b=-i/2$, i.e., to the element

$$\frac{1}{2}-\frac{i}{2}x\in\mathbf{Q}(i)[x]/(x^2+1).$$

This element in turn corresponds to

$$\frac{1}{2}\otimes 1-\frac{i}{2}\otimes i\in\mathbf{Q}(i)\otimes_{\mathbf{Q}}\mathbf{Q}(i).$$

Similarly the other element $(0,1)$ corresponds to

$$\frac{1}{2}\otimes 1+\frac{i}{2}\otimes i\in\mathbf{Q}(i)\otimes_{\mathbf{Q}}\mathbf{Q}(i).$$

As a double check, observe that

$$\begin{aligned}\left(\frac{1}{2}\otimes 1-\frac{i}{2}\otimes i\right)\cdot\left(\frac{1}{2}\otimes 1+\frac{i}{2}\otimes i\right)&=\frac{1}{4}\otimes 1+\frac{i}{4}\otimes i-\frac{i}{4}\otimes i-\frac{i^2}{4}\otimes i^2\\&=\frac{1}{4}\otimes 1-\frac{1}{4}\otimes 1=0\in\mathbf{Q}(i)\otimes_{\mathbf{Q}}\mathbf{Q}(i).\end{aligned}$$

Clearing the denominator of 2 and writing $1\otimes 1=1$, we have $(1-i\otimes i)(1+i\otimes i)=0$, so $i\otimes i$ is a root of the polynomimal x^2-1, and $i\otimes i$ is not ± 1, so x^2-1 has more than 2 roots.

In general, to understand $A\otimes_K B$ explicitly is the same as factoring either the defining polynomial of B over the field A, or factoring the defining polynomial of A over B.

Corollary 18.2.4. *Let $a\in B$ be any element and let $f(x)\in K[x]$ be the characteristic polynomials of a over K and let $g_j(x)\in A[x]$ (for $1\le j\le J$) be the characteristic polynomials of the images of a under $B\to A\otimes_K B\to K_j$ over A, respectively. Then*

$$f(x)=\prod_{j=1}^{J}g_j(X). \tag{18.2.2}$$

Proof. We show that both sides of (18.2.2) are the characteristic polynomial $T(x)$ of the image of a in $A\otimes_K B$ over A. That $f(x)=T(x)$ follows at once by computing the characteristic polynomial in terms of a basis $\underline{w}_1,\ldots,\underline{w}_N$ of $A\otimes_K B$, where $w_1,\ldots,w_N$ is a basis for B over K (this is because the matrix of left multiplication by b on $A \otimes_K B$ is exactly the same as the matrix of left multiplication on B, so the characteristic polynomial doesn't change). To see that $T(X) = \prod g_j(X)$, compute the action of the image of a in $A \otimes_K B$ with respect to a basis of

$$A \otimes_K B \cong \bigoplus_{j=1}^{J} K_j \tag{18.2.3}$$

composed of basis of the individual extensions K_j of A. The resulting matrix will be a block direct sum of submatrices, each of whose characteristic polynomials is one of the $g_j(X)$. Taking the product gives the claimed identity (18.2.2). □

Corollary 18.2.5. *For $a \in B$ we have*

$$\mathrm{Norm}_{B/K}(a) = \prod_{j=1}^{J} \mathrm{Norm}_{K_j/A}(a),$$

and

$$\mathrm{Tr}_{B/K}(a) = \sum_{j=1}^{J} \mathrm{Tr}_{K_j/A}(a),$$

Proof. This follows from Corollary 18.2.4. First, the norm is ± the constant term of the characteristic polynomial, and the constant term of the product of polynomials is the product of the constant terms (and one sees that the sign matches up correctly). Second, the trace is minus the second coefficient of the characteristic polynomial, and second coefficients add when one multiplies polynomials:

$$(x^n+a_{n-1}x^{n-1}+\cdots)\cdot(x^m+a_{m-1}x^{m-1}+\cdots) = x^{n+m}+x^{n+m-1}(a_{m-1}+a_{n-1})+\cdots .$$

One could also see both the statements by considering a matrix of left multiplication by a first with respect to the basis of $\underline{w}_n$ and second with respect to the basis coming from the left side of (18.2.3).

□

Chapter 19

Extensions and Normalizations of Valuations

19.1 Extensions of Valuations

In this section we continue to tacitly assume that all valuations are nontrivial. We do not assume all our valuations satisfy the triangle

Suppose $K \subset L$ is a finite extension of fields, and that $|\cdot|$ and $\|\cdot\|$ are valuations on K and L, respectively.

Definition 19.1.1 (Extends). We say that $\|\cdot\|$ *extends* $|\cdot|$ if $|a| = \|a\|$ for all $a \in K$.

Theorem 19.1.2. *Suppose that K is a field that is complete with respect to $|\cdot|$ and that L is a finite extension of K of degree $N = [L : K]$. Then there is precisely one extension of $|\cdot|$ to K, namely*

$$\|a\| = \left|\mathrm{Norm}_{L/K}(a)\right|^{1/N}, \tag{19.1.1}$$

where the Nth root is the non-negative real Nth root of the nonnegative real number $\left|\mathrm{Norm}_{L/K}(a)\right|$.

Proof. We may assume that $|\cdot|$ is normalized so as to satisfy the triangle inequality. Otherwise, normalize $|\cdot|$ so that it does, prove the theorem for the normalized valuation $|\cdot|^c$, then raise both sides of (19.1.1) to the power $1/c$. In the uniqueness proof, by the same argument we may assume that $\|\cdot\|$ also satisfies the triangle inequality.

Uniqueness. View L as a finite-dimensional vector space over K. Then $\|\cdot\|$ is a norm in the sense defined earlier (Definition 18.1.1). Hence any two extensions $\|\cdot\|_1$ and $\|\cdot\|_2$ of $|\cdot|$ are equivalent as norms, so induce the same topology on K. But as we have seen (Proposition 16.1.4), two valuations which induce the same topology are equivalent valuations, i.e., $\|\cdot\|_1 = \|\cdot\|_2^c$, for some positive real c. Finally $c = 1$ since $\|a\|_1 = |a| = \|a\|_2$ for all $a \in K$.

Existence. We do not give a proof of existence in the general case. Instead we give a proof, which was suggested by Dr. Geyer at the conference out of which [Cas67] arose. It is valid when K is locally compact, which is the only case we will use later.

We see at once that the function defined in (19.1.1) satisfies the condition (i) that $\|a\| \geq 0$ with equality only for $a = 0$, and (ii) $\|ab\| = \|a\| \cdot \|b\|$ for all $a, b \in L$. The difficult part of the proof is to show that there is a constant $C > 0$ such that

$$\|a\| \leq 1 \implies \|1 + a\| \leq C.$$

Note that we do not know (and will not show) that $\|\cdot\|$ as defined by (19.1.1) is a norm as in Definition 18.1.1, since showing that $\|\cdot\|$ is a norm would entail showing that it satisfies the triangle inequality, which is not obvious.

Choose a basis $b_1, \ldots, b_N$ for L over K. Let $\|\cdot\|_0$ be the max norm on L, so for $a = \sum_{i=1}^N c_i b_i$ with $c_i \in K$ we have

$$\|a\|_0 = \left\|\sum_{i=1}^N c_i b_i\right\|_0 = \max\{|c_i| : i = 1, \ldots, N\}.$$

(Note: in Cassels's original article he let $\|\cdot\|_0$ be *any* norm, but we don't because the rest of the proof does not work, since we can't use homogeneity as he claims to do. This is because it need not be possible to find, for any nonzero $a \in L$ some element $c \in K$ such that $\|ac\|_0 = 1$. This would fail, e.g., if $\|a\|_0 \neq |c|$ for any $c \in K$.) The rest of the argument is very similar to our proof from Lemma 18.1.3 of uniqueness of norms on vector spaces over complete fields.

With respect to the $\|\cdot\|_0$-topology, L has the product topology as a product of copies of K. The function $a \mapsto \|a\|$ is a composition of continuous functions on L with respect to this topology (e.g., $\mathrm{Norm}_{L/K}$ is the determinant, hence polynomial), hence $\|\cdot\|$ defines nonzero continuous function on the compact set

$$S = \{a \in L : \|a\|_0 = 1\}.$$

By compactness, there are real numbers $\delta, \Delta \in \mathbf{R}_{>0}$ such that

$$0 < \delta \leq \|a\| \leq \Delta \qquad \text{for all } a \in S.$$

For any nonzero $a \in L$ there exists $c \in K$ such that $\|a\|_0 = |c|$; to see this take c to be a c_i in the expression $a = \sum_{i=1}^N c_i b_i$ with $|c_i| \geq |c_j|$ for any j. Hence $\|a/c\|_0 = 1$, so $a/c \in S$ and

$$0 \leq \delta \leq \frac{\|a/c\|}{\|a/c\|_0} \leq \Delta.$$

Then by homogeneity

$$0 \leq \delta \leq \frac{\|a\|}{\|a\|_0} \leq \Delta.$$

Suppose now that $\|a\| \leq 1$. Then $\|a\|_0 \leq \delta^{-1}$, so

$$\begin{aligned} \|1+a\| &\leq \Delta \cdot \|1+a\|_0 \\ &\leq \Delta \cdot (\|1\|_0 + \|a\|_0) \\ &\leq \Delta \cdot \left(\|1\|_0 + \delta^{-1}\right) \\ &= C \quad \text{(say)}, \end{aligned}$$

as required. □

Example 19.1.3. Consider the extension $\mathbf{C}$ of $\mathbf{R}$ equipped with the archimedean valuation. The unique extension is the ordinary absolute value on $\mathbf{C}$:

$$\|x+iy\| = \left(x^2+y^2\right)^{1/2}.$$

Example 19.1.4. Consider the extension $\mathbf{Q}_2(\sqrt{2})$ of $\mathbf{Q}_2$ equipped with the 2-adic absolute value. Since x^2-2 is irreducible over $\mathbf{Q}_2$ we can do some computations by working in the subfield $\mathbf{Q}(\sqrt{2})$ of $\mathbf{Q}_2(\sqrt{2})$.

```
> K<a> := NumberField(x^2-2);
> K;
Number Field with defining polynomial x^2 - 2 over the Rational Field
> function norm(x) return Sqrt(2^(-Valuation(Norm(x),2))); end function;
> norm(1+a);
1.000000000000000000000000000
> norm(1+a+1);
0.7071067811865475244008443621 
> z := 3+2*a;
> norm(z);
1.000000000000000000000000000
> norm(z+1);
0.3535533905932737622004221810 
```

Remark 19.1.5. Geyer's existence proof gives (19.1.1). But it is perhaps worth noting that in any case (19.1.1) is a consequence of unique existence, as follows. Suppose L/K is as above. Suppose M is a finite Galois extension of K that contains L. Then by assumption there is a unique extension of $|\cdot|$ to M, which we shall also denote by $\|\cdot\|$. If $\sigma \in \mathrm{Gal}(M/K)$, then

$$\|a\|_\sigma := \|\sigma(a)\|$$

is also an extension of $|\cdot|$ to M, so $\|\cdot\|_\sigma = \|\cdot\|$, i.e.,

$$\|\sigma(a)\| = \|a\| \qquad \text{for all } a \in M.$$

But now

$$\mathrm{Norm}_{L/K}(a) = \sigma_1(a) \cdot \sigma_2(a) \cdots \sigma_N(a)$$

for $a \in K$, where $\sigma_1, \ldots, \sigma_N \in \mathrm{Gal}(M/K)$ extend the embeddings of L into M. Hence

$$\begin{aligned} \left|\mathrm{Norm}_{L/K}(a)\right| &= \left\|\mathrm{Norm}_{L/K}(a)\right\| \\ &= \prod_{1\leq n\leq N} \|\sigma_n(a)\| \\ &= \|a\|^N, \end{aligned}$$

as required.

Corollary 19.1.6. *Let $w_1, \ldots, w_N$ be a basis for L over K. Then there are positive constants c_1 and c_2 such that*

$$c_1 \leq \frac{\left\|\sum_{n=1}^{N} b_n w_n\right\|}{\max\{|b_n| : n = 1, \ldots, N\}} \leq c_2$$

for any $b_1, \ldots, b_N \in K$ not all 0.

Proof. For $\left|\sum_{n=1}^{N} b_n w_n\right|$ and $\max |b_n|$ are two norms on L considered as a vector space over K.

I don't believe this proof, which I copied from Cassels's article. My problem with it is that the proof of Theorem 19.1.2 does not give that $C \leq 2$, i.e., that the triangle inequality holds for $\|\cdot\|$. By changing the basis for L/K one can make any nonzero vector $a \in L$ have $\|a\|_0 = 1$, so if we choose a such that $|a|$ is very large, then the Δ in the proof will also be very large. One way to fix the corollary is to only claim that there are positive constants c_1, c_2, c_3, c_4 such that

$$c_1 \leq \frac{\left\|\sum_{n=1}^{N} b_n w_n\right\|^{c_3}}{\max\{|b_n|^{c_4} : n = 1, \ldots, N\}} \leq c_2.$$

Then choose c_3, c_4 such that $\|\cdot\|^{c_3}$ and $|\cdot|^{c_4}$ satisfies the triangle inequality, and prove the modified corollary using the proof suggested by Cassels. □

Corollary 19.1.7. *A finite extension of a completely valued field K is complete with respect to the extended valuation.*

Proof. By the proceeding corollary it has the topology of a finite-dimensional vector space over K. (The problem with the proof of the previous corollary is not an issue, because we can replace the extended valuation by an inequivalent one that satisfies the triangle inequality and induces the same topology.) □

When K is no longer complete under $|\cdot|$ the position is more complicated:

Theorem 19.1.8. *Let L be a separable extension of K of finite degree $N = [L : K]$. Then there are at most N extensions of a valuation $|\cdot|$ on K to L, say $\|\cdot\|_j$, for $1 \leq j \leq J$. Let K_v be the completion of K with respect to $|\cdot|$, and for each j let L_j be the completion of L with respect to $\|\cdot\|_j$. Then*

$$K_v \otimes_K L \cong \bigoplus_{1\leq j\leq J} L_j \tag{19.1.2}$$

algebraically and topologically, where the right hand side is given the product topology.

Proof. We already know (Lemma 18.2.1) that $K_v \otimes_K L$ is of the shape (19.1.2), where the L_j are finite extensions of K_v. Hence there is a unique extension $|\cdot|_j^*$ of $|\cdot|$ to the L_j, and by Corollary 19.1.7 the L_j are complete with respect to the extended valuation. Further, the ring homomorphisms

$$\lambda_j : L \to K_v \otimes_K L \to L_j$$

are injections. Hence we get an extension $\|\cdot\|_j$ of $|\cdot|$ to L by putting

$$\|b\|_j = |\lambda_j(b)|_j^* .$$

Further, $L \cong \lambda_j(L)$ is dense in L_j with respect to $\|\cdot\|_j$ because $L = K \otimes_K L$ is dense in $K_v \otimes_K L$ (since K is dense in K_v). Hence L_j is exactly the completion of L.

It remains to show that the $\|\cdot\|_j$ are distinct and that they are the only extensions of $|\cdot|$ to L.

Suppose $\|\cdot\|$ is any valuation of L that extends $|\cdot|$. Then $\|\cdot\|$ extends by continuity to a real-valued function on $K_v \otimes_K L$, which we also denote by $\|\cdot\|$. (We are again using that L is dense in $K_v \otimes_K L$.) By continuity we have for all $a, b \in K_v \otimes_K L$,

$$\|ab\| = \|a\| \cdot \|b\|$$

and if C is the constant in axiom (iii) for L and $\|\cdot\|$, then

$$\|a\| \leq 1 \implies \|1 + a\| \leq C.$$

(In Cassels, he inexplicable assume that $C = 1$ at this point in the proof.)

We consider the restriction of $\|\cdot\|$ to one of the L_j. If $\|a\| \neq 0$ for some $a \in L_j$, then $\|a\| = \|b\| \cdot \left\|ab^{-1}\right\|$ for every $b \neq 0$ in L_j so $\|b\| \neq 0$. Hence either $\|\cdot\|$ is identically 0 on L_j or it induces a valuation on L_j.

Further, $\|\cdot\|$ cannot induce a valuation on two of the L_j. For

$$(a_1, 0, \ldots, 0) \cdot (0, a_2, 0, \ldots, 0) = (0, 0, 0, \ldots, 0),$$

so for any $a_1 \in L_1$, $a_2 \in L_2$,

$$\|a_1\| \cdot \|a_2\| = 0.$$

Hence $\|\cdot\|$ induces a valuation in precisely one of the L_j, and it extends the given valuation $|\cdot|$ of K_v. Hence $\|\cdot\| = \|\cdot\|_j$ for precisely one j.

It remains only to show that (19.1.2) is a topological homomorphism. For

$$(b_1, \ldots, b_J) \in L_1 \oplus \cdots \oplus L_J$$

put

$$\|(b_1, \ldots, b_J)\|_0 = \max_{1 \le j \le J} \|b_j\|_j \,.$$

Then $\|\cdot\|_0$ is a norm on the right hand side of (19.1.2), considered as a vector space over K_v and it induces the product topology. On the other hand, any two norms are equivalent, since K_v is complete, so $\|\cdot\|_0$ induces the tensor product topology on the left hand side of (19.1.2). □

Corollary 19.1.9. *Suppose $L = K(a)$, and let $f(x) \in K[x]$ be the minimal polynomial of a. Suppose that*

$$f(x) = \prod_{1 \le j \le J} g_j(x)$$

in $K_v[x]$, where the g_j are irreducible. Then $L_j = K_v(b_j)$, where b_j is a root of g_j.

19.2 Extensions of Normalized Valuations

Let K be a complete field with valuation $|\cdot|$. We consider the following three cases:

(1) $|\cdot|$ is discrete non-archimedean and the residue class field is finite.

(2i) The completion of K with respect to $|\cdot|$ is **R**.

(2ii) The completion of K with respect to $|\cdot|$ is **C**.

(Alternatively, these cases can be subsumed by the hypothesis that the completion of K is locally compact.)

In case (1) we defined the normalized valuation to be the one such that if Haar measure of the ring of integers $\mathcal{O}$ is 1, then $\mu(a\mathcal{O}) = |a|$ (see Definition 17.1.11). In case (2i) we say that $|\cdot|$ is normalized if it is the ordinary absolute value, and in (2ii) if it is the *square* of the ordinary absolute value:

$$|x + iy| = x^2 + y^2 \qquad \text{(normalized)}.$$

In every case, for every $a \in K$, the map

$$a : x \mapsto ax$$

on K^+ multiplies any choice of Haar measure by $|a|$, and this characterizes the normalized valuations among equivalent ones.

We have already verified the above characterization for non-archimedean valuations, and it is clear for the ordinary absolute value on **R**, so it remains to verify it for $\mathbf{C}$. The additive group $\mathbf{C}^+$ is topologically isomorphic to $\mathbf{R}^+ \oplus \mathbf{R}^+$, so a choice of Haar measure of $\mathbf{C}^+$ is the usual area measure on the Euclidean plane. Multiplication by $x+iy \in \mathbf{C}$ is the same as rotation followed by scaling by a factor of $\sqrt{x^2+y^2}$, so if we rescale a region by a factor of $x+iy$, the area of the region changes by a factor of the square of $\sqrt{x^2+y^2}$. This explains why the normalized valuation on $\mathbf{C}$ is the square of the usual absolute value. Note that the normalized valuation on $\mathbf{C}$ does not satisfy the triangle inequality:

$$|1+(1+i)| = |2+i| = 2^2+1^2 = 5 \not\leq 3 = 1^2 + (1^2+1^2) = |1| + |1+i|\,.$$

The constant C in axiom (3) of a valuation for the ordinary absolute value on $\mathbf{C}$ is 2, so the constant for the normalized valuation $|\cdot|$ is $C \leq 4$:

$$|x+iy| \leq 1 \implies |x+iy+1| \leq 4.$$

Note that $x^2+y^2 \leq 1$ implies

$$(x+1)^2+y^2 = x^2+2x+1+y^2 \leq 1+2x+1 \leq 4$$

since $x \leq 1$.

Lemma 19.2.1. *Suppose K is a field that is complete with respect to a normalized valuation $|\cdot|$ and let L be a finite extension of K of degree $N = [L : K]$. Then the normalized valuation $\|\cdot\|$ on L which is equivalent to the unique extension of $|\cdot|$ to L is given by the formula*

$$\|a\| = \left|\mathrm{Norm}_{L/K}(a)\right| \qquad \textit{all } a \in L. \tag{19.2.1}$$

Proof. Let $\|\cdot\|$ be the normalized valuation on L that extends $|\cdot|$. Our goal is to identify $\|\cdot\|$, and in particular to show that it is given by (19.2.1).

By the preceding section there is a positive real number c such that for all $a \in L$ we have

$$\|a\| = \left|\mathrm{Norm}_{L/K}(a)\right|^c.$$

Thus all we have to do is prove that $c = 1$. In case 2 the only nontrivial situation is $L = \mathbf{C}$ and $K = \mathbf{R}$, in which case $\left|\mathrm{Norm}_{\mathbf{C}/\mathbf{R}}(x+iy)\right| = \left|x^2+y^2\right|$, which is the normalized valuation on $\mathbf{C}$ defined above.

One can argue in a unified way in all cases as follows. Let $w_1, \ldots, w_N$ be a basis for L/K. Then the map

$$\varphi : L^+ \to \bigoplus_{n=1}^{N} K^+, \qquad \sum a_n w_n \mapsto (a_1, \ldots, a_N)$$

is an isomorphism between the additive group L^+ and the direct sum $\oplus_{n=1}^N K^+$, and this is a homeomorphism if the right hand side is given the product topology. In particular, the Haar measures on L^+ and on $\oplus_{n=1}^N K^+$ are the same up to a multiplicative constant in $\mathbf{Q}^*$.

Let $b \in K$. Then the left-multiplication-by-b map

$$b : \sum a_n w_n \mapsto \sum b a_n w_n$$

on L^+ is the same as the map

$$(a_1, \ldots, a_N) \mapsto (ba_1, \ldots, ba_N)$$

on $\oplus_{n=1}^{N} K^+$, so it multiplies the Haar measure by $|b|^N$, since $|\cdot|$ on K is assumed normalized (the measure of each factor is multiplied by $|b|$, so the measure on the product is multiplied by $|b|^N$). Since $\|\cdot\|$ is assumed normalized, so multiplication by b rescales by $\|b\|$, we have

$$\|b\| = |b|^N .$$

But $b \in K$, so $\mathrm{Norm}_{L/K}(b) = b^N$. Since $|\cdot|$ is nontrivial and for $a \in K$ we have

$$\|a\| = |a|^N = \left|a^N\right| = \left|\mathrm{Norm}_{L/K}(a)\right| ,$$

so we must have $c = 1$ in (19.2.1), as claimed. □

In the case when K need not be complete with respect to the valuation $|\cdot|$ on K, we have the following theorem.

Theorem 19.2.2. *Suppose $|\cdot|$ is a (nontrivial as always) normalized valuation of a field K and let L be a finite extension of K. Then for any $a \in L$,*

$$\prod_{1 \leq j \leq J} \|a\|_j = \left|\mathrm{Norm}_{L/K}(a)\right|$$

where the $\|\cdot\|_j$ are the normalized valuations equivalent to the extensions of $|\cdot|$ to K.

Proof. Let K_v denote the completion of K with respect to $|\cdot|$. Write

$$K_v \otimes_K L = \bigoplus_{1 \leq j \leq J} L_j .$$

Then Theorem 19.2.2 asserts that

$$\mathrm{Norm}_{L/K}(a) = \prod_{1 \leq j \leq J} \mathrm{Norm}_{L_j/K_v}(a). \qquad (19.2.2)$$

By Theorem 19.1.8, the $\|\cdot\|_j$ are exactly the normalizations of the extensions of $|\cdot|$ to the L_j (i.e., the L_j are in bijection with the extensions of valuations, so there are no other valuations missed). By Lemma 19.1.1, the normalized valuation $\|\cdot\|_j$ on L_j is $|a| = \left|\mathrm{Norm}_{L_J/K_v}(a)\right|$. The theorem now follows by taking absolute values of both sides of (19.2.2). □

What next?! We'll building up to giving a new proof of finiteness of the class group that uses that the class group naturally has the discrete topology and is the continuous image of a compact group.

Chapter 20

Global Fields and Adeles

20.1 Global Fields

Definition 20.1.1 (Global Field). A *global field* is a number field or a finite separable extension of $\mathbf{F}(t)$, where $\mathbf{F}$ is a finite field, and t is transcendental over $\mathbf{F}$.

Below we will focus attention on number fields leaving the function field case to the reader.

The following lemma essentially says that the denominator of an element of a global field is only "nontrivial" at a finite number of valuations.

Lemma 20.1.2. *Let $a \in K$ be a nonzero element of a global field K. Then there are only finitely many inequivalent valuations $|\cdot|$ of K for which*

$$|a| > 1.$$

Proof. If $K = \mathbf{Q}$ or $\mathbf{F}(t)$ then the lemma follows by Ostrowski's classification of all the valuations on K (see Theorem 15.3.2). For example, when $a = \frac{n}{d} \in \mathbf{Q}$, with $n, d \in \mathbf{Z}$, then the valuations where we could have $|a| > 1$ are the archimedean one, or the p-adic valuations $|\cdot|_p$ for which $p \mid d$.

Suppose now that K is a finite extension of $\mathbf{Q}$, so a satisfies a monic polynomial

$$a^n + c_{n-1}a^{n-1} + \cdots + c_0 = 0,$$

for some n and $c_0, \ldots, c_{n-1} \in \mathbf{Q}$. If $|\cdot|$ is a non-archimedean valuation on K, we have

$$\begin{aligned} |a|^n &= \left|-(c_{n-1}a^{n-1} + \cdots + c_0)\right| \\ &\leq \max(1, |a|^{n-1}) \cdot \max(|c_0|, \ldots, |c_{n-1}|). \end{aligned}$$

Dividing each side by $|a|^{n-1}$, we have that

$$|a| \leq \max(|c_0|, \ldots, |c_{n-1}|),$$

so in all cases we have

$$|a| \le \max(1, |c_0|, \ldots, |c_{n-1}|)^{1/(n-1)}. \tag{20.1.1}$$

We know the lemma for $\mathbf{Q}$, so there are only finitely many valuations $|\cdot|$ on $\mathbf{Q}$ such that the right hand side of (20.1.1) is bigger than 1. Since each valuation of $\mathbf{Q}$ has finitely many extensions to K, and there are only finitely many archimedean valuations, it follows that there are only finitely many valuations on K such that $|a| > 1$. □

Any valuation on a global field is either archimedean, or discrete non-archimedean with finite residue class field, since this is true of $\mathbf{Q}$ and $\mathbf{F}(t)$ and is a property preserved by extending a valuation to a finite extension of the base field. Hence it makes sense to talk of normalized valuations. Recall that the normalized p-adic valuation on $\mathbf{Q}$ is $|x|_p = p^{-\operatorname{ord}_p(x)}$, and if v is a valuation on a number field K equivalent to an extension of $|\cdot|_p$, then the normalization of v is the composite of the sequence of maps

$$K \hookrightarrow K_v \xrightarrow{\text{Norm}} \mathbf{Q}_p \xrightarrow{|\cdot|_p} \mathbf{R},$$

where K_v is the completion of K at v.

Example 20.1.3. Let $K = \mathbf{Q}(\sqrt{2})$, and let $p = 2$. Because $\sqrt{2} \notin \mathbf{Q}_2$, there is exactly one extension of $|\cdot|_2$ to K, and it sends $a = 1/\sqrt{2}$ to

$$\left|\text{Norm}_{\mathbf{Q}_2(\sqrt{2})/\mathbf{Q}_2}(1/\sqrt{2})\right|_2^{1/2} = \sqrt{2}.$$

Thus the normalized valuation of a is 2.

There are two extensions of $|\cdot|_7$ to $\mathbf{Q}(\sqrt{2})$, since $\mathbf{Q}(\sqrt{2}) \otimes_{\mathbf{Q}} \mathbf{Q}_7 \cong \mathbf{Q}_7 \oplus \mathbf{Q}_7$, as $x^2 - 2 = (x-3)(x-4) \pmod 7$. The image of $\sqrt{2}$ under each embedding into $\mathbf{Q}_7$ is a unit in $\mathbf{Z}_7$, so the normalized valuation of $a = 1/\sqrt{2}$ is, in both cases, equal to 1. More generally, for any valuation of K of characteristic an odd prime p, the normalized valuation of a is 1.

Since $K = \mathbf{Q}(\sqrt{2}) \hookrightarrow \mathbf{R}$ in two ways, there are exactly two normalized archimedean valuations on K, and both of their values on a equal $1/\sqrt{2}$. Notice that the product of the absolute values of a with respect to all normalized valuations is

$$2 \cdot \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \cdot 1 \cdot 1 \cdot 1 \cdots = 1.$$

This "product formula" holds in much more generality, as we will now see.

Theorem 20.1.4 (Product Formula). *Let $a \in K$ be a nonzero element of a global field K. Let $|\cdot|_v$ run through the normalized valuations of K. Then $|a|_v = 1$ for almost all v, and*

$$\prod_{\text{all } v} |a|_v = 1 \qquad \text{(the product formula).}$$

We will later give a more conceptual proof of this using Haar measure (see Remark 20.3.9).

Proof. By Lemma 20.1.2, we have $|a|_v \leq 1$ for almost all v. Likewise, $1/|a|_v = |1/a|_v \leq 1$ for almost all v, so $|a|_v = 1$ for almost all v.

Let w run through all normalized valuations of $\mathbf{Q}$ (or of $\mathbf{F}(t)$), and write $v \mid w$ if the restriction of v to $\mathbf{Q}$ is equivalent to w. Then by Theorem 19.2.2,

$$\prod_v |a|_v = \prod_w \left(\prod_{v|w} |a|_v \right) = \prod_w \left|\mathrm{Norm}_{K/\mathbf{Q}}(a)\right|_w,$$

so it suffices to prove the theorem for $K = \mathbf{Q}$.

By multiplicativity of valuations, if the theorem is true for b and c then it is true for the product bc and quotient b/c (when $c \neq 0$). The theorem is clearly true for -1, which has valuation 1 at all valuations. Thus to prove the theorem for $\mathbf{Q}$ it suffices to prove it when $a = p$ is a prime number. Then we have $|p|_\infty = p$, $|p|_p = 1/p$, and for primes $q \neq p$ that $|p|_q = 1$. Thus

$$\prod_v |p|_v = p \cdot \frac{1}{p} \cdot 1 \cdot 1 \cdot 1 \cdots = 1,$$

as claimed. □

If v is a valuation on a field K, recall that we let K_v denote the completion of K with respect to v. Also when v is non-archimedean, let

$$\mathcal{O}_v = \mathcal{O}_{K,v} = \{x \in K_v : |x| \leq 1\}$$

be the ring of integers of the completion.

Definition 20.1.5 (Almost All). We say a condition holds for *almost all* elements of a set if it holds for all but finitely many elements.

We will use the following lemma later (see Lemma 20.3.3) to prove that formation of the adeles of a global field is compatible with base change.

Lemma 20.1.6. *Let $\omega_1, \ldots, \omega_n$ be a basis for L/K, where L is a finite separable extension of the global field K of degree n. Then for almost all normalized non-archimedean valuations v on K we have*

$$\omega_1 \mathcal{O}_v \oplus \cdots \oplus \omega_n \mathcal{O}_v = \mathcal{O}_{w_1} \oplus \cdots \oplus \mathcal{O}_{w_g} \subset K_v \otimes_K L, \tag{20.1.2}$$

where $w_1, \ldots, w_g$ are the extensions of v to L. Here we have identified $a \in L$ with its canonical image in $K_v \otimes_K L$, and the direct sum on the left is the sum taken inside the tensor product (so directness means that the intersections are trivial).

Proof. The proof proceeds in two steps. First we deduce easily from Lemma 20.1.2 that for almost all v the left hand side of (20.1.2) is contained in the right hand side. Then we use a trick involving discriminants to show the opposite inclusion for all but finitely many primes.

Since $\mathcal{O}_v \subset \mathcal{O}_{w_i}$ for all i, the left hand side of (20.1.2) is contained in the right hand side if $|\omega_i|_{w_j} \leq 1$ for $1 \leq i \leq n$ and $1 \leq j \leq g$. Thus by Lemma 20.1.2, for all but finitely many v the left hand side of (20.1.2) is contained in the right hand side. We have just eliminated the finitely many primes corresponding to "denominators" of some ω_i, and now only consider v such that $\omega_1, \ldots, \omega_n \in \mathcal{O}_w$ for all $w \mid v$.

For any elements $a_1, \ldots, a_n \in K_v \otimes_K L$, consider the discriminant

$$D(a_1, \ldots, a_n) = \mathrm{Det}(\mathrm{Tr}(a_i a_j)) \in K_v,$$

where the trace is induced from the L/K trace. Since each ω_i is in each $\mathcal{O}_w$, for $w \mid v$, the traces lie in $\mathcal{O}_v$, so

$$d = D(\omega_1, \ldots, \omega_n) \in \mathcal{O}_v.$$

Also note that $d \in K$ since each ω_i is in L. Now suppose that

$$\alpha = \sum_{i=1}^{n} a_i \omega_i \in \mathcal{O}_{w_1} \oplus \cdots \oplus \mathcal{O}_{w_g},$$

with $a_i \in K_v$. Then by properties of determinants for any m with $1 \leq m \leq n$, we have

$$D(\omega_1, \ldots, \omega_{m-1}, \alpha, \omega_{m+1}, \ldots, \omega_n) = a_m^2 D(\omega_1, \ldots, \omega_n). \tag{20.1.3}$$

The left hand side of (20.1.3) is in $\mathcal{O}_v$, so the right hand side is well, i.e.,

$$a_m^2 \cdot d \in \mathcal{O}_v, \qquad (\text{for } m = 1, \ldots, n),$$

where $d \in K$. Since $\omega_1, \ldots, \omega_n$ are a basis for L over K and the trace pairing is nondegenerate, we have $d \neq 0$, so by Theorem 20.1.4 we have $|d|_v = 1$ for all but finitely many v. Then for all but finitely many v we have that $a_m^2 \in \mathcal{O}_v$. For these v, that $a_m^2 \in \mathcal{O}_v$ implies $a_m \in \mathcal{O}_v$ since $a_m \in K_v$, i.e., α is in the left hand side of (20.1.2). □

Example 20.1.7. Let $K = \mathbf{Q}$ and $L = \mathbf{Q}(\sqrt{2})$. Let $\omega_1 = 1/3$ and $\omega_2 = 2\sqrt{2}$. In the first stage of the above proof we would eliminate $|\cdot|_3$ because ω_2 is not integral at 3. The discriminant is

$$d = D\left(\frac{1}{3}, 2\sqrt{2}\right) = \mathrm{Det}\begin{pmatrix} \frac{2}{9} & 0 \\ 0 & 16 \end{pmatrix} = \frac{32}{9}.$$

As explained in the second part of the proof, as long as $v \neq 2, 3$, we have equality of the left and right hand sides in (20.1.2).

20.2 Restricted Topological Products

In this section we describe a topological tool, which we need in order to define adeles (see Definition 20.3.1).

Definition 20.2.1 (Restricted Topological Products). Let X_λ, for $\lambda \in \Lambda$, be a family of topological spaces, and for almost all λ let $Y_\lambda \subset X_\lambda$ be an open subset of X_λ. Consider the space X whose elements are sequences $\mathbf{x} = \{x_\lambda\}_{\lambda\in\Lambda}$, where $x_\lambda \in X_\lambda$ for every λ, and $x_\lambda \in Y_\lambda$ for almost all λ. We give X a topology by taking as a basis of open sets the sets $\prod U_\lambda$, where $U_\lambda \subset X_\lambda$ is open for all λ, and $U_\lambda = Y_\lambda$ for almost all λ. We call X with this topology the *restricted topological product* of the X_λ with respect to the Y_λ.

Corollary 20.2.2. *Let S be a finite subset of Λ, and let X_S be the set of $\mathbf{x} \in X$ with $x_\lambda \in Y_\lambda$ for all $\lambda \notin S$, i.e.,*

$$X_S = \prod_{\lambda\in S} X_\lambda \times \prod_{\lambda\notin S} Y_\lambda \subset X.$$

Then X_S is an open subset of X, and the topology induced on X_S as a subset of X is the same as the product topology.

The restricted topological product depends on the totality of the Y_λ, but not on the individual Y_λ:

Lemma 20.2.3. *Let $Y'_\lambda \subset X_\lambda$ be open subsets, and suppose that $Y_\lambda = Y'_\lambda$ for almost all λ. Then the restricted topological product of the X_λ with respect to the Y'_λ is canonically isomorphic to the restricted topological product with respect to the Y_λ.*

Lemma 20.2.4. *Suppose that the X_λ are locally compact and that the Y_λ are compact. Then the restricted topological product X of the X_λ is locally compact.*

Proof. For any finite subset S of Λ, the open subset $X_S \subset X$ is locally compact, because by Lemma 20.2.2 it is a product of finitely many locally compact sets with an infinite product of compact sets. (Here we are using Tychonoff's theorem from topology, which asserts that an arbitrary product of compact topological spaces is compact (see Munkres's *Topology, a first course*, chapter 5).) Since $X = \cup_S X_S$, and the X_S are open in X, the result follows. □

The following measure will be extremely important in deducing topological properties of the ideles, which will be used in proving finiteness of class groups. See, e.g., the proof of Lemma 20.4.1, which is a key input to the proof of strong approximation (Theorem 20.4.4).

Definition 20.2.5 (Product Measure). For all $\lambda \in \Lambda$, suppose μ_λ is a measure on X_λ with $\mu_\lambda(Y_\lambda) = 1$ when Y_λ is defined. We define the *product measure* μ on X to be that for which a basis of measurable sets is

$$\prod_\lambda M_\lambda$$

where each $M_\lambda \subset X_\lambda$ has finite μ_λ-measure and $M_\lambda = Y_\lambda$ for almost all λ, and where

$$\mu\left(\prod_\lambda M_\lambda\right) = \prod_\lambda \mu_\lambda(M_\lambda).$$

20.3 The Adele Ring

Let K be a global field. For each normalization $|\cdot|_v$ of K, let K_v denote the completion of K. If $|\cdot|_v$ is non-archimedean, let $\mathcal{O}_v$ denote the ring of integers of K_v.

Definition 20.3.1 (Adele Ring). The *adele ring* $\mathbb{A}_K$ of K is the topological ring whose underlying topological space is the restricted topological product of the K_v with respect to the $\mathcal{O}_v$, and where addition and multiplication are defined componentwise:

$$(\mathbf{x}\mathbf{y})_v = \mathbf{x}_v\mathbf{y}_v \qquad (\mathbf{x}+\mathbf{y})_v = \mathbf{x}_v + \mathbf{y}_v \qquad \text{for } \mathbf{x}, \mathbf{y} \in \mathbb{A}_K. \tag{20.3.1}$$

It is readily verified that (i) this definition makes sense, i.e., if $\mathbf{x}, \mathbf{y} \in \mathbb{A}_K$, then $\mathbf{x}\mathbf{y}$ and $\mathbf{x}+\mathbf{y}$, whose components are given by (20.3.1), are also in $\mathbb{A}_K$, and (ii) that addition and multiplication are continuous in the $\mathbb{A}_K$-topology, so $\mathbb{A}_K$ is a topological ring, as asserted. Also, Lemma 20.2.4 implies that $\mathbb{A}_K$ is locally compact because the K_v are locally compact (Corollary 17.1.6), and the $\mathcal{O}_v$ are compact (Theorem 17.1.4).

There is a natural continuous ring inclusion

$$K \hookrightarrow \mathbb{A}_K \tag{20.3.2}$$

that sends $x \in K$ to the adele every one of whose components is x. This is an adele because $x \in \mathcal{O}_v$ for almost all v, by Lemma 20.1.2. The map is injective because each map $K \to K_v$ is an inclusion.

Definition 20.3.2 (Principal Adeles). The image of (20.3.2) is the ring of *principal adeles.*

It will cause no trouble to identify K with the principal adeles, so we shall speak of K as a subring of $\mathbb{A}_K$.

Formation of the adeles is compatibility with base change, in the following sense.

Lemma 20.3.3. *Suppose L is a finite (separable) extension of the global field K. Then*

$$\mathbb{A}_K \otimes_K L \cong \mathbb{A}_L \tag{20.3.3}$$

both algebraically and topologically. Under this isomorphism,

$$L \cong K \otimes_K L \subset \mathbb{A}_K \otimes_K L$$

maps isomorphically onto $L \subset \mathbb{A}_L$.

Proof. Let $\omega_1, \ldots, \omega_n$ be a basis for L/K and let v run through the normalized valuations on K. The left hand side of (20.3.3), with the tensor product topology, is the restricted product of the tensor products

$$K_v \otimes_K L \cong K_v \cdot \omega_1 \oplus \cdots \oplus K_v \cdot \omega_n$$

with respect to the integers

$$\mathcal{O}_v \cdot \omega_1 \oplus \cdots \oplus \mathcal{O}_v \cdot \omega_n. \tag{20.3.4}$$

(An element of the left hand side is a finite linear combination $\sum \mathbf{x}_i \otimes a_i$ of adeles $\mathbf{x}_i \in \mathbb{A}_K$ and coefficients $a_i \in L$, and there is a natural isomorphism from the ring of such formal sums to the restricted product of the $K_v \otimes_K L$.)

We proved before (Theorem 19.1.8) that

$$K_v \otimes_K L \cong L_{w_1} \oplus \cdots \oplus L_{w_g},$$

where $w_1, \ldots, w_g$ are the normalizations of the extensions of v to L. Furthermore, as we proved using discriminants (see Lemma 20.1.6), the above identification identifies (20.3.4) with

$$\mathcal{O}_{L_{w_1}} \oplus \cdots \oplus \mathcal{O}_{L_{w_g}},$$

for almost all v. Thus the left hand side of (20.3.3) is the restricted product of the $L_{w_1} \oplus \cdots \oplus L_{w_g}$ with respect to the $\mathcal{O}_{L_{w_1}} \oplus \cdots \oplus \mathcal{O}_{L_{w_g}}$. But this is canonically isomorphic to the restricted product of all completions L_w with respect to $\mathcal{O}_w$, which is the right hand side of (20.3.3). This establishes an isomorphism between the two sides of (20.3.3) as topological spaces. The map is also a ring homomorphism, so the two sides are algebraically isomorphic, as claimed. □

Corollary 20.3.4. *Let $\mathbb{A}_K^+$ denote the topological group obtained from the additive structure on $\mathbb{A}_K$. Suppose L is a finite seperable extension of K. Then*

$$\mathbb{A}_L^+ = \mathbb{A}_K^+ \oplus \cdots \oplus \mathbb{A}_K^+, \qquad (\text{[}L:K\text{] } \textit{summands}).$$

In this isomorphism the additive group $L^+ \subset \mathbb{A}_L^+$ of the principal adeles is mapped isomorphically onto $K^+ \oplus \cdots \oplus K^+$.

Proof. For any nonzero $\omega \in L$, the subgroup $\omega \cdot \mathbb{A}_K^+$ of $\mathbb{A}_L^+$ is isomorphic as a topological group to $\mathbb{A}_K^+$ (the isomorphism is multiplication by $1/\omega$). By Lemma 20.3.3, we have isomorphisms

$$\mathbb{A}_L^+ = \mathbb{A}_K^+ \otimes_K L \cong \omega_1 \cdot \mathbb{A}_K^+ \oplus \cdots \oplus \omega_n \cdot \mathbb{A}_K^+ \cong \mathbb{A}_K^+ \oplus \cdots \oplus \mathbb{A}_K^+.$$

If $a \in L$, write $a = \sum b_i \omega_i$, with $b_i \in K$. Then a maps via the above map to

$$x = (\omega_1 \cdot \{b_1\}, \ldots, \omega_n \cdot \{b_n\}),$$

where $\{b_i\}$ denotes the principal adele defined by b_i. Under the final map, x maps to the tuple

$$(b_1, \ldots, b_n) \in K \oplus \cdots \oplus K \subset \mathbb{A}_K^+ \oplus \cdots \oplus \mathbb{A}_K^+.$$

The dimensions of L and of $K \oplus \cdots \oplus K$ over K are the same, so this proves the final claim of the corollary. □

Theorem 20.3.5. *The global field K is discrete in $\mathbb{A}_K$ and the quotient $\mathbb{A}_K^+/K^+$ of additive groups is compact in the quotient topology.*

At this point Cassels remarks

> "It is impossible to conceive of any other uniquely defined topology on K. This metamathematical reason is more persuasive than the argument that follows!"

Proof. Corollary 20.3.4, with K for L and $\mathbf{Q}$ or $\mathbf{F}(t)$ for K, shows that it is enough to verify the theorem for $\mathbf{Q}$ or $\mathbf{F}(t)$, and we shall do it here for $\mathbf{Q}$.

To show that $\mathbf{Q}^+$ is discrete in $\mathbb{A}_{\mathbf{Q}}^+$ it is enough, because of the group structure, to find an open set U that contains $0 \in \mathbb{A}_{\mathbf{Q}}^+$, but which contains no other elements of $\mathbf{Q}^+$. (If $\alpha \in \mathbf{Q}^+$, then $U + \alpha$ is an open subset of $\mathbb{A}_{\mathbf{Q}}^+$ whose intersection with $\mathbf{Q}^+$ is $\{\alpha\}$.) We take for U the set of $\mathbf{x} = \{x_v\}_v \in \mathbb{A}_{\mathbf{Q}}^+$ with

$$|x_\infty|_\infty < 1 \qquad \text{and} \qquad |x_p|_p \leq 1 \quad (\text{all } p),$$

where $|\cdot|_p$ and $|\cdot|_\infty$ are respectively the p-adic and the usual archimedean absolute values on $\mathbf{Q}$. If $b \in \mathbf{Q} \cap U$, then in the first place $b \in \mathbf{Z}$ because $|b|_p \leq 1$ for all p, and then $b = 0$ because $|b|_\infty < 1$. This proves that K^+ is discrete in $\mathbb{A}_{\mathbf{Q}}^+$. (If we leave out one valuation, as we will see later (Theorem 20.4.4), this theorem is false—what goes wrong with the proof just given?)

Next we prove that the quotient $\mathbb{A}_{\mathbf{Q}}^+/\mathbf{Q}^+$ is compact. Let $W \subset \mathbb{A}_{\mathbf{Q}}^+$ consist of the $\mathbf{x} = \{x_v\}_v \in \mathbb{A}_{\mathbf{Q}}^+$ with

$$|x_\infty|_\infty \leq \frac{1}{2} \qquad \text{and} \qquad |x_p|_p \leq 1 \qquad \text{for all primes } p.$$

We show that every adele $\mathbf{y} = \{y_v\}_v$ is of the form

$$\mathbf{y} = a + \mathbf{x}, \qquad a \in \mathbf{Q}, \quad \mathbf{x} \in W,$$

which will imply that the compact set W maps surjectively onto $\mathbb{A}_{\mathbf{Q}}^{+}/\mathbf{Q}^{+}$. Fix an adele $\mathbf{y} = \{y_v\} \in \mathbb{A}_{\mathbf{Q}}^{+}$. Since $\mathbf{y}$ is an adele, for each prime p we can find a rational number

$$r_p = \frac{z_p}{p^{n_p}} \qquad \text{with} \quad z_p \in \mathbf{Z} \quad \text{and} \quad n_p \in \mathbf{Z}_{\geq 0}$$

such that

$$|y_p - r_p|_p \leq 1,$$

and

$$r_p = 0 \qquad \text{almost all } p.$$

More precisely, for the finitely many p such that

$$y_p = \sum_{n \geq -|s|} a_n p^n \notin \mathbf{Z}_p,$$

choose r_p to be a rational number that is the value of an appropriate truncation of the p-adic expansion of y_p, and when $y_p \in \mathbf{Z}_p$ just choose $r_p = 0$. Hence $r = \sum_p r_p \in \mathbf{Q}$ is well defined. The r_q for $q \neq p$ do not mess up the inequality $|y_p - r|_p \leq 1$ since the valuation $|\cdot|_p$ is non-archimedean and the r_q do not have any p in their denominator:

$$|y_p - r|_p = \left| y_p - r_p - \sum_{q \neq p} r_q \right|_p \leq \max\left(|y_p - r_p|_p, \left| \sum_{q \neq p} r_q \right|_p \right) \leq \max(1, 1) = 1.$$

Now choose $s \in \mathbf{Z}$ such that

$$|b_\infty - r - s| \leq \frac{1}{2}.$$

Then $a = r + s$ and $\mathbf{x} = \mathbf{y} - a$ do what is required, since $\mathbf{y} - a = \mathbf{y} - r - s$ has the desired property (since $s \in \mathbf{Z}$ and the p-adic valuations are non-archimedean).

Hence the continuous map $W \to \mathbb{A}_{\mathbf{Q}}^{+}/\mathbf{Q}^{+}$ induced by the quotient map $\mathbb{A}_{\mathbf{Q}}^{+} \to \mathbb{A}_{\mathbf{Q}}^{+}/\mathbf{Q}^{+}$ is surjective. But W is compact (being the topological product of the compact spaces $|x_\infty|_\infty \leq 1/2$ and the $\mathbf{Z}_p$ for all p), hence $\mathbb{A}_{\mathbf{Q}}^{+}/\mathbf{Q}^{+}$ is also compact. □

Corollary 20.3.6. *There is a subset W of $\mathbb{A}_K$ defined by inequalities of the type $|x_v|_v \leq \delta_v$, where $\delta_v = 1$ for almost all v, such that every $\mathbf{y} \in \mathbb{A}_K$ can be put in the form*

$$\mathbf{y} = a + \mathbf{x}, \qquad a \in K, \quad \mathbf{x} \in W,$$

i.e., $\mathbb{A}_K = K + W$.

Proof. We constructed such a set for $K = \mathbf{Q}$ when proving Theorem 20.3.5. For general K the W coming from the proof determines compenent-wise a subset of $\mathbb{A}_K^{+} \cong \mathbb{A}_{\mathbf{Q}}^{+} \oplus \cdots \oplus \mathbb{A}_{\mathbf{Q}}^{+}$ that is a subset of a set with the properties claimed by the corollary. □

As already remarked, $\mathbb{A}_K^+$ is a locally compact group, so it has an invariant Haar measure. In fact one choice of this Haar measure is the product of the Haar measures on the K_v, in the sense of Definition 20.2.5.

Corollary 20.3.7. *The quotient $\mathbb{A}_K^+/K^+$ has finite measure in the quotient measure induced by the Haar measure on $\mathbb{A}_K^+$.*

Remark 20.3.8. This statement is independent of the particular choice of the multiplicative constant in the Haar measure on $\mathbb{A}_K^+$. We do not here go into the question of finding the measure $\mathbb{A}_K^+/K^+$ in terms of our explicitly given Haar measure. (See Tate's thesis, [Cp86, Chapter XV].)

Proof. This can be reduced similarly to the case of $\mathbf{Q}$ or $\mathbf{F}(t)$ which is immediate, e.g., the W defined above has measure 1 for our Haar measure.

Alternatively, finite measure follows from compactness. To see this, cover $\mathbb{A}_K/K^+$ with the translates of U, where U is a nonempty open set with finite measure. The existence of a finite subcover implies finite measure. □

Remark 20.3.9. We give an alternative proof of the product formula $\prod |a|_v = 1$ for nonzero $a \in K$. We have seen that if $x_v \in K_v$, then multiplication by x_v magnifies the Haar measure in K_v^+ by a factor of $|x_v|_v$. Hence if $\mathbf{x} = \{x_v\} \in \mathbb{A}_K$, then multiplication by $\mathbf{x}$ magnifies the Haar measure in $\mathbb{A}_K^+$ by $\prod |x_v|_v$. But now multiplication by $a \in K$ takes $K^+ \subset \mathbb{A}_K^+$ into K^+, so gives a well-defined bijection of $\mathbb{A}_K^+/K^+$ onto $\mathbb{A}_K^+/K^+$ which magnifies the measure by the factor $\prod |a|_v$. Hence $\prod |a|_v = 1$ Corollary 20.3.7. (The point is that if μ is the measure of $\mathbb{A}_K^+/K^+$, then $\mu = \prod |a|_v \cdot \mu$, so because μ is finite we must have $\prod |a|_v = 1$.)

20.4 Strong Approximation

We first prove a technical lemma and corollary, then use them to deduce the strong approximation theorem, which is an extreme generalization of the Chinese Remainder Theorem; it asserts that K^+ is dense in the analogue of the adeles with one valuation removed.

The proof of Lemma 20.4.1 below will use in a crucial way the normalized Haar measure on $\mathbb{A}_K$ and the induced measure on the compact quotient $\mathbb{A}_K^+/K^+$. Since I am not formally developing Haar measure on locally compact groups, and since I didn't explain induced measures on quotients well in the last chapter, hopefully the following discussion will help clarify what is going on.

The real numbers $\mathbf{R}^+$ under addition is a locally compact topological group. Normalized Haar measure μ has the property that $\mu([a, b]) = b - a$, where $a \leq b$ are real numbers and $[a, b]$ is the closed interval from a to b. The subset $\mathbf{Z}^+$ of $\mathbf{R}^+$ is discrete, and the quotient $S^1 = \mathbf{R}^+/\mathbf{Z}^+$ is a compact topological group, which thus has a Haar measure. Let $\overline{\mu}$ be the Haar measure on S^1 normalized so that the natural quotient $\pi : \mathbf{R}^+ \to S^1$ preserves the measure, in the sense that if $X \subset \mathbf{R}^+$ is a measurable set that maps injectively into S^1, then $\mu(X) = \overline{\mu}(\pi(X))$. This determine $\overline{\mu}$ and we have $\overline{\mu}(S^1) = 1$ since $X = [0, 1)$ is a measurable set that maps bijectively onto S^1 and has measure 1. The situation for the map $\mathbb{A}_K \to \mathbb{A}_K/K^+$ is pretty much the same.

Lemma 20.4.1. *There is a constant $C > 0$ that depends only on the global field K with the following property:*

Whenever $\mathbf{x} = \{x_v\}_v \in \mathbb{A}_K$ is such that

$$\prod_v |x_v|_v > C, \tag{20.4.1}$$

then there is a nonzero principal adele $a \in K \subset \mathbb{A}_K$ such that

$$|a|_v \leq |x_v|_v \qquad \text{for all } v.$$

Proof. This proof is modelled on Blichfeldt's proof of Minkowski's Theorem in the Geometry of Numbers, and works in quite general circumstances.

First we show that (20.4.1) implies that $|x_v|_v = 1$ for almost all v. Because $\mathbf{x}$ is an adele, we have $|x_v|_v \leq 1$ for almost all v. If $|x_v|_v < 1$ for infinitely many v, then the product in (20.4.1) would have to be 0. (We prove this only when K is a finite extension of $\mathbf{Q}$.) Excluding archimedean valuations, this is because the normalized valuation $|x_v|_v = |\mathrm{Norm}(x_v)|_p$, which if less than 1 is necessarily $\leq 1/p$. Any infinite product of numbers $1/p_i$ must be 0, whenever p_i is a sequence of primes.

Let c_0 be the Haar measure of $\mathbb{A}_K^+/K^+$ induced from normalized Haar measure on $\mathbb{A}_K^+$, and let c_1 be the Haar measure of the set of $\mathbf{y} = \{y_v\}_v \in \mathbb{A}_K^+$ that satisfy

$$\begin{aligned} |y_v|_v &\leq \frac{1}{2} && \text{if } v \text{ is real archimedean,} \\ |y_v|_v &\leq \frac{1}{2} && \text{if } v \text{ is complex archimedean,} \\ |y_v|_v &\leq 1 && \text{if } v \text{ is non-archimedean.} \end{aligned}$$

(As we will see, any positive real number $\leq 1/2$ would suffice in the definition of c_1 above. For example, in Cassels's article he uses the mysterious 1/10. He also doesn't discuss the subtleties of the complex archimedean case separately.)

Then $0 < c_0 < \infty$ since $\mathbb{A}_K/K^+$ is compact, and $0 < c_1 < \infty$ because the number of archimedean valuations v is finite. We show that

$$C = \frac{c_0}{c_1}$$

will do. Thus suppose $\mathbf{x}$ is as in (20.4.1).

The set T of $\mathbf{t} = \{t_v\}_v \in \mathbb{A}_K^+$ such that

$$\begin{aligned} |t_v|_v &\leq \frac{1}{2}\,|x_v|_v && \text{if } v \text{ is real archimedean,} \\ |t_v|_v &\leq \frac{1}{2}\sqrt{|x_v|_v} && \text{if } v \text{ is complex archimedean,} \\ |t_v|_v &\leq |x_v|_v && \text{if } v \text{ is non-archimedean} \end{aligned}$$

has measure

$$c_1 \cdot \prod_v |x_v|_v > c_1 \cdot C = c_0. \tag{20.4.2}$$

(Note: If there are complex valuations, then the some of the $|x_v|_v$'s in the product must be squared.)

Because of (20.4.2), in the quotient map $\mathbb{A}_K^+ \to \mathbb{A}_K^+/K^+$ there must be a pair of distinct points of T that have the same image in $\mathbb{A}_K^+/K^+$, say

$$\mathbf{t}' = \{t'_v\}_v \in T \quad \text{and} \quad \mathbf{t}'' = \{t''_v\}_v \in T$$

and

$$a = \mathbf{t}' - \mathbf{t}'' \in K^+$$

is nonzero. Then

$$|a|_v = \left|t'_v - t''_v\right|_v \leq \begin{cases} |t'_v| + |t''_v| \leq 2 \cdot \frac{1}{2} |x_v|_v \leq |x_v|_v & \text{if } v \text{ is real archimedean, or} \\ \max(|t'_v|, |t''_v|) \leq |x_v|_v & \text{if } v \text{ is non-archimedean,} \end{cases}$$

for all v. In the case of complex archimedean v, we must be careful because the normalized valuation $|\cdot|_v$ is the *square* of the usual archimedean complex valuation $|\cdot|_\infty$ on $\mathbf{C}$, so e.g., it does not satisfy the triangle inequality. In particular, the quantity $|t'_v - t''_v|_v$ is at most the square of the maximum distance between two points in the disc in $\mathbf{C}$ of radius $\frac{1}{2}\sqrt{|x_v|_v}$, where by distance we mean the usual distance. This maximum distance in such a disc is at most $\sqrt{|x_v|_v}$, so $|t'_v - t''_v|_v$ is at most $|x_v|_v$, as required. Thus a satisfies the requirements of the lemma. □

Corollary 20.4.2. *Let v_0 be a normalized valuation and let $\delta_v > 0$ be given for all $v \neq v_0$ with $\delta_v = 1$ for almost all v. Then there is a nonzero $a \in K$ with*

$$|a|_v \leq \delta_v \qquad \textit{(all } v \neq v_0\textit{)}.$$

Proof. This is just a degenerate case of Lemma 20.4.1. Choose $x_v \in K_v$ with $0 < |x_v|_v \leq \delta_v$ and $|x_v|_v = 1$ if $\delta_v = 1$. We can then choose $x_{v_0} \in K_{v_0}$ so that

$$\prod_{\text{all } v \text{ including } v_0} |x_v|_v > C.$$

Then Lemma 20.4.1 does what is required. □

Remark 20.4.3. The character group of the locally compact group $\mathbb{A}_K^+$ is isomorphic to $\mathbb{A}_K^+$ and K^+ plays a special role. See Chapter XV of [Cp86], Lang's [Lan64], Weil's [Wei82], and Godement's Bourbaki seminars 171 and 176. This duality lies behind the functional equation of ζ and L-functions. Iwasawa has shown [Iwa53] that the rings of adeles are characterized by certain general topologico-algebraic properties.

We proved before that K is discrete in $\mathbb{A}_K$. If one valuation is removed, the situation is much different.

Theorem 20.4.4 (Strong Approximation). *Let v_0 be any normalized nontrivial valuation of the global field K. Let $\mathbb{A}_{K,v_0}$ be the restricted topological product of the K_v with respect to the $\mathcal{O}_v$, where v runs through all normalized valuations $v \neq v_0$. Then K is dense in $\mathbb{A}_{K,v_0}$.*

Proof. This proof was suggested by Prof. Kneser at the Cassels-Frohlich conference.

Recall that if $\mathbf{x} = \{x_v\}_v \in \mathbb{A}_{K,v_0}$ then a basis of open sets about $\mathbf{x}$ is the collection of products

$$\prod_{v \in S} B(x_v, \varepsilon_v) \times \prod_{v \notin S,\ v \neq v_0} \mathcal{O}_v,$$

where $B(x_v, \varepsilon_v)$ is an open ball in K_v about x_v, and S runs through finite sets of normalized valuations (not including v_0). Thus denseness of K in $\mathbb{A}_{K,v_0}$ is equivalent to the following statement about elements. Suppose we are given (i) a finite set S of valuations $v \neq v_0$, (ii) elements $x_v \in K_v$ for all $v \in S$, and (iii) an $\varepsilon > 0$. Then there is an element $b \in K$ such that $|b - x_v|_v < \varepsilon$ for all $v \in S$ and $|b|_v \leq 1$ for all $v \notin S$ with $v \neq v_0$.

By the corollary to our proof that $\mathbb{A}_K^+/K^+$ is compact (Corollary 20.3.6), there is a $W \subset \mathbb{A}_K$ that is defined by inequalities of the form $|y_v|_v \leq \delta_v$ (where $\delta_v = 1$ for almost all v) such that ever $\mathbf{z} \in \mathbb{A}_K$ is of the form

$$\mathbf{z} = \mathbf{y} + c, \qquad \mathbf{y} \in W, \quad c \in K. \tag{20.4.3}$$

By Corollary 20.4.2, there is a nonzero $a \in K$ such that

$$\begin{aligned} |a|_v &< \frac{1}{\delta_v} \cdot \varepsilon && \text{for } v \in S, \\ |a|_v &\leq \frac{1}{\delta_v} && \text{for } v \notin S,\ v \neq v_0. \end{aligned}$$

Hence on putting $\mathbf{z} = \frac{1}{a} \cdot \mathbf{x}$ in (20.4.3) and multiplying by a, we see that every $\mathbf{x} \in \mathbb{A}_K$ is of the shape

$$\mathbf{x} = \mathbf{w} + b, \qquad \mathbf{w} \in a \cdot W, \quad b \in K,$$

where $a \cdot W$ is the set of $a\mathbf{y}$ for $\mathbf{y} \in W$. If now we let $\mathbf{x}$ have components the given x_v at $v \in S$, and (say) 0 elsewhere, then $b = \mathbf{x} - \mathbf{w}$ has the properties required. □

Remark 20.4.5. The proof gives a quantitative form of the theorem (i.e., with a bound for $|b|_{v_0}$). For an alternative approach, see [Mah64].

In the next chapter we'll introduce the ideles $\mathbb{A}_K^*$. Finally, we'll relate ideles to ideals, and use everything so far to give a new interpretation of class groups and their finiteness.

Chapter 21

Ideles and Ideals

In this chapter, we introduce the ideles $\mathbb{I}_K$, and relate ideles to ideals, and use what we've done so far to give an alternative interpretation of class groups and their finiteness, thus linking the adelic point of view with the classical point of view of the first part of this course.

21.1 The Idele Group

The invertible elements of any commutative topological ring R are a group R^* under multiplication. In general R^* is not a topological group if it is endowed with the subset topology because inversion need not be continuous (only multiplication and addition on R are required to be continuous). It is usual therefore to give R^* the following topology. There is an injection

$$x \mapsto \left(x, \ \frac{1}{x}\right) \tag{21.1.1}$$

of R^* into the topological product $R \times R$. We give R^* the corresponding subset topology. Then R^* with this topology is a topological group and the inclusion map $R^* \hookrightarrow R$ is continous. To see continuity of inclusion, note that this topology is finer (has at least as many open sets) than the subset topology induced by $R^* \subset R$, since the projection maps $R \times R \to R$ are continuous.

Example 21.1.1. This is a "non-example". The inverse map on $\mathbf{Z}_p^*$ is continuous with respect to the p-adic topology. If $a, b \in \mathbf{Z}_p^*$, then $|a| = |b| = 1$, so if $|a - b| < \varepsilon$, then

$$\left|\frac{1}{a} - \frac{1}{b}\right| = \left|\frac{b-a}{ab}\right| = \frac{|b-a|}{|ab|} < \frac{\varepsilon}{1} = \varepsilon.$$

Definition 21.1.2 (Idele Group). The *idele group* $\mathbb{I}_K$ of K is the group $\mathbb{A}_K^*$ of invertible elements of the adele ring $\mathbb{A}_K$.

We shall usually speak of $\mathbb{I}_K$ as a subset of $\mathbb{A}_K$, and will have to distinguish between the $\mathbb{I}_K$ and $\mathbb{A}_K$-topologies.

Example 21.1.3. For a rational prime p, let $\mathbf{x}_p \in \mathbb{A}_{\mathbf{Q}}$ be the adele whose pth component is p and whose vth component, for $v \neq p$, is 1. Then $\mathbf{x}_p \to 1$ as $p \to \infty$ in $\mathbb{A}_{\mathbf{Q}}$, for the following reason. We must show that if U is a basic open set that contains the adele $1 = \{1\}_v$, the $\mathbf{x}_p$ for all sufficiently large p are contained in U. Since U contains 1 and is a basic open set, it is of the form

$$\prod_{v \in S} U_v \times \prod_{v \notin S} \mathbf{Z}_v,$$

where S if a finite set, and the U_v, for $v \in S$, are arbitrary open subsets of $\mathbf{Q}_v$ that contain 1. If q is a prime larger than any prime in S, then $\mathbf{x}_p$ for $p \geq q$, is in U. This proves convergence. If the inverse map were continuous on $\mathbb{I}_K$, then the sequence of $\mathbf{x}_p^{-1}$ would converge to $1^{-1} = 1$. However, if U is an open set as above about 1, then for sufficiently large p, *none* of the adeles $\mathbf{x}_p$ are contained in U.

Lemma 21.1.4. *The group of ideles $\mathbb{I}_K$ is the restricted topological project of the K_v^* with respect to the units $U_v = \mathcal{O}_v^* \subset K_v$, with the restricted product topology.*

We omit the proof of Lemma 21.1.4, which is a matter of thinking carefully about the definitions. The main point is that inversion is continuous on $\mathcal{O}_v^*$ for each v. (See Example 21.1.1.)

We have seen that K is naturally embedded in $\mathbb{A}_K$, so K^* is naturally embedded in $\mathbb{I}_K$.

Definition 21.1.5 (Principal Ideles). We call K^*, considered as a subgroup of $\mathbb{I}_K$, the *principal ideles.*

Lemma 21.1.6. *The principal ideles K^* are discrete as a subgroup of $\mathbb{I}_K$.*

Proof. For K is discrete in $\mathbb{A}_K$, so K^* is embedded in $\mathbb{A}_K \times \mathbb{A}_K$ by (21.1.1) as a discrete subset. (Alternatively, the subgroup topology on $\mathbb{I}_K$ is finer than the topology coming from $\mathbb{I}_K$ being a subset of $\mathbb{A}_K$, and K is already discrete in $\mathbb{A}_K$.) □

Definition 21.1.7 (Content of an Idele). The *content* of $\mathbf{x} = \{x_v\}_v \in \mathbb{I}_K$ is

$$c(\mathbf{x}) = \prod_{\text{all } v} |x_v|_v \in \mathbf{R}_{>0}.$$

Lemma 21.1.8. *The map $\mathbf{x} \to c(\mathbf{x})$ is a continuous homomorphism of the topological group $\mathbb{I}_K$ into $\mathbf{R}_{>0}$, where we view $\mathbf{R}_{>0}$ as a topological group under multiplication. If K is a number field, then c is surjective.*

Proof. That the content map c satisfies the axioms of a homomorphisms follows from the multiplicative nature of the defining formula for c. For continuity, suppose (a, b) is an open interval in $\mathbf{R}_{>0}$. Suppose $\mathbf{x} \in \mathbb{I}_K$ is such that $c(\mathbf{x}) \in (a, b)$. By considering small intervals about each non-unit component of $\mathbf{x}$, we find an open neighborhood $U \subset \mathbb{I}_K$ of $\mathbf{x}$ such that $c(U) \subset (a, b)$. It follows the $c^{-1}((a, b))$ is open.

For surjectivity, use that each archimedean valuation is surjective, and choose an idele that is 1 at all but one archimedean valuation. □

Remark 21.1.9. Note also that the $\mathbb{I}_K$-topology is that appropriate to a group of operators on $\mathbb{A}_K^+$: a basis of open sets is the $S(C, U)$, where $C, U \subset \mathbb{A}_K^+$ are, respectively, $\mathbb{A}_K$-compact and $\mathbb{A}_K$-open, and S consists of the $\mathbf{x} \in \mathbb{I}_J$ such that $(1 - \mathbf{x})C \subset U$ and $(1 - \mathbf{x}^{-1})C \subset U$.

Definition 21.1.10 (1-Ideles). The subgroup $\mathbb{I}_K^1$ of 1-*ideles* is the subgroup of ideles $\mathbf{x} = \{x_v\}$ such that $c(\mathbf{x}) = 1$. Thus $\mathbb{I}_K^1$ is the kernel of c, so we have an exact sequence

$$1 \to \mathbb{I}_K^1 \to \mathbb{I}_K \xrightarrow{c} \mathbf{R}_{>0} \to 1,$$

where the surjectivity on the right is only if K is a number field.

Lemma 21.1.11. *The subset $\mathbb{I}_K^1$ of $\mathbb{A}_K$ is closed as a subset, and the $\mathbb{A}_K$-subset topology on $\mathbb{I}_K^1$ coincides with the $\mathbb{I}_K$-subset topology on $\mathbb{I}_K^1$.*

Proof. Let $\mathbf{x} \in \mathbb{A}_K$ with $\mathbf{x} \notin \mathbb{I}_K^1$. To prove that $\mathbb{I}_K^1$ is closed in $\mathbb{A}_K$, we find an $\mathbb{A}_K$-neighborhood W of $\mathbf{x}$ that does not meet $\mathbb{I}_K^1$.

1st Case. Suppose that $\prod_v |x_v|_v < 1$ (possibly $= 0$). Then there is a finite set S of v such that

1. S contains all the v with $|x_v|_v > 1$, and

2. $\prod_{v \in S} |x_v|_v < 1$.

Then the set W can be defined by

$$\begin{aligned} |w_v - x_v|_v &< \varepsilon \qquad v \in S \\ |w_v|_v &\leq 1 \qquad v \notin S \end{aligned}$$

for sufficiently small ε.

2nd Case. Suppose that $C := \prod_v |x_v|_v > 1$. Then there is a finite set S of v such that

1. S contains all the v with $|x_v|_v > 1$, and

2. if $v \notin S$ an inequality $|w_v|_v < 1$ implies $|w_v|_v < \frac{1}{2C}$. (This is because for a non-archimedean valuation, the largest absolute value less than 1 is $1/p$, where p is the residue characteristic. Also, the upper bound in Cassels's article is $\frac{1}{2}C$ instead of $\frac{1}{2C}$, but I think he got it wrong.)

We can choose ε so small that $|w_v - x_v|_v < \varepsilon$ (for $v \in S$) implies $1 < \prod_{v \in S} |w_v|_v < 2C$. Then W may be defined by

$$\begin{aligned} |w_v - x_v|_v &< \varepsilon \qquad v \in S \\ |w_v|_v &\leq 1 \qquad v \notin S. \end{aligned}$$

This works because if $\mathbf{w} \in W$, then either $|w_v|_v = 1$ for all $v \notin S$, in which case $1 < c(\mathbf{w}) < 2c$, so $\mathbf{w} \notin \mathbb{I}_K^1$, or $|w_{v_0}|_{v_0} < 1$ for some $v_0 \notin S$, in which case

$$c(\mathbf{w}) = \left(\prod_{v \in S} |w_v|_v\right) \cdot |w_{v_0}| \cdots < 2C \cdot \frac{1}{2C} \cdots < 1,$$

so again $\mathbf{w} \notin \mathbb{I}_K^1$.

We next show that the $\mathbb{I}_K$- and $\mathbb{A}_K$-topologies on $\mathbb{I}_K^1$ are the same. If $\mathbf{x} \in \mathbb{I}_K^1$, we must show that every $\mathbb{A}_K$-neighborhood of $\mathbf{x}$ contains an $\mathbb{A}_K$-neighborhood and vice-versa.

Let $W \subset \mathbb{I}_K^1$ be an $\mathbb{A}_K$-neighborhood of $\mathbf{x}$. Then it contains an $\mathbb{A}_K$-neighborhood of the type

$$|w_v - x_v|_v < \varepsilon \qquad v \in S \tag{21.1.2}$$
$$|w_v|_v \leq 1 \qquad v \notin S \tag{21.1.3}$$

where S is a finite set of valuations v. This contains the $\mathbb{I}_K$-neighborhood in which $\leq$ in (21.1.2) is replaced by $=$.

Next let $H \subset \mathbb{I}_K^1$ be an $\mathbb{I}_K$-neighborhood. Then it contains an $\mathbb{I}_K$-neighborhood of the form

$$|w_v - x_v|_v < \varepsilon \qquad v \in S \tag{21.1.4}$$
$$|w_v|_v = 1 \qquad v \notin S, \tag{21.1.5}$$

where the finite set S contains at least all archimedean valuations v and all valuations v with $|x_v|_v \neq 1$. Since $\prod |x_v|_v = 1$, we may also suppose that ε is so small that (21.1.4) implies

$$\prod_v |w_v|_v < 2.$$

Then the intersection of (21.1.4) with $\mathbb{I}_K^1$ is the same as that of (21.1.2) with $\mathbb{I}_K^1$, i.e., (21.1.4) defines an $\mathbb{A}_K$-neighborhood. □

By the product formula we have that $K^* \subset \mathbb{I}_K^1$. The following result is of vital importance in class field theory.

Theorem 21.1.12. *The quotient $\mathbb{I}_K^1/K^*$ with the quotient topology is compact.*

Proof. After the preceeding lemma, it is enough to find an $\mathbb{A}_K$-compact set $W \subset \mathbb{A}_K$ such that the map

$$W \cap \mathbb{I}_K^1 \to \mathbb{I}_K^1/K^*$$

is surjective. We take for W the set of $\mathbf{w} = \{w_v\}_v$ with

$$|w_v|_v \leq |x_v|_v\,,$$

where $\mathbf{x} = \{x_v\}_v$ is any idele of content greater than the C of Lemma 20.4.1.

Let $\mathbf{y} = \{y_v\}_v \in \mathbb{I}_K^1$. Then the content of $\mathbf{x}/\mathbf{y}$ equals the content of $\mathbf{x}$, so by Lemma 20.4.1 there is an $a \in K^*$ such that

$$|a|_v \leq \left|\frac{x_v}{y_v}\right|_v \qquad \text{all } v.$$

Then $a\mathbf{y} \in W$, as required. □

Remark 21.1.13. The quotient $\mathbb{I}_K^1/K^*$ is totally disconnected in the function field case. For the structure of its connected component in the number field case, see papers of Artin and Weil in the "Proceedings of the Tokyo Symposium on Algebraic Number Theory, 1955" (Science Council of Japan) or [AT90]. The determination of the character group of $\mathbb{I}_K/K^*$ is global class field theory.

21.2 Ideals and Divisors

Suppose that K is a finite extension of $\mathbf{Q}$. Let F_K be the the free abelian group on a set of symbols in bijection with the non-archimedean valuation v of K. Thus an element of F_K is a formal linear combination

$$\sum_{v \text{ non arch.}} n_v \cdot v$$

where $n_v \in \mathbf{Z}$ and all but finitely many n_v are 0.

Lemma 21.2.1. *There is a natural bijection between F_K and the group of nonzero fractional ideals of $\mathcal{O}_K$. The correspondence is induced by*

$$v \mapsto \wp_v = \{x \in \mathcal{O}_K : v(x) < 1\},$$

where v is a non-archimedean valuation.

Endow F_K with the discrete topology. Then there is a natural continuous map $\pi : \mathbb{I}_K \to F_K$ given by

$$\mathbf{x} = \{x_v\}_v \mapsto \sum_v \mathrm{ord}_v(x_v) \cdot v.$$

This map is continuous since the inverse image of a valuation v (a point) is the product

$$\pi^{-1}(v) = \pi\mathcal{O}_v^* \quad \times \prod_{w \text{ archimedean}} K_w^* \quad \times \prod_{w \neq v \text{ non-arch.}} \mathcal{O}_w^*,$$

which is an open set in the restricted product topology on $\mathbb{I}_K$. Moreover, the image of K^* in F_K is the group of nonzero principal fractional ideals.

Recall that the *class group* C_K of the number field K is by definition the quotient of F_K by the image of K^*.

Theorem 21.2.2. *The class group C_K of a number field K is finite.*

Proof. We first prove that the map $\mathbb{I}_K^1 \to F_K$ is surjective. Let ∞ be an archimedean valuation on K. If v is a non-archimedean valuation, let $\mathbf{x} \in \mathbb{I}_K^1$ be a 1-idele such that $x_w = 1$ at ever valuation w except v and ∞. At v, choose $x_v = \pi$ to be a generator for the maximal ideal of $\mathcal{O}_v$, and choose x_∞ to be such that $|x_\infty|_\infty = 1/|x_v|_v$. Then $\mathbf{x} \in \mathbb{I}_K$ and $\prod_w |x_w|_w = 1$, so $\mathbf{x} \in \mathbb{I}_K^1$. Also $\mathbf{x}$ maps to $v \in F_K$.

Thus the group of ideal classes is the continuous image of the compact group $\mathbb{I}_K^1/K^*$ (see Theorem 21.1.12), hence compact. But a compact discrete group is finite. □

21.2.1 The Function Field Case

When K is a finite separable extension of $\mathbf{F}(t)$, we define the divisor group D_K of K to be the free abelian group on all the valuations v. For each v the number of elements of the residue class field $\mathbf{F}_v = \mathcal{O}_v/\wp_v$ of v is a power, say q^{n_v}, of the number q of elements in $\mathbf{F}_v$. We call n_v the degree of v, and similarly define $\sum n_v d_v$ to be the degree of the divisor $\sum n_v \cdot v$. The divisors of degree 0 form a group D_K^0. As before, the principal divisor attached to $a \in K^*$ is $\sum \mathrm{ord}_v(a) \cdot v \in D_K$. The following theorem is proved in the same way as Theorem 21.2.2.

Theorem 21.2.3. *The quotient of D_K^0 modulo the principal divisors is a finite group.*

21.2.2 Jacobians of Curves

For those familiar with algebraic geometry and algebraic curves, one can prove Theorem 21.2.3 from an alternative point of view. There is a bijection between nonsingular geometrically irreducible projective curves over $\mathbf{F}$ and function fields K over $\mathbf{F}$ (which we assume are finite separable extensions of $\mathbf{F}(t)$ such that $\overline{\mathbf{F}} \cap K = \mathbf{F}$). Let X be the curve corresponding to K. The group D_K^0 is in bijection with the divisors of degree 0 on X, a group typically denoted $\mathrm{Div}^0(X)$. The quotient of $\mathrm{Div}^0(X)$ by principal divisors is denoted $\mathrm{Pic}^0(X)$. The *Jacobian* of X is an abelian variety $J = \mathrm{Jac}(X)$ over the finite field $\mathbf{F}$ whose dimension is equal to the genus of X. Moreover, assuming X has an $\mathbf{F}$-rational point, the elements of $\mathrm{Pic}^0(X)$ are in natural bijection with the $\mathbf{F}$-rational points on J. In particular, with these hypothesis, the class group of K, which is isomorphic to $\mathrm{Pic}^0(X)$, is in bijection with the group of $\mathbf{F}$-rational points on an algebraic variety over a finite field. This gives an alternative more complicated proof of finiteness of the degree 0 class group of a function field.

Without the degree 0 condition, the divisor class group won't be finite. It is an extension of $\mathbf{Z}$ by a finite group.

$$0 \to \mathrm{Pic}^0(X) \to \mathrm{Pic}(X) \xrightarrow{\deg} n\mathbf{Z} \to 0,$$

where n is the greatest common divisor of the degrees of elements of $\mathrm{Pic}(X)$, which is 1 when X has a rational point.

Chapter 22

Exercises

1. Let $A = \begin{pmatrix} 4 & 7 & 2 \\ 2 & 4 & 6 \\ 0 & 0 & 0 \end{pmatrix}$.

   (a) Find invertible integer matrices P and Q such that PAQ is in Smith normal form.

   (b) What is the group structure of the cokernel of the map $\mathbf{Z}^3 \to \mathbf{Z}^3$ defined by multiplication by A?

2. Let G be the abelian group generated by x, y, z with relatoins $2x + y = 0$ and $x - y + 3z = 0$. Find a product of cyclic groups that is isomorphic to G.

3. Prove that each of the following rings have infinitely many prime ideals:

   (a) The integers $\mathbf{Z}$. [Hint: Euclid gave a famous proof of this long ago.]

   (b) The ring $\mathbf{Q}[x]$ of polynomials over $\mathbf{Q}$.

   (c) The ring $\mathbf{Z}[x]$ of polynomials over $\mathbf{Z}$.

   (d) The ring $\overline{\mathbf{Z}}$ of all algebraic integers. [Hint: Use Zorn's lemma, which implies that every ideal is contained in a maximal ideal. See, e.g., Prop 1.12 on page 589 of Artin's *Algebra*.]

4. (This problem was on the graduate qualifying exam on Tuesday.) Let $\overline{\mathbf{Z}}$ denote the subset of all elements of $\overline{\mathbf{Q}}$ that satisfy a monic polynomial with coefficients in the ring $\mathbf{Z}$ of integers. We proved in class that $\overline{\mathbf{Z}}$ is a ring.

   (a) Show that the ideals (2) and $(\sqrt{2})$ in $\overline{\mathbf{Z}}$ are distinct.

   (b) Prove that $\overline{\mathbf{Z}}$ is not Noetherian.

5. Show that neither $\mathbf{Z}[\sqrt{-6}]$ nor $\mathbf{Z}[\sqrt{5}]$ is a unique factorization domain. [Hint: Consider the factorization into irreducible elements of 6 in the first case and 4 in the second. A nonzero element a in a ring R is an *irreducible element* if it is not a unit and if whenever $a = qr$, then one of q or r is a unit.]

6. Find the ring of integers of each of the following number fields:

   (a) $\mathbf{Q}(\sqrt{-3})$,

   (b) $\mathbf{Q}(\sqrt{3})$, and

   (c) $\mathbf{Q}(\sqrt[3]{2})$.

   Do not use a computer for the first two.

7. Find the discriminants of the rings of integers of the numbers fields in the previous problem. (Do not use a computer.)

8. Let R be a finite integral domain. Prove that R is a field. [Hint: Show that if x is a nonzero element, then x has an inverse by considering powers of x.]

9. Suppose $K \subset L \subset M$ is a tower of number fields and let $\sigma : L \hookrightarrow \overline{\mathbf{Q}}$ be a field embedding of L into $\overline{\mathbf{Q}}$ that fixes K elementwise. Show that σ extends in exactly $[M : L]$ ways to a field embedding $M \hookrightarrow \overline{\mathbf{Q}}$.

10. (a) Suppose I and J are principal ideals in a ring R. Show that the set $\{ab : a \in I,\, b \in J\}$ is an ideal.

    (b) Give an example of ideals I and J in the polynomial ring $\mathbf{Q}[x, y]$ in two variables such that $\{ab : a \in I,\, b \in J\}$ is not an ideal. Your example illustrates why it is necessary to define the product of two ideals to be the ideal generated by $\{ab : a \in I,\, b \in J\}$.

    (c) Give an example of a ring of integers $\mathcal{O}_K$ of a number field, and ideals I and J such that $\{ab : a \in I,\, b \in J\}$ is not an ideal.

11. (a) Let k be a field. Prove that $k[x]$ is a Dedekind domain.

    (b) (Problem 1.12 from Swinnerton-Dyer) Let x be an indeterminate. Show that the ring $\mathbf{Z}[x]$ is Noetherian and integrally closed in its field of fractions, but is not a Dedekind domain.

12. Use Magma to write each of the following (fractional) ideals as a product of explicitly given prime ideals:

    (a) The ideal (2004) in $\mathbf{Q}(\sqrt{-1})$.

    (b) The ideals $I = (7)$ and $J = (3)$ in the ring of integers of $\mathbf{Q}(\zeta_7)$, where ζ_7 is a root of the irreducible polynomial $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$. (The field $\mathbf{Q}(\zeta_7)$ is called the 7th cyclotomic field.)

    (c) The principal fractional ideal $(3/8)$ in $\mathbf{Q}(\sqrt{5})$.

13. Suppose R is an order in the ring $\mathcal{O}_K$ of integers of a number field. (Recall that an order is a subring of finite index in $\mathcal{O}_K$.) For each of the following questions, either explain why the answer is yes for any possible order R in any $\mathcal{O}_K$, or find one specific counterexample:

(a) Is R necessarily Noetherian?

(b) Is R necessarily integrally closed in its field of fractions?

(c) Is every nonzero prime ideal of R necessarily maximal?

(d) Is it always possible to write every ideal of R uniquely as a product of prime ideals of R?

14. Let $\mathcal{O}_K$ be the ring of integers of a number field K. Prove that the group of fractional ideals of $\mathcal{O}_K$, under multiplication is (non-canonically) isomorphic to the group of positive rational numbers under multiplication.

15. (a) Suppose K is a number field of degree 2. Prove that $\mathcal{O}_K = \mathbf{Z}[a]$ for some $a \in \mathcal{O}_K$.

    (b) Prove that if K and K' are two number fields of degree 2 and $\mathrm{Disc}(\mathcal{O}_K) = \mathrm{Disc}(\mathcal{O}_{K'})$ then $K = K'$.

16. (*) Does there exist a number field K of degree 4 such that $\mathcal{O}_K \neq \mathbf{Z}[a]$ for all $a \in \mathcal{O}_K$? If so, give an explicit example.

17. Let K be the quintic number field generated by a root of $x^5+7x^4+3x^2-x+1$. Draw a diagram (be creative) that illustrates the factorization of every prime $p \in \mathbf{Z}$, with $p < 100$, in $\mathcal{O}_K$.

18. (Problem 1.9 in Swinnerton-Dyer) Show that the only solutions $x, y \in \mathbf{Z}$ to $y^2 = x^3 - 13$ are given by $x = 17, y = \pm 70$, as follows. Factor the equation $y^2 + 13 = x^3$ in the number field $\mathbf{Q}(\sqrt{-13})$, which has class number 2. Show that if x, y is an integer solution then the ideal $(y + \sqrt{-13})$ must be the cube of an ideal, and hence $y + \sqrt{-13} = (a + b\sqrt{-13})^3$; thus $1 = b(3a^2 - 13b^2)$.

19. Suppose I and J are ideals in the ring $\mathcal{O}_K$ of integers of a number field K. Does $IJ = I \cap J$? Prove or give a counterexample.

20. Let $\mathcal{O}_K$ be the ring of integers $\mathbf{Q}(\sqrt{5})$, and let

$$I = (5, 2+\sqrt{5}) \qquad \text{and} \qquad J = (209, (389+\sqrt{5})/2)$$

be integral ideals of $\mathcal{O}_K$.

(a) Find an element of $\mathcal{O}_K$ that is congruent to $\sqrt{5}$ modulo I and is congruent to $1-\sqrt{5}$ modulo J.

(b) What is the cardinality of $(\mathcal{O}_K/I) \oplus (\mathcal{O}_K/J)$?

(c) Find an element $a \in I$ such that $(a)/I$ is coprime to J.

21. Let $\mathcal{O}_K$ be the ring of integers of a number field K, and suppose K has exactly $2s$ complex embeddings. Prove that the sign of $\mathrm{Disc}(\mathcal{O}_K)$ is $(-1)^s$.

22. (*) Suppose $\mathcal{O}$ is an order in the ring of integers $\mathcal{O}_K$ of a number field. Is every ideal in $\mathcal{O}$ necessarily generated by two elements? (Answer: No. Challenge: Given an example.)

23. Find representative ideals for each element of the class group of $\mathbf{Q}(\sqrt{-23})$. Illustrate how to use the Minkowski bound to prove that your list of representatives is complete.

24. Suppose $\mathcal{O}$ is an order in the ring of integers $\mathcal{O}_K$ of a number field. Is every ideal in $\mathcal{O}$ necessarily generated by two elements?

25. Let K be a number field of degree $n > 1$ with s pairs of complex conjugate embeddings. Prove that
$$\left(\frac{\pi}{4}\right)^s \cdot \frac{n^n}{n!} > 1.$$

26. Do the exercise on page 19 of Swinnerton-Dyer, which shows that the quantity $C_{r,s}$ in the finiteness of class group theorem can be taken to be $\left(\frac{4}{\pi}\right)^s \frac{n!}{n^n}$.

27. Let α denote a root of $x^3 - x + 2$ and let $K = \mathbf{Q}(\alpha)$. Show that $\mathcal{O}_K = \mathbf{Z}[\alpha]$ and that K has class number 1 (don't just read this off from the output of the Magma commands `MaximalOrder` and `ClassNumber`). [Hint: consider the square factors of the discriminant of $x^3 - x + 2$ and show that $\frac{1}{2}(a + b\alpha + c\alpha^2)$ is an algebra integer if and only if a, b, and c are all even.]

28. If S is a closed, bounded, convex, symmetric set in $\mathbf{R}^n$ with $\mathrm{Vol}(S) \geq m2^n$, for some positive integer m, show that S contains at least $2m$ nonzero points in $\mathbf{Z}^n$.

29. Prove that any finite subgroup of the multiplicative group of a field is cyclic.

30. For a given number field K, which seems more difficult for Magma to compute, the class groups or explicit generators for the group of units? It is very difficult (but not impossible) to not get full credit on this problem. Play around with some examples, see what seems more difficult, and *justify* your response with examples. (This problem might be annoying to do using the Magma web page, since it kills your Magma job after 30 seconds. Feel free to request a binary of Magma from me, or an account on MECCAH (Mathematics Extreme Computation Cluster at Harvard).)

31. (a) Prove that there is no number field K such that $U_K \cong \mathbf{Z}/10\mathbf{Z}$.

    (b) Is there a number field K such that $U_K \cong \mathbf{Z} \times \mathbf{Z}/6\mathbf{Z}$?

32. Prove that the rank of U_K is unbounded as K varies over all number fields.

33. Let $K = \mathbf{Q}(\zeta_5)$.

    (a) Show that $r = 0$ and $s = 2$.

(b) Find explicitly generators for the group of units of U_K (you can use Magma for this).

(c) Draw an illustration of the log map $\varphi : U_K \to \mathbf{R}^2$, including the hyperplane $x_1+x_2 = 0$ and the lattice in the hyperplane spanned by the image of U_K.

34. Find the group of units of $\mathbf{Q}(\zeta_n)$ as an abstract group as a function of n. (I.e., find the number of cyclic factors and the size of the torsion subgroup. You do not have to find explicit generators!)

35. Let $K = \mathbf{Q}(a)$, where a is a root $x^3 - 3x + 1$.

    (a) Show that $r = 3$.

    (b) Find explicitly the log embedding of U_K into a 2-dimensional hyperplane in $\mathbf{R}^3$, and draw a picture.

36. Prove that if K is a quadratic field and the torsion subgroup of U_K has order bigger than 2, then $K = \mathbf{Q}(\sqrt{-3})$ or $K = \mathbf{Q}(\sqrt{-1})$.

37. A *Salem number* is a real algebraic integer, greater than 1, with the property that all of its conjugates lie on or within the unit circle, and at least one conjugate lies on the unit circle. By any method (including "google"), give two examples of Salem numbers.

38. Let $p \in \mathbf{Z}$ and let K be a number field. Show that $\mathrm{Norm}_{K/\mathbf{Q}}(p\mathcal{O}_K) = p^{[K:\mathbf{Q}]}$.

39. A totally real number field is a number field in which all embeddings into $\mathbf{C}$ have image in $\mathbf{R}$. Prove there are totally real number fields of degree p, for every prime p. [Hint: Let ζ_n denote a primitive nth root of unity. For $n \geq 3$, show that $\mathbf{Q}(\zeta_n + 1/\zeta_n)$ is totally real of degree $\varphi(n)/2$. Now prove that $\varphi(n)/2$ can be made divisible by any prime.]

40. Give an example of a number field $K/\mathbf{Q}$ and a prime p such that the e_i in the factorization of $p\mathcal{O}_K$ are not all the same.

41. Let K be a number field. Give the "simplest" proof you can think of that there are only finitely many primes that ramify (i.e., have some $e_i > 1$) in K. [The meaning of "simplest" is a matter of taste.]

42. Give examples to show that for $K/\mathbf{Q}$ a Galois extension, the quantity e can be arbirarily large and f can be arbitrarily large.

43. Suppose $K/\mathbf{Q}$ is Galois and p is a prime such that $p\mathcal{O}_K$ is also prime (i.e., p is inert in K). Show that $\mathrm{Gal}(K/\mathbf{Q})$ is a cyclic group.

44. (Problem 7, page 116, from Marcus *Number Fields*) For each of the following, find a prime p and quadratic extensions K and L of $\mathbf{Q}$ that illustrates the assertion:

(a) The prime p can be totally ramified in K and L without being totally ramified in KL.

(b) The fields K and L can each contain unique primes lying over p while KL does not.

(c) The prime p can be inert in K and L without being inert in KL.

(d) The residue field extensions of $\mathbf{F}_p$ can be trivial for K and L without being trivial for KL.

45. Let S_3 by the symmetric group on three symbols, which has order 6.

(a) Observe that $S_3 \cong D_3$, where D_3 is the dihedral group of order 6, which is the group of symmetries of an equilateral triangle.

(b) Use (45a) to write down an explicit embedding $S_3 \hookrightarrow \mathrm{GL}_2(\mathbf{C})$.

(c) Let K be the number field $\mathbf{Q}(\sqrt[3]{2}, \omega)$, where $\omega^3 = 1$ is a nontrivial cube root of unity. Show that K is a Galois extension with Galois group isomorphic to S_3.

(d) We thus obtain a 2-dimensional irreducible complex Galois representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Gal}(K/\mathbf{Q}) \cong S_3 \subset \mathrm{GL}_2(\mathbf{C}).$$

Compute a representative matrix of Frob_p and the characteristic polynomial of Frob_p for $p = 5, 7, 11, 13$.

46. Let $K = \mathbf{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7})$. Show that K is Galois over $\mathbf{Q}$, compute the Galois group of K, and compute Frob_{37}.

47. Let k be any field. Prove that the only nontrivial valuations on $k(t)$ which are trivial on k are equivalent to the valuation (15.3.3) or (15.3.4) of page 115.

48. A field with the topology induced by a valuation is a topological field, i.e., the operations sum, product, and reciprocal are continuous.

49. Give an example of a non-archimedean valuation on a field that is not discrete.

50. Prove that the field $\mathbf{Q}_p$ of p-adic numbers is uncountable.

51. Prove that the polynomial $f(x) = x^3 - 3x^2 + 2x + 5$ has all its roots in $\mathbf{Q}_5$, and find the 5-adic valuations of each of these roots. (You might need to use Hensel's lemma, which we don't discuss in detail in this book. See [Cas67, App. C].)

52. In this problem you will compute an example of weak approximation, like I did in the Example 16.3.3. Let $K = \mathbf{Q}$, let $|\cdot|_7$ be the 7-adic absolute value, let $|\cdot|_{11}$ be the 11-adic absolute value, and let $|\cdot|_\infty$ be the usual archimedean absolute value. Find an element $b \in \mathbf{Q}$ such that $|b - a_i|_i < \frac{1}{10}$, where $a_7 = 1$, $a_{11} = 2$, and $a_\infty = -2004$.

53. Prove that -9 has a cube root in $\mathbf{Q}_{10}$ using the following strategy (this is a special case of Hensel's Lemma, which you can read about in an appendix to Cassel's article).

    (a) Show that there is an element $\alpha \in \mathbf{Z}$ such that $\alpha^3 \equiv 9 \pmod{10^3}$.

    (b) Suppose $n \geq 3$. Use induction to show that if $\alpha_1 \in \mathbf{Z}$ and $\alpha^3 \equiv 9 \pmod{10^n}$, then there exists $\alpha_2 \in \mathbf{Z}$ such that $\alpha_2^3 \equiv 9 \pmod{10^{n+1}}$. (Hint: Show that there is an integer b such that $(\alpha_1 + b \cdot 10^n)^3 \equiv 9 \pmod{10^{n+1}}$.)

    (c) Conclude that 9 has a cube root in $\mathbf{Q}_{10}$.

54. Compute the first 5 digits of the 10-adic expansions of the following rational numbers:

$$\frac{13}{2}, \quad \frac{1}{389}, \quad \frac{17}{19}, \quad \text{the 4 square roots of 41.}$$

55. Let $N > 1$ be an integer. Prove that the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} n! = 1! - 2! + 3! - 4! + 5! - 6! + \cdots .$$

    converges in $\mathbf{Q}_N$.

56. Prove that -9 has a cube root in $\mathbf{Q}_{10}$ using the following strategy (this is a special case of "Hensel's Lemma").

    (a) Show that there is $\alpha \in \mathbf{Z}$ such that $\alpha^3 \equiv 9 \pmod{10^3}$.

    (b) Suppose $n \geq 3$. Use induction to show that if $\alpha_1 \in \mathbf{Z}$ and $\alpha^3 \equiv 9 \pmod{10^n}$, then there exists $\alpha_2 \in \mathbf{Z}$ such that $\alpha_2^3 \equiv 9 \pmod{10^{n+1}}$. (Hint: Show that there is an integer b such that $(\alpha_1 + b10^n)^3 \equiv 9 \pmod{10^{n+1}}$.)

    (c) Conclude that 9 has a cube root in $\mathbf{Q}_{10}$.

57. Let $N > 1$ be an integer.

    (a) Prove that $\mathbf{Q}_N$ is equipped with a natural ring structure.

    (b) If N is prime, prove that $\mathbf{Q}_N$ is a field.

58. (a) Let p and q be distinct primes. Prove that $\mathbf{Q}_{pq} \cong \mathbf{Q}_p \times \mathbf{Q}_q$.

    (b) Is $\mathbf{Q}_{p^2}$ isomorphic to either of $\mathbf{Q}_p \times \mathbf{Q}_p$ or $\mathbf{Q}_p$?

59. Prove that every finite extension of $\mathbf{Q}_p$ "comes from" an extension of $\mathbf{Q}$, in the following sense. Given an irreducible polynomial $f \in \mathbf{Q}_p[x]$ there exists an irreducible polynomial $g \in \mathbf{Q}[x]$ such that the fields $\mathbf{Q}_p[x]/(f)$ and $\mathbf{Q}_p[x]/(g)$ are isomorphic. [Hint: Choose each coefficient of g to be sufficiently close to the corresponding coefficient of f, then use Hensel's lemma to show that g has a root in $\mathbf{Q}_p[x]/(f)$.]

60. Find the 3-adic expansion to precision 4 of each root of the following polynomial over $\mathbf{Q}_3$:

$$f = x^3 - 3x^2 + 2x + 3 \in \mathbf{Q}_3[x].$$

Your solution should conclude with three expressions of the form

$$a_0 + a_1 \cdot 3 + a_2 \cdot 3^2 + a_3 \cdot 3^3 + O(3^4).$$

61. (a) Find the normalized Haar measure of the following subset of $\mathbf{Q}_7^+$:

$$U = B\left(28, \frac{1}{50}\right) = \left\{x \in \mathbf{Q}_7 : |x - 28| < \frac{1}{50}\right\}.$$

(b) Find the normalized Haar measure of the subset $\mathbf{Z}_7^*$ of $\mathbf{Q}_7^*$.

62. Suppose that K is a finite extension of $\mathbf{Q}_p$ and L is a finite extension of $\mathbf{Q}_q$, with $p \neq q$ and assume that K and L have the same degree. Prove that there is a polynomial $g \in \mathbf{Q}[x]$ such that $\mathbf{Q}_p[x]/(g) \cong K$ and $\mathbf{Q}_q[x]/(g) \cong L$. [Hint: Combine your solution to 59 with the weak approximation theorem.]

63. Prove that the ring C defined in Section 9 really is the tensor product of A and B, i.e., that it satisfies the defining universal mapping property for tensor products. Part of this problem is for you to look up a functorial definition of tensor product.

64. Find a zero divisor pair in $\mathbf{Q}(\sqrt{5}) \otimes_{\mathbf{Q}} \mathbf{Q}(\sqrt{5})$.

65. (a) Is $\mathbf{Q}(\sqrt{5}) \otimes_{\mathbf{Q}} \mathbf{Q}(\sqrt{-5})$ a field?

(b) Is $\mathbf{Q}(\sqrt[4]{5}) \otimes_{\mathbf{Q}} \mathbf{Q}(\sqrt[4]{-5}) \otimes_{\mathbf{Q}} \mathbf{Q}(\sqrt{-1})$ a field?

66. Suppose ζ_5 denotes a primitive 5th root of unity. For any prime p, consider the tensor product $\mathbf{Q}_p \otimes_{\mathbf{Q}} \mathbf{Q}(\zeta_5) = K_1 \oplus \cdots \oplus K_{n(p)}$. Find a simple formula for the number $n(p)$ of fields appearing in the decomposition of the tensor product $\mathbf{Q}_p \otimes_{\mathbf{Q}} \mathbf{Q}(\zeta_5)$. To get full credit on this problem your formula must be correct, but you do *not* have to prove that it is correct.

67. Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms on a finite-dimensional vector space V over a field K (with valuation $|\cdot|$). Carefully prove that the topology induced by $\|\cdot\|_1$ is the same as that induced by $\|\cdot\|_2$.

68. Suppose K and L are number fields (i.e., finite extensions of $\mathbf{Q}$). Is it possible for the tensor product $K \otimes_{\mathbf{Q}} L$ to contain a nilpotent element? (A nonzero element a in a ring R is *nilpotent* if there exists $n > 1$ such that $a^n = 0$.)

69. Let K be the number field $\mathbf{Q}(\sqrt[5]{2})$.

(a) In how many ways does the 2-adic valuation $|\cdot|_2$ on $\mathbf{Q}$ extend to a valuation on K?

(b) Let $v = |\cdot|$ be a valuation on K that extends $|\cdot|_2$. Let K_v be the completion of K with respect to v. What is the residue class field $\mathbf{F}$ of K_v?

70. Prove that the product formula holds for $\mathbf{F}(t)$ similar to the proof we gave in class using Ostrowski's theorem for $\mathbf{Q}$. You may use the analogue of Ostrowski's theorem for $\mathbf{F}(t)$, which you had on a previous homework assignment. (Don't give a measure-theoretic proof.)

71. Prove Theorem 20.3.5, that "The global field K is discrete in $\mathbb{A}_K$ and the quotient $\mathbb{A}_K^+/K^+$ of additive groups is compact in the quotient topology." in the case when K is a finite extension of $\mathbf{F}(t)$, where $\mathbf{F}$ is a finite field.

Bibliography

[ABC$^+$] B. Allombert, K. Belabas, H. Cohen, X. Roblot, and I. Zakharevitch, `PARI/GP`, `http://pari.math.u-bordeaux.fr/`.

[Art59] E. Artin, *Theory of algebraic numbers*, Notes by Gerhard Würges from lectures held at the Mathematisches Institut, Göttingen, Germany, in the Winter Semester, vol. 1956/7, George Striker, Schildweg 12, Göttingen, 1959. MR 24 #A1884

[Art91] M. Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991. MR 92g:00001

[AT90] E. Artin and J. Tate, *Class field theory*, second ed., Advanced Book Classics, Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1990. MR 91b:11129

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[Cas67] J. W. S. Cassels, *Global fields*, Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965), Thompson, Washington, D.C., 1967, pp. 42–84.

[Cas91] ———, *Lectures on elliptic curves*, London Mathematical Society Student Texts, vol. 24, Cambridge University Press, Cambridge, 1991. MR 92k:11058

[Coh93] H. Cohen, *A course in computational algebraic number theory*, Springer-Verlag, Berlin, 1993. MR 94i:11105

[Cp86] J. W. S. Cassels and A. Fröhlich (eds.), *Algebraic number theory*, London, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1986, Reprint of the 1967 original.

[EH00] D. Eisenbud and J. Harris, *The geometry of schemes*, Springer-Verlag, New York, 2000. MR 2001d:14002

[Fre94] G. Frey (ed.), *On Artin's conjecture for odd 2-dimensional representations*, Springer-Verlag, Berlin, 1994, 1585. MR 95i:11001

[Iwa53] K. Iwasawa, *On the rings of valuation vectors*, Ann. of Math. (2) **57** (1953), 331–356. MR 14,849a

[Lan64] S. Lang, *Algebraic numbers*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-Palo Alto-London, 1964. MR 28 #3974

[Lan80] R. P. Langlands, *Base change for* GL(2), Princeton University Press, Princeton, N.J., 1980.

[Len02] H. W. Lenstra, Jr., *Solving the Pell equation*, Notices Amer. Math. Soc. **49** (2002), no. 2, 182–192. MR 2002i:11028

[LL93] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The development of the number field sieve*, Springer-Verlag, Berlin, 1993. MR 96m:11116

[Mah64] K. Mahler, *Inequalities for ideal bases in algebraic number fields*, J. Austral. Math. Soc. **4** (1964), 425–448. MR 31 #1243

[SD01] H. P. F. Swinnerton-Dyer, *A brief guide to algebraic number theory*, London Mathematical Society Student Texts, vol. 50, Cambridge University Press, Cambridge, 2001. MR 2002a:11117

[Ser73] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[Wei82] A. Weil, *Adeles and algebraic groups*, Progress in Mathematics, vol. 23, Birkhäuser Boston, Mass., 1982, With appendices by M. Demazure and Takashi Ono. MR 83m:10032

Index

23 Visibility of Mordell-Weil Groups

Visibility of Mordell-Weil Groups

William A. Stein[1]

Abstract.

We introduce a notion of visibility for Mordell-Weil groups, make a conjecture about visibility, and support it with theoretical evidence and data. These results shed new light on relations between Mordell-Weil and Shafarevich-Tate groups.

1 Introduction

Consider an exact sequence $0 \to C \to B \to A \to 0$ of abelian varieties over a number field K. We say that the covering $B \to A$ is *optimal* since its kernel C is connected. As introduced in [LT58], there is a corresponding long exact sequence of Galois cohomology

$$0 \to C(K) \to B(K) \to A(K) \xrightarrow{\delta} \mathrm{H}^1(K, C) \to \mathrm{H}^1(K, B) \to \mathrm{H}^1(K, A) \to \cdots$$

The study of the Mordell-Weil group $A(K)$ is central in arithmetic geometry. For example, the Birch and Swinnerton-Dyer conjecture (BSD conjecture) of [Bir71, Tat66]), which is one of the Clay Math Problems [Wil00], asserts that the rank r of $A(K)$ equals the ordering vanishing of $L(A, s)$ at $s = 1$, and also gives a conjectural formula for $L^{(r)}(A, 1)$ in terms of the invariants of A.

The group $\mathrm{H}^1(K, A)$ is also of interest in connection with the BSD conjecture, because it contains the Shafarevich-Tate group

$$Ш(A/K) = \mathrm{Ker}\left(\mathrm{H}^1(K, A) \to \bigoplus_v \mathrm{H}^1(K_v, A)\right),$$

which is the most mysterious object appearing in the BSD conjecture.

[1] This material is based upon work supported by the National Science Foundation under Grant No. 0400386.

Definition 1.0.1 (Visibility). The *visible subgroup* of $\mathrm{H}^1(K,C)$ relative to the embedding $C \hookrightarrow B$ is

$$\begin{aligned}\mathrm{Vis}_B\,\mathrm{H}^1(K,C) &= \mathrm{Ker}(\mathrm{H}^1(K,C) \to \mathrm{H}^1(K,B))\\ &\cong \mathrm{Coker}(B(K) \to A(K)).\end{aligned}$$

The *visible quotient* of $A(K)$ relative to the optimal covering $B \to A$ is

$$\begin{aligned}\mathrm{Vis}^B(A(K)) &= \mathrm{Coker}(B(K) \to A(K))\\ &\cong \mathrm{Vis}_B\,\mathrm{H}^1(K,C).\end{aligned}$$

We say an abelian variety over $\mathbb{Q}$ is *modular* if it is a quotient of the modular Jacobian $J_1(N) = \mathrm{Jac}(X_1(N))$, for some N. For example, every elliptic curve over $\mathbb{Q}$ is modular [BCDT01].

This paper gives evidence toward the following conjecture that Mordell-Weil groups should give rise to many visible Shafarevich-Tate groups.

Conjecture 1.0.2. *Let A be an abelian variety over a number field K. For every integer m, there is an exact sequence $0 \to C \to B \to A \to 0$ such that:*

1. *The image of $B(K)$ in $A(K)$ is contained in $mA(K)$, so $A(K)/mA(K)$ is a quotient of $\mathrm{Vis}^B(A(K))$.*
2. *If $K = \mathbb{Q}$ and A is modular, then B is modular.*
3. *The rank of C is zero.*
4. *We have $\mathrm{Coker}(B(K) \to A(K)) \subset Ш(C/K)$, via the connecting homomorphism.*

In [Ste04] we give the following computational evidence for this conjecture.

Theorem 1.0.3. *Let E be the rank 1 elliptic curve $y^2+y = x^3-x$ of conductor 37. Then Conjecture 1.0.2 is true for all primes $m = p < 25000$ with $p \neq 2, 37$.*

Let $f = \sum a_n q^n$ be the newform associated to the elliptic curve E of Theorem 1.0.3. Suppose p is one of the primes in the theorem. Then there is an $\ell \equiv 1 \pmod{p}$ and a surjective Dirichlet character $\chi : (\mathbb{Z}/\ell\mathbb{Z})^* \to \mu_p$ such that $L(f \otimes \chi, 1) \neq 0$. The C of the theorem is, up to isogeny, the abelian variety associated to f^χ, which has dimension $p-1$.

In general, we expect the construction of [Ste04] to work for any elliptic curve and any odd prime p of good reduction. The main obstruction to proving that it does work is proving a nonvanishing result for the special values $L(f^\chi, 1)$. In [Ste04], we verified this hypothesis using modular symbols for $p < 25000$.

A surprising observation that comes out of the construction of [Ste04] is that $\#Ш(A) = p \cdot n^2$, where n^2 is an integer square. We thus obtained the first ever examples of abelian varieties whose Shafarevich-Tate groups have order neither a square nor twice a square.

1.1 Contents

In Section 2, we give a brief review of results about visibility of Shafarevich-Tate groups. In Section 3, we give evidence for Conjecture 1.0.2 using results of Kato, Lichtenbaum and Mazur. Section 4 is about bounding the dimension of the abelian varieties in which Mordell-Weil groups are visible. We prove that every Mordell-Weil group is 2-visible relative to an abelian surface. In Section 5, we describe how to construct visible quotients of Mordell-Weil groups, and carry out a computational study of relations between Mordell-Weil groups of elliptic curves and the arithmetic of rank 0 factors of $J_0(N)$.

1.2 Acknowledgement

The author had extremely helpful conversations with Barry Mazur and Grigor Grigorov. Proposition 3.0.3 was proved jointly with Ken Ribet. The author was supported by NSF grant DMS-0400386. He used Magma [BCP97] and Python [Ros] for computing the data in Section 5.

2 Review of Visibility of Galois Cohomology

In this section, we briefly review visibility of elements of $\mathrm{H}^1(K, A)$, as first introduced by Mazur in [CM00, Maz99], and later developed by Agashe and Stein in [Aga99a, AS05, AS02]. We describe two basic results about visibility, and in Section 2.2 we discuss modularity of elements of $\mathrm{H}^1(K, A)$.

Consider an exact sequence of abelian varieties

$$0 \to A \to B \to C \to 0$$

over a number field K. Elements of $\mathrm{H}^0(K, C)$ are points, so they are relatively easy to "visualize", but elements of $\mathrm{H}^1(K, A)$ are mysterious.

There is a geometric way to view elements of $\mathrm{H}^1(K, A)$. The Weil-Chatalet group $\mathrm{WC}(A/K)$ of A over K is the group of isomorphism classes of principal homogeneous spaces for A, where a principal homogeneous space is a variety X and a simply-transitive action $A \times X \to X$. Thus X is a twist of A as a variety, but $X(K) = \emptyset$, unless X is isomorphic to A. Also, the elements of $Ш(A)$ correspond to the classes of X that have a K_v-rational point for all places v. By [LT58, Prop. 4], there is an isomorphism between $\mathrm{H}^1(K, A)$ and $\mathrm{WC}(A/K)$.

In [CM00], Mazur introduced the visible subgroup of H^1 as in Definition 1.0.1 in order to help unify diverse constructions of principal homogeneous spaces. Many papers were subsequently written about visibility, including [Aga99b, Maz99, Kle01, AS02, MO03, DWS03, AS05, Dum01].

Remark 2.0.1. Note that $\mathrm{Vis}_B\,\mathrm{H}^1(K, A)$ depends on the embedding of A into B. For example, if $B = B_1 \times A$. Then there could be nonzero visible elements if A is embedded into the first factor, but there will be no nonzero visible elements if A is embedded into the second factor.

A connection between visibility and $\mathrm{WC}(A/K)$ is as follows. Suppose

$$0 \to A \to B \xrightarrow{\pi} C \to 0$$

is an exact sequence of abelian varieties and that $c \in \mathrm{H}^1(K, A)$ is visible in B. Thus there exists $x \in C(K)$ such that $\delta(x) = c$, where $\delta : C(K) \to \mathrm{H}^1(K, A)$ is the connecting homomorphism. Then $X = \pi^{-1}(x) \subset B$ is a translate of A in B, so the group law on B gives X the structure of principal homogeneous space for A, and this homogeneous space in $\mathrm{WC}(A/K)$ corresponds to c.

2.1 Basic Facts

Two basic facts about visibility are that the visible subgroup of $\mathrm{H}^1(K, A)$ in B is finite, and that each element of $\mathrm{H}^1(K, A)$ is visible in some B.

Lemma 2.1.1. *The group* $\mathrm{Vis}_B\,\mathrm{H}^1(K, A)$ *is finite.*

Proof. Let $C = B/A$. By the Mordell-Weil theorem $C(K)$ is finitely generated. The group $\mathrm{Vis}_B\,\mathrm{H}^1(K, A)$ is a homomorphic image of $C(K)$ so it is finitely generated. On the other hand, it is a subgroup of $\mathrm{H}^1(K, A)$, so it is a torsion group. But a finitely generated torsion abelian group is finite. □

Proposition 2.1.2. *Let $c \in \mathrm{H}^1(K, A)$. Then there exists an abelian variety B and an embedding $A \hookrightarrow B$ such that c is visible in B. Moreover, B can be chosen to be a twist of a power of A.*

Proof. See [AS02, Prop. 1.3] for a cohomological proof or [JS05, §5] for an equivalent geometric proof. Johan de Jong also proved that everything is visible somewhere in the special case $\dim(A) = 1$ using Azumaya algebras, Néron models, and étale cohomology, as explained in [CM00, pg. 17–18], but his proof gives no (obvious) specific information about the structure of B. □

2.2 Modularity

Usually one focuses on visibility of elements in $\text{Ш}(A) \subset \mathrm{H}^1(K, A)$. The papers [CM00, AS02, AS05] contain a number of results about visibility in various special cases, and tables involving elliptic curves and modular abelian varieties.

For example, if $A \subset J_0(389)$ is the 20-dimensional simple newform abelian variety, then we show that

$$\mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \cong E(\mathbb{Q})/5E(\mathbb{Q}) \subset \text{Ш}(A),$$

where E is the elliptic curve of conductor 389. The divisibility $5^2 \mid \#\text{Ш}(A)$ is as predicted by the BSD conjecture. The paper [AS05] contains a few dozen other examples like this; in most cases, explicit computational construction of the Shafarevich-Tate group seems hopeless using any other known techniques.

The author has conjectured that if A is a modular abelian variety, then every element of $\text{Ш}(A)$ is modular, i.e., visible in a modular abelian variety. It is a theorem that if $c \in \text{Ш}(A)$ has order either 2 or 3 and A is an elliptic curve, then c is modular (see [JS05]).

3 Results Toward Conjecture 1.0.2

The main result of this section is a proof of parts 1 and 2 of Conjecture 1.0.2 for elliptic curves over $\mathbb{Q}$. We prove more generally that Mazur's conjecture on finite generatedness of Mordell-Weil groups over cyclotomic $\mathbb{Z}_p$-extensions implies part 1 of Conjecture 1.0.2. Then we observe that for elliptic curves over $\mathbb{Q}$, Mazur's conjecture is known, and prove that the abelian varieties that appear in our visibility construction are modular, so parts 1 and 2 of Conjecture 1.0.2 are true for elliptic curves over $\mathbb{Q}$.

For a prime p, the cyclotomic $\mathbb{Z}_p$-extension of $\mathbb{Q}$ is an extension $\mathbb{Q}_{p^\infty}$ of $\mathbb{Q}$ with Galois group $\mathbb{Z}_p$; also $\mathbb{Q}_{p^\infty}$ is contained in the cyclotomic field $\mathbb{Q}(\mu_{p^\infty})$. We let $\mathbb{Q}_{p^n}$ denote the unique subfield of $\mathbb{Q}_{p^\infty}$ of degree p^n over $\mathbb{Q}$. If K is an arbitrary number field, the cyclotomic $\mathbb{Z}_p$-extension of K is $K_{p^\infty} = K\cdot\mathbb{Q}_{p^\infty}$. We denote by K_{p^n} the unique subfield of K_{p^∞} of degree p^n over K. The extension K_{p^∞} of K decomposes as a tower

$$K = K_{p^0} \subset K_{p^1} \subset \cdots \subset K_{p^n} \subset \cdots \subset K_{p^\infty} = \bigcup_{n=0}^{\infty} K_{p^n}.$$

Mazur hints at the following conjecture in [Maz78] and [RM05, §3]:

Conjecture 3.0.1 (Mazur). *If A is an abelian variety over a number field K and p is a prime, then $A(K_{p^\infty})$ is a finitely generated abelian group.*

Let L/K be a finite extension of number fields and A an abelian variety over K. In much of the rest of this paper we will use the *restriction of scalars* $R = \mathrm{Res}_{L/K}(A_L)$ of A viewed as an abelian variety over L. Thus R is an abelian variety over K of dimension $[L : K]$, and R represents the following functor on the category of K-schemes:

$$S \mapsto E_L(S_L).$$

If L/K is Galois, then we have an isomorphism of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-modules

$$R(\overline{\mathbb{Q}}) = A(\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Z}[\mathrm{Gal}(L/K)],$$

where $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ acts on $\sum P_\sigma \otimes \sigma$ by

$$\tau\left(\sum P_\sigma \otimes \sigma\right) = \sum \tau(P_\sigma) \otimes \tau_{|L} \cdot \sigma,$$

where $\tau_{|L}$ is the image of τ in $\mathrm{Gal}(L/K)$.

Theorem 3.0.2. *Conjecture 3.0.1 implies part 1 of Conjecture 1.0.2. More precisely, if A/K is an abelian variety, m is a positive integer, and $A(K_{p^\infty})$ is finitely generated for each $p \mid m$, then there is an optimal covering of the form $B = \mathrm{Res}_{L/K}(A_L) \to A$ such that L is abelian over K and the image of $B(K)$ in $A(K)$ is contained in $mA(K)$.*

Proof. Fix a prime $p \mid m$. Let $M = K_{p^\infty}$. Because $A(M)$ is finitely generated, some finite set of generators must be in a single sufficiently large $A(K_{p^n})$, and for this n we have $A(M) = A(K_{p^n})$. For any integer $j > 0$ let

$$R_j = \operatorname{Res}_{K_{p^j}/K}(A_{K_{p^j}}).$$

Then, as explained in [Ste04], the trace map induces an exact sequence

$$0 \to B_j \to R_j \xrightarrow{\pi_j} A \to 0,$$

with B_j an abelian variety. Then for any $j \geq n$, $A(K_{p^j}) = A(K_{p^n})$, so

$$\begin{aligned}
\operatorname{Vis}^{B_j}(A(K)) &\cong A(K)/\pi_j(R_j(K)) \\
&= A(K)/\operatorname{Tr}_{K_{p^j}/K}(A(K_{p^j})) \\
&= A(K)/\operatorname{Tr}_{K_{p^n}/K}(\operatorname{Tr}_{K_{p^j}/K_{p^n}}(A(K_{p^j}))) \\
&= A(K)/\operatorname{Tr}_{K_{p^n}/K}(\operatorname{Tr}_{K_{p^j}/K_{p^n}}(A(K_{p^n}))) \\
&= A(K)/\operatorname{Tr}_{K_{p^n}/K}(p^{j-n}A(K_{p^n})) \\
&= A(K)/p^{j-n}\operatorname{Tr}_{K_{p^n}/K}(A(K_{p^n})) \\
&\to A(K)/p^{j-n}A(K),
\end{aligned}$$

where the last map is surjective since

$$\operatorname{Tr}_{K_{p^n}/K}(A(K_{p^n})) \subset A(K).$$

Arguing as above, for each prime $p \mid m$, we find an extension L_p of K of degree a power of p such that $\operatorname{Tr}_{L_p/K}(A(L_p)) \subset p^{\nu_p}A(K)$, where $\nu_p = \operatorname{ord}_p(m)$. Let L be the compositum of the fields L_p. Then for each $p \mid m$,

$$\operatorname{Tr}_{L/K}(A(L)) = \operatorname{Tr}_{L_p/K}(\operatorname{Tr}_{L/L_p}(A(L))) \subset \operatorname{Tr}_{L_p/K}(A(L_p)) \subset p^{\nu_p}A(K).$$

Thus

$$\operatorname{Tr}_{L/K}(A(L)) \subset \bigcap_{p \mid m} p^{\nu_p}A(K) = mA(K), \tag{1}$$

where for the last equality we view $A(K)$ as a finite direct sum of cyclic groups.

Let $R = \operatorname{Res}_{L/K}(A_L)$. Then trace induces an optimal cover $R \to A$, and (1) implies that we have the required surjective map

$$\operatorname{Vis}^R(A(K)) = A(K)/\operatorname{Tr}_{L/K}(A(L)) \to A(K)/mA(K).$$

□

We will next prove parts 1 and 2 of Conjecture 1.0.2 for elliptic curves over $\mathbb{Q}$ by observing that Conjecture 3.0.1 is a theorem of Kato in this case. We first prove a modularity property for restriction of scalars. Recall that a modular abelian variety is a quotient of $J_1(N)$.

Proposition 3.0.3. *If A is a modular abelian variety over $\mathbb{Q}$ and K is an abelian extension of $\mathbb{Q}$, then $\mathrm{Res}_{K/\mathbb{Q}}(A_K)$ is also a modular abelian variety.*

Proof. Since A is modular, A is isogenous to a product of abelian varieties A_f attached to newforms in $S_2(\Gamma_1(N))$, for various N. Since the formation of restriction of scalars commutes with products, it suffices to prove the proposition under the hypothesis that $A = A_f$ for some newform f. Let $R = \mathrm{Res}_{K/\mathbb{Q}}(A_f)$. As discussed in [Mil72, pg. 178], for any prime p there is an isomorphism of $\mathbb{Q}_p$-adic Tate modules

$$V_p(R) \cong \mathrm{Ind}_{G_K}^{G_{\mathbb{Q}}} V_p(A_K).$$

The induced representation on the right is the direct sum of twists of $V_p(A_K)$ by characters of $\mathrm{Gal}(K/\mathbb{Q})$. This is isomorphic to the $\mathbb{Q}_p$-adic Tate module of some abelian variety $P = \prod_\chi A_{g^\chi}$, where χ runs through certain Dirichlet characters corresponding to the abelian extension $K/\mathbb{Q}$, and g runs through certain $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f, and g^χ denotes the twist of g by χ. Falting's theorem (see e.g., [Fal86, §5]) then gives us the desired isogeny $R \to P$.

It is not necessary to use the full power of Falting's theorem to prove this proposition, since Ribet [Rib80] gave a more elementary proof of Falting's theorem in the case of modular abelian varieties. However, we must work some to apply Ribet's theorem, since we do not know yet that R is modular.

Let R and P be as above. Over $\overline{\mathbb{Q}}$, the abelian variety A is isogenous to a power of a simple abelian variety B, since if more than one non-isogenous simple occurred in the decomposition of $A/\overline{\mathbb{Q}}$, then $\mathrm{End}(A/\overline{\mathbb{Q}})$ would not be a matrix ring over a (possibly skew) field (see [Rib92, §5]). For any character χ, by the (3) $\implies$ (2) assertion of [Rib80, Thm. 4.7], the abelian varieties A_f and A_{f^χ} are isogenous over $\overline{\mathbb{Q}}$ to powers of the same abelian variety A', hence to powers of the simple B. A basic property of restriction of scalars is that R_K is isomorphic to a power of $(A_f)_K$, hence R_K is isogenous over $\overline{\mathbb{Q}}$ to a power of B. Thus R and P are both isogenous over $\overline{\mathbb{Q}}$ to a power of B, so R is isogenous to P over $\overline{\mathbb{Q}}$, since they have the same dimension, as their Tate modules are isomorphic. Let L be a Galois number field over which such an isogeny is defined. Consider the natural $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-equivariant inclusion

$$\mathrm{Hom}(R_{\mathbb{Q}}, P_{\mathbb{Q}}) \otimes_{\mathbb{Q}_p} \hookrightarrow \mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}(V_p(R), V_p(P)). \tag{2}$$

By Ribet's proof of the Tate conjecture for modular abelian varieties [Rib80], the inclusion

$$\mathrm{Hom}(R_L, P_L) \otimes_{\mathbb{Q}_p} \hookrightarrow \mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}(V_p(R), V_p(P)) \tag{3}$$

is an isomorphism, since there is an isogeny $P_L \to R_L$ and P is modular. But then (2) must also be an isomorphism, since (2) is the result of taking $\mathrm{Gal}(L/\mathbb{Q})$-invariants of both sides of (3).

By construction of P, there is an isomorphism $V_p(R) \cong V_p(P)$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules, so by (2) there is an isomorphism in $\mathrm{Hom}(R_{\mathbb{Q}}, P_{\mathbb{Q}}) \otimes \mathbb{Q}_p$. Thus there is a $\mathbb{Q}_p$-linear combination of elements of $\mathrm{Hom}(R_{\mathbb{Q}}, P_{\mathbb{Q}})$ that has nonzero determinant. However, if a $\mathbb{Q}_p$-linear combination of matrices has nonzero determinant, then some $\mathbb{Q}$-linear combination does, since the determinant is a polynomial function of the coefficients and $\mathbb{Q}$ is dense in $\mathbb{Q}_p$. Thus there is an isogeny $R \to P$ defined over $\mathbb{Q}$, so R is modular. □

Corollary 3.0.4. *Parts 1 and 2 of Conjecture 1.0.2 are true for every elliptic curve E over $\mathbb{Q}$.*

Proof. Suppose p is a prime, and let $\mathbb{Q}_{p^\infty}$ be the cyclotomic $\mathbb{Z}_p$ extension of $\mathbb{Q}$. By [BCDT01], E is a modular elliptic curve, so Rohrlich [Roh84] implies that all but finitely many special values $L(E, \chi, 1)$ are nonzero, where χ runs over all Dirichlet characters of p-power order. Kato proved (see, e.g., [Kat04, Sch98]) that if $L(E, \chi, 1) \neq 0$, then the χ part of $E(\mathbb{Q}_{p^\infty}) \otimes \mathbb{Q}$ vanishes. Combining these results, we see that $E(\mathbb{Q}_{p^\infty})$ is finitely generated, so we can apply Theorem 3.0.2 to conclude that if $x \in E(\mathbb{Q})$ and $m \mid \mathrm{order}(x)$, then x is m-visible relative to an optimal cover of E by a restriction of scalars B from an abelian extension. Then Proposition 3.0.3 implies that B is modular. □

4 The Visibility Dimension

The visibility dimension is analogous to the visibility dimension for elements of $\mathrm{H}^1(K, A)$ introduced in [AS02, §2]. We prove below that elements of order 2 in Mordell-Weil groups of elliptic curves over $\mathbb{Q}$ are 2-visible relative to an abelian surface. Along the way, we make a general conjecture about stability of rank and show that it implies a general bound on the visibility dimension.

Definition 4.0.5 (Visibility Dimension). Let A be an abelian variety over a number field K and suppose m is an integer. Then A has *m-visibility dimension n* if there is an optimal cover $B \to A$ with $n = \dim(B)$ and the image of $B(K)$ in $A(K)$ is contained in $mA(K)$, so $A(K)/mA(K)$ is a quotient of $\mathrm{Vis}^B(A(K))$.

The following rank-stability conjecture is motivated by its usefulness for proving a result about m-visibility.

Conjecture 4.0.6. *Suppose A is an abelian variety over a number field K, that L is a finite extension of K, and $m > 0$ is an integer. Then there is an extension M of K of degree m such that* $\mathrm{rank}(A(K)) = \mathrm{rank}(A(M))$ *and $M \cap L = K$.*

The following proposition describes how Conjecture 4.0.6 can be used to find an extension where the index of $A(K)$ in $A(M)$ is coprime to m.

Proposition 4.0.7. *Let A be an abelian variety over a number field K and suppose m is a positive integer. If Conjecture 4.0.6 is true for A and m, then there is an extension M of K of degree m such that $A(M)/A(K)$ is of order coprime to m.*

Proof. Choose a finite set $P_1, \dots, P_n$ of generators for $A(K)$. Let

$$L = K\left(\frac{1}{m}P_1, \dots, \frac{1}{m}P_n\right)$$

be the extension of K generated by *all* mth roots of each P_i. Since the set of mth roots of a point is closed under the action of $\mathrm{Gal}(\overline{K}/K)$, the extension L/K is Galois. Note also that the m torsion of A is defined over L, since the differences of conjugates of a given $\frac{1}{m}P_i$ are exactly the elements of $A[m]$. Let S be the set of primes of K that ramify in L.

By our hypothesis that Conjecture 4.0.6 is true for A and m, there is an extension M of K of degree m such that

$$\mathrm{rank}(A(K)) = \mathrm{rank}(A(M))$$

and $M \cap L = K$. In particular, $C = A(M)/A(K)$ is a finite group. Suppose, for the sake of contradiction, that $\gcd(m, \#C) \neq 1$, so there is some prime divisor $p \mid m$ and an element $[Q] \in C$ of exact order p. Here $Q \in A(M)$ is such that $pQ \in A(K)$ but $Q \notin A(K)$. Because $P_1, \dots, P_n$ generate $A(K)$ and $pQ \in A(K)$, there are integers $a_1, \dots a_n$ such that

$$pQ = \sum_{i=1}^{n} a_i P_i.$$

Then for any fixed choice of the $\frac{1}{p}P_i$, we have

$$Q - \sum_{i=1}^{n} a_i \cdot \frac{1}{p}P_i \in A[p],$$

since

$$p\left(Q - \sum_{i=1}^{n} a_i \cdot \frac{1}{p}P_i\right) = pQ - \sum_{i=1}^{n} a_i \cdot P_i = 0.$$

Thus $Q \in A(L)$. But then since $L \cap M = K$, so we obtain a contradiction from

$$Q \in A(L) \cap A(M) = A(K).$$

□

With Proposition 4.0.7 in hand, we show that Conjecture 4.0.6 bounds the visibility dimension of Mordell-Weil groups. In particular, we see that Conjecture 4.0.6 implies that for any abelian variety A over a number field K, and any m, there is an embedding $A(K)/mA(K) \hookrightarrow \mathrm{H}^1(K, C)$ coming from a δ map, where C is an abelian variety over K of rank 0.

THEOREM 4.0.8. *Let A be an abelian variety over a number field K and suppose m is a positive integer. If Conjecture 4.0.6 is true for A and m, then there is an optimal covering $B \to A$ with B of dimension m such that*

$$\mathrm{Vis}^B(A(K)) \cong A(K)/mA(K).$$

Proof. By Proposition 4.0.7, there is an extension M of K of degree m such that the quotient $A(M)/A(K)$ is finite of order coprime to m. Then, as in [Ste04], the restriction of scalars $B = \mathrm{Res}_{M/K}(A_M)$ is an optimal cover of A and

$$\mathrm{Vis}^B(A(K)) \cong A(K)/\mathrm{Tr}(A(M)).$$

However, there is also an inclusion $A \hookrightarrow B$ from which one sees that

$$mA(M) \subset \mathrm{Tr}(A(M)),$$

so $\mathrm{Vis}^B(A(K))$ is an m-torsion group.

We have

$$[\mathrm{Tr}(A(M)) : \mathrm{Tr}(A(K))] \bigm| [A(M) : A(K)].$$

We showed above that $\gcd([A(M) : A(K)], m) = 1$, so since

$$\mathrm{Tr}(A(M))/\mathrm{Tr}(A(K))$$

is killed by m, it follows that $\mathrm{Tr}(A(M)) = \mathrm{Tr}(A(K))$. We conclude that

$$\mathrm{Vis}^B(A(K)) = A(K)/mA(K).$$

□

Proposition 4.0.9. *If E is an elliptic curve over $\mathbb{Q}$ and $m = 2$, then Conjecture 4.0.6 is true for E and m.*

Proof. Let L be as in Conjecture 4.0.6, so L is an extension of $\mathbb{Q}$ of possibly large degree. Let D be the discriminant of L. By [MM97, BFH90] there are infinitely many quadratic imaginary extensions M of $\mathbb{Q}$ such that $L(E^M, 1) \neq 0$, where E^M is the quadratic twist of E by M. By [Kol91, Kol88] all these curves have rank 0. Since there are only finitely many quadratic fields ramified only at the primes that divide D, there must be some field M that is ramified at a prime $p \nmid D$. If M is contained in L, then all the primes that ramify in M divide D, so M is not contained in L. Since M is quadratic, it follows that $M \cap L = \mathbb{Q}$, as required. Since the image of $E(\mathbb{Q}) + E^M(\mathbb{Q})$ in $E(M)$ has finite index, it follows that $E(M)/E(\mathbb{Q})$ is finite. □

Corollary 4.0.10. *If E is an elliptic curve over $\mathbb{Q}$, then there is an optimal cover $B \to E$, with B a 2-dimension modular abelian variety, such that*

$$\mathrm{Vis}^B(E(\mathbb{Q})) \cong E(\mathbb{Q})/2E(\mathbb{Q}).$$

Proof. Combine Proposition 4.0.9 with Theorem 4.0.8. Also B is modular since it is isogenous to $E \times E'$, where E' is a quadratic twist of E. □

Note that the B of Corollary 4.0.10 is isomorphic to $(E \times E^D)/\Phi$, where E^D is a rank 0 quadratic imaginary twist of E and $\Phi \cong E[2]$ is embedded antidiagonally in $E \times E^D$. Note that E^D also has analytic rank 0, since it was constructed using the theorems of [Kol91, Kol88] and [MM97, BFH90]. Thus our construction is compatible with the one of Proposition 5.1.1 below.

5 Some Data About Visibility and Modularity

This section contains a computational investigation of modularity of Mordell-Weil groups of elliptic curves relative to abelian varieties that are quotients of $J_0(N)$. One reason that we restrict to $J_0(N)$ is so that computations are more tractable. Also, for $m > 2$, the twisting constructions that we have given in previous sections are no longer allowed since they take place in $J_1(N)$. Furthermore, the work of [KL89] suggests that we understand the arithmetic of $J_0(N)$ better than that of $J_1(N)$.

5.1 A Visibility Construction for Mordell-Weil Groups

The following proposition is an analogue of [AS02, Thm. 3.1] but for visibility of Mordell-Weil groups (compare also [CM00, pg. 19]).

Proposition 5.1.1. *Let E be an elliptic curve over a number field K, and let $\Phi = E[m]$ as a $\mathrm{Gal}(\overline{K}/K)$-module. Suppose A is an abelian variety over K such that $\Phi \subset A$, as $G_{\mathbb{Q}}$-modules. Let $B = (A \times E)/\Phi$, where Φ is embedded anti-diagonally. Then there is an exact sequence*

$$0 \to B(K)/(A(K) + E(K)) \to E(K)/mE(K) \to \mathrm{Vis}^B(E(K)) \to 0.$$

Moreover, if $(A/E[m])(K)$ is finite of order coprime to m, then the first term of the sequence is 0, so

$$\mathrm{Vis}^B(E(K)) \cong E(K)/mE(K).$$

Proof. Using the definition of B and multiplication by m on E, we obtain the following commutative diagram, whose rows and columns are exact:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & E[m] \arrow[r] \arrow[d] & E \arrow[r, "m"] \arrow[d] & E \arrow[r] \arrow[d, "\cong"] & 0 \\
0 \arrow[r] & A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & E \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A/E[m] \arrow[r, "\cong"] \arrow[d] & B/E \arrow[r] \arrow[d] & 0 & \\
 & 0 & 0 & &
\end{tikzcd}$$

Taking K-rational points we arrive at the following diagram with exact rows and columns:

$$\begin{tikzcd}
0 \arrow[r] & E(K)/E(K)[m] \arrow[r, "m"] \arrow[d] & E(K) \arrow[r] \arrow[d, "\cong"] & E(K)/mE(K) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & B(K)/A(K) \arrow[r] \arrow[d] & E(K) \arrow[r] \arrow[d] & \mathrm{Vis}^B(E(K)) \arrow[r] & 0 \\
& B(K)/(A(K)+E(K)) & 0 & &
\end{tikzcd}$$

The snake lemma and the fact that the middle vertical map is an isomorphism implies that the right vertical map is a surjection with kernel isomorphic to $B(K)/(A(K)+E(K))$. Thus we obtain an exact sequence

$$0 \to B(K)/(A(K)+E(K)) \to E(K)/mE(K) \to \mathrm{Vis}^B(E(K)) \to 0.$$

This proves the first statement of the proposition. For the second, note that we have an exact sequence $0 \to E \to B \to A/E[m] \to 0$. Taking Galois cohomology yields an exact sequence

$$0 \to E(K) \to B(K) \to (A/E[m])(K) \to \cdots,$$

so $\#(B(K)/E(K)) \mid \#(A/E[m])(K)$. If $(A/E[m])(K)$ is finite of order coprime to m, then $B(K)/(A(K)+E(K))$ has order dividing $\#(A/E[m])(K)$, so the quotient $B(K)/(A(K)+E(K))$ is trivial, since it injects into $E(K)/mE(K)$. □

5.2 Tables

The data in this section suggests the following conjecture.

Conjecture 5.2.1. *Suppose E is an elliptic curve over $\mathbb{Q}$ and p is a prime such that $E[p]$ is irreducible. Then there exists infinitely many newforms $g \in S_2(\Gamma_0(N))$, for various integers N, such that $L(g,1) \neq 0$ and $E[p] \subset A_g$ and $\mathrm{Vis}^B(E(\mathbb{Q})) = E(\mathbb{Q})/pE(\mathbb{Q})$, where $B = (A_g \times E)/E[p]$.*

Let E be the elliptic curve $y^2 + y = x^3 - x$. This curve has conductor 37 and Mordell-Weil group free of rank 1. According to [Cre97], E is isolated in its isogeny class, so each $E[p]$ is irreducible.

Table 1 gives for each N the *odd* primes p such that there is a mod p congruence between f_E and some newform g in $S_2(\Gamma_0(37N))$ such that A_g has rank 0 and the isogeny class of A_g contains no abelian variety with rational p torsion. The first time a p occurs, it is in bold. We bound the torsion in the isogeny class using the algorithm from [AS05, §3.5] with primes up to 17. Thus by Proposition 5.1.1, the Mordell-Weil group of E is p-modular of level $37N$. A $-$ means there are no such p. Table 2, which was derived directly from Table 1, gives for a prime p, all integers N such that $E(\mathbb{Q})$ is p-modular of level $37N$.

Table 1: Visibility of Mordell-Weil for $y^2 + y = x^3 - x$

N	$p's$
2	5
3	7
4	–
5	–
6	–
7	3
8	–
9	–
10	–
11	17
12	–
13	–
14	–
15	–
16	–
17	3
18	–
19	5
20	–
21	7
22	–
23	11
24	–
25	–
26	–
27	3
28	–
29	3
30	–
31	3
32	–
33	7
34	–
35	–
36	–
37	–
38	5
39	–
40	–
41	$3, 17$
42	–
43	7
44	–
45	–
46	–
47	3
48	–
49	–
50	5
51	–
52	–
53	53
54	–
55	–
56	–
57	–
58	–
59	13
60	–
61	$5, 7$
62	–
63	3
64	–
65	–
66	–
67	$3, 5$
68	–
69	–
70	–
71	$3, 7$
72	–
73	$3, 5$
74	–
75	–
76	–
77	–
78	–
79	–
80	–
81	3
82	–
83	$3, 11$
84	–
85	–
86	–
87	–
88	–
89	43
90	–
91	3
92	–
93	7
94	–
95	–
96	–
97	47
98	–
99	–
100	–
101	$3, 11$
102	–
103	43
104	–
105	–
106	5
107	$3, 5$
108	–
109	$3, 7$
110	–
111	–
112	–
113	$3, 11$
114	–
115	–
116	–
117	–
118	–
119	3
120	–
121	–
122	–
123	–
124	–
125	5
126	–
127	127
128	–
129	–
130	–
131	3
132	–
133	–
134	–
135	–
136	–
137	3
138	–
139	17
140	–
141	–
142	–
143	–
144	–
145	–
146	–
147	7
148	–
149	$5, 31$
150	–
151	17
152	–
153	3
154	–
155	–
156	–
157	$3, 5$
158	–
159	–
160	–
161	–
162	–
163	$7, 13$
164	–
165	–
166	–
167	$3, 5$
168	–
169	–
170	–
171	–
172	–
173	$3, 5, 11$
174	–
175	–
176	–
177	–
178	–
179	3
180	–
181	$3, 59$
182	–
183	–
184	–
185	–
186	–
187	–
188	–
189	3
190	–
191	7
192	–
193	$5, 11$
194	–
195	–
196	–
197	$3, 5, 13$
198	–
199	$3, 11$
200	–
201	–
202	5
203	3
204	–
205	–
206	–
207	–
208	–
209	–

Table 2: Levels Where Mordell-Weil is p-Visible for $y^2 + y = x^3 - x$

p	N such that $37N$ is a level of p-modularity of $E(\mathbb{Q})$
3	7, 17, 27, 29, 31, 41, 47, 63, 67, 71, 73, 81, 83, 91, 101, 107, 109, 113, 119, 131, 137, 153, 157, 167, 173, 179, 181, 189, 197, 199, 203
5	2, 19, 38, 50, 61, 67, 73, 106, 107, 125, 149, 157, 167, 173, 193, 197, 202
7	3, 21, 33, 43, 61, 71, 93, 109, 147, 163, 191
11	23, 83, 101, 113, 173, 193, 199
13	59, 163, 197
17	11, 41, 139, 151
$19 - 29$	-
31	149
$37 - 41$	-
43	89, 103
47	97
53	53
59	181
$61 - 113$	-
127	127

Table 3: Visibility of Mordell-Weil for $y^2 + y = x^3 + x^2$

N	$p's$	N	$p's$	N	$p's$	N	$p's$	N	$p's$	N	$p's$	N	$p's$
2	5	17	$3,7$	32	–	47	–	62	–	77	–	92	–
3	3	18	–	33	3	48	–	63	–	78	–	93	–
4	–	19	–	34	5	49	–	64	–	79	–	94	–
5	5	20	–	35	–	50	5	65	–	80	–	95	–
6	–	21	–	36	–	51	3	66	–	81	3	96	–
7	–	22	5	37	19	52	–	67	71	82	–	97	$7,13$
8	–	23	5	38	–	53	59	68	–	83	$3,23$	98	–
9	–	24	–	39	3	54	–	69	–	84	–	99	3
10	–	25	–	40	–	55	5	70	–	85	5	100	–
11	3	26	–	41	37	56	–	71	$5,7$	86	–		
12	–	27	3	42	–	57	3	72	–	87	3		
13	19	28	–	43	–	58	–	73	3	88	–		
14	–	29	3	44	–	59	3	74	–	89	47		
15	–	30	–	45	–	60	–	75	–	90	–		
16	–	31	–	46	–	61	5	76	–	91	–		

Ribet's level raising theorem [Rib90] gives necessary and sufficient conditions on a prime N for there to be a newform g of level $37N$ that is congruent to f_E modulo p. Note that the form g is new rather than just p-new since 37 is prime and there are no modular forms of level 1 and weight 2. If, moreover, we impose the condition $L(g,1) \neq 0$, then Ribet's condition requires that p divides $N + 1 + \varepsilon a_N$, where ε is the root number of E. Since E has odd analytic rank, in this case $\varepsilon = -1$. For each primes $p \leq 127$ and each $N \leq 203$, were find the levels of such g. The *only* cases in which we don't already find a congruence level already listed in Table 2 corresponding to a newform with torsion multiple coprime to p are

$$p = 3, \quad N = 43 \qquad \text{and} \qquad p = 19, \quad N = 47, 79.$$

In all other cases in which Ribet's theorem produces a congruent g with $\operatorname{ord} L(g,s)$ even (hence possibly 0), we actually find a g with $L(g,1) \neq 0$ and can show that $\#A_g(\mathbb{Q})_{\text{tor}}$ is coprime to p.

For $p = 3$ and $N = 43$ we find a unique newform $g \in S_2(\Gamma_0(1591))$ that is congruent to f_E modulo 3. This form is attached to the elliptic curve $y^2 + y = x^3 - 71x + 552$ of conductor 1591, which has Mordell-Weil groups $\mathbb{Z} \oplus \mathbb{Z}$. Thus this is an example of a congruence relating a rank 1 curve to a rank 2 curve. For $p = 19$ and $N = 47$, the g has degree 43, so A_g has dimension 43, we have $L(g,1) \neq 0$, but the torsion multiple is $76 = 19 \cdot 4$, which is divisible by 19. For $p = 19$ and $N = 79$, the A_g has dimension 57, we have $L(g,1) \neq 0$, but the torsion multiple is 76 again.

Tables 3–4 are the analogues of Tables 1–2 but for the elliptic curve $y^2 + y =$

Table 4: Levels Where Mordell-Weil is p-Visible for $y^2 + y = x^3 + x^2$

p	N such that $43N$ is a level of p-modularity of $E(\mathbb{Q})$
3	3, 11, 17, 27, 29, 33, 39, 51, 57, 59, 73, 81, 83, 87, 99
5	2, 5, 22, 23, 34, 50, 55, 61, 71, 85
7	17, 71, 97
11	-
13	97
17	-
19	13, 37
23	83
$29, 31$	-
37	41
$41, 43$	-
47	89
53	-
59	53
$61, 67$	-
71	67

Table 5: Visibility of Mordell-Weil for $y^2 + y = x^3 + x^2 - 2x$

N	$p's$
1	5
2	–
3	–
4	–
5	3
6	–
7	3
8	–
9	3
10	–
11	–
12	–
13	11
14	–
15	3
16	–
17	–
18	–
19	–
20	–
21	–
22	–
23	5
24	–
25	–
26	–
27	3
28	–
29	3

Table 6: Levels Where Mordell-Weil is p-Visible for $y^2 + y = x^3 + x^2 - 2x$

p	N such that $389N$ is a level of p-modularity of $E(\mathbb{Q})$
3	5, 7, 9, 15, 27, 29
5	1, 23
7	-
11	13

$x^3 + x^2$ of conductor 43. This elliptic curve also has rank 1 and all mod p representations are irreducible. The primes p and N such that Ribet's theorem produces a congruent g with $\mathrm{ord}_{s=1} L(g, s)$ even, yet we do not find one with $L(g, 1) \neq 0$ and the torsion multiple coprime to p are

$$p = 3, \quad N = 31, 61 \qquad \text{and} \qquad p = 11, \quad N = 19, 31, 47, 79.$$

The situation for $p = 11$ is interesting since in this case all the g with $\mathrm{ord}_{s=1} L(g, s)$ even fail to satisfy our hypothesis. At level $19 \cdot 43$ we find that g has degree 18 and $L(g, 1) \neq 0$, but the torsion multiple is divisible by 11.

Let E be the elliptic curve $y^2 + y = x^3 + x^2 - 2x$ of conductor 389. This curve has Mordell-Weil group free of rank 2. Tables 5–6 are the analogues of Tables 1–2 but for E. The primes p and N such that Ribet's theorem produces a congruent g with $\mathrm{ord}_{s=1} L(g, s)$ even, yet we do not find one with $L(g, 1) \neq 0$ and the torsion multiple coprime to p are

$$p = 3, \quad N = 17 \qquad \text{and} \qquad p = 5, \quad N = 19.$$

For $p = 3$, there is a unique g of level $6613 = 37 \cdot 17$ with $\mathrm{ord}_{s=1} L(g, s)$ even and $E[3] \subset A_g$. This form has degree 5 and $L(g, 1) = 0$, so this is another example where the rank 0 hypothesis of Proposition 5.1.1 is not satisfied. Note that the torsion multiple in this case is 1. For $p = 5$, there is a unique g of level $7391 = 37 \cdot 19$, with $\mathrm{ord}_{s=1} L(g, s)$ even and $E[5] \subset A_g$. This form has degree 4 and $L(g, 1) \neq 0$, but the torsion multiple is divisible by 5.

References

[Aga99a] A. Agashe, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. 328 (1999), no. 5, 369–374. MR 1 678 131

[Aga99b] Amod Agashé, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. 328 (1999), no. 5, 369–374. MR 2000e:11083

[AS02] A. Agashe and W. A. Stein, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory 97 (2002), no. 1, 171–185. MR 2003h:11070

[AS05] A. Agashe and W. Stein, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank zero*, Math. Comp. 74 (2005), no. 249, 455–484 (electronic), With an appendix by J. Cremona and B. Mazur. MR 2085902

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. 14 (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. 24 (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[BFH90] D. Bump, S. Friedberg, and J. Hoffstein, *Eisenstein series on the metaplectic group and nonvanishing theorems for automorphic L-functions and their derivatives*, Ann. of Math. (2) 131 (1990), no. 1, 53–127.

[Bir71] B. J. Birch, *Elliptic curves over* $\mathbf{Q}$*: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, pp. 396–400.

[CM00] J. E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. 9 (2000), no. 1, 13–28. MR 1 758 797

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, http://www.maths.nott.ac.uk/personal/jec/book/.

[Dum01] N. Dummigan, *Congruences of modular forms and Selmer groups*, Math. Res. Lett. 8 (2001), no. 4, 479–494. MR MR1849264 (2002k:11064)

[DWS03] N. Dummigan, M. Watkins, and W. A. Stein, *Constructing Elements in Shafarevich-Tate Groups of Modular Motives*, Number theory and algebraic geometry, ed. by Miles Reid and Alexei Skorobogatov 303 (2003), 91–118.

[Fal86] G. Faltings, *Finiteness theorems for abelian varieties over number fields*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, Translated from the German original [Invent. Math. 73 (1983), no. 3, 349–366; ibid. 75 (1984), no. 2, 381; MR 85g:11026ab] by Edward Shipz, pp. 9–27. MR 861 971

[JS05] D. Jetchev and W. Stein, *Visibility of Shafarevich-Tate Groups at Higher Level*, in preparation.

[Kat04] Kazuya Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, Astérisque (2004), no. 295, ix, 117–290, Cohomologies p-adiques et applications arithmétiques. III. MR MR2104361

[KL89] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Algebra i Analiz 1 (1989), no. 5, 171–196.

[Kle01] T. Klenke, *Modular Varieties and Visibility*, Ph.D. thesis, Harvard University (2001).

[Kol88] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E,\mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. 52 (1988), no. 3, 522–540, 670–671. MR 89m:11056

[Kol91] V. A. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 429–436. MR 93c:11046

[LT58] S. Lang and J. Tate, *Principal homogeneous spaces over abelian varieties*, Amer. J. Math. 80 (1958), 659–684.

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. 44 (1978), no. 2, 129–162.

[Maz99] ______, *Visualizing elements of order three in the Shafarevich-Tate group*, Asian J. Math. 3 (1999), no. 1, 221–232, Sir Michael Atiyah: a great mathematician of the twentieth century. MR 2000g:11048

[Mil72] J. S. Milne, *On the arithmetic of abelian varieties*, Invent. Math. 17 (1972), 177–190. MR 48 #8512

[MM97] M. R. Murty and V. K. Murty, *Non-vanishing of L-functions and applications*, Birkhäuser Verlag, Basel, 1997.

[MO03] William J. McGraw and Ken Ono, *Modular form congruences and Selmer groups*, J. London Math. Soc. (2) 67 (2003), no. 2, 302–318. MR MR1956137 (2004d:11033)

[Rib80] K. A. Ribet, *Twists of modular forms and endomorphisms of abelian varieties*, Math. Ann. 253 (1980), no. 1, 43–62. MR 82e:10043

[Rib90] ______, *Raising the levels of modular representations*, Séminaire de Théorie des Nombres, Paris 1987–88, Birkhäuser Boston, Boston, MA, 1990, pp. 259–271.

[Rib92] ______, *Abelian varieties over $\mathbf{Q}$ and modular forms*, Algebra and topology 1992 (Taejŏn), Korea Adv. Inst. Sci. Tech., Taejŏn, 1992, pp. 53–79. MR 94g:11042

[RM05] K. Rubin and B. Mazur, *Finding large selmer groups*, in preparation.

[Roh84] D. E. Rohrlich, *On L-functions of elliptic curves and cyclotomic towers*, Invent. Math. 75 (1984), no. 3, 409–423. MR 86g:11038b

[Ros] Guido van Rossum, *Python,* http://www.python.org.

[Sch98] A. J. Scholl, *An introduction to Kato's Euler systems*, Galois Representations in Arithmetic Algebraic Geometry, Cambridge University Press, 1998, pp. 379–460.

[Ste04] W. A. Stein, *Shafarevich-Tate Groups of Nonsquare Order*, Modular Curves and Abelian Varieties, Progress of Mathematics (2004), 277–289.

[Tat66] J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1965/66, pp. Exp. No. 306, 415–440.

[Wil00] A. J. Wiles, *The Birch and Swinnerton-Dyer Conjecture*, http://www.claymath.org/prize_problems/birchsd.htm.

William A. Stein
Department of Mathematics
Harvard University
Cambridge, MA 02138
was@math.harvard.edu

24 Visibility of the Shafarevich-Tate Group at Higher Level, with D. Jetchev

Visibility of the Shafarevich-Tate Group at Higher Level

Dimitar P. Jetchev   William A. Stein

Abstract. We study visibility of Shafarevich-Tate groups of modular abelian varieties in Jacobians of modular curves of higher level. We prove a theorem about the existence of visible elements at a specific higher level under hypotheses that can be verified explicitely. We also provide a table of examples of visible subgroups at higher level and state conjectures inspired by our data.

1 Introduction

1.1 Motivation

Mazur suggested that the Shafarevich-Tate group $\text{Ш}(K, E)$ of an abelian variety A over a number field K could be studied via a collection of finite subgroups (the *visible subgroups*) corresponding to different embeddings of the variety into larger abelian varieties C over K (see [Maz99] and [CM00]). The advantage of this approach is that the isomorphism classes of principal homogeneous spaces, for which one has *à priori* little geometric information, can be given a much more explicit description as K-rational points on the quotient abelian variety C/A (the reason why they are called *visible elements*).

Agashe, Cremona, Klenke and the second author built upon the ideas of Mazur and developed a systematic theory of visibility of Shafarevich-Tate groups of abelian varieties over number fields (see [Aga99b, AS02, AS05, CM00, Kle01, Ste00]). More precisely, Agashe and Stein provided sufficient conditions for the existence of visible sugroups of certain order in the Shafarevich-Tate group and applied their general theory to the case of newform subvarieties $A_{f_{/\mathbb{Q}}}$ of the Jacobian $J_0(N)_{/\mathbb{Q}}$ of the modular curve $X_0(N)_{/\mathbb{Q}}$ (here, f is a newform of level N and weight 2 which is an eigenform for the Hecke operators acting on the space $S_2(\Gamma_0(N))$ of cuspforms of level N and weight 2). Unfortunately, there is no guarantee that a non-trivial element of $\text{Ш}(\mathbb{Q}, A_f)$ is visible for the embedding $A_f \hookrightarrow J_0(N)$.

In this paper we consider the case of modular abelian varieties over $\mathbb{Q}$ and make use of the algebraic and arithmetic properties of the corresponding newforms to provide sufficient conditions for the existence of visible elements of $Ш(\mathbb{Q}, A_f)$ in modular Jacobians of level a multiple of the base level N. More precisely, we consider morphism of the form $A_f \hookrightarrow J_0(N) \xrightarrow{\phi} J_0(MN)$, where ϕ is a suitable linear combination of degeneracy maps which makes the kernel of the composition morphism almost trivial (i.e., trivial away from the 2-part). For specific examples, the sufficient conditions can be verified explicitely. We also provide a table of examples where certain elements of $Ш(\mathbb{Q}, A_f)$ which are invisible in $J_0(N)$ become visible at a suitably chosen higher level. At the end, we state some general conjectures inspired by our results.

1.2 Organization of the paper

Section 2 discusses the basic definitions and notation for modular abelian varieties, modular forms, Hecke algebras, the Shimura construction and modular degrees. Section 3 is a brief introduction to visibility theory for Shafarevich-Tate groups. In Section 4 we state and prove an equivariant version of a theorem of Agashe-Stein (see [AS05, Thm 3.1]) which guarantees existence of visible elements. The theorem is more general because it makes use of the action of the Hecke algebra on the modular Jacobian.

In Section 5 we introduce the notion of *strong visibility* which is relevant for visualizing cohomology classes in Jacobians of modular curves whose level is a multiple of the level of the original abelian variety. Theorem 5.1.3 guarantees existence of strongly visible elements of the Shafarevich-Tate group under some hypotheses on the component groups, a congruence condition between modular forms, and irreducibility of the Galois representation. In Section 5.4 we prove a variant of the same theorem (Theorem 5.4.2) with more stringent hypotheses that are easier to verify in specific cases.

Section 6 discusses in detail two computational examples for which strongly visible elements of certain order exist which provides evidence for the Birch and Swinnerton-Dyer conjecture. We state a general conjecture (Conjecture 7.1.1) in Section 7 according to which every element of the Shafarevich-Tate group of a modular abelian variety becomes visible at higher level. We provide evidence for the the conjecture in Section 7.2 and tables of computational data in Section 7.4.

Acknowledgement: The authors would like to thank David Helm, Ben Howard, Barry Mazur, Bjorn Poonen and Ken Ribet for discussions and comments on the paper.

2 Notation

1. Abelian varieties. For a number field K, $A_{/K}$ denotes an abelian variety over K. We denote the dual of A by $A^\vee_{/K}$. If $\varphi : A \to B$ is an isogeny of degree n, we denote the *complementary isogeny* by φ'; this is the isogeny $\phi' : B \to A$, such that $\varphi \circ \varphi' = \varphi' \circ \varphi = [n]$, the multiplication-by-$n$ map on A. Unless otherwise specified, Néron models of abelian varieties will be denoted by the corresponding caligraphic letters, e.g., $\mathcal{A}$ denotes the Néron model of A.

2. Galois cohomology. For a fixed algebraic closure $\overline{K}$ of K, G_K will be the Galois group $\mathrm{Gal}(\overline{K}/K)$. If v is any non-archimedean place of K, K_v and k_v will always mean the completion and the residue field of K at v, respectively. By K_v^{ur} we always mean the maximal unramified extension of the completion K_v. Given a G_K-module M, we let $\mathrm{H}^1(K, M)$ denote the Galois cohomology group $\mathrm{H}^1(G_K, M)$.

3. Component groups. The *component group* of A at v is the finite group $\Phi_{A,v} = \mathcal{A}_{k_v}/\mathcal{A}^0_{k_v}$ which also has a structure of a finite group scheme over k_v. The *Tamagawa number* of A at v is $c_{A,v} = \#\Phi_{A,v}(k_v)$, and the *component group order* of A at v is $\overline{c}_{A,v} = \#\Phi_{A,v}(\overline{k}_v)$.

4. Modular abelian varieties. Let $h = 0$ or 1. A *J_h-modular abelian variety* is an abelian variety $A_{/K}$ which is a quotient of $J_h(N)$ for some N, i.e., there exists a surjective morphism $J_h(N) \twoheadrightarrow A$ defined over K. We define the *level* of a modular abelian variety A to be the minimal N, such that A is a quotient of $J_h(N)$. The modularity theorem of Wiles et al. (see [BCDT01]) implies that all elliptic curves over $\mathbb{Q}$ are modular. Serre's modularity conjecture implies that the modular abelian varieties over $\mathbb{Q}$ are precisely the abelian varieties over $\mathbb{Q}$ of GL_2-type (see [Rib92, §4]).

5. Shimura construction. Let $f = \displaystyle\sum_{n=1}^{\infty} a_n q^n \in S_2(\Gamma_0(N))$ be a newform of level N and weight 2 for $\Gamma_0(N)$ which is an eigenform for all Hecke operators in the Hecke algebra $\mathbb{T}(N)$. Shimura (see [Shi94, Thm. 7.14]) associated to f an abelian subvariety $A_{f/\mathbb{Q}}$ of $J_0(N)$, simple over $\mathbb{Q}$, of dimension $d = [K : \mathbb{Q}]$, where $K = \mathbb{Q}(\dots, a_n, \dots)$ is the Hecke eigenvalue field. More precisely, if $I_f = \mathrm{Ann}_{\mathbb{T}(N)}(f)$ then A_f is the connected component containing the identity of the I_f-torsion subgroup of $J_0(N)$, i.e., $A_f = J_0(N)[I_f]^0 \subset J_0(N)$. The quotient $\mathbb{T}(N)/I_f$ of the Hecke algebra $\mathbb{T}(N)$ is a subalgebra of the endomorphism ring $\mathrm{End}_{\mathbb{Q}}(A_{/\mathbb{Q}})$. Also $L(A_f, s) = \displaystyle\prod_{i=1}^{d} L(f_i, s)$, where the f_i are the $G_{\mathbb{Q}}$-conjugates of f. We also consider the dual abelian variety $A_f^{\vee}$ which is a quotient variety of $J_0(N)$.

6. I-torsion submodules. If M is a module over a commutative ring R and I is an ideal of R, let

$$M[I] = \{x \in M : mx = 0 \text{ all } m \in I\}$$

be the *I-torsion submodule* of M.

7. Hecke algebras. Let $S_2(\Gamma)$ denote the space of cusp forms of weight 2 for any congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$. Let

$$\mathbb{T}(N) = \mathbb{Z}[\ldots, T_n, \ldots] \subseteq \mathrm{End}_{\mathbb{Q}}(J_0(N))$$

be the Hecke algebra, where T_n is the nth Hecke operator. $\mathbb{T}(N)$ also acts on $S_2(\Gamma_0(N))$ and the integral homology $H_1(X_0(N), \mathbb{Z})$.

8. Modular degree. If A is an abelian subvariety of $J_0(N)$, let

$$\theta : A \to J_0(N) \cong J_0(N)^\vee \to A^\vee$$

be the induced polarization. The *modular degree* of A is

$$m_A = \sqrt{\#\,\mathrm{Ker}(A \xrightarrow{\theta} A^\vee)}.$$

See [AS02] for why m_A is an integer and for an algorithm to compute it.

3 Visible Subgroups of Shafarevich-Tate Groups

Let K be a number field and $\iota : A_{/K} \hookrightarrow C_{/K}$ be an embedding of an abelian variety into another abelian variety over K.

Definition 3.0.1. The *visible subgroup* of $\mathrm{H}^1(K, A)$ relative to ι is

$$\mathrm{Vis}_C\,\mathrm{H}^1(K, A) = \mathrm{Ker}\left(\iota_* : \mathrm{H}^1(K, A) \to \mathrm{H}^1(K, C)\right).$$

The visible subgroup of Ш(K, A) relative to the embedding ι is

$$\begin{aligned}\mathrm{Vis}_C\,Ш(K, A) &= Ш(K, A) \cap \mathrm{Vis}_C\,\mathrm{H}^1(K, A)\\ &= \mathrm{Ker}\left(Ш(K, A) \to Ш(K, C)\right)\end{aligned}$$

Let Q be the abelian variety $C/\iota(A)$, which is defined over K. The long exact sequence of Galois cohomology corresponding to the short exact sequence $0 \to A \to C \to Q \to 0$ gives rise to the following exact sequence

$$0 \to A(K) \to C(K) \to Q(K) \to \mathrm{Vis}_C\,\mathrm{H}^1(K, A) \to 0.$$

The last map being surjective means that the cohomology classes of $\mathrm{Vis}_C\,\mathrm{H}^1(K, A)$ are images of K-rational points on Q, which explains the meaning of the word *visible* in the definition. The group $\mathrm{Vis}_C\,\mathrm{H}^1(K, A)$ is finite since it is torsion and since the Mordell-Weil group $Q(K)$ is finitely generated.

Remark 3.0.2. If $A_{/K}$ is an abelian variety and $c \in \mathrm{H}^1(K, A)$ is any cohomology class, there exists an abelian variety $C_{/K}$ and an embedding $\iota : A \hookrightarrow C$ defined over K, such that $c \in \mathrm{Vis}_C\,\mathrm{H}^1(K, A)$, i.e., c is visible in C (see [AS02, Prop. 1.3]). The C of [AS02, Prop. 1.3] is the restriction of scalars of $A_L = A \times_K L$ down to K, where L is any finite extension of K such that c has trivial image in $\mathrm{H}^1(L, A)$.

4 Equivariant Visibility

Let K be a number field, let $A_{/K}$ and $B_{/K}$ be abelian subvarieties of an abelian variety $C_{/K}$, such that $C = A + B$ and $A \cap B$ is finite. Let $Q_{/K}$ denotes the quotient C/B. Let N be a positive integer divisible by all primes of bad reduction for C.

Let ℓ be a prime such that $B[\ell] \subset A$ and $e < \ell - 1$, where e is the largest ramification index of any prime of K lying over ℓ. Suppose that

$$\ell \nmid N \cdot \#B(K)_{\mathrm{tor}} \cdot \#Q(K)_{\mathrm{tor}} \cdot \prod_{v \mid N} c_{A,v} c_{B,v}.$$

Under those conditions, Agashe and Stein (see [AS02, Thm. 3.1]) construct a homomorphism $B(K)/\ell B(K) \to \text{Ш}(K, A)[\ell]$ whose kernel has $\mathbb{F}_\ell$-dimension bounded by the Mordell-Weil rank of $A(K)$.

In this paper, we refine [AS02, Prop. 1.3] by taking into account the algebraic structure coming from the endomorphism ring $\mathrm{End}_K(C)$. In particular, when we apply the theory to modular abelian varieties, we would like to use the additional structure coming from the Hecke algebra. There are numerous example (see [AS05]) where [AS02, Prop. 1.3] does not apply, but nevertheless, we can use our refinement to prove existence of visible elements of $\text{Ш}(\mathbb{Q}, A_f)$ at higher level (e.g., see Propositions 6.1.3 and 6.2.1 below).

4.1 The main theorem

Let $A_{/K}$, $B_{/K}$, $C_{/K}$, $Q_{/K}$, N and ℓ be as above. Let R be a commutative subring of $\mathrm{End}_K(C)$ that leaves A and B stable and let $\mathfrak{m}$ be a maximal ideal of R of residue characteristic ℓ. By the Néron mapping property, the subgroups $\Phi_{A,v}(k_v)$ and $\Phi_{B,v}(k_v)$ of k_v-points of the corresponding component groups can be viewed as R-modules.

Theorem 4.1.1 (Equivariant Visibility Theorem). *Suppose that $A(K)$ has rank zero and that the groups $Q(K)[\mathfrak{m}]$, $B(K)[\mathfrak{m}]$, $\Phi_{A,v}(k_v)[\mathfrak{m}]$ and $\Phi_{B,v}(k_v)[\ell]$ are all trivial for all nonarchimedean places v of K. Then there is an injective homomorphism of $R/\mathfrak{m}$-vector spaces*

$$(B(K)/\ell B(K))[\mathfrak{m}] \hookrightarrow \mathrm{Vis}_C(\text{Ш}(K, A))[\mathfrak{m}]. \tag{1}$$

Remark 4.1.2. Applying the above result for $R = \mathbb{Z}$, we recover the result of Agashe and Stein in the case when $A(K)$ has Mordell-Weil rank zero. We could relax the hypothesis that $A(K)$ is finite and instead give a bound on the dimension of the kernel of (1) in terms of the rank of $A(K)$ similar to the bound in [AS02, Thm. 3.1]. We will not need this stronger result in our paper.

4.2 Some commutative algebra

Before proving Theorem 4.1.1 we recall some well-known lemmas from commutative algebra. Let M be a module over a commutative ring R and let $\mathfrak{m}$ be a finitely generated prime ideal of R.

Lemma 4.2.1. *If $M_{\mathfrak{m}}$ is Artinian, then* $M_{\mathfrak{m}} \neq 0 \iff M[\mathfrak{m}] \neq 0$.

Proof. ($\Longleftarrow$) We first prove that $M_{\mathfrak{m}} = 0$ implies $M[\mathfrak{m}] = 0$ by a slight modification of the proof of [AM69, Prop. I.3.8]. Suppose $M_{\mathfrak{m}} = 0$, yet there is a nonzero $x \in M[\mathfrak{m}]$. Let $I = \mathrm{Ann}_R(x)$. Then $I \neq (1)$ is an ideal that contains $\mathfrak{m}$, so $I = \mathfrak{m}$. Consider $\frac{x}{1} \in M_{\mathfrak{m}}$. Since $M_{\mathfrak{m}} = 0$, we have $x/1 = 0$, hence by definition of localization, x is killed by some element of $R - \mathfrak{m}$ (set-theoretic difference). But this is impossible since $\mathrm{Ann}_R(x) = \mathfrak{m}$.

($\Longrightarrow$) Next we prove that $M_{\mathfrak{m}} \neq 0$ implies $M[\mathfrak{m}] \neq 0$. Since $M_{\mathfrak{m}}$ is an Artinian module over the (local) ring $R_{\mathfrak{m}}$, by [AM69, Prop. 6.8], $M_{\mathfrak{m}}$ has a composition series:

$$M_{\mathfrak{m}} = M_0 \supset M_1 \supset \cdots \supset M_{n-1} \supset M_n = 0,$$

where by definition each quotient M_i/M_{i+1} is a simple $R_{\mathfrak{m}}$-module. In particular, M_{n-1} is a simple $R_{\mathfrak{m}}$-module. Suppose $x \in M_{n-1}$ is nonzero, and let $I = \mathrm{Ann}_{R_{\mathfrak{m}}}(x)$. Then

$$R_{\mathfrak{m}}/I \cong R_{\mathfrak{m}} \cdot x \subset M_{n-1},$$

so by simplicity $R_{\mathfrak{m}}/I \cong M_{n-1}$ is simple. Thus $I = \mathfrak{m}$, otherwise $R_{\mathfrak{m}}/I$ would have $\mathfrak{m}/I$ as a proper submodule. Thus $x \in M_{n-1}[\mathfrak{m}]$ is nonzero.

Write $x = [y, a]$ with $y \in M$ and $a \in R - \mathfrak{m}$, where $[y/a]$ means the class of y/a in the localization (same as (y, a) on page 36 of [AM69]). Since $a \in R - \mathfrak{m}$, the element a acts as a unit on $M_{\mathfrak{m}}$, hence $ax = [y/1] \in M_{n-1}$ is nonzero and also still annihilated by $\mathfrak{m}$ (by commutativity).

To say that $[y/1]$ is annihilated by $\mathfrak{m}$ means that for all $\alpha \in \mathfrak{m}$ there exists $t \in R - \mathfrak{m}$ such that $t\alpha y = 0$ in M. Since $\mathfrak{m}$ is finitely generated, we can write $\mathfrak{m} = (\alpha_1, \ldots, \alpha_n)$ and for each α_i we get corresponding elements $t_1, \ldots, t_n$ and a product $t = t_1 \cdots t_n$. Also $t \notin \mathfrak{m}$ since $\mathfrak{m}$ is a prime ideal and each $t_i \notin \mathfrak{m}$. Let $z = ty$. Then for all $\alpha \in \mathfrak{m}$ we have $\alpha z = t\alpha y = 0$. Also $z \neq 0$ since t acts as a unit on M_{n-1}. Thus $z \in M[\mathfrak{m}]$, and is nonzero, which completes the proof of the lemma. □

Lemma 4.2.2. *Suppose* $0 \to M_1 \to N \to M_2 \to 0$ *is an exact sequence of R-modules each of whose localization at $\mathfrak{m}$ is Artinian. Then* $N[\mathfrak{m}] \neq 0 \iff (M_1 \oplus M_2)[\mathfrak{m}] \neq 0$.

Proof. By Lemma 4.2.1 we have $N[\mathfrak{m}] \neq 0$ if and only if $N_{\mathfrak{m}} \neq 0$. By Proposition 3.3 on page 39 of [AM69], the localized sequence

$$0 \to (M_1)_{\mathfrak{m}} \to N_{\mathfrak{m}} \to (M_2)_{\mathfrak{m}} \to 0$$

is exact. Thus $N_{\mathfrak{m}} \neq 0$ if and only if at least one of $(M_1)_{\mathfrak{m}}$ or $(M_2)_{\mathfrak{m}}$ is nonzero. Again by Lemma 4.2.1, at least one of $(M_1)_{\mathfrak{m}}$ or $(M_2)_{\mathfrak{m}}$ is nonzero if and only if at least one of $M_1[\mathfrak{m}]$ or $M_2[\mathfrak{m}]$ is nonzero. The latter is the case if and only if $(M_1 \oplus M_2)[\mathfrak{m}] \neq 0$. □

Remark 4.2.3. One could also prove the lemmas using the isomorphism $M[\mathfrak{m}] \cong \mathrm{Hom}_R(R/\mathfrak{m}, M)$ and exactness properties of Hom, but even with this approach many of the details in Lemma 4.2.1 still have to be checked.

Remark 4.2.4. In Theorem 4.1.1, we have $R \subset \mathrm{End}(C)$, hence R is finitely generated as a $\mathbb{Z}$-module, so R is noetherian.

Lemma 4.2.5. *Let G be a finite cyclic group, M be a finite G-module that is also a module over a commutative ring R such that the action of G and R commute (i.e., M is an $R[G]$-module). Suppose $\mathfrak{p}$ is a finitely-generated prime ideal of R, and $H^0(G, M)[\mathfrak{p}] = 0$. Then $H^1(G, M)[\mathfrak{p}] = 0$.*

Proof. Argue as in [Se79, Prop. VIII.4.8], but noting that all modules are modules over R and maps are morphisms of R-modules. □

4.3 Proof of Theorem 4.1.1

Proof of Theorem 4.1.1. We argue as in the proof of [AS02, Thm. 3.1]. The construction of the map (1) is similar to the one in the proof of [AS02, Lem. 3.6]. We have the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & B[\ell] \arrow[r] \arrow[d] & B \arrow[r, "\ell"] \arrow[d] \arrow[dr, "\psi"] & B \arrow[r] \arrow[d, "\pi"] & 0 \\
0 \arrow[r] & A \arrow[r] & C \arrow[r] & Q \arrow[r] & 0,
\end{tikzcd}$$

where $\psi : B \to Q$ is the composition of the inclusion $B \hookrightarrow C$ with the quotient map $C \to Q$, and the existence of the morphism $\pi : B \to Q$ follows from the inclusion $B[\ell] \subset \mathrm{Ker}(\psi) = A \cap B$. By naturality for the long exact sequence of Galois cohomology we obtain the following commutative diagram with exact rows and columns

$$\begin{tikzcd}
 & M_0 \arrow[d] & M_1 \arrow[d] & M_2 \arrow[d] & \\
0 \arrow[r] & B(K)/(B(K)[\ell]) \arrow[r, "\ell"] \arrow[d] & B(K) \arrow[r] \arrow[d] \arrow[dr, "\pi"] & B(K)/\ell B(K) \arrow[r] \arrow[d, "\varphi"] & 0 \\
0 \arrow[r] & C(K)/A(K) \arrow[r] \arrow[d] & Q(K) \arrow[r] & \mathrm{Vis}_C(\mathrm{H}^1(K, A)) \arrow[r] & 0 \\
 & M_3. & & &
\end{tikzcd}$$

Here, M_0, M_1 and M_2 denote the kernels of the corresponding vertical maps and M_3 denotes the cokernel of the first map. Since R preserves A, B, and

$B[\ell]$, all objects in the diagram are R-module and the morphisms of abelian varieties are also R-module homomorphisms.

The snake lemma yields an exact sequence

$$0 \to M_0 \to M_1 \to M_2 \to M_3.$$

By hypothesis, $B(K)[\mathfrak{m}] = 0$, so $N_0 = \mathrm{Ker}(B(K) \to C(K)/A(K))$ has no $\mathfrak{m}$ torsion. Noting that $B(K)[\ell] \subset N_0$, it follows that $M_0 = N_0/(B(K)[\ell])$ has no $\mathfrak{m}$ torsion either, by Lemma 4.2.2. Also, $M_1[\mathfrak{m}] = 0$ again since $B(K)[\mathfrak{m}] = 0$.

By the long exact sequence on Galois cohomology, the quotient $C(K)/B(K)$ is isomorphic to a subgroup of $Q(K)$ and by hypothesis $Q(K)[\mathfrak{m}] = 0$, so $(C(K)/B(K))[\mathfrak{m}] = 0$. Since Q is isogenous to A and $A(K)$ is finite and $C(K)/B(K) \hookrightarrow Q(K)$, we see that $C(K)/B(K)$ is finite. Thus M_3 is a quotient of the finite R-module $C(K)/B(K)$, which has no $\mathfrak{m}$-torsion, so Lemma 4.2.2 implies that $M_3[\mathfrak{m}] = 0$. The same lemma implies that M_1/M_0 has no $\mathfrak{m}$-torsion, since it is a quotient of the finite module M_1, which has no $\mathfrak{m}$-torsion. Thus, we have an exact sequence

$$0 \to M_1/M_0 \to M_2 \to M_3 \to 0,$$

and both of M_1/M_0 and M_3 have trivial $\mathfrak{m}$-torsion. It follows by Lemma 4.2.2, that $M_2[\mathfrak{m}] = 0$. Therefore, we have an injective morphism of $R/\mathfrak{m}$-vector spaces

$$\varphi : (B(K)/\ell B(K))[\mathfrak{m}] \hookrightarrow \mathrm{Vis}_C(H^1(K, A))[\mathfrak{m}].$$

It remains to show that for any $x \in B(K)$, we have $\varphi(x) \in \mathrm{Vis}_C(Ш(K, A))$, i.e., that $\varphi(x)$ is locally trivial.

We proceed exactly as in Section 3.5 of [AS05]. In both cases $\mathrm{char}(v) \neq \ell$ and $\mathrm{char}(v) = \ell$ we arrive at the conclusion that the restriction of $\varphi(x)$ to $\mathrm{H}^1(K_v, A)$ is an element $c \in \mathrm{H}^1(K_v^{\mathrm{ur}}/K_v, A(K_v^{\mathrm{ur}}))$. (Note that in the case $\mathrm{char}(v) \neq \ell$ the proof uses our hypothesis that $\ell \nmid \#\Phi_{B,v}(k_v)$.) By [Mil86, Prop I.3.8], there is an isomorphism

$$\mathrm{H}^1(K_v^{\mathrm{ur}}/K_v, A(K_v^{\mathrm{ur}})) \cong \mathrm{H}^1(\overline{k}_v/k_v, \Phi_{A,v}(\overline{k}_v)). \tag{2}$$

We will use our hypothesis that

$$\Phi_{A,v}(k_v)[\mathfrak{m}] = \Phi_{B,v}(k_v)[\ell] = 0$$

for all v of bad reduction to deduce that the image of φ lies in $\mathrm{Vis}_C(Ш(K, A))[\mathfrak{m}]$. Let d denote the image of c in $\mathrm{H}^1(\overline{k}_v/k_v, \Phi_{A,v}(\overline{k}_v))$. The construction of d is compatible with the action of R on Galois cohomology, since (as is explained in the proof of [Mil86, Prop. I.3.8]) the isomorphism (2) is induced from the exact sequence of $\mathrm{Gal}(K_v^{\mathrm{ur}}/K_v)$-modules

$$0 \to \mathcal{A}^0(K_v^{\mathrm{ur}}) \to \mathcal{A}(K_v^{\mathrm{ur}}) \to \Phi_{A,v}(\overline{k}_v) \to 0,$$

where $\mathcal{A}$ is the Néron model of A and $\mathcal{A}^0$ is the subgroup scheme whose generic fiber is A and whose closed fiber is the identity component of $\mathcal{A}_{k_v}$. Since $\varphi(x) \in \mathrm{H}^1(K, A)[\mathfrak{m}]$, it follows that

$$d \in \mathrm{H}^1(\overline{k}_v/k_v, \Phi_{A,v}(\overline{k}_v))[\mathfrak{m}].$$

Lemma 4.2.5, our hypothesis that $\Phi_{A,v}(k_v)[\mathfrak{m}] = 0$, and that

$$\mathrm{H}^1(\overline{k}_v/k_v, \Phi_{A,v}(\overline{k}_v)) = \varinjlim \mathrm{H}^1(\mathrm{Gal}(k_v'/k_v), \Phi_{A,v}(k_v'))),$$

together imply that $\mathrm{H}^1(\overline{k}_v/k_v, \Phi_{A,v}(\overline{k}_v))[\mathfrak{m}] = 0$, hence $d = 0$. Thus $c = 0$, so $\varphi(x)$ is locally trivial, which completes the proof. □

5 Strong Visibility at Higher Level

5.1 Strongly visible subgroups

Let $A_{/\mathbb{Q}}$ be an abelian subvariety of $J_0(N)_{/\mathbb{Q}}$ and let $p \nmid N$ be a prime. Let

$$\varphi = \delta_1^* + \delta_p^* : J_0(N) \to J_0(pN), \tag{3}$$

where δ_1^* and δ_p^* are the pullback maps on equivalence classes of degree-zero divisors of the degeneracy maps $\delta_1, \delta_p : X_0(pN) \to X_0(N)$. Let $\mathrm{H}^1(\mathbb{Q}, A)^{\mathrm{odd}}$ be the prime-to-2-part of the group $\mathrm{H}^1(\mathbb{Q}, A)$.

Definition 5.1.1 (Strongly Visibility). The *strongly visible* subgroup of $\mathrm{H}^1(\mathbb{Q}, A)$ for $J_0(pN)$ is

$$\mathrm{Vis}_{pN}\, \mathrm{H}^1(\mathbb{Q}, A) = \mathrm{Ker}\left(\mathrm{H}^1(\mathbb{Q}, A)^{\mathrm{odd}} \xrightarrow{\varphi_*} \mathrm{H}^1(\mathbb{Q}, J_0(pN))\right) \subset \mathrm{H}^1(\mathbb{Q}, A).$$

Also,

$$\mathrm{Vis}_{pN}\, Ш(\mathbb{Q}, A) = Ш(\mathbb{Q}, A) \cap \mathrm{Vis}_{pN}\, \mathrm{H}^1(\mathbb{Q}, A).$$

The reason we replace $\mathrm{H}^1(\mathbb{Q}, A)$ by $\mathrm{H}^1(\mathbb{Q}, A)^{\mathrm{odd}}$ is that the kernel of φ is a 2-group (see [Rib90b]).

Remark 5.1.2. We could obtain more visible subgroups by considering the map $\delta_1^* - \delta_p^*$ in Definition 5.1.1. However, the methods of this paper do not apply to this map.

For a positive integer N, let

$$\nu(N) = \frac{1}{6} \cdot \prod_{q^r \| N} (q^r + q^{r-1}) = \frac{1}{6} \cdot [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)].$$

We call the number $\nu(N)$ the *Sturm bound* (see [Stu87]).

Theorem 5.1.3. *Let $A_{/\mathbb{Q}} = A_f$ be a newform abelian subvariety of $J_0(N)$ for which $L(A_{/\mathbb{Q}}, 1) \neq 0$ and let $p \nmid N$ be a prime. Suppose that there is a maximal ideal $\lambda \subset \mathbb{T}(N)$ and an elliptic curve $E_{/\mathbb{Q}}$ of conductor pN such that:*

1. *[Nondivisibility] The residue characteristic ℓ of λ satisfies*
$$\ell \nmid 2 \cdot N \cdot p \cdot \prod_{q \mid N} c_{E,q}.$$
2. *[Component Groups] For each prime $q \mid N$,*
$$\Phi_{A,q}(\mathbb{F}_q)[\lambda] = 0.$$
3. *[Fourier Coefficients] Let $a_n(E)$ be the n-th Fourier coefficient of the modular form attached to E, and $a_n(f)$ the n-th Fourier coefficient of f. Assume that $a_p(E) = -1$,*
$$a_p(f) \equiv -(p+1) \pmod{\lambda} \quad \text{and} \quad a_q(f) \equiv a_q(E) \pmod{\lambda},$$
for all primes $q \neq p$ with $q \leq \nu(pN)$.
4. *[Irreducibility] The mod ℓ representation $\overline{\rho}_{E,\ell}$ is irreducible.*

Then there is an injective homomorphism
$$E(\mathbb{Q})/\ell E(\mathbb{Q}) \hookrightarrow \mathrm{Vis}_{pN}(\text{Ш}(\mathbb{Q}, A_f))[\lambda].$$

Remark 5.1.4. In fact, we have
$$E(\mathbb{Q})/\ell E(\mathbb{Q}) \hookrightarrow \mathrm{Ker}(\text{Ш}(\mathbb{Q}, A_f) \to \text{Ш}(\mathbb{Q}, C))[\lambda] \subset \mathrm{Vis}_{pN}(\text{Ш}(\mathbb{Q}, A_f))[\lambda],$$
where $C \subset J_0(pN)$ is isogenous to $A_f \times E$.

5.2 Some auxiliary lemmas

We will use the following lemmas in the proof of Theorem 5.1.3. The notation is as in the previous section. In addition, if $f \in S_2(\Gamma_0(N))$, we denote by $a_n(f)$ the n-th Fourier coefficient of f and by K_f and $\mathcal{O}_f$ the Hecke eigenvalue field and its ring of integers, respectively.

Lemma 5.2.1. *Suppose $A_f \subset J_0(N)$ and $A_g \subset J_0(pN)$ are attached to newforms f and g of level N and pN, respectively, with $p \nmid N$. Suppose that there is a prime ideal λ of residue characteristic $\ell \nmid 2pN$ in an integrally closed subring $\mathcal{O}$ of $\overline{\mathbb{Q}}$ that contains the ring of integers of the composite field $K = K_f K_g$ such that for $q \leq \nu(pN)$,*
$$a_q(f) \equiv \begin{cases} a_q(g) \pmod{\lambda} & \text{if } q \neq p, \\ (p+1)a_p(g) \pmod{\lambda} & \text{if } q = p. \end{cases}$$
Assume that $a_p(g) = -1$. Let $\lambda_f = \mathcal{O}_f \cap \lambda$ and $\lambda_g = \mathcal{O}_g \cap \lambda$ and assume that $A_f[\lambda_f]$ is an irreducible $G_{\mathbb{Q}}$-module. Then we have an equality
$$\varphi(A_f[\lambda_f]) = A_g[\lambda_g]$$
of subgroups of $J_0(pN)$, where φ is as in (3).

Proof. Our hypothesis that $a_p(f) \equiv -(p+1) \pmod{\lambda_f}$ implies, by the proofs in [Rib90b], that

$$\varphi(A_f[\lambda_f]) \subset \varphi(A_f) \cap J_0(pN)_{p\text{-new}},$$

where $J_0(pN)_{p\text{-new}}$ is the p-new abelian subvariety of $J_0(N)$.

By [Rib90b, Lem. 1], the operator $U_p = T_p$ on $J_0(pN)$ acts as -1 on $\varphi(A_f[\lambda_f])$. Consider the action of U_p on the 2-dimensional vector space spanned by $\{f(q), f(q^p)\}$. The matrix of U_p with respect to this basis is

$$U_p = \begin{pmatrix} a_p(f) & p \\ -1 & 0 \end{pmatrix}.$$

In particular, neither of $f(q)$ and $f(q^p)$ is an eigenvector for U_p. The characteristic polynomial of U_p acting on the span of $f(q)$ and $f(q^p)$ is $x^2 - a_p(f)x + p$. Using our hypothesis on $a_p(f)$ again, we have

$$x^2 - a_p(f)x + p \equiv x^2 + (p+1)x + p \equiv (x+1)(x+p) \pmod{\lambda}.$$

Thus we can choose an algebraic integer α such that

$$f_1(q) = f(q) + \alpha f(q^p)$$

is an eigenvector of U_p with eigenvalue congruent to -1 modulo λ. (It does not matter for our purposes whether $x^2 + a_p(f)x + p$ has distinct roots; nonetheless, since $p \nmid N$, [CV92, Thm. 2.1] implies that it does have distinct roots.) The cusp form f_1 has the same prime-indexed Fourier coefficients as f at primes other than p. Enlarge $\mathcal{O}$ if necessary so that $\alpha \in \mathcal{O}$. The p-th coefficient of f_1 is congruent modulo λ to -1 and f_1 is an eigenvector for the full Hecke algebra. It follows from the recurrence relation for coefficients of the eigenforms that

$$a_n(g) \equiv a_n(f_1) \pmod{\lambda}$$

for all integers $n \leq \nu(pN)$.

By [Stu87], we have $g \equiv f_1 \pmod{\lambda}$, so $a_q(g) \equiv a_q(f) \pmod{\lambda}$ for all primes $q \neq p$. Thus by the Brauer-Nesbitt theorem [CR62], the 2-dimensional $G_{\mathbb{Q}}$-representations $\varphi(A_f[\lambda_f])$ and $A_g[\lambda_g]$ are isomorphic.

Let $\mathfrak{m}$ be a maximal ideal of the Hecke algebra $\mathbb{T}(pN)$ that annihilates the module $A_g[\lambda_g]$. Note that $A_g[\mathfrak{m}] = A_g[\lambda_g]$ since $A_g[\mathfrak{m}] \subset A_g[\lambda_g]$ and $A_g[\lambda_g] \cong \varphi(A_f[\lambda_f])$ is irreducible as a $G_{\mathbb{Q}}$-module. The maximal ideal $\mathfrak{m}$ gives rise to a Galois representation $\overline{\rho}_{\mathfrak{m}} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T}(pN)/\mathfrak{m})$ isomorphic to $A_g[\lambda_g]$, which is irreducible since the Galois module $A_f[\lambda_f]$ is irreducible. Finally, we apply [Wil95, Thm. 2.1(i)] for $H = (\mathbb{Z}/N\mathbb{Z})^\times$ (i.e., $J_H = J_0(N)$) to conclude that $J_0(N)(\overline{\mathbb{Q}})[\mathfrak{m}] \cong (\mathbb{T}(pN)/\mathfrak{m})^2$, i.e., the representation $\overline{\rho}_{\mathfrak{m}}$ occurs with multiplicity one in $J_0(pN)$. Thus

$$A_g[\lambda_g] = \varphi(A_f[\lambda_f]).$$

□

Lemma 5.2.2. *Suppose $\varphi : A \to B$ and $\psi : B \to C$ are homomorphisms of abelian varieties over a number field K, with φ an isogeny and ψ injective. Suppose n is an integer that is relatively prime to the degree of φ. If $G = \mathrm{Vis}_C(\text{Ш}(\mathbb{Q}, B))[n^\infty]$, then there is some injective homomorphism*

$$f : G \hookrightarrow \mathrm{Ker}\left\{(\psi \circ \varphi)_* : \text{Ш}(\mathbb{Q}, A) \longrightarrow \text{Ш}(\mathbb{Q}, C)\right\},$$

such that $\varphi_*(f(G)) = G$.

Proof. Let m be the degree of the isogeny $\varphi : A \to B$. Consider the complementary isogeny $\varphi' : B \to A$, which satisfies $\varphi \circ \varphi' = \varphi' \circ \varphi = [m]$. By hypothesis m is coprime to n, so $\gcd(m, \#G) = \gcd(m, n^\infty) = 1$, hence

$$\varphi_*(\varphi'_*(G)) = [m]G = G.$$

Thus $\varphi'_*(G)$ maps, via φ_*, to $G \subset \text{Ш}(\mathbb{Q}, B)$, which in turn maps to 0 in $\text{Ш}(\mathbb{Q}, C)$. □

Lemma 5.2.3. *Let M be an odd integer coprime to N and let R be the subring of $\mathbb{T}(N)$ generated by all Hecke operators T_n with $\gcd(n, M) = 1$. Then $R = \mathbb{T}(N)$.*

Proof. See the lemma on page 491 of [Wil95]. (The condition that M is odd is necessary, as there is a counterexample when $N = 23$ and $M = 2$.) □

Lemma 5.2.4. *Suppose λ is a maximal ideal of $\mathbb{T}(N)$ with generators a prime ℓ and $T_n - a_n$ (for all $n \in \mathbb{Z}$), with $a_n \in \mathbb{Z}$. For each integer $p \nmid N$, let λ_p be the ideal in $\mathbb{T}(N)$ generated by ℓ and all $T_n - a_n$, where n varies over integers coprime to p. Then $\lambda = \lambda_p$.*

Proof. Since $\lambda_p \subset \lambda$ and λ is maximal, it suffices to prove that λ_p is maximal. Let R be the subring of $\mathbb{T}(N)$ generated by Hecke operators T_n with $p \nmid n$. The quotient R/λ_p is a quotient of $\mathbb{Z}$ since each generator T_n is equivalent to an integer. Also, $\ell \in \lambda_p$, so $R/\lambda_p = \mathbb{F}_\ell$. But by Lemma 5.2.3, $R = \mathbb{T}(N)$, so $\mathbb{T}(N)/\lambda_p = \mathbb{F}_\ell$, hence λ_p is a maximal ideal. □

Lemma 5.2.5. *Suppose that A is an abelian variety over a field K. Let R be a commutative subring of $\mathrm{End}(A)$ and I an ideal of R. Then*

$$(A/A[I])[I] \cong A[I^2]/A[I],$$

where the isomorphism is an isomorphism of $R[G_K]$-modules.

Proof. Let $a + A[I]$, for some $a \in A$, be an I-torsion element of $A/A[I]$. Then by definition, $xa \in A[I]$ for each $x \in I$. Therefore, $a \in A[I^2]$. Thus $a + A[I] \mapsto a + A[I]$ determines a well-defined homomorphism of $R[G_K]$-modules

$$\varphi : (A/A[I])[I] \to A[I^2]/A[I].$$

Clearly this homomorphism is injective. It is also surjective as every element $a + A[I] \in A[I^2]/A[I]$ is I-torsion as an element of $A/A[I]$, as $Ia \in A[I]$. Therefore, φ is an isomorphism of $R[G_K]$-modules. □

Lemma 5.2.6. *Let ℓ be a prime and let $\phi : E \to E'$ be an isogeny of elliptic curves of degree coprime to ℓ defined over a number field K. If v is any place of K then $\ell \mid c_{E,v}$ if and only if $\ell \mid c_{E',v}$.*

Proof. Consider the complementary isogeny $\phi' : E' \to E$. Both ϕ and ϕ' induce homomorphisms $\phi : \Phi_{E,v}(k_v) \to \Phi_{E',v}(k_v)$ and $\phi' : \Phi_{E',v}(k_v) \to \Phi_{E,v}(k_v)$ and $\phi \circ \phi'$ and $\phi' \circ \phi$ are multiplication-by-n maps. Since $(n, \ell) = 1$ then $\# \ker \phi$ and $\# \ker \phi'$ must be coprime to ℓ which implies the statement. □

5.3 Proof of Theorem 5.1.3

Proof of Theorem 5.1.3. By [BCDT01] E is modular, so there is a rational newform $f \in S_2^{\mathrm{new}}(pN)$ which is an eigenform for the Hecke operators and an isogeny $E \to E_f$ defined over $\mathbb{Q}$, which by Hypothesis 4 can be chosen to have degree coprime to ℓ. Indeed, every cyclic rational isogeny is a composition of rational isogenies of prime degree, and E admits no rational ℓ-isogeny since $\overline{\rho}_{E,\ell}$ is irreducible.

By Hypothesis 1 the Tamagawa numbers of E are coprime to ℓ. Since E and E_f are related by an isogeny of degree coprime to ℓ, the Tamagawa numbers of E_f are also not divisible by ℓ by Lemma 5.2.6. Moreover, note that

$$E(\mathbb{Q}) \otimes \mathbb{F}_\ell \cong E_f(\mathbb{Q}) \otimes \mathbb{F}_\ell.$$

Let $\mathfrak{m}$ be the ideal of $\mathbb{T}(pN)$ generated by ℓ and $T_n - a_n(E)$ for all integers n coprime to p. Note that $\mathfrak{m}$ is maximal by Lemma 5.2.4.

Let φ be as in (3), and let $A = \varphi(A_f)$. Note that if $T_n \in \mathbb{T}(pN)$ then $T_n(E_f) \subset E_f$ since E_f is attached to a newform, and if, moreover $p \nmid n$, then $T_n(A) \subset A$ since the Hecke operators with index coprime to p commute with the degeneracy maps. Lemma 5.2.1 implies that

$$E_f[\ell] = E_f[\mathfrak{m}] = \varphi(A_f[\lambda]) \subset A,$$

so $\Psi = E_f[\ell]$ is a subgroup of A as a $G_{\mathbb{Q}}$-module. Let

$$C = (A \times E_f)/\Psi,$$

where we embed Ψ in $A \times E_f$ anti-diagonally, i.e., by the map $x \mapsto (x, -x)$. The antidiagonal map $\Psi \to A \times E_f$ commutes with the Hecke operators T_n for $p \nmid n$, so $(A \times E_f)/\Psi$ is preserved by the T_n with $p \nmid n$. Let R be the subring of $\mathrm{End}(C)$ generated by the action of all Hecke operators T_n, with $p \nmid n$. Also note that $T_p \in \mathrm{End}(J_0(pN))$ acts by Hypothesis 3 as -1 on E_f, but T_p need *not* preserve A.

Suppose for the moment that we have verified that the hypothesis of Theorem 4.1.1 are satisfied with A, $B = E_f$, C, $Q = C/B$, R as above and $K = \mathbb{Q}$. Then we obtain an injective homomorphism

$$E(\mathbb{Q})/\ell E(\mathbb{Q}) \cong E_f(\mathbb{Q})/\ell E_f(\mathbb{Q}) \hookrightarrow \mathrm{Ker}(Ш(\mathbb{Q}, A) \to Ш(\mathbb{Q}, C))[\mathfrak{m}].$$

We then apply Lemma 5.2.2 with $n = \ell$, A_f, A, and C, respectively, to see that

$$E_f(\mathbb{Q})/\ell E_f(\mathbb{Q}) \subset \mathrm{Ker}(\text{Ш}(\mathbb{Q}, A_f) \to \text{Ш}(\mathbb{Q}, C))[\lambda].$$

That $E_f(\mathbb{Q})/\ell E_f(\mathbb{Q})$ lands in the λ-torsion is because the subgroup of $\mathrm{Vis}_C(\text{Ш}(\mathbb{Q}, E_f))$ that we constructed is $\mathfrak{m}$-torsion.

Finally, consider $A \times E_f \to J_0(pN)$ given by $(x, y) \mapsto x + y$. Note that Ψ maps to 0, since $(x, -x) \mapsto 0$ and the elements of Ψ are of the form $(x, -x)$. We have a (not-exact!) sequence of maps

$$\text{Ш}(\mathbb{Q}, A_f) \to \text{Ш}(\mathbb{Q}, C) \to \text{Ш}(\mathbb{Q}, J_0(pN)),$$

hence inclusions

$$\begin{aligned} E_f(\mathbb{Q})/\ell E_f(\mathbb{Q}) &\subseteq \mathrm{Ker}(\text{Ш}(\mathbb{Q}, A_f) \to \text{Ш}(\mathbb{Q}, C)) \\ &\subseteq \mathrm{Ker}(\text{Ш}(\mathbb{Q}, A_f) \to \text{Ш}(\mathbb{Q}, J_0(pN))), \end{aligned}$$

which gives the conclusion of the theorem.

It remains to verify the hypotheses of Theorem 4.1.1. That $C = A + B$ is clear from the definition of C. Also, $A \cap E_f = E_f[\ell]$, which is finite. We explained above when defining R that each of A and E_f is preserved by R. Since $K = \mathbb{Q}$ and ℓ is odd the condition $1 = e < \ell - 1$ is satisfied. That $A(\mathbb{Q})$ is finite follows from our hypothesis that $L(A_f, 1) \neq 0$ (by [KL89]).

It remains is to verify that the groups

$$Q(\mathbb{Q})[\mathfrak{m}], \quad E_f(\mathbb{Q})[\mathfrak{m}], \quad \Phi_{A,q}(\mathbb{F}_q)[\mathfrak{m}], \quad \text{and } \Phi_{E_f,q}(\mathbb{F}_q)[\ell],$$

are 0 for all primes $q \mid pN$. Since $\ell \in \mathfrak{m}$, we have by Hypothesis 4 that

$$E_f(\mathbb{Q})[\mathfrak{m}] = E_f(\mathbb{Q})[\ell] = 0.$$

We will now verify that $Q(\mathbb{Q})[\mathfrak{m}] = 0$. From the definition of C and Ψ we have $Q \cong A/\Psi$. Let λ_p be as in Lemma 5.2.4 with $a_n = a_n(E)$. The map φ induces an isogeny of 2-power degree

$$A_f/(A_f[\lambda]) \to A/\Psi.$$

Thus there is λ_p-torsion in $(A_f/(A_f[\lambda]))(\mathbb{Q})$ if and only if there is $\mathfrak{m}$-torsion in $(A/\Psi)(\mathbb{Q})$. Thus it suffices to prove that $(A_f/A_f[\lambda])(\mathbb{Q})[\lambda_p] = 0$.

By Lemma 5.2.4, we have $\lambda_p = \lambda$, and by Lemma 5.2.5,

$$(A_f/A_f[\lambda])[\lambda] \cong A_f[\lambda^2]/A_f[\lambda].$$

By [Maz77, §II.14], the quotient $A_f[\lambda^2]/A_f[\lambda]$ injects into a direct sum of copies of $A_f[\lambda]$ as Galois modules. But $A_f[\lambda] \cong E[\ell]$ is irreducible, so $(A_f[\lambda^2]/A_f[\lambda])(\mathbb{Q}) = 0$, as required.

By Hypothesis 2, we have $\Phi_{A_f,q}(\mathbb{F}_q)[\lambda] = 0$ for each prime divisor $q \mid N$.Since A is 2-power isogenous to A_f and ℓ is odd, this verifies the Tamagawa number hypothesis for A. Our hypothesis that $a_p(E) = -1$ implies that Frob_p acts on $\Phi_{E_f,p}(\overline{\mathbb{F}}_p)$ as -1. Thus $\Phi_{E_f,p}(\mathbb{F}_p)[\ell] = 0$ since ℓ is odd. This completes the proof. □

Remark 5.3.1. An essential ingrediant in the proof of the above theorem is the multiplicity one result used in the paper of Wiles (see [Wil95, Thm. 2.1.]). Since this result holds for Jacobians J_H of the curves $X_H(N)$ that are intermediate covers for the covering $X_1(N) \to X_0(N)$ corresponding to subgroups $H \subseteq (\mathbb{Z}/N\mathbb{Z})^\times$ (i.e., the Galois group of $X_1(N) \to X_H$ is H), one should be able to give a generalization of Theorem 5.1.3 which holds for newform subvarieties of J_H. This requires generalizing some results from [Rib90b] to arbitrary H.

5.4 A Variant of Theorem 5.1.3 with Simpler Hypothesis

Proposition 5.4.1. *Suppose $A = A_f \subset J_0(N)$ is a newform abelian variety and q is a prime that exactly divides N. Suppose $\mathfrak{m} \subset \mathbb{T}(N)$ is a non-Eisenstein maximal ideal of residue characteristic ℓ and that $\ell \nmid m_A$, where m_A is the modular degree of A. Then $\Phi_{A,q}(\overline{\mathbb{F}}_q)[\mathfrak{m}] = 0$.*

Proof. The component group of $\Phi_{J_0(N),q}(\overline{\mathbb{F}}_q)$ is Eisenstein by [Rib87], so

$$\Phi_{J_0(N),q}(\overline{\mathbb{F}}_q)[\mathfrak{m}] = 0.$$

By Lemma 4.2.2, the image of $\Phi_{J_0(N),q}(\overline{\mathbb{F}}_q)$ in $\Phi_{A^\vee,q}(\overline{\mathbb{F}}_q)$ has no $\mathfrak{m}$ torsion. By the main theorem of [CS01], the cokernel $\Phi_{J_0(N),q}(\overline{\mathbb{F}}_q)$ in $\Phi_{A^\vee,q}(\overline{\mathbb{F}}_q)$ has order that divides m_A. Since $\ell \nmid m_A$, it follows that the cokernel also has no $\mathfrak{m}$ torsion. Thus Lemma 4.2.2 implies that $\Phi_{A^\vee,q}(\overline{\mathbb{F}}_q)[\mathfrak{m}] = 0$. Finally, the modular polarization $A \to A^\vee$ has degree m_A, which is coprime to ℓ, so the induced map $\Phi_{A,q}(\overline{\mathbb{F}}_q) \to \Phi_{A^\vee,q}(\overline{\mathbb{F}}_q)$ is an isomorphism on ℓ primary parts. In particular, that $\Phi_{A^\vee,q}(\overline{\mathbb{F}}_q)[\mathfrak{m}] = 0$ implies that $\Phi_{A,q}(\overline{\mathbb{F}}_q)[\mathfrak{m}] = 0$. □

If E is a semistable elliptic curve over $\mathbb{Q}$ with discriminant Δ, then we see using Tate curves that $\overline{c}_p = \mathrm{ord}_p(\Delta)$.

Theorem 5.4.2. *Suppose $A = A_f \subset J_0(N)$ is a newform abelian variety with $L(A_{/\mathbb{Q}}, 1) \neq 0$ and N square free, and let ℓ be a prime. Suppose that $p \nmid N$ is a prime, and that there is an elliptic curve E of conductor pN such that:*

1. *[Rank] The Mordell-Weil rank of $E(\mathbb{Q})$ is positive.*
2. *[Divisibility] We have $a_p(E) = -1$, $\ell \mid \overline{c}_{E,p}$, and*
$$\ell \nmid 2 \cdot N \cdot p \cdot c_{E,p} \cdot \prod_{q \mid N} \overline{c}_{E,q}.$$
3. *[Irreducibility] The mod ℓ representation $\overline{\rho}_{E,\ell}$ is irreducible.*
4. *[Noncongruence] The representation $\overline{\rho}_{E,\ell}$ is not isomorphic to any representation $\overline{\rho}_{g,\lambda}$ where $g \in S_2(\Gamma_0(N))$ is a newform of level dividing N that is not conjugate to f.*

Then there is an element of order ℓ in $Ш(\mathbb{Q}, A_f)$ that is not visible in $J_0(N)$ but is strongly visible in $J_0(pN)$. More precisely, there is an inclusion

$$E(\mathbb{Q})/\ell E(\mathbb{Q}) \hookrightarrow \mathrm{Ker}(Ш(\mathbb{Q}, A_f) \to Ш(\mathbb{Q}, C))[\lambda] \subset \mathrm{Vis}_{pN}(Ш(\mathbb{Q}, A_f))[\lambda],$$

where $C \subset J_0(pN)$ is isogenous to $A_f \times E$, the homomorphism $A_f \to C$ has degree a power of 2, and λ is the maximal ideal of $\mathbb{T}(N)$ corresponding to $\overline{\rho}_{E,\ell}$.

Proof. The divisibility assumptions of Hypothesis 2 on the $\overline{c}_{E,q}$ imply that the Serre level of $\overline{\rho}_{E,\ell}$ is N and since $\ell \nmid N$, the Serre weight is 2 (see [RS01, Thm. 2.10]). Since ℓ is odd, Ribet's level lowering theorem [Rib91] implies that there is *some* newform $h = \sum b_n q^n \in S_2(\Gamma_0(N))$ and a maximal ideal λ over ℓ such that $a_q(E) \equiv b_q \pmod{\lambda}$ for all primes $q \neq p$. By our non-congruence hypothesis, the only possibility is that h is a $G_{\mathbb{Q}}$-conjugate of f. Since we can replace f by any Galois conjugate of f without changing A_f, we may assume that $f = h$. Also $a_p(f) \equiv -(p+1) \pmod{\lambda}$, as explained in [Rib83, pg. 506].

Hypothesis 3 implies that λ is not Eisenstein, and by assumption $\ell \nmid m_A$, so Proposition 5.4.1 implies that $\Phi_{A,q}(\overline{\mathbb{F}}_q)[\lambda] = 0$ for each $q \mid N$.

The theorem now follows from Theorem 5.1.3. □

Remark 5.4.3. The condition $a_p(E) = -1$ is redundant. Indeed, we have $\overline{c}_{E,p} \neq c_{E,p}$ since $\overline{c}_{E,p}$ is divisible by ℓ and $c_{E,p}$ is not. By studying the action of Frobenius on the component group at p one can show that this implies that E has nonsplit multiplicative reduction, so $a_p(E) = -1$.

Remark 5.4.4. The non-congruence hypothesis of Theorem 5.4.2 can be verified using modular symbols as follows. Let $W \subset H_1(X_0(N), \mathbb{Z})_{\mathrm{new}}$ be the saturated submodule of $H_1(X_0(N), \mathbb{Z})$ that corresponds to all newforms in $S_2(\Gamma_0(N))$ that are not Galois conjugate to f. Let $\overline{W} = W \otimes \mathbb{F}_\ell$. We require that the intersection of the kernels of $T_q|_{\overline{W}} - a_q(E)$, for $q \neq p$, has dimension 0.

6 Computational Examples

In this section we give examples that illustrate how to use Theorem 5.4.2 to prove existence of elements of the Shafarevich-Tate group of a newform subvariety of $J_0(N)$ (for 767 and 959) which are invisible at the base level, but become visible in a modular Jacobian of higher level.

Hypothesis 6.0.5. The statements in this section all make the hypothesis that certain commands of the computer algebra system Magma [BCP97] produce correct output.

6.1 Level 767

Consider the modular Jacobian $J_0(767)$. Using the modular symbols package in Magma, one decomposes $J_0(767)$ (up to isogeny) into a product of six optimal quotients of dimensions 2, 3, 4, 10, 17 and 23. The duals of these quotients are subvarieties $A_2, A_3, A_4, A_{10}, A_{17}$ and A_{23} defined over $\mathbb{Q}$, where A_d has dimension d. Consider the subvariety A_{23}.

We first show that the Birch and Swinnerton-Dyer conjectural formula predicts that the orders of the groups $Ш(\mathbb{Q}, A_{23})$ and $Ш(\mathbb{Q}, A_{23}^{\vee})$ are both divisible by 9.

PROPOSITION 6.1.1. *Assume [AS05, Conj. 2.2]. Then*

$$3^2 \mid \#Ш(\mathbb{Q}, A_{23}) \quad \textit{and} \quad 3^2 \mid \#Ш(\mathbb{Q}, A_{23}^{\vee}).$$

Proof. Let $A = A_{23}^{\vee}$. We use [AS05, §3.5 and §3.6] (see also [Ka81]) to compute a multiple of the order of the torsion subgroup $A(\mathbb{Q})_{\mathrm{tor}}$. This multiple is obtained by injecting the torsion subgroup into the group of $\mathbb{F}_p$-rational points on the reduction of A for odd primes p of good reduction and then computing the order of that group. Hence, the multiple is an isogeny invariant, so one gets the same multiple for $A^{\vee}(\mathbb{Q})_{\mathrm{tor}}$. For producing a divisor of $\#A(\mathbb{Q})_{\mathrm{tor}}$, we use the injection of the subgroup of rational cuspidal divisor classes of degree 0 into $A(\mathbb{Q})_{\mathrm{tor}}$. Using the implementation in Magma we obtain $120 \mid \#A(\mathbb{Q})_{\mathrm{tor}} \mid 240$. To compute a divisor of $A^{\vee}(\mathbb{Q})_{\mathrm{tor}}$, we use the algorithm described in [AS05, §3.3] to find that the modular degree $m_A = 2^{34}$, which is not divisible by any odd primes, hence $15 \mid \#A^{\vee}(\mathbb{Q})_{\mathrm{tor}} \mid 240$.

Next, we use [AS05, §4] to compute the ratio of the special value of the L-function of $A_{/\mathbb{Q}}$ at 1 over the real Néron period Ω_A. We obtain $\dfrac{L(A_{/\mathbb{Q}}, 1)}{\Omega_A} = c_A \cdot \dfrac{2^9 \cdot 3}{5}$, where $c_A \in \mathbb{Z}$ is the Manin constant. Since $c_A \mid 2^{\dim(A)}$ by [ARS06] then

$$\frac{L(A_{/\mathbb{Q}}, 1)}{\Omega_A} = \frac{2^{n+2} \cdot 3}{5},$$

for some $0 \leq n \leq 23$. In particular, the modular abelian variety $A_{/\mathbb{Q}}$ has rank zero over $\mathbb{Q}$.

Next, using the algorithms from [CS01, KS00] we compute the Tamagawa number $c_{A,13} = 1920 = 2^3 \cdot 3 \cdot 5$. We also find that $2 \mid c_{A,59}$ is a power of 2 because W_{59} acts as 1 on A, and on the component group $\mathrm{Frob}_{59} = -W_{59}$, so the fixed subgroup $\Phi_{A,59}(\mathbb{F}_{59})$ of Frobenius is a 2-group (for more details, see [Rib90a, Prop. 3.7–8]).

Finally, the Birch and Swinnerton-Dyer conjectural formula for abelian varieties of Mordell-Weil rank zero (see [AS05, Conj. 2.2]) asserts that

$$\frac{L(A_{/\mathbb{Q}}, 1)}{\Omega_A} = \frac{\#Ш(\mathbb{Q}, A) \cdot c_{A,13} \cdot c_{A,59}}{\#A(\mathbb{Q})_{\mathrm{tor}} \cdot \#A^{\vee}(\mathbb{Q})_{\mathrm{tor}}}.$$

By substituting what we computed above, we obtain $3^2 \mid \#Ш(\mathbb{Q}, A)$. Since $L(A_{/\mathbb{Q}}, 1) \neq 0$, [KL89] implies that $Ш(\mathbb{Q}, A)$ is finite. By the nondegeneracy of the Cassels-Tate pairing, $\#Ш(\mathbb{Q}, A) = \#Ш(A^{\vee}/\mathbb{Q})$. Thus, if the BSD conjectural formula is true then $3^2 \mid \#Ш(\mathbb{Q}, A) = \#Ш(\mathbb{Q}, A^{\vee})$. □

We next observe that there are no visible elements of odd order for the embedding $A_{23/\mathbb{Q}} \hookrightarrow J_0(767)_{/\mathbb{Q}}$.

Lemma 6.1.2. *Any element of* $Ш(\mathbb{Q}, A_{23})$ *which is visible in* $J_0(767)$ *has order a power of* 2.

Proof. Since $m_{A_{23}} = 2^{34}$, [AS05, Prop. 3.15] implies that any element of $Ш(\mathbb{Q}, A_{23})$ that is visible in $J_0(767)$ has order a power of 2. □

Finally, we use Theorem 5.4.2 to prove the existence of non-trivial elements of order 3 in $Ш(\mathbb{Q}, A_{23})$ which are invisible at level 767, but become visible at higher level. In particular, we prove unconditionally that $3 \mid \#Ш(\mathbb{Q}, A_{23})$ which provides evidence for the Birch and Swinnerton-Dyer conjectural formula.

Proposition 6.1.3. *There is an element of order 3 in* $Ш(\mathbb{Q}, A_{23})$ *which is not visible in* $J_0(767)$ *but is strongly visible in* $J_0(2 \cdot 767)$.

Proof. Let $A = A_{23}$, and note that A has rank 0, since $L(A_{/\mathbb{Q}}, 1) \neq 0$. Using Cremona's database [Cre] we find that the elliptic curve

$$E : \qquad y^2 + xy = x^3 - x^2 + 5x + 37$$

has conductor $2 \cdot 767$ and Mordell-Weil group $E(\mathbb{Q}) = \mathbb{Z} \oplus \mathbb{Z}$. Also

$$c_2 = 2, c_{13} = 2, c_{59} = 1, \bar{c}_2 = 6, \bar{c}_{13} = 2, \bar{c}_{59} = 1.$$

We apply Theorem 5.4.2 with $\ell = 3$ and $p = 2$. Since E does not admit any rational 3-isogeny (by [Cre]), Hypothesis 3 is satisfied. The level is square free and the modular degree of A is a power of 2, so Hypothesis 2 is satisfied.

We have $a_3(E) = -3$. Using Magma we find

$$\det(T_3|_{\overline{W}} - (-3)) \equiv 1 \pmod{3},$$

which verifies the noncongruence hypothesis and completes the proof.

□

6.2 Level 959

We do similar computations for a 24-dimensional abelian subvariety of $J_0(959)$. We have $959 = 7 \cdot 137$, which is square free. There are five newform abelian subvarieties of the Jacobian, A_2, A_7, A_{10}, A_{24} and A_{26}, whose dimensions are the corresponding subscripts. Let $A_f = A_{24}$ be the 24-dimensional newform abelian subvariety.

Proposition 6.2.1. *There is an element of order 3 in* $Ш(A_f/\mathbb{Q})$ *which is not visible in* $J_0(959)$ *but is strongly visible in* $J_0(2 \cdot 959)$.

Proof. Using Magma we find that $m_A = 2^{32} \cdot 583673$, which is coprime to 3. Thus we apply Theorem 5.4.2 with $\ell = 3$ and $p = 2$. Consulting [Cre] we find the curve E=1918C1, with Weierstrass equation

$$y^2 + xy + y = x^3 - 22x - 24,$$

with Mordell-Weil group $E(\mathbb{Q}) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$, and

$$c_2 = 2, c_7 = 2, c_{137} = 1, \bar{c}_2 = 6, \bar{c}_7 = 2, \bar{c}_{137} = 1.$$

Using [Cre] we find that E has no rational 3-isogeny. The modular form attached to E is

$$g = q - q^2 - 2q^3 + q^4 - 2q^5 + \cdots ,$$

and we have

$$\det(T_2|_{\overline{W}} - (-2)) = 2177734400 \equiv 2 \pmod 3,$$

which completes the verification. □

7 Conjecture, evidence and more computational data

We state several conjectures, provide some evidence and finally, provide a table that we computed using similar techniques to those in Section 6

7.1 The conjecture

The two examples computed in Section 6 show that for an abelian subvariety A of $J_0(N)$ an invisible element of $\text{Ш}(\mathbb{Q}, A)$ at the base level N might become visible at a multiple level NM. We state a general conjecture according to which any element of $\text{Ш}(\mathbb{Q}, A)$ should have such a property.

Conjecture 7.1.1. *Let $h = 0$ or 1. Suppose A is a J_h-modular abelian variety and $c \in \text{Ш}(\mathbb{Q}, A)$. Then there is a J_h-modular abelian variety C and an inclusion $\iota : A \to C$ such that $\iota_* c = 0$.*

Remark 7.1.2. For any prime ℓ, the Jacobian $J_h(N)$ comes equipped with two morphisms $\alpha^*, \beta^* : J_h(N) \to J_h(N\ell)$ induced by the two degeneracy maps $\alpha, \beta : X_h(\ell N) \to X_h(N)$ between the modular curves of levels ℓN and N, and it is natural to consider visibility of $\text{Ш}(\mathbb{Q}, A)$ in $J_h(N\ell)$ via morphisms ι constructed from these degeneracy maps.

Remark 7.1.3. It would be interesting to understand the set of all levels N of J_h-modular abelian varieties C that satisfy the conclusion of the conjecture.

7.2 Theoretical Evidence for the Conjectures

The first piece of theoretical evidence for Conjecture 7.1.1 is Remark 3.0.2, according to which any cohomology class $c \in \mathrm{H}^1(K, A)$ is visible in some abelian variety $C_{/K}$.

The next proposition gives evidence for elements of $Ш(\mathbb{Q}, E)$ for an elliptic curve E and elements of order 2 or 3.

Proposition 7.2.1. *Suppose E is an elliptic curve over $\mathbb{Q}$. Then Conjecture 7.1.1 for $h = 0$ is true for all elements of order 2 and 3 in $Ш(\mathbb{Q}, E)$.*

Proof. We first show that there is an abelian variety C of dimension 2 and an injective homomorphism $i : E \hookrightarrow C$ such that $c \in \mathrm{Vis}_C(Ш(\mathbb{Q}, E))$. If c has order 2, this follows from [AS02, Prop. 2.4] or [Kle01], and if c has order 3, this follows from [Maz99, Cor. pg. 224]. The quotient C/E is an elliptic curve, so C is isogenous to a product of two elliptic curves. Thus by [BCDT01], C is a quotient of $J_0(N)$, for some N. □

We also prove that Conjecture 7.1.1 is true with $h = 1$ for all elements of $Ш(\mathbb{Q}, A)$ which split over abelian extensions.

Proposition 7.2.2. *Suppose $A_{/\mathbb{Q}}$ is a J_1-modular abelian variety over $\mathbb{Q}$ and $c \in Ш(\mathbb{Q}, A)$ splits over an abelian extension of $\mathbb{Q}$. Then Conjecture 7.1.1 is true for c with $h = 1$.*

Proof. Suppose K is an abelian extension such that $\mathrm{res}_K(c) = 0$ and let $C = \mathrm{Res}_{K/\mathbb{Q}}(A_K)$. Then c is visible in C (see Section 3.0.2). It remains to verify that C is modular. As discussed in [Mil72, pg. 178], for any abelian variety B over K, we have an isomorphism of Tate modules

$$\mathrm{Tate}_\ell(\mathrm{Res}_{K/\mathbb{Q}}(B_K)) \cong \mathrm{Ind}_{G_K}^{G_\mathbb{Q}} \mathrm{Tate}_\ell(B_K),$$

and by Faltings's isogeny theorem [Fal86], the Tate module determines an abelian variety up to isogeny. Thus if $B = A_f$ is an abelian variety attached to a newform, then $\mathrm{Res}_{K/\mathbb{Q}}(B_K)$ is isogenous to a product of abelian varieties A_{f^χ}, where χ runs through Dirichlet characters attached to the abelian extension $K/\mathbb{Q}$. Since A is isogenous to a product of abelian varieties of the form A_f (for various f), it follows that the restriction of scalars C is modular. □

Remark 7.2.3. Suppose that E is an elliptic curve and $c \in Ш(\mathbb{Q}, E)$. Is there an abelian extension $K/\mathbb{Q}$ such that $\mathrm{res}_K(c) = 0$? The answer is "yes" if and only if there is a K-rational point (with K-abelian) on the locally trivial principal homogeneous space corresponding to c (this homogenous space is a genus one curve). Recently, M. Ciperiani and A. Wiles proved that any genus one curve over $\mathbb{Q}$ which has local points everywhere and whose Jacobian is a semistable elliptic curve admits a point over a solvable extension of $\mathbb{Q}$ (see [CW06]). Unfortunately, this paper does not answer our question about the existence of abelian points.

Remark 7.2.4. As explained in [Ste04], if $K/\mathbb{Q}$ is an abelian extension of prime degree then there is an exact sequence

$$0 \to A \to \mathrm{Res}_{K/\mathbb{Q}}(E_K) \xrightarrow{\mathrm{Tr}} E \to 0,$$

where A is an abelian variety with $L(A_{/\mathbb{Q}}, s) = \prod L(f_i, s)$ (here, the f_i's are the $G_{\mathbb{Q}}$-conjugates of the twist of the newform f_E attached to E by the Dirichlet character associated to $K/\mathbb{Q}$). Thus one could approach the question in the previous remark by investigating whether or not $L(f_E, \chi, 1) = 0$ which one could do using modular symbols (see [CFK06]). The authors expect that L-functions of twists of degree larger than three are very unlikely to vanish at $s = 1$ (see [CFK06]), which suggests that in general, the question might have a negative answer for cohomology classes of order larger than 3.

7.3 Visibility of Kolyvagin cohomology classes

It would also be interesting to study visibility at higher level of Kolyvagin cohomology classes. The following is a first "test question" in this direction.

Question 7.3.1. Suppose $E \subset J_0(N)$ is an elliptic curve with conductor N, and fix a prime ℓ such that $\overline{\rho}_{E,\ell}$ is surjective. Fix a quadratic imaginary field K that satisfies the Heegner hypothesis for E. For any prime p satisfying the conditions of [Rub89, Prop. 5], let $c_p \in \mathrm{H}^1(\mathbb{Q}, E)[\ell]$ be the corresponding Kolyvagin cohomology class. There are two natural homomorphisms $\delta_1^*, \delta_p^* : E \to J_0(Np)$. When is

$$(\delta_1^* \pm \delta_\ell^*)_*(c_\ell) = 0 \in \mathrm{H}^1(\mathbb{Q}, J_0(Np))?$$

When is

$$\mathrm{res}_v((\delta_1^* \pm \delta_\ell^*)_*(c_\ell)) = 0 \in \mathrm{H}^1(\mathbb{Q}_v, J_0(Np))?$$

7.4 Table of Strong Visibility at Higher Level

The following is a table that gives the known examples of $A_{f_{/\mathbb{Q}}}$ with square free conductor $N \leq 1339$, such that the Birch and Swinnerton-Dyer conjectural formula predicts an odd prime divisor ℓ of $Ш(\mathbb{Q}, A_f)$, but ℓ does not divide the modular degree of A_f. These were taken from [AS05]. If there is an entry in the fourth column, this means we have verified the hypothesis of Theorem 5.4.2, hence there really is a nonzero element in $Ш(\mathbb{Q}, A_f)$ that is not visible in $J_0(N)$, but is strongly visible in $J_0(pN)$. The notation in the fourth column is (p, E, q), where p is the prime used in Theorem 5.4.2, E is an elliptic curve, denoted using a Cremona label, and $q \neq p$ is a prime such that

$$\bigcap_{q' \leq q} \mathrm{Ker}(T'_q|_{\overline{W}} - a_{q'}(E)) = 0.$$

A_f	dim	$\ell \mid Ш(A_f)_?$	moddeg	(p, E, q)'s
551H	18	3	$2^? \cdot 13^2$	(2, 1102A1, -)
767E	23	3	2^{34}	(2, 1534B1, 3)
959D	24	3	$2^{32} \cdot 583673$	(2, 1918C1, 5), (7, 5369A1,2)
1337E	33	3	$2^{59} \cdot 71$	(2, 2674A1, 5)
1339G	30	3	$2^{48} \cdot 5776049$	(2, 2678B1, 3), (11, 14729A1,2)

REFERENCES

[Aga99b] A. Agashe, *On invisible elements of the Tate-Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. 328 (1999), no. 5, 369–374.

[AS02] A. Agashe and W. A. Stein, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory 97 (2002), no. 1, 171–185.

[AS05] A. Agashe and W. Stein, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank zero*, Math. Comp. 74 (2005), no. 249, 455–484 (electronic), With an appendix by J. Cremona and B. Mazur.

[ARS06] A. Agashe, K. A. Ribet and W. Stein, *The Manin constant*, to appear in Quarterly J. of Pure and Applied Math. volume in honor of J. Coates.

[AM69] M.F. Atiyah and I. G.Macdonald, *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., (1969).

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. 14 (2001), no. 4, 843–939 (electronic).

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. 24 (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993).

[CW06] M. Ciperiani, A. Wiles, *Solvable points on genus one curves*, preprint (2006).

[CFK06] C. David, J. Fearnly, H. Kisilevsky, *Vanishing of twisted L-functions of elliptic curves*, to appear in Experiment. Math.

[CM00] J. E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. 9 (2000), no. 1, 13–28.

[CV92] R. F. Coleman, J. F. Voloch, *Companion forms and Kodaira-Spencer theory*, Invent. Math., 110, (1992), 2, 263–281.

[CS01] B.Conrad, W. A. Stein, *Component groups of purely toric quotients*, Math. Res. Lett., 8, 5–6, (2001), 745–766.

[Cre] J. E. Cremona, *Tables of Elliptic Curves,* `http://www.maths.nott.ac.uk/personal/jec/ftp/data/`

[CR62] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1962, Pure and Applied Mathematics, Vol. XI.

[Fal86] G. Faltings, *Finiteness theorems for abelian varieties over number fields*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, Translated from the German original [Invent. Math. 73 (1983), no. 3, 349–366; ibid. 75 (1984), no. 2, 381] by Edward Shipz, pp. 9–27.

[Ka81] N. M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. 62 (1981), no. 3, 481–502.

[Kle01] T. Klenke, *Modular Varieties and Visibility*, Ph.D. thesis, Harvard University (2001).

[KS00] D. R. Kohel and W. A. Stein, *Component Groups of Quotients of* $J_0(N)$, Proc. ANTS-IV, Springer, 2000.

[KL89] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Algebra i Analiz 1 (1989), no. 5, 171–196.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math., 47, (1977), 33–186.

[Maz99] ———, *Visualizing elements of order three in the Shafarevich-Tate group*, Asian J. Math. 3 (1999), no. 1, 221–232, Sir Michael Atiyah: a great mathematician of the twentieth century.

[Mil72] J. S. Milne, *On the arithmetic of abelian varieties*, Invent. Math. 17 (1972), 177–190.

[Mil86] ———, *Arithmetic duality theorems*, Academic Press Inc., Boston, Mass., (1986), x+421.

[Rib83] K. A. Ribet, *Congruence relations between modular forms*, Proc. International Congress of Mathematicians, 503–514, (1983).

[Rib87] ———, *On the component groups and the Shimura subgroup of* $J_0(N)$, Séminaire de Théorie des Nombres, 1987–1988 (Talence, 1987–1988), Exp. No. 6, 10, Univ. Bordeaux I.

[Rib90a] ———, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math., 100 1990, no. 2, 431–476.

[Rib90b] ———, *Raising the levels of modular representations*, Séminaire de Théorie des Nombres, Paris 1987–88, Birkhäuser Boston, Boston, MA, 1990, pp. 259–271.

[Rib91] ———, *Lowering the levels of modular representations without multiplicity one*, International Mathematics Research Notices, (1991), 15–19.

[Rib92] ———, *Abelian varieties over* $\mathbf{Q}$ *and modular forms*, Algebra and topology 1992 (Taejŏn), Korea Adv. Inst. Sci. Tech., Taejŏn, 1992, pp. 53–79.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., 9, 143–232, Amer. Math. Soc., Providence, RI, (2001).

[Rub89] K. Rubin, *The work of Kolyvagin on the arithmetic of elliptic curves*, Arithmetic of complex manifolds (Erlangen, 1988), 128–136, Springer, Berlin, (1989).

[Se79] J-P. Serre, *Local fields*, Springer-Verlag, New York, (1979).

[Shi94] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, reprint of the 1971 original, Kan Memorial Lectures, 1, Princeton University Press, (1994).

[Ste00] W. A. Stein, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley (2000).

[Ste04] W. A. Stein, *Shafarevich-Tate Groups of Nonsquare Order*, Modular Curves and Abelian Varieties, Progress of Mathematics (2004), 277–289.

[Stu87] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin (1987), 275–280.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2), 141(3), (1995), 443–551.

Dimitar P. Jetchev
Department of Mathematics
University of California
Berkeley, CA 94720-3840
jetchev@math.berkeley.edu

William A. Stein
Department of Mathematics
University of Washington
Seattle, WA 98195-4350
wstein@u.washington.edu

25 Computation of p-Adic Heights and Log Convergence, with B. Mazur and J. Tate

Computation of p-Adic Heights and Log Convergence

In celebration of John Coates' 60th birthday

Barry Mazur, William Stein[1], John Tate

Received: September 9, 2005
Revised: May 2, 2006

Abstract.

This paper is about computational and theoretical questions regarding p-adic height pairings on elliptic curves over a global field K. The main stumbling block to computing them efficiently is in calculating, for each of the completions K_v at the places v of K dividing p, a *single quantity*: the value of the p-adic modular form $\mathbf{E}_2$ associated to the elliptic curve. Thanks to the work of Dwork, Katz, Kedlaya, Lauder and Monsky-Washnitzer we offer an efficient algorithm for computing these quantities, i.e., for computing the value of $\mathbf{E}_2$ of an elliptic curve. We also discuss the p-adic convergence rate of canonical expansions of the p-adic modular form $\mathbf{E}_2$ on the Hasse domain. In particular, we introduce a new notion of log convergence and prove that $\mathbf{E}_2$ is log convergent.

2000 Mathematics Subject Classification: 11F33, 11Y40, 11G50

Keywords and Phrases: p-adic heights, algorithms, p-adic modular forms, Eisenstein series, sigma-functions

[1] This material is based upon work supported by the National Science Foundation under Grant No. 0555776.

1 Introduction

Let p be an odd prime number, and E an elliptic curve over a global field K that has good ordinary reduction at p. Let L be any (infinite degree) Galois extension with a continuous injective homomorphism ρ of its Galois group to $\mathbf{Q}_p$. To the data (E, K, ρ), one associates[2] a canonical (bilinear, symmetric) (p-adic) height pairing

$$(\ ,\ )_\rho : E(K) \times E(K) \longrightarrow \mathbf{Q}_p.$$

Such pairings are of great interest for the arithmetic of E over K, and they arise specifically in p-adic analogues of the Birch and Swinnerton-Dyer conjecture.[3]

The goal of this paper is to discuss some computational questions regarding p-adic height pairings. The main stumbling block to computing them efficiently is in calculating, for each of the completions K_v at the places v of K dividing p, the value of the p-adic modular form $\mathbf{E}_2$ associated to the elliptic curve with a chosen Weierstrass form of good reduction over K_v.

We shall offer an algorithm for computing these quantities, i.e., for computing the value of $\mathbf{E}_2$ of an elliptic curve (that builds on the works of Katz and Kedlaya listed in our bibliography) and we also discuss the p-adic convergence rate of canonical expansions of the p-adic modular form $\mathbf{E}_2$ on the Hasse domain, where for $p \geq 5$ we view $\mathbf{E}_2$ as an infinite sum of classical modular forms divided by powers of the (classical) modular form $\mathbf{E}_{p-1}$, while for $p \leq 5$ we view it as a sum of classical modular forms divided by powers of $\mathbf{E}_4$.

We were led to our fast method of computing $\mathbf{E}_2$ by our realization that the more naive methods, of computing it by integrality or by approximations to it as function on the Hasse domain, were not practical, because the convergence is "logarithmic" in the sense that the nth convergent gives only an accuracy of $\log_p(n)$. We make this notion of log convergence precise in Part II, where we also prove that $\mathbf{E}_2$ is log convergent.

The reason why this constant $\mathbf{E}_2$ enters the calculation is because it is needed for the computation of the p-adic sigma function [MT91], which in turn is the critical element in the formulas for height pairings.

For example, let us consider the *cyclotomic* p-adic height pairing in the special case where $K = \mathbf{Q}$ and $p \geq 5$.

If $G_{\mathbf{Q}}$ is the Galois group of an algebraic closure of $\mathbf{Q}$ over $\mathbf{Q}$, we have the natural surjective continuous homomorphism $\chi : G_{\mathbf{Q}} \to \mathbf{Z}_p^*$ pinned down by the standard formula $g(\zeta) = \zeta^{\chi(g)}$ where $g \in G_{\mathbf{Q}}$ and ζ is any p-power root of unity. The p-adic logarithm $\log_p : \mathbf{Q}_p^* \to (\mathbf{Q}_p, +)$ is the unique group homomorphism with $\log_p(p) = 0$ that extends the homomorphism $\log_p : 1 + p\mathbf{Z}_p \to \mathbf{Q}_p$ defined by the usual power series of $\log(x)$ about 1. Explicitly, if $x \in \mathbf{Q}_p^*$, then

$$\log_p(x) = \frac{1}{p-1} \cdot \log_p(u^{p-1}),$$

[2] See [MT83], [Sch82] [Sch85], [Zar90], [Col91], [Nek93], [Pla94], [IW03], and [Bes04].

[3] See [Sch82], [Sch85] [MT83], [MT87], [PR03a]. See also the important recent work of Jan Nekovář [Nek03].

where $u = p^{-\operatorname{ord}_p(x)} \cdot x$ is the unit part of x, and the usual series for log converges at u^{p-1}.

The composition $(\frac{1}{p} \cdot \log_p) \circ \chi$ is a cyclotomic linear functional $G_{\mathbf{Q}} \to \mathbf{Q}_p$ which, in the body of our text, will be dealt with (thanks to class field theory) as the idele class functional that we denote $\rho_{\mathbf{Q}}^{\text{cycl}}$.

Let $\mathcal{E}$ denote the Néron model of E over $\mathbf{Z}$. Let $P \in E(\mathbf{Q})$ be a non-torsion point that reduces to $0 \in E(\mathbf{F}_p)$ and to the connected component of $\mathcal{E}_{\mathbf{F}_\ell}$ at all primes ℓ of bad reduction for E. Because $\mathbf{Z}$ is a unique factorization domain, any nonzero point $P = (x(P), y(P)) \in E(\mathbf{Q})$ can be written uniquely in the form $(a/d^2, b/d^3)$, where $a, b, d \in \mathbf{Z}$, $\gcd(a, d) = \gcd(b, d) = 1$, and $d > 0$. The function $d(P)$ assigns to P this square root d of the denominator of $x(P)$.

Here is the formula for the *cyclotomic* p-adic height of P, i.e., the value of

$$h_p(P) := -\frac{1}{2}(P, P)_p \in \mathbf{Q}_p$$

where $(\ ,\ )_p$ is the height pairing attached to $G_{\mathbf{Q}} \to \mathbf{Q}_p$, the cyclotomic linear functional described above:

$$h_p(P) = \frac{1}{p} \cdot \log_p \left(\frac{\sigma(P)}{d(P)} \right) \in \mathbf{Q}_p. \tag{1.1}$$

Here $\sigma = \sigma_p$ is the p-adic sigma function of [MT91] associated to the pair (E, ω). The σ-function depends only on (E, ω) and not on a choice of Weierstrass equation, and behaves like a modular form of weight -1, that is $\sigma_{E,c\omega} = c \cdot \sigma_{E,\omega}$. It is "quadratic" the sense that for any $m \in \mathbf{Z}$ and point Q in the formal group $E^f(\overline{\mathbf{Z}}_p)$, we have

$$\sigma(mQ) = \sigma(Q)^{m^2} \cdot f_m(Q), \tag{1.2}$$

where f_m is the mth division polynomial of E relative to ω (as in [MT91, App. 1]). The σ-function is "bilinear" in that for any $P, Q \in E^f(\mathbf{Z}_p)$, we have

$$\frac{\sigma(P - Q) \cdot \sigma(P + Q)}{\sigma^2(P) \cdot \sigma^2(Q)} = x(Q) - x(P). \tag{1.3}$$

See [MT91, Thm. 3.1] for proofs of the above properties of σ.

The height function h_p of (1.1) extends uniquely to a function on the full Mordell-Weil group $E(\mathbf{Q})$ that satisfies $h_p(nQ) = n^2 h_p(Q)$ for all integers n and $Q \in E(\mathbf{Q})$. For $P, Q \in E(\mathbf{Q})$, setting

$$(P, Q)_p = h_p(P) + h_p(Q) - h_p(P + Q),$$

we obtain a pairing on $E(\mathbf{Q})$. The *p-adic regulator* of E is the discriminant of the induced pairing on $E(\mathbf{Q})_{/\text{tor}}$ (well defined up to sign), and we have the following standard conjecture about this height pairing.

Conjecture 1.1. *The cyclotomic height pairing $(\ ,\ )_p$ is nondegenerate; equivalently, the p-adic regulator is nonzero.*

Remark 1.2. Height pairings attached to other p-adic linear functionals can be degenerate; in fact, given an elliptic curve defined over $\mathbf{Q}$ with good ordinary reduction at p, and K a quadratic imaginary field over which the Mordell-Weil group $E(K)$ is of odd rank, the p-adic anticyclotomic height pairing for E over K is *always* degenerate.

The p-adic σ function is the most mysterious quantity in (1.1). There are many ways to define σ, e.g., [MT91] contains 11 different characterizations of σ! We now describe a characterization that leads directly to an algorithm (see Algorithm 3.3) to compute $\sigma(t)$. Let

$$x(t) = \frac{1}{t^2} + \cdots \in \mathbf{Z}_p((t)) \tag{1.4}$$

be the formal power series that expresses x in terms of the local parameter $t = -x/y$ at infinity. The following theorem, which is proved in [MT91], uniquely determines σ and c.

Theorem 1.3. *There is exactly one odd function $\sigma(t) = t + \cdots \in t\mathbf{Z}_p[[t]]$ and constant $c \in \mathbf{Z}_p$ that together satisfy the differential equation*

$$x(t) + c = -\frac{d}{\omega}\left(\frac{1}{\sigma}\frac{d\sigma}{\omega}\right), \tag{1.5}$$

where ω is the invariant differential $dx/(2y + a_1x + a_3)$ associated with our chosen Weierstrass equation for E.

Remark 1.4. The condition that σ is odd and that the coefficient of t is 1 are essential.

In (1.1), by $\sigma(P)$ we mean $\sigma(-x/y)$, where $P = (x, y)$. We have thus given a complete definition of $h_p(Q)$ for any point $Q \in E(\mathbf{Q})$ and a prime $p \geq 5$ of good ordinary reduction for E.

1.1 The p-adic σ-function

The differential equation (1.5) leads to a slow algorithm to compute $\sigma(t)$ to any desired precision. This is Algorithm 3.3 below, which we now summarize. If we expand (1.5), we can view c as a formal variable and solve for $\sigma(t)$ as a power series with coefficients that are polynomials in c. Each coefficient of $\sigma(t)$ must be in $\mathbf{Z}_p$, so we obtain conditions on c modulo powers of p. Taking these together for many coefficients must eventually yield enough information to compute $c \pmod{p^n}$, for a given n, hence $\sigma(t) \pmod{p^n}$. This integrality algorithm is hopelessly slow in general.

Another approach to computing σ is to observe that, up to a constant, c is closely related to the value of a certain p-adic modular form. More precisely, suppose that E is given by a (not necessarily minimal) Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.6}$$

and let $\omega = dx/(2y + a_1 x + a_3)$. Let $x(t)$ be as in (1.4). Then the series

$$\wp(t) = x(t) + \frac{a_1^2 + 4a_2}{12} \in \mathbf{Q}((t)) \tag{1.7}$$

satisfies $(\wp')^2 = 4\wp^3 - g_2\wp - g_3$. In [MT91] we find[4] that

$$x(t) + c = \wp(t) - \frac{1}{12} \cdot \mathbf{E}_2(E, \omega), \tag{1.8}$$

where $\mathbf{E}_2(E, \omega)$ is the value of the Katz p-adic weight 2 Eisenstein series at (E, ω), and the equality is of elements of $\mathbf{Q}_p((t))$. Using the definition of $\wp(t)$ and solving for c, we find that

$$c = \frac{a_1^2 + 4a_2}{12} - \frac{1}{12}\mathbf{E}_2(E, \omega). \tag{1.9}$$

Thus computing c is equivalent to computing the p-adic number $\mathbf{E}_2(E, \omega)$. Having computed c to some precision, we then solve for σ in (1.5) using Algorithm 3.1 below.

1.2 p-adic analogues of the Birch and Swinnerton-Dyer conjecture

One motivation for this paper is to provide tools for doing computations in support of p-adic analogues of the BSD conjectures (see [MTT86]), especially when $E/\mathbf{Q}$ has rank at least 2. For example, in [PR03b], Perrin-Riou uses her results about the p-adic BSD conjecture in the supersingular case to prove that $Ш(E/\mathbf{Q})[p] = 0$ for certain p and elliptic curves E of rank > 1, for which the work of Kolyvagin and Kato does not apply.

Another motivation for this work comes from the study of the fine structure of Selmer modules. Let K be a number field and Λ the p-adic integral group ring of the Galois group of the maximal $\mathbf{Z}_p$-power extension of K. Making use of fundamental results of Nekovář [Nek03] and of Greenberg [Gre03] one can construct (see [RM05]) for certain elliptic curves defined over K, a skew-Hermitian matrix with coefficients in Λ from which one can read off a free Λ-resolution of the canonical Selmer Λ-module of the elliptic curve in question over K. To compute the entries of this matrix modulo the square of the augmentation ideal in Λ one must know *all* the p-adic height pairings of the elliptic curve over K. Fast algorithms for doing this provide us with an important first stage in the computation of free Λ-resolutions of Selmer Λ-modules.

The paper [GJP$^+$05] is about computational verification of the full Birch and Swinnerton-Dyer conjecture for specific elliptic curves E. There are many cases in which the rank of E is 1 and the upper bound on $\#Ш(E/\mathbf{Q})$ coming from Kolyvagin's Euler system is divisible by a prime $p \geq 5$ that also divides a Tamagawa number. In such cases, theorems of Kolyvagin and Kato combined

[4] There is a sign error in [MT91].

with explicit computation do not give a sufficiently sharp upper bound on $\#\text{Ш}(E/\mathbf{Q})$. However, it should be possible in these cases to compute p-adic heights and p-adic L-functions, and use results of Kato, Schneider, and others to obtain better bounds on $\#\text{Ш}(E/\mathbf{Q})$. Wuthrich and the second author (Stein) are writing a paper on this.

1.3 Sample computations

In Section 4 we illustrate our algorithms with curves of ranks $1, 2, 3, 4$ and 5, and two twists of $X_0(11)$ of rank 2.

Acknowledgement: It is a pleasure to thank Nick Katz for feedback that led to Section 3. We would also like to thank Mike Harrison for discussions about his implementation of Kedlaya's algorithm in Magma, Kiran Kedlaya for conversations about his algorithm, Christian Wuthrich for feedback about computing p-adic heights, Alan Lauder for discussions about computing $\mathbf{E}_2$ in families, and Fernando Gouvea for remarks about non-overconvergence of $\mathbf{E}_2$. We would also like to thank all of the above people for comments on early drafts of the paper. Finally, we thank Jean-Pierre Serre for the proof of Lemma 6.6.

Part I
Heights, σ-functions, and $\mathbf{E}_2$

2 The Formulas

In this section we give formulas for the p-adic height pairing in terms of the σ function. We have already done this over $\mathbf{Q}$ in Section 1. Let p be an (odd) prime number, K a number field, and E an elliptic curve over K with good ordinary reduction at all places of K above p. For any non-archimedean place w of K, let k_w denote the residue class field at w.

2.1 General global height pairings

By the *idele class* $\mathbf{Q}_p$*-vector space* of K let us mean

$$I(K) = \mathbf{Q}_p \otimes_{\mathbf{Z}} \left\{ \mathbf{A}_K^*/ \left(K^* \cdot \prod_{v \nmid p} \mathcal{O}_v^* \cdot \mathrm{C} \right) \right\},$$

where $\mathbf{A}_K^*$ is the group of ideles of K, and C denotes its connected component containing the identity. Class field theory gives us an identification $I(K) = \Gamma(K) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p$, where $\Gamma(K)$ is the Galois group of the maximal $\mathbf{Z}_p$-power extension of K. For every (nonarchimedean) place v of K, there is a natural homomorphism $\iota_v : K_v^* \to I(K)$.

For K-rational points $\alpha, \beta \in E(K)$ we want to give explicit formulas for an element that we might call the “universal” p-adic height pairing of α and β; denote it $(\alpha, \beta) \in I(K)$. If $\rho : I(K) \to \mathbf{Q}_p$ is any linear functional, then the *ρ-height pairing* is a symmetric bilinear pairing

$$(\ , \ )_\rho : E(K) \times E(K) \to \mathbf{Q}_p,$$

defined as the composition of the universal pairing with the linear functional ρ:

$$(\alpha, \beta)_\rho = \rho(\alpha, \beta) \in \mathbf{Q}_p.$$

We define the *ρ-height* of a point $\alpha \in E(K)$ by:

$$h_\rho(\alpha) = -\frac{1}{2}(\alpha, \alpha)_\rho \in \mathbf{Q}_p.$$

Of course, any such (nontrivial) linear functional ρ uniquely determines a $\mathbf{Z}_p$-extension, and we sometimes refer to the ρ-height pairing in terms of this $\mathbf{Z}_p$-extension. E.g., if ρ cuts out the cyclotomic $\mathbf{Z}_p$-extension, then the ρ-height pairing is a normalization of the *cyclotomic height pairing* that has, for the rational field, already been discussed in the introduction.

If K is quadratic imaginary, and ρ is the anti-cyclotomic linear functional, meaning that it is the unique linear functional (up to normalization) that has the property that $\rho(\bar{x}) = -\rho(x)$ where $\bar{x}$ is the complex conjugate of x, then we will be presently obtaining explicit formulas for this anti-cyclotomic height pairing.

We will obtain a formula for $(\alpha, \beta) \in I(K)$ by defining, for every nonarchimedean place, v, of K a “local height pairing,” $(\alpha, \beta)_v \in K_v^*$. These local pairings will be very sensitive to some auxiliary choices we make along the way, but for a fixed α and β the local height pairings $(\alpha, \beta)_v$ will vanish for all but finitely many places v; the global height is the sum of the local ones and will be independent of all the choices we have made.

2.2 Good representations

Let $\alpha, \beta \in E(K)$. By a *good representation* of the pair α, β we mean that we are given a four-tuple of points (P, Q, R, S) in $E(K)$ (or, perhaps, in $E(K')$ where K'/K is a number field extension of K) such that

- α is the divisor class of the divisor $[P] - [Q]$ of E, and β is the divisor class of the divisor $[R] - [S]$,
- P, Q, R, S are four distinct points,
- for each $v \mid p$ all four points P, Q, R, S specialize to the same point on the fiber at v of the Néron model of E.
- at all places v of K the points P, Q, R, S specialize to the same component of the fiber at v of the Néron model of E.

We will show how to erase these special assumptions later, but for now, let us assume all this, fix a choice of a good representation, P, Q, R, S, of (α, β) as above, and give the formulas in this case.

2.3 Local height pairings when $v \mid p$

Let σ_v be the canonical p-adic σ-function attached to the elliptic curve E over K_v given in Weierstrass form. We may view σ_v as a mapping from $E_1(K_v)$ to K_v^*, where $E_1(K_v)$ is the kernel of the reduction map $E(K_v) \to E(k_v)$, and $E(k_v)$ denotes the group of points on the reduction of E modulo v. Define $(\alpha, \beta)_v \in K_v^*$ by the formula,

$$(\alpha, \beta)_v = \frac{\sigma_v(P-R)\sigma_v(Q-S)}{\sigma_v(P-S)\sigma_v(Q-R)} \ \in \ K_v^*.$$

The dependence of σ on the Weierstrass equation is through the differential $\omega = dx/(2y + a_1 x + a_3)$, and $\sigma_{c\omega} = c\sigma_\omega$, so this depends upon the choice of P, Q, R, S, but does not depend on the choice of Weierstrass equation for E.

2.4 Local height pairings when $v \nmid p$

First let x denote the "x-coordinate" in some minimal Weierstrass model for A at v. Define for a point T in $E(K_v)$ the rational number $\lambda_v(T)$ to be *zero* if $x(T) \in \mathcal{O}_v$, and to be $-\frac{1}{2}v(x(T))$ if $x(T) \notin \mathcal{O}_v$.

Next, choose a uniformizer π_v of K_v and define:

$$\tilde{\sigma}_v(T) = \pi_v^{\lambda_v(T)},$$

the square of which is in K_v^*. We think of $\tilde{\sigma}_v$ as a rough replacement for σ_v in the following sense. The v-adic valuation of $\tilde{\sigma}_v$ is the same as v-adic valuation of the v-adic sigma function (if such a function is definable at v) and therefore, even if σ_v cannot be defined, $\tilde{\sigma}_v$ is a perfectly serviceable substitute at places v at which our p-adic idele class functionals ρ are necessarily unramified, and therefore sensitive only to the v-adic valuation.

For $v \nmid p$, put:

$$(\alpha, \beta)_v = \frac{\tilde{\sigma}_v(P-R)\tilde{\sigma}_v(Q-S)}{\tilde{\sigma}_v(P-S)\tilde{\sigma}_v(Q-R)}.$$

The square of this is in K_v^*. However, note that $\pi_v^{\lambda_v(T)}$ really means $\sqrt{\pi_v}^{2\lambda_v(T)}$, for a fixed choice of $\sqrt{\pi_v}$ and that the definition of $(\alpha, \beta)_v$ is independent of the choice of square root and therefore that $(\alpha, \beta)_v$, not only its square, is in K_v^*.

Our local height $(\alpha, \beta)_v$, depends upon the choice of P, Q, R, S and of the uniformizer π_v.

2.5 How the local heights change, when we change our choice of divisors

Let $\beta \in E(K)$ be represented by both $[R]-[S]$ and $[R']-[S']$. Let $\alpha \in E(K)$ be represented by $[P]-[Q]$. Moreover let both four-tuples P,Q,R,S and P,Q,R',S' satisfy the *good representation* hypothesis described at the beginning of Section 2.2. Since, by hypothesis, $[R]-[S]-[R']+[S']$ is linearly equivalent to zero, there is a rational function f whose divisor of zeroes and poles is

$$(f) = [R]-[S]-[R']+[S'].$$

If v is a nonarchimedean place of K define $(\alpha,\beta)_v$ to be as defined in the previous sections using the choice of four-tuple of points P,Q,R,S, (and of uniformizer π_v when $v \nmid p$). Similarly, define $(\alpha,\beta)'_v$ to be as defined in the previous sections using the choice of four-tuple of points P,Q,R',S', (and of uniformizer π_v when $v \nmid p$).

Proposition 2.1. *1. If $v \mid p$ then*

$$(\alpha,\beta)_v = \frac{f(P)}{f(Q)} \cdot (\alpha,\beta)'_v \;\in\; K_v^*.$$

2. If $v \nmid p$ then there is a unit u in the ring of integers of K_v such that

$$(\alpha,\beta)_v^2 = u \cdot \left(\frac{f(P)}{f(Q)} \cdot (\alpha,\beta)'_v\right)^2 \;\in\; K_v^*.$$

2.6 The global height pairing more generally

We can then form the sum of local terms to define the global height

$$(\alpha,\beta) = \quad \frac{1}{2}\sum_v \iota_v((\alpha,\beta)_v^2) \;\in\; I(K).$$

This definition is independent of any of the (good representation) choices P,Q,R,S and the π_v's made. It is independent of the choice of π_v's because the units in the ring of integers of K_v is in the kernel of ι_v if $v \nmid p$. It is independent of the choice of P,Q,R,S because by the previous proposition, a change (an allowable one, given our hypotheses) of P,Q,R,S changes the value of (α,β) by a factor that is a principal idele, which is sent to zero in $I(K)$.

What if, though, our choice of P,Q,R,S does *not* have the property that α and β reduce to the same point in the Néron fiber at v for all $v \mid p$, or land in the same connected component on each fiber of the Néron model? In this case the pair α,β do not have a *good representation.* But replacing α,β by $m\cdot\alpha, n\cdot\beta$ for sufficiently large positive integers m,n we can guarantee that the pair $m\cdot\alpha, n\cdot\beta$ does possess a good representation, and obtain formulas for (α,β) by:

$$(\alpha,\beta) = \frac{1}{mn}(m\cdot\alpha, n\cdot\beta).$$

Note in passing that to compute the global height pairing (α,α) for a non-torsion point $\alpha \in E(K)$ that specializes to 0 in the Néron fiber at v for all $v \mid p$, and that lives in the connected component containing the identity in all Néron fibers, we have quite a few natural choices of *good representations*. For example, for positive integers $m \neq n$, take

$$P = (m+1)\cdot\alpha;\ Q = m\cdot\alpha;\ R = (n+1)\cdot\alpha;\ S = n\cdot\alpha.$$

Then for any p-adic idele class functional ρ the global ρ-height pairing $(\alpha,\alpha)_\rho$ is given by

$$\sum_{v\ \mid\ p} \rho_v \left\{ \frac{\sigma_v((m-n)\alpha)^2}{\sigma_v((m-n+1)\alpha)\cdot\sigma_v((m-n-1)\alpha)} \right\} \\ + \sum_{v\ \nmid\ p} \rho_v \left\{ \frac{\tilde{\sigma}_v((m-n)\alpha)^2}{\tilde{\sigma}_v((m-n+1)\alpha)\cdot\tilde{\sigma}_v((m-n-1)\alpha)} \right\},$$

which simplifies to

$$(2(m-n)^2 - (m-n+1)^2 - (m-n-1)^2)\cdot \left\{ \sum_{v\ \mid\ p} \rho_v\sigma_v(\alpha) + \sum_{v\ \nmid\ p} \rho_v\tilde{\sigma}_v(\alpha) \right\}.$$

Since $(2(m-n)^2 - (m-n+1)^2 - (m-n-1)^2) = -2$ we have the formula

$$h_\rho(\alpha) = -\frac{1}{2}(\alpha,\alpha)_\rho$$

quoted earlier.

2.7 Formulas for the ρ-height

For each v, let σ_v be the canonical p-adic σ-function of E over K_v given in Weierstrass form. Suppose $P \in E(K)$ is a (non-torsion) point that reduces to 0 in $E(k_v)$ for each $v \mid p$, and to the connected component of all special fibers of the Néron model of E. Locally at each place w of K, we have a denominator $d_w(P)$, well defined up to units.

We have

$$h_\rho(P) = \sum_{v\ \mid\ p} \rho_v(\sigma_v(P)) - \sum_{w\ \nmid\ p} \rho_w(d_w(P)).$$

Note that h_ρ is quadratic because of the quadratic property of σ from (1.2), and the h_ρ-pairing is then visibly bilinear. See also property (1.3).

2.8 Cyclotomic p-adic heights

The idele class $\mathbf{Q}_p$-vector space $I(\mathbf{Q})$ attached to $\mathbf{Q}$ is canonically isomorphic to $\mathbf{Q}_p \otimes \mathbf{Z}_p^*$. Composition of this canonical isomorphism with the mapping $1 \times \frac{1}{p}\log_p$ induces an isomorphism

$$\rho_{\mathrm{cycl}}^{\mathbf{Q}} : I(\mathbf{Q}) = \mathbf{Q}_p \otimes \mathbf{Z}_p^* \xrightarrow{\cong} \mathbf{Q}_p.$$

For K any number field, consider the homomorphism on idele class $\mathbf{Q}_p$-vector spaces induced by the norm $N_{K/\mathbf{Q}} : I(K) \to I(\mathbf{Q})$, and define

$$\rho_{\mathrm{cycl}}^{K} : I(K) \to \mathbf{Q}_p$$

as the composition

$$\rho_{\mathrm{cycl}}^{K} = \rho_{\mathrm{cycl}}^{\mathbf{Q}} \circ N_{K/\mathbf{Q}}.$$

By the *cyclotomic height pairing* for an elliptic curve E over K (of good ordinary reduction at all places v of K above p) we mean the ρ_{cycl}^{K}-height pairing $E(K) \times E(K) \to \mathbf{Q}_p$. We put

$$h_p(P) = h_{\rho_{\mathrm{cycl}}^{K}}(P)$$

for short. Here is an explicit formula for it.

$$h_p(P) = \frac{1}{p} \cdot \left(\sum_{v|p} \log_p(N_{K_v/\mathbf{Q}_p}(\sigma_v(P))) - \sum_{w\nmid p} \mathrm{ord}_w(d_w(P)) \cdot \log_p(\#k_w) \right).$$

If we assume that P lies in a sufficiently small (finite index) subgroup of $E(K)$ (see [Wut04, Prop. 2]), then there will be a global choice of denominator $d(P)$, and the formula simplifies to

$$h_p(P) = \frac{1}{p} \cdot \log_p \left(\prod_{v|p} N_{K_v/\mathbf{Q}_p} \left(\frac{\sigma_v(P)}{d(P)} \right) \right).$$

2.9 Anti-cyclotomic p-adic heights

Let K be a quadratic imaginary field in which p splits as $(p) = \pi \cdot \bar{\pi}$. Suppose $\rho : \mathbf{A}_K^*/K^* \to \mathbf{Z}_p$ is a nontrivial *anti-cyclotomic* idele class character, meaning that if $\mathbf{c} : \mathbf{A}_K^*/K^* \to \mathbf{A}_K^*/K^*$ denotes the involution of the idele class group induced by complex conjugation $x \mapsto \bar{x}$ in K, then $\rho \cdot \mathbf{c} = -\rho$. Then the term

$$\sum_{v \ | \ p} \rho_v(\sigma_v(P))$$

in the formula for the ρ-height at the end of Section 2.7 is just

$$\sum_{v \ | \ p} \rho_v(\sigma_v(P)) = \rho_\pi(\sigma_\pi(P)) - \rho_\pi(\sigma_\pi(\bar{P})),$$

so we have the following formula for the ρ-height of P:

$$h_\rho(P) = \rho_\pi(\sigma_\pi(P)) - \rho_\pi(\sigma_\pi(\bar{P})) - \sum_{w \nmid p} \rho_w(d_w(P)).$$

Remark 2.2. The Galois equivariant property of the p-adic height pairing implies that if P is a $\mathbf{Q}$-rational point, its anti-cyclotomic height is 0. Specifically, let K/k be any Galois extension of number fields, with Galois group $G = \mathrm{Gal}(K/k)$. Let $V = V(K)$ be the $\mathbf{Q}_p$-vector space (say) defined as $(G_K)^{\mathrm{ab}} \otimes \mathbf{Q}_p$, so that V is naturally a G-representation space. Let E be an elliptic curve over k and view the Mordell-Weil group $E(K)$ as equipped with its natural G-action. Then (if p is a good ordinary prime for E) we have the p-adic height pairing

$$\langle P, Q\rangle \in V,$$

for $P, Q \in E(K)$ and we have Galois equivariance,

$$\langle g \cdot P, g \cdot Q\rangle = g \cdot \langle P, Q\rangle,$$

for any g in the Galois group.

Put $k = \mathbf{Q}$, K/k a quadratic imaginary field. Then V is of dimension two, with $V = V^+ \oplus V^-$ each of the $V^\pm$ being of dimension one, with the action of complex conjugation, $g \in G$ on $V^\pm$ being given by the sign; so that V^+ corresponds to the cyclotomic $\mathbf{Z}_p$-extension and V^- corresponds to the anticyclotomic $\mathbf{Z}_p$- extension. In the notation above, the anticyclotomic height of P and Q is just $\langle g \cdot P, g \cdot Q\rangle^-$ where the superscript $-$ means projection to V^-. Suppose that $P \in E(\mathbf{Q})$, so that $g \cdot P = P$. Then we have by Galois equivariance

$$\langle P, P\rangle^- = \langle g \cdot P, g \cdot P\rangle^- = -\langle P, P\rangle^-,$$

so $\langle P, P\rangle^- = 0$. More generally, the anticyclotomic height is zero as a pairing on either $E(K)^+ \times E(K)^+$ or $E(K)^- \times E(K)^-$ and can only be nonzero on $E(K)^+ \times E(K)^-$. If $E(K)$ is of odd rank, then the ranks of $E(K)^+$ and $E(K)^-$ must be different, which obliges the pairing on $E(K)^+ \times E(K)^-$ to be either left-degenerate or right-degenerate (or, of course, degenerate on both sides). Rubin and the first author conjecture that it is nondegenerate on one side (the side, of course having smaller rank); for more details see, e.g., [MR04, Conj. 11].

3 The Algorithms

Fix an elliptic curve E over $\mathbf{Q}$ and a good ordinary prime $p \geq 5$. In this section we discuss algorithms for computing the cyclotomic p-adic height of elements of $E(\mathbf{Q})$.

3.1 Computing the p-adic σ-function

First we explicitly solve the differential equation (1.5). Let $z(t)$ be the formal logarithm on E, which is given by $z(t) = \int \frac{\omega}{dt} = t + \cdots$ (here the symbol $\int$

means formal integration with 0 constant term). There is a unique function $F(z) \in \mathbf{Q}((z))$ such that $t = F(z(t))$. Set $x(z) = x(F(z))$. Rewrite (1.5) as

$$x(z) + c = -\frac{d}{\omega}\left(\frac{d\log(\sigma)}{\omega}\right). \tag{3.1}$$

A crucial observation is that

$$x(z) + c = \frac{1}{z^2} - \frac{a_1^2 + 4a_2}{12} + c + \cdots ;$$

in particular, the coefficient of $1/z$ in the expansion of $g(z) = x(z) + c$ is 0.

Since $z = \int(\omega/dt)$ we have $dz = (\omega/dt)dt = \omega$, hence $dz/\omega = 1$, so

$$-\frac{d}{\omega}\left(\frac{d\log(\sigma)}{\omega}\right) = -\frac{dz}{\omega}\frac{d}{dz}\left(\frac{d\log(\sigma)}{\omega}\right) = -\frac{d}{dz}\left(\frac{d\log(\sigma)}{dz}\right). \tag{3.2}$$

Write $\sigma(z) = z\sigma_0(z)$ where $\sigma_0(z)$ has nonzero constant term. Then

$$-\frac{d}{dz}\left(\frac{d\log(\sigma)}{dz}\right) = \frac{1}{z^2} - \frac{d}{dz}\left(\frac{d\log(\sigma_0)}{dz}\right). \tag{3.3}$$

Thus combining (3.1)–(3.3) and changing sign gives

$$\frac{1}{z^2} - x(z) - c = \frac{d}{dz}\left(\frac{d\log(\sigma_0)}{dz}\right).$$

This is particularly nice, since $g(z) = \frac{1}{z^2} - x(z) - c \in \mathbf{Q}[[z]]$. We can thus solve for $\sigma_0(z)$ by formally integrating twice and exponentiating:

$$\sigma_0(z) = \exp\left(\int\int g(z)dzdz\right),$$

where we choose the constants in the double integral to be 0, so $\int\int g = 0 + 0z + \cdots$. Using (1.8) we can rewrite $g(z)$ in terms of $e_2 = \mathbf{E}_2(E,\omega)$ and $\wp(z)$ as

$$g(z) = \frac{1}{z^2} - (x(z) + c) = \frac{1}{z^2} - \wp(z) + \frac{e_2}{12}.$$

Combining everything and using that $\sigma(z) = z\sigma_0(z)$ yields

$$\sigma(z) = z \cdot \exp\left(\int\int\left(\frac{1}{z^2} - \wp(z) + \frac{e_2}{12}\right)dzdz\right),$$

Finally, to compute $\sigma(t)$ we compute $\sigma(z)$ and obtain $\sigma(t)$ as $\sigma(z(t))$.

We formalize the resulting algorithm below.

Algorithm 3.1 (The Canonical p-adic Sigma Function). Given an elliptic curve E over $\mathbf{Q}$, a good ordinary prime p for E, and an approximation e_2 for $\mathbf{E}_2(E,\omega)$, this algorithm computes an approximation to $\sigma(t) \in \mathbf{Z}_p[[t]]$.

1. [Compute Formal Log] Compute the formal logarithm $z(t) = t + \cdots \in \mathbf{Q}((t))$ using that
$$z(t) = \int \frac{dx/dt}{2y(t) + a_1x(t) + a_3}, \qquad \text{(0 constant term)} \qquad (3.4)$$
where $x(t) = t/w(t)$ and $y(t) = -1/w(t)$ are the local expansions of x and y in terms of $t = -x/y$, and $w(t) = \sum_{n\geq 0} s_n t^n$ is given by the following explicit inductive formula (see, e.g., [Blu, pg. 18]):
$$s_0 = s_1 = s_2 = 0, \qquad s_3 = 1, \qquad \text{and for } n \geq 4,$$
$$s_n = a_1 s_{n-1} + a_2 s_{n-2} + a_3 \sum_{i+j=n} s_i s_j + a_4 \sum_{i+j=n-1} s_i s_j + a_6 \sum_{i+j+k=n} s_i s_j s_k.$$
2. [Reversion] Using a power series "reversion" (functional inverse) algorithm, find the unique power series $F(z) \in \mathbf{Q}[[z]]$ such that $t = F(z)$. Here F is the reversion of z, which exists because $z(t) = t + \cdots$.
3. [Compute $\wp$] Compute $\alpha(t) = x(t) + (a_1^2 + 4a_2)/12 \in \mathbf{Q}[[t]]$, where the a_i are as in (1.6). Then compute the series $\wp(z) = \alpha(F(z)) \in \mathbf{Q}((z))$.
4. [Compute $\sigma(z)$] Set $g(z) = \frac{1}{z^2} - \wp(z) + \frac{e_2}{12} \in \mathbf{Q}_p((z))$, and compute
$$\sigma(z) = z \cdot \exp\left(\int\int g(z)dzdz\right) \in \mathbf{Q}_p[[z]].$$
5. [Compute $\sigma(t)$] Set $\sigma(t) = \sigma(z(t)) \in t \cdot \mathbf{Z}_p[[t]]$, where $z(t)$ is the formal logarithm computed in Step 1. Output $\sigma(t)$ and terminate.

3.2 Computing $\mathbf{E}_2(E, \omega)$ using cohomology

This section is about a fast method of computation of $\mathbf{E}_2(E, \omega)$ for individual ordinary elliptic curves, "one at a time". The key input is [Kat73, App. 2] (see also [Kat76]), which gives an interpretation of $\mathbf{E}_2(E, \omega)$ as the "direction" of the unit root eigenspace (cf. formula A.2.4.1 of [Kat73, App. 2]) of Frobenius acting on the one-dimensional de Rham cohomology of E.

Concretely, consider an elliptic curve E over $\mathbf{Z}_p$ with good ordinary reduction. Assume that $p \geq 5$. Fix a Weierstrass equation for E of the form $y^2 = 4x^3 - g_2x - g_3$, The differentials $\omega = dx/y$ and $\eta = xdx/y$ form a $\mathbf{Z}_p$-basis for the first p-adic de Rham cohomology group H^1 of E, and we wish to compute the matrix F of absolute Frobenius with respect to this basis. Frobenius is $\mathbf{Z}_p$-linear, since we are working over $\mathbf{Z}_p$; if we were working over the Witt vectors of $\mathbf{F}_q$, then Frobenius would only be semi-linear.

We explicitly calculate F (to a specified precision) using Kedlaya's algorithm, which makes use of Monsky-Washnitzer cohomology of the affine curve $E - \mathcal{O}$. Kedlaya designed his algorithm for computation of zeta functions of hyperelliptic curves over finite fields. An intermediate step in Kedlaya's algorithm is computation of the matrix of absolute Frobenius on p-adic de Rham cohomology, via Monsky-Washnitzer cohomology. For more details see [Ked01] and [Ked03]. For recent formulations and applications of fast algorithms to compute Frobenius eigenvalues, see [LW02].

Now that we have computed F, we deduce $\mathbf{E}_2(E,\omega)$ as follows. The unit root subspace is a direct factor, call it U, of H^1, and we know that a complementary direct factor is the $\mathbf{Z}_p$ span of ω. We also know that $F(\omega)$ lies in $p\,\mathrm{H}^1$, and this tells us that, mod p^n, the subspace U is the span of $F^n(\eta)$. Thus if for each n, we write $F^n(\eta) = a_n\omega + b_n\eta$, then b_n is a unit (congruent (mod p) to the nth power of the Hasse invariant) and $\mathbf{E}_2(E,\omega) \equiv -12a_n/b_n \pmod{p^n}$. Note that a_n and b_n are the entries of the second column of the matrix F^n.

Algorithm 3.2 (Evaluation of $\mathbf{E}_2(E,\omega)$). Given an elliptic curve over $\mathbf{Q}$ and a good ordinary prime $p \geq 5$, this algorithm approximates $\mathbf{E}_2(E,\omega) \in \mathbf{Z}_p$ modulo p^n.

1. [Invariants] Let c_4 and c_6 be the c-invariants of a minimal model of E. Set
$$a_4 = -\frac{c_4}{2^4\cdot 3} \qquad \text{and} \qquad a_6 = -\frac{c_6}{2^5\cdot 3^3}.$$
2. [Kedlaya] Apply Kedlaya's algorithm to the hyperelliptic curve $y^2 = x^3 + a_4x + a_6$ (which is isomorphic to E) to obtain the matrix F (modulo p^n) of the action of absolute Frobenius on the basis
$$\omega = \frac{dx}{y}, \qquad \eta = \frac{xdx}{y}.$$
We view F as acting from the left.
3. [Iterate Frobenius] Compute the second column $\begin{pmatrix} a \\ b \end{pmatrix}$ of F^n, so $\mathrm{Frob}^n(\eta) = a\omega + b\eta$.
4. [Finished] Output $-12a/b$ (which is a number modulo p^n, since b is a unit).

3.3 Computing $\mathbf{E}_2(E,\omega)$ using integrality

The algorithm in this section is more elementary than the one in Section 3.2, and is directly motivated by Theorem 1.3. In practice it is very slow, except if p is small (e.g., $p=5$) and we only require $\mathbf{E}_2(E,\omega)$ to very low precision. Our guess is that it should be exponentially hard to compute a quantity using a log convergent series for it, and that this "integrality" method is essentially the same as using log convergent expansions.

Let c be an indeterminate and in view of (1.9), write $e_2 = -12c + a_1^2 + 4a_2 \in \mathbf{Q}[c]$. If we run Algorithm 3.1 with this (formal) value of e_2, we obtain a series $\sigma(t,c) \in \mathbf{Q}[c][[t]]$. For each prime $p \geq 5$, Theorem 1.3 implies that there is a unique choice of $c_p \in \mathbf{Z}_p$ such that $\sigma(t,c_p) = t + \cdots \in t\mathbf{Z}_p[[t]]$

is odd. Upon fixing a prime p, we compute the coefficients of $\sigma(t,c)$, which are polynomials in $\mathbf{Q}[c]$; integrality of $\sigma(t,c_p)$ then imposes conditions that together must determine c_p up to some precision, which depends on the number of coefficients that we consider. Having computed c_p to some precision, we recover $\mathbf{E}_2(E,\omega)$ as $-12c_p + a_1^2 + 4a_2$. We formalize the above as an algorithm.

Algorithm 3.3 (Integrality). Given an elliptic curve over $\mathbf{Q}$ and a good ordinary prime $p \geq 5$, this algorithm approximates the associated p-adic σ-function.

1. [Formal Series] Use Algorithm 3.1 with $e_2 = -12c + a_1^2 + 4a_2$ to compute $\sigma(t) \in \mathbf{Q}[c][[t]]$ to some precision.
2. [Approximate c_p] Obtain constraints on c using that the coefficients of σ must be in $\mathbf{Z}_p$. These determine c to some precision. (For more details see the example in Section 4.1).

3.4 Computing cyclotomic p-adic heights

Finally we give an algorithm for computing the cyclotomic p-adic height $h_p(P)$ that combines Algorithm 3.2 with the discussion elsewhere in this paper. We have computed σ and h_p in numerous cases using the algorithm described below, and implementations of the "integrality" algorithm described above, and the results match.

Algorithm 3.4 (The p-adic Height). Given an elliptic curve E over $\mathbf{Q}$, a good ordinary prime p, and a non-torsion element $P \in E(\mathbf{Q})$, this algorithm approximates the p-adic height $h_p(P) \in \mathbf{Q}_p$.

1. [Prepare Point] Compute a positive integer m such that mP reduces to $\mathcal{O} \in E(\mathbf{F}_p)$ and to the connected component of $\mathcal{E}_{\mathbf{F}_\ell}$ at all bad primes ℓ. For example, m could be the least common multiple of the Tamagawa numbers of E and $\#E(\mathbf{F}_p)$. Set $Q = mP$ and write $Q = (x, y)$.
2. [Denominator] Let d be the positive integer square root of the denominator of x.
3. [Compute σ] Approximate $\sigma(t)$ using Algorithm 3.1 together with either Algorithm 3.2 or Algorithm 3.3, and set $s = \sigma(-x/y) \in \mathbf{Q}_p$.
4. [Height] Compute $h_p(Q) = \frac{1}{p}\log_p\left(\frac{s}{d}\right)$, then $h_p(P) = \frac{1}{m^2} \cdot h_p(Q)$. Output $h_p(P)$ and terminate.

4 Sample Computations

We did the calculations in this section using SAGE [SJ05] and Magma [BCP97]. In particular, SAGE includes an optimized implementation due to J. Balakrishnan, R. Bradshaw, D. Harvey, Y. Qiang, and W. Stein of our algorithm for computing p-adic heights for elliptic curves over $\mathbf{Q}$. This implementation includes further tricks, e.g., for series manipulation, which are not described in this paper.

4.1 The rank one curve of conductor 37

Let E be the rank 1 curve $y^2 + y = x^3 - x$ of conductor 37. The point $P = (0, 0)$ is a generator for $E(\mathbf{Q})$. We illustrate the above algorithms in detail by computing the p-adic height of P for the good ordinary prime $p = 5$. The steps of Algorithm 3.4 are as follows:

1. [Prepare Point] The component group of $\mathcal{E}_{\mathbf{F}_{37}}$ is trivial. The group $E(\mathbf{F}_5)$ has order 8 and the reduction of P to $E(\mathbf{F}_5)$ also has order 8, so let
$$Q = 8P = \left(\frac{21}{25}, -\frac{69}{125}\right).$$
2. [Denominator] We have $d = 5$.
3. [Compute σ] We illustrate computation of $\sigma(t)$ using both Algorithm 3.2 and Algorithm 3.3.
   (a) [Compute $\sigma(t, c)$] We use Algorithm 3.1 with $e_2 = 12c - a_1^2 - 4a_2$ to compute σ as a series in t with coefficients polynomials in c, as follows:
      i. [Compute Formal Log] Using the recurrence, we find that
      $$w(t) = t^3 + t^6 - t^7 + 2t^9 - 4t^{10} + 2t^{11} + 5t^{12} - 5t^{13} + 5t^{14} + \cdots$$
      Thus
      $$x(t) = t^{-2} - t + t^2 - t^4 + 2t^5 - t^6 - 2t^7 + 6t^8 - 6t^9 - 3t^{10} + \cdots$$
      $$y(t) = -t^{-3} + 1 - t + t^3 - 2t^4 + t^5 + 2t^6 - 6t^7 + 6t^8 + 3t^9 + \cdots$$
      so integrating (3.4) we see that the formal logarithm is
      $$z(t) = t + \frac{1}{2}t^4 - \frac{2}{5}t^5 + \frac{6}{7}t^7 - \frac{3}{2}t^8 + \frac{2}{3}t^9 + 2t^{10} - \frac{60}{11}t^{11} + 5t^{12} + \cdots$$
      ii. [Reversion] Using reversion, we find F with $F(z(t)) = t$:
      $$F(z) = z - \frac{1}{2}z^4 + \frac{2}{5}z^5 + \frac{1}{7}z^7 - \frac{3}{10}z^8 + \frac{2}{15}z^9 - \frac{1}{28}z^{10} + \frac{54}{385}z^{11} + \cdots$$
      iii. [Compute $\wp$] We have $a_1 = a_2 = 0$, so
      $$\alpha(t) = x(t) + (a_1^2 + 4a_2)/12 = x(t),$$
      so
      $$\wp(z) = x(F(z)) = z^{-2} + \frac{1}{5}z^2 - \frac{1}{28}z^4 + \frac{1}{75}z^6 - \frac{3}{1540}z^8 + \cdots$$
      Note that the coefficient of z^{-1} is 0 and all exponents are even.

iv. [Compute $\sigma(t,c)$] Noting again that $a_1 = a_2 = 0$, we have

$$\begin{aligned} g(z,c) &= \frac{1}{z^2} - \wp(z) + \frac{12c - a_1^2 - 4a_2}{12} \\ &= c - \frac{1}{5}z^2 + \frac{1}{28}z^4 - \frac{1}{75}z^6 + \frac{3}{1540}z^8 - \frac{1943}{3822000}z^{10} + \cdots \end{aligned}$$

Formally integrating twice and exponentiating, we obtain

$$\begin{aligned} \sigma(z,c) &= z \cdot \exp\left(\int\int g(z,c)dzdz\right) \\ &= z \cdot \exp\Big(\frac{c}{2} \cdot z^2 - \frac{1}{60}z^4 + \frac{1}{840}z^6 - \frac{1}{4200}z^8 + \frac{1}{46200}z^{10} \\ &\qquad - \frac{1943}{504504000}z^{12} + \cdots\Big) \\ &= z + \frac{1}{2}cz^3 + \left(\frac{1}{8}c^2 - \frac{1}{60}\right)z^5 + \left(\frac{1}{48}c^3 - \frac{1}{120}c + \frac{1}{840}\right)z^7 + \\ &\qquad \left(\frac{1}{384}c^4 - \frac{1}{480}c^2 + \frac{1}{1680}c - \frac{1}{10080}\right)z^9 + \cdots \end{aligned}$$

Finally,

$$\begin{aligned} \sigma(t) = \sigma(z(t)) = t + \frac{1}{2}ct^3 + \frac{1}{2}t^4 + \left(\frac{1}{8}c^2 - \frac{5}{12}\right)t^5 + \frac{3}{4}ct^6 + \\ \left(\frac{1}{48}c^3 - \frac{73}{120}c + \frac{103}{120}\right)t^7 + \cdots \end{aligned}$$

(b) [Approximate] The first coefficient of $\sigma(t)$ that gives integrality information is the coefficient of t^7. Since

$$\frac{1}{48}c^3 - \frac{73}{120}c + \frac{103}{120} \in \mathbf{Z}_5,$$

multiplying by 5 we see that

$$\frac{5}{48}c^3 - \frac{73}{24}c + \frac{103}{24} \equiv 0 \pmod{5}.$$

Thus

$$c \equiv \frac{103}{24} \cdot \frac{24}{73} \equiv 1 \pmod{5}.$$

The next useful coefficient is the coefficient of t^{11}, which is

$$\frac{1}{3840}c^5 - \frac{169}{2880}c^3 + \frac{5701}{6720}c^2 + \frac{127339}{100800}c - \frac{40111}{7200}$$

Multiplying by 25, reducing coefficients, and using integrality yields the congruence

$$10c^5 + 5c^3 + 20c^2 + 2c + 3 \equiv 0 \pmod{25}.$$

Writing $c = 1+5d$ and substituting gives the equation $10d+15 \equiv 0 \pmod{25}$, so $2d + 3 \equiv 0 \pmod 5$. Thus $d \equiv 1 \pmod 5$, hence $c = 1+5+O(5^2)$. Repeating the procedure above with more terms, we next get new information from the coefficient of t^{31}, where we deduce that $c = 1 + 5 + 4 \cdot 5^2 + O(5^3)$.

Using Algorithm 3.2: Using Kedlaya's algorithm (as implemented in [BCP97]) we find almost instantly that

$$\mathbf{E}_2(E,\omega) = 2+4\cdot 5+2\cdot 5^3+5^4+3\cdot 5^5+2\cdot 5^6+5^8+3\cdot 5^9+4\cdot 5^{10}+\cdots .$$

Thus

$$c = \frac{1}{12}\mathbf{E}_2(E,\omega) = 1+5+4\cdot 5^2+5^3+5^4+5^6+4\cdot 5^7+3\cdot 5^8+2\cdot 5^9+4\cdot 5^{10}+\cdots ,$$

which is consistent with what we found above using integrality.

4. [Height] For $Q = (x,y) = 8(0,0)$ as above, we have

$$s = \sigma\left(-\frac{x}{y}\right) = \sigma\left(\frac{35}{23}\right) = 4\cdot 5 + 5^2 + 5^3 + 5^4 + \cdots ,$$

so

$$\begin{aligned} h_5(Q) &= \frac{1}{5}\cdot \log_5\left(\frac{s}{5}\right) = \frac{1}{5}\cdot \log_5(4+5+5^2+5^3+2\cdot 5^5+\cdots) \\ &= 3+5+2\cdot 5^3+3\cdot 5^4+\cdots . \end{aligned}$$

Finally,

$$h_5(P) = \frac{1}{8^2}\cdot h_5(Q) = 2+4\cdot 5+5^2+2\cdot 5^3+2\cdot 5^4+\cdots .$$

Remark 4.1. A *very* good check to see whether or not any implementation of the algorithms in this paper is really correct, is just to make control experiments every once in a while, by computing $h(P)$ and then comparing it with $h(2P)/4$, $h(3P)/9$, etc. In particular, compute $h(P) - h(nP)/n^2$ for several n and check that the result is p-adically small. We have done this in many cases for the implementation used to compute the tables in this section.

4.2 Curves of ranks 1, 2, 3, 4, and 5

4.2.1 Rank 1

The first (ordered by conductor) curve of rank 1 is the curve with Cremona label 37A, which we considered in Section 4.1 above.

p	p-adic regulator of 37A
5	$1 + 5 + 5^2 + 3 \cdot 5^5 + 4 \cdot 5^6 + O(5^7)$
7	$1 + 7 + 3 \cdot 7^2 + 7^3 + 6 \cdot 7^4 + 2 \cdot 7^5 + 4 \cdot 7^6 + O(7^7)$
11	$7 + 9 \cdot 11 + 7 \cdot 11^2 + 8 \cdot 11^3 + 9 \cdot 11^4 + 2 \cdot 11^5 + 7 \cdot 11^6 + O(11^7)$
13	$12 \cdot 13 + 5 \cdot 13^2 + 9 \cdot 13^3 + 10 \cdot 13^4 + 4 \cdot 13^5 + 2 \cdot 13^6 + O(13^7)$
23	$20 + 10 \cdot 23 + 18 \cdot 23^2 + 16 \cdot 23^3 + 13 \cdot 23^4 + 4 \cdot 23^5 + 15 \cdot 23^6 + O(23^7)$
29	$19 + 4 \cdot 29 + 26 \cdot 29^2 + 2 \cdot 29^3 + 26 \cdot 29^4 + 26 \cdot 29^5 + 17 \cdot 29^6 + O(29^7)$
31	$15 + 10 \cdot 31 + 13 \cdot 31^2 + 2 \cdot 31^3 + 24 \cdot 31^4 + 9 \cdot 31^5 + 8 \cdot 31^6 + O(31^7)$
41	$30 + 2 \cdot 41 + 23 \cdot 41^2 + 15 \cdot 41^3 + 27 \cdot 41^4 + 8 \cdot 41^5 + 17 \cdot 41^6 + O(41^7)$
43	$30 + 30 \cdot 43 + 22 \cdot 43^2 + 38 \cdot 43^3 + 11 \cdot 43^4 + 29 \cdot 43^5 + O(43^6)$
47	$11 + 37 \cdot 47 + 27 \cdot 47^2 + 23 \cdot 47^3 + 22 \cdot 47^4 + 34 \cdot 47^5 + 3 \cdot 47^6 + O(47^7)$
53	$26 \cdot 53^{-2} + 30 \cdot 53^{-1} + 20 + 47 \cdot 53 + 10 \cdot 53^2 + 32 \cdot 53^3 + O(53^4)$

Note that when $p = 53$ we have $\#E(\mathbf{F}_p) = p$, i.e., p is anomalous.

4.3 Rank 2

The first curve of rank 2 is the curve 389A of conductor 389. The p-adic regulators of this curve are as follows:

p	p-adic regulator of 389A
5	$1 + 2 \cdot 5 + 2 \cdot 5^2 + 4 \cdot 5^3 + 3 \cdot 5^4 + 4 \cdot 5^5 + 3 \cdot 5^6 + O(5^7)$
7	$6 + 3 \cdot 7^2 + 2 \cdot 7^3 + 6 \cdot 7^4 + 7^5 + 2 \cdot 7^6 + O(7^7)$
11	$4 + 7 \cdot 11 + 6 \cdot 11^2 + 11^3 + 9 \cdot 11^4 + 10 \cdot 11^5 + 3 \cdot 11^6 + O(11^7)$
13	$9 + 12 \cdot 13 + 10 \cdot 13^2 + 5 \cdot 13^3 + 5 \cdot 13^4 + 13^5 + 9 \cdot 13^6 + O(13^7)$
17	$4 + 8 \cdot 17 + 15 \cdot 17^2 + 11 \cdot 17^3 + 13 \cdot 17^4 + 16 \cdot 17^5 + 6 \cdot 17^6 + O(17^7)$
19	$3 + 5 \cdot 19 + 8 \cdot 19^2 + 16 \cdot 19^3 + 13 \cdot 19^4 + 14 \cdot 19^5 + 11 \cdot 19^6 + O(19^7)$
23	$17 + 23 + 22 \cdot 23^2 + 16 \cdot 23^3 + 3 \cdot 23^4 + 15 \cdot 23^5 + O(23^7)$
29	$9 + 14 \cdot 29 + 22 \cdot 29^2 + 29^3 + 22 \cdot 29^4 + 29^5 + 20 \cdot 29^6 + O(29^7)$
31	$1 + 17 \cdot 31 + 4 \cdot 31^2 + 16 \cdot 31^3 + 18 \cdot 31^4 + 21 \cdot 31^5 + 8 \cdot 31^6 + O(31^7)$
37	$28 + 37 + 11 \cdot 37^2 + 7 \cdot 37^3 + 3 \cdot 37^4 + 24 \cdot 37^5 + 17 \cdot 37^6 + O(37^7)$
41	$20 + 26 \cdot 41 + 41^2 + 29 \cdot 41^3 + 38 \cdot 41^4 + 31 \cdot 41^5 + 23 \cdot 41^6 + O(41^7)$
43	$40 + 25 \cdot 43 + 15 \cdot 43^2 + 18 \cdot 43^3 + 36 \cdot 43^4 + 35 \cdot 43^5 + O(43^6)$
47	$25 + 24 \cdot 47 + 7 \cdot 47^2 + 11 \cdot 47^3 + 35 \cdot 47^4 + 3 \cdot 47^5 + 9 \cdot 47^6 + O(47^7)$

4.4 Rank 3

The first curve of rank 3 is the curve 5077A of conductor 5077. The p-adic regulators of this curve are as follows:

p	p-adic regulator of 5077A
5	$5^{-2} + 5^{-1} + 4 + 2 \cdot 5 + 2 \cdot 5^2 + 2 \cdot 5^3 + 4 \cdot 5^4 + 2 \cdot 5^5 + 5^6 + O(5^7)$
7	$1 + 3 \cdot 7 + 3 \cdot 7^2 + 4 \cdot 7^3 + 4 \cdot 7^5 + O(7^7)$
11	$6 + 11 + 5 \cdot 11^2 + 11^3 + 11^4 + 8 \cdot 11^5 + 3 \cdot 11^6 + O(11^7)$
13	$2 + 6 \cdot 13 + 13^3 + 6 \cdot 13^4 + 13^5 + 4 \cdot 13^6 + O(13^7)$
17	$11 + 15 \cdot 17 + 8 \cdot 17^2 + 16 \cdot 17^3 + 9 \cdot 17^4 + 5 \cdot 17^5 + 11 \cdot 17^6 + O(17^7)$
19	$17 + 9 \cdot 19 + 10 \cdot 19^2 + 15 \cdot 19^3 + 6 \cdot 19^4 + 13 \cdot 19^5 + 17 \cdot 19^6 + O(19^7)$
23	$7 + 17 \cdot 23 + 19 \cdot 23^3 + 21 \cdot 23^4 + 19 \cdot 23^5 + 22 \cdot 23^6 + O(23^7)$
29	$8 + 16 \cdot 29 + 11 \cdot 29^2 + 20 \cdot 29^3 + 9 \cdot 29^4 + 8 \cdot 29^5 + 24 \cdot 29^6 + O(29^7)$
31	$17 + 11 \cdot 31 + 28 \cdot 31^2 + 3 \cdot 31^3 + 17 \cdot 31^5 + 29 \cdot 31^6 + O(31^7)$
43	$9 + 13 \cdot 43 + 15 \cdot 43^2 + 32 \cdot 43^3 + 28 \cdot 43^4 + 18 \cdot 43^5 + 3 \cdot 43^6 + O(43^7)$
47	$29 + 3 \cdot 47 + 46 \cdot 47^2 + 4 \cdot 47^3 + 23 \cdot 47^4 + 25 \cdot 47^5 + 37 \cdot 47^6 + O(47^7)$

For $p = 5$ and E the curve 5077A, we have $\#E(\mathbf{F}_5) = 10$, so $a_p \equiv 1 \pmod{5}$, hence p is anamolous.

4.5 Rank 4

Next we consider the curve of rank 4 with smallest known conductor ($234446 = 2 \cdot 117223$):

$$y^2 + xy = x^3 - x^2 - 79x + 289.$$

Note that computation of the p-adic heights is just as fast for this curve as the above curves, i.e., our algorithm for computing heights is insensitive to the conductor, only the prime p (of course, computing the Mordell-Weil group could take much longer if the conductor is large).

p	p-adic regulator of rank 4 curve
5	$2 \cdot 5^{-2} + 2 \cdot 5^{-1} + 3 \cdot 5 + 5^2 + 4 \cdot 5^3 + 4 \cdot 5^4 + 3 \cdot 5^5 + 3 \cdot 5^6 + O(5^7)$
7	$6 \cdot 7 + 4 \cdot 7^2 + 5 \cdot 7^3 + 5 \cdot 7^5 + 3 \cdot 7^6 + O(7^7)$
11	$5 + 10 \cdot 11 + 5 \cdot 11^2 + 11^3 + 3 \cdot 11^5 + 11^6 + O(11^7)$
13	$12 + 2 \cdot 13 + 4 \cdot 13^2 + 10 \cdot 13^3 + 3 \cdot 13^4 + 5 \cdot 13^5 + 7 \cdot 13^6 + O(13^7)$
17	$15 + 8 \cdot 17 + 13 \cdot 17^2 + 5 \cdot 17^3 + 13 \cdot 17^4 + 7 \cdot 17^5 + 14 \cdot 17^6 + O(17^7)$
19	$14 + 16 \cdot 19 + 15 \cdot 19^2 + 6 \cdot 19^3 + 10 \cdot 19^4 + 7 \cdot 19^5 + 13 \cdot 19^6 + O(19^7)$
23	$3 + 15 \cdot 23 + 15 \cdot 23^2 + 12 \cdot 23^4 + 20 \cdot 23^5 + 7 \cdot 23^6 + O(23^7)$
29	$25 + 4 \cdot 29 + 18 \cdot 29^2 + 5 \cdot 29^3 + 27 \cdot 29^4 + 23 \cdot 29^5 + 27 \cdot 29^6 + O(29^7)$
31	$21 + 26 \cdot 31 + 22 \cdot 31^2 + 25 \cdot 31^3 + 31^4 + 3 \cdot 31^5 + 14 \cdot 31^6 + O(31^7)$
37	$34 + 14 \cdot 37 + 32 \cdot 37^2 + 25 \cdot 37^3 + 28 \cdot 37^4 + 36 \cdot 37^5 + O(37^6)$
41	$33 + 38 \cdot 41 + 9 \cdot 41^2 + 35 \cdot 41^3 + 25 \cdot 41^4 + 15 \cdot 41^5 + 30 \cdot 41^6 + O(41^7)$
43	$14 + 34 \cdot 43 + 12 \cdot 43^2 + 26 \cdot 43^3 + 32 \cdot 43^4 + 26 \cdot 43^5 + O(43^6)$
47	$43 + 47 + 17 \cdot 47^2 + 28 \cdot 47^3 + 40 \cdot 47^4 + 6 \cdot 47^5 + 7 \cdot 47^6 + O(47^7)$

4.6 Rank 5

Next we consider the curve of rank 5 with smallest known conductor, which is the prime 19047851. The curve is

$$y^2 + y = x^3 - 79x + 342$$

p	p-adic regulator of rank 5 curve
5	$2 \cdot 5 + 5^2 + 5^3 + 2 \cdot 5^4 + 5^5 + 5^6 + O(5^7)$
7	$2 + 6 \cdot 7 + 4 \cdot 7^2 + 3 \cdot 7^3 + 6 \cdot 7^4 + 2 \cdot 7^5 + 4 \cdot 7^6 + O(7^7)$
11	$10 + 11 + 6 \cdot 11^2 + 2 \cdot 11^3 + 6 \cdot 11^4 + 7 \cdot 11^5 + 5 \cdot 11^6 + O(11^7)$
13	$11 + 8 \cdot 13 + 3 \cdot 13^2 + 4 \cdot 13^3 + 10 \cdot 13^4 + 5 \cdot 13^5 + 6 \cdot 13^6 + O(13^7)$
17	$4 + 11 \cdot 17 + 4 \cdot 17^2 + 5 \cdot 17^3 + 13 \cdot 17^4 + 5 \cdot 17^5 + 2 \cdot 17^6 + O(17^7)$
19	$11 + 7 \cdot 19 + 11 \cdot 19^2 + 7 \cdot 19^3 + 9 \cdot 19^4 + 6 \cdot 19^5 + 10 \cdot 19^6 + O(19^7)$
23	$14 + 14 \cdot 23 + 20 \cdot 23^2 + 6 \cdot 23^3 + 19 \cdot 23^4 + 9 \cdot 23^5 + 15 \cdot 23^6 + O(23^7)$
29	$3 + 5 \cdot 29 + 20 \cdot 29^3 + 21 \cdot 29^4 + 18 \cdot 29^5 + 11 \cdot 29^6 + O(29^7)$
31	$4 + 26 \cdot 31 + 11 \cdot 31^2 + 12 \cdot 31^3 + 3 \cdot 31^4 + 15 \cdot 31^5 + 22 \cdot 31^6 + O(31^7)$
37	$3 + 20 \cdot 37 + 11 \cdot 37^2 + 17 \cdot 37^3 + 33 \cdot 37^4 + 5 \cdot 37^5 + O(37^7)$
41	$3 + 41 + 35 \cdot 41^2 + 29 \cdot 41^3 + 22 \cdot 41^4 + 27 \cdot 41^5 + 25 \cdot 41^6 + O(41^7)$
43	$35 + 41 \cdot 43 + 43^2 + 11 \cdot 43^3 + 32 \cdot 43^4 + 11 \cdot 43^5 + 18 \cdot 43^6 + O(43^7)$
47	$25 + 39 \cdot 47 + 45 \cdot 47^2 + 25 \cdot 47^3 + 42 \cdot 47^4 + 13 \cdot 47^5 + O(47^6)$

Note that the regulator for $p = 5$ is not a unit, and $\#E(F_5) = 9$. This is the only example of a regulator in our tables with positive valuation.

Part II
Computing expansions for $\mathbf{E}_2$ in terms of classical modular forms

We next study convergence of $\mathbf{E}_2$ in the general context of p-adic and overconvergent modular forms. Coleman, Gouvea, and Jochnowitz prove in [CGJ95] that $\mathbf{E}_2$ is *transcendental* over the ring of overconvergent modular forms, so $\mathbf{E}_2$ is certainly non-overconvergent. However, $\mathbf{E}_2$ is *log convergent* in a sense that we make precise in this part of the paper.

5 Questions about rates of convergence

Fix p a prime number, which, in this section, we will assume is ≥ 5. We only consider modular forms of positive even integral weight, on $\Gamma_0(M)$ for some M, and with Fourier coefficients in $\mathbf{C}_p$. By a *classical modular form* we will mean one with these properties, and by a *Katz modular form* we mean a p-adic modular form in the sense of Katz ([Kat73]), again with these properties, i.e., of integral weight $k \geq 0$, of tame level N for a positive integer N prime to p, and with Fourier coefficients in $\mathbf{C}_p$. A *p-integral modular form* is a modular form with Fourier coefficients in $\mathbf{Z}_p$. Note that throughout Sections 5 and 6, all our modular forms can be taken to be with coefficients in $\mathbf{Q}_p$.

If f is a classical, or Katz, modular form, we will often simply identify the form f with its Fourier expansion, $f = \sum_{n\geq 0} c_f(n)q^n$. By $\mathrm{ord}_p(f)$ we mean the greatest lower bound of the non-negative integers $\mathrm{ord}_p(c_f(n))$ for $n \geq 0$. The valuation ord_p on $\mathbf{C}_p$ here is given its natural normalization, i.e., $\mathrm{ord}_p(p) = 1$.

We say two p-integral modular forms are *congruent* modulo p^n, denoted

$$f \equiv g \pmod{p^n},$$

if their corresponding Fourier coefficients are congruent modulo p^n. Equivalently, $f \equiv g \pmod{p^n}$ if $\mathrm{ord}_p(f - g) \geq n$.

Recall the traditional notation,

$$\sigma_{k-1}(n) = \sum_{0\,<\,d\,\mid\,n} d^{k-1},$$

and put $\sigma(n) = \sigma_1(n)$.

Let $E_k = -b_k/2k + \sum_{n=0}^{\infty} \sigma_{k-1}(n)q^n$ be the Eisenstein series of even weight $k \geq 2$, and denote by $\mathcal{E}_k$ the "other natural normalization" of the Eisenstein series,

$$\mathcal{E}_k = 1 - \frac{2k}{b_k} \cdot \sum_{n=0}^{\infty} \sigma_{k-1}(n)q^n,$$

for $k \geq 2$. We have

$$\mathcal{E}_{p-1} \equiv 1 \pmod{p}.$$

(Note that $\mathcal{E}_k$ is the q-expansion of the Katz modular form that we denote by $\mathbf{E}_k$ elsewhere in this paper.)

For $k > 2$ these are classical modular forms of level 1, while the Fourier series $E_2 = -1/24 + \sum_{n=0}^{\infty} \sigma(n)q^n$, and the corresponding $\mathcal{E}_2$, are not; nevertheless, they may all be viewed as Katz modular forms of tame level 1.

Put

$$\sigma^{(p)}(n) = \sum_{0\,<\,d\,\mid\,n;\ (p,d)=1} d,$$

so that we have:

$$\sigma(n) = \sigma^{(p)}(n) + p\sigma^{(p)}(n/p) + p^2\sigma^{(p)}(n/p^2) + \cdots \tag{5.1}$$

where the convention is that $\sigma^{(p)}(r) = 0$ if r is not an integer.

Let $V = V_p$ be the operator on power series given by the rule:

$$V\left(\sum_{n\geq 0} c_n q^n\right) = \sum_{n\geq 0} c_n q^{pn}.$$

If $F = \sum_{n\geq 0} c_n q^n$ is a classical modular form of weight k on $\Gamma_0(M)$, then $V(F)$ is (the Fourier expansion of) a classical modular form of weight k on $\Gamma_0(Mp)$ (cf. [Lan95, Ch. VIII]).

The Fourier series

$$E_2^{(p)} = (1 - pV)E_2 = \frac{p-1}{24} + \sum \sigma_1^{(p)}(n)q^n$$

is, in contrast to E_2, a classical modular form (of weight 2 on $\Gamma_0(p)$) and we can invert the formula of its definition to give the following equality of Fourier series:

$$E_2 = \sum_{\nu \geq 0} p^\nu V^\nu E_2^{(p)}, \tag{5.2}$$

this equality being, for the corresponding Fourier coefficients other than the constant terms, another way of phrasing (5.1).

Definition 5.1 (Convergence Rate). We call a function $\alpha(\nu)$ taking values that are either positive integers or $+\infty$ on integers $\nu = 0, \pm 1, \pm 2, \dots$ a *convergence rate* if $\alpha(\nu)$ is a non-decreasing function such that $\alpha(\nu) = 0$ for $\nu \leq 0$, $\alpha(\nu+\mu) \leq \alpha(\nu) + \alpha(\mu)$, and $\alpha(\nu)$ tends to $+\infty$ as ν does.

A simple nontrivial example of a convergence rate is

$$\alpha(\nu) = \begin{cases} 0 & \text{for } \nu \leq 0, \\ \nu & \text{for } \nu \geq 0. \end{cases}$$

If $\alpha(\nu)$ is a convergence rate, put $T\alpha(\nu) = \alpha(\nu - 1)$; note that $T\alpha(\nu)$ is also a convergence rate (T translates the graph of α one to the right). Given a collection $\{\alpha_j\}_{j \in J}$ of convergence rates, the "max" function $\alpha(\nu) = \max_{j \in J} \alpha_j(\nu)$ is again a convergence rate.

Definition 5.2 (α-Convergent). Let α be a convergence rate. A Katz modular form f is *α-convergent* if there is a function $a : \mathbf{Z}_{\geq 0} \to \mathbf{Z}_{\geq 0}$ such that

$$f = \sum_{\nu=0}^{\infty} p^{a(\nu)} f_\nu \mathcal{E}_{p-1}^{-\nu} \tag{5.3}$$

with f_ν a classical p-integral modular form (of weight $k + \nu(p-1)$ and level N) and $a(\nu) \geq \alpha(\nu)$ for all $\nu \geq 0$.

If $\alpha' \leq \alpha$ are convergence rates and a modular form f is *α-convergent* then it is also *α'-convergent.* As formulated, an expansion of the shape of (5.3) for a given f is not unique but [Kat73] and [Gou88] make a certain sequence of choices that enable them to get canonical expansions of the type (5.3), dependent on those initial choices. Specifically, let $M_{\text{classical}}(N, k, \mathbf{Z}_p)$ denote the $\mathbf{Z}_p$-module of classical modular forms on $\Gamma_0(N)$ of weight k and with Fourier coefficients in $\mathbf{Z}_p$. Multiplication by $\mathcal{E}_{p-1}$ allows one to identify $M_{\text{classical}}(N, k, \mathbf{Z}_p)$ with a saturated $\mathbf{Z}_p$-lattice in $M_{\text{classical}}(N, k + p - 1, \mathbf{Z}_p)$. (The lattice is saturated because multiplication by $E_{p-1}(\text{mod } p)$ is injective, since it is the identity map on q-expansions.) *Fix*, for each k, a $\mathbf{Z}_p$-module,

$$C(N, k + p - 1, \mathbf{Z}_p) \subset M_{\text{classical}}(N, k + p - 1, \mathbf{Z}_p)$$

that is complementary to $\mathcal{E}_{p-1}\cdot M_{\text{classical}}(N,k,\mathbf{Z}_p)\subset M_{\text{classical}}(N,k+p-1,\mathbf{Z}_p)$. Requiring the classical modular forms f_ν of the expansion (5.3) to lie in these complementary submodules, i.e., $f_\nu\in C(N,k+\nu(p-1),\mathbf{Z}_p)$ for all ν, pins down the expansion uniquely. Let us call an expansion of the form

$$f=\sum_{\nu=0}^{\infty}p^{\alpha(\nu)}f_\nu\mathcal{E}_{p-1}^{-\nu}$$

pinned down by a choice of complementary submodules as described above a *Katz expansion* of f.

A *classical* p-integral modular form is, of course, α-convergent for every α. For any given convergence rate α, the α-convergent Katz modular forms of tame level N are closed under multiplication, and the collection of them forms an algebra over the ring of classical modular forms of level N (with Fourier coefficients in $\mathbf{Z}_p$). Any Katz p-integral modular form is α-convergent, for some convergence rate α (see [Gou88]).

Proposition 5.3. *A Katz p-integral modular form f of weight k and tame level N as above is α-convergent if and only if the Fourier series of $f\mathcal{E}_{p-1}^{\nu}$ is congruent to the Fourier series of a classical p-integral modular form (of weight $k+\nu(p-1)$ and level N) modulo $p^{\alpha(\nu+1)}$ for every integer $\nu\geq 0$.*

Proof. We use the q-expansion principle. Specifically, if G_ν is a classical modular form such that $f\mathcal{E}_{p-1}^{\nu}\equiv G_\nu \pmod{p^{\alpha(\nu+1)}}$ then $g_\nu=p^{-\alpha(\nu+1)}(f\mathcal{E}_{p-1}^{\nu}-G_\nu)$ is again a Katz modular form, and we can produce the requisite α-convergent Katz expansion by inductive consideration of these g_ν's. (Note that the other implication is trivial. Also note our running hypothesis that $p\geq 5$.) □

In view of this, we may define, for any f as in Proposition 5.3, the function $a_f(\nu)$ (for $\nu\geq 0$) as follows: $a_f(0)=0$, and for $\nu\geq 1$, $a_f(\nu)$ is the largest integer a such that $f\mathcal{E}_{p-1}^{\nu-1}$ is congruent to a classical p-integral modular form (of weight $k+(\nu-1)(p-1)$ and level N) modulo p^a.

Corollary 5.4. *The Katz p-integral modular form f is α-convergent for any convergence rate α that is majorized by the function a_f. (I.e., for which $\alpha(\nu)\leq a_f(\nu)$ for all $\nu\geq 0$.)*

Definition 5.5 (Overconvergent of Radius r). Let $r\in\mathbf{Q}$ be a positive rational number. A Katz p-integral modular form f of tame level N is *overconvergent of radius r* if and only if it is α-convergent for some function α such that $\alpha(\nu)\geq r\cdot\nu$ for all ν, and $\alpha(\nu)-r\cdot\nu$ tends to infinity with ν.

Remarks 5.6. It is convenient to say, for two function $\alpha(\nu)$ and $\alpha'(\nu)$, that

$$\alpha(\nu)\ \gg\ \alpha'(\nu)$$

if $\alpha(\nu)\geq\alpha'(\nu)$ and $\alpha(\nu)-\alpha'(\nu)$ tends to infinity with ν. So, we may rephrase the above definition as saying that f is overconvergent with radius r if it is

α-convergent with $\alpha(\nu) \gtrdot r \cdot \nu$. The above definition is equivalent to the definition of [Kat73, Gou88] except for the fact that the word *radius* in these references does not denote the rational number r above, but rather a choice of p-adic number whose ord_p is r. We may think of our manner of phrasing the definition as being a *definition by Katz expansion convergence rate* as opposed to what one might call the *definition by rigid analytic geometric behavior*, meaning the equivalent, and standard, formulation (cf. [Kat73]) given by considering f as a rigid analytic function on an appropriate extension of the Hasse domain in the (rigid analytic space associated to) $X_0(N)$.

Definition 5.7 ((Precisely) Log Convergent). A Katz p-integral modular form f is *log-convergent* if $c \cdot \log(\nu) \leq a_f(\nu)$ for some positive constant c and all but finitely many ν (equivalently: if it is α-convergent for $\alpha(\nu) = c \cdot \log(\nu)$ for some positive constant c). We will say that f is *precisely log-convergent* if there are positive constants c, C such that $c \cdot \log(\nu) \leq a_f(\nu) \leq C \cdot \log(\nu)$ for all but finitely many ν.

Remark 5.8. As in Definition 5.1 above, we may think of this manner of phrasing the definition as being a *definition by Katz expansion convergence rate*. This seems to us to be of some specific interest in connection with the algorithms that we present in this article for the computation of $\mathbf{E}_2$. For more theoretical concerns, however, we think it would be interesting to give, if possible, an equivalent *definition by rigid analytic geometric behavior*: is there some explicit behavior at the "rim" of the Hasse domain that characterizes log-convergence?

Proposition 5.9. *Let $p \geq 5$. Let f be a Katz p-integral modular form of weight k and tame level N that admits an expansion of the type*

$$f = \sum_{\nu=0}^{\infty} p^{\nu} \mathcal{F}_{\nu} \mathcal{E}_{p-1}^{-\nu}$$

where, for all $\nu \geq 0$, $\mathcal{F}_\nu$ is a classical p-integral modular form (of weight $k + \nu(p-1)$) on $\Gamma_0(p^{\nu+1})$. Then f is log-convergent and

$$\liminf_{n\to\infty} \frac{a_f(n)}{\log(n)} \geq \frac{1}{\log(p)}.$$

Proof. The classical modular form $\mathcal{F}_\nu$ on $\Gamma_0(p^{\nu+1})$ is an overconvergent Katz modular form of radius r for any r such that $r < \frac{1}{p^{\nu-1}(p+1)}$ (cf. [Kat73], [Gou88, Cor. II.2.8]). Let

$$\mathcal{F}_\nu = \sum_{\mu=0}^{\infty} f_\mu^{(\nu)} \mathcal{E}_{p-1}^{-\mu}$$

be its Katz expansion. So,

$$\mathrm{ord}_p(f_\mu^{(\nu)}) \ \gtrdot \ \left(\frac{1}{p^{\nu-1}(p+1)} - \epsilon_{\mu,\nu}\right) \cdot \mu$$

for any choice of positive $\epsilon_{\mu,\nu}$. We have

$$f = \sum_{\nu=0}^{\infty} p^{\nu} \sum_{\mu=0}^{\infty} f_{\mu}^{(\nu)} \mathcal{E}_{p-1}^{-(\mu+\nu)},$$

or (substituting $\gamma = \mu + \nu$)

$$f = \sum_{\gamma=0}^{\infty} \left\{ \sum_{\nu=0}^{\gamma} p^{\nu} \ f_{\gamma-\nu}^{(\nu)} \right\} \mathcal{E}_{p-1}^{-\gamma}.$$

Putting $G_{\gamma} = \sum_{\nu=0}^{\gamma} p^{\nu} \ f_{\gamma-\nu}^{(\nu)}$ we may write the above expansion as

$$f = \sum_{\gamma=0}^{\infty} G_{\gamma} \mathcal{E}_{p-1}^{-\gamma},$$

and we must show that

$$\operatorname{ord}_p(G_{\gamma}) \geq c \cdot \log(\gamma)$$

for some positive constant c.

For any $\nu \leq \gamma$ we have

$$\operatorname{ord}_p \left(p^{\nu} \ f_{\gamma-\nu}^{(\nu)} \right) \ \geqq \ \nu + \left(\frac{1}{p^{\nu-1}(p+1)} - \epsilon_{\gamma-\nu,\nu} \right) (\gamma - \nu).$$

We need to find a lower bound for the minimum value achieved by the right-hand side of this equation. To prepare for this, first note that at the extreme value $\nu = 0$ we compute $\operatorname{ord}_p(\ f_{\gamma}^{(0)}) \geq \left(\frac{p}{(p+1)} - \epsilon_{\gamma,0} \right) \cdot \gamma$, and to study the remaining cases, $\nu = 1, \ldots, \gamma$, we look at the function

$$R(t) = t + \left(\frac{1}{p^{t-1}(p+1)} \right) (\gamma - t)$$

in the range $1 \leq t \leq \gamma$. This, by calculus, has a unique minimum at $t = t_{\gamma} \in (1, \gamma)$ given by the equation

$$\frac{p+1}{p} \cdot p^{t_{\gamma}} = \log(p) \cdot (\gamma - t_{\gamma}) + 1. \tag{5.4}$$

Define $e_{\gamma} = t_{\gamma} - \log_p(\gamma)$ and substituting, we get:

$$p^{e_{\gamma}} = \frac{p \log(p)}{p+1} - \frac{p \log(p)}{p+1} \frac{e_{\gamma}}{\gamma} + A_{\gamma} \tag{5.5}$$

where A_{γ} goes to zero, as γ goes to ∞.

If e_{γ} is positive we get that

$$p^{e_{\gamma}} \leq \frac{p \log(p)}{p+1} + A_{\gamma}$$

and so e_γ is bounded from above, independent of γ, while if $e_\gamma = -d_\gamma$ with d_γ positive, we have

$$\frac{1}{p^{d_\gamma}} = \frac{p\log(p)}{p+1} + \frac{p\log(p)}{p+1}\frac{d_\gamma}{\gamma} + A_\gamma.$$

Recall that since $t_\gamma > 0$ we also have $d_\gamma < \log_p(\gamma)$, so that the right hand side of the displayed equation tends to $\frac{p\log(p)}{p+1}$ as γ goes to ∞, so the equation forces d_γ to be bounded from above, as γ tends to ∞.

This discussion gives:

Lemma 5.10. *The quantity $|t_\gamma - \log_p(\gamma)|$ is bounded independent of γ.*

Substituting $t_\gamma = \log_p(\gamma) + e_\gamma$ in the defining equation for $R(t)$ and noting the boundedness of $|e_\gamma|$, we get that $|R(t_\gamma) - \log_p(\gamma)|$ is bounded as γ goes to ∞, thereby establishing our proposition.

□

Corollary 5.11. *For all $p \geq 5$, the Katz modular form $f = E_2$ is log-convergent and*

$$\liminf_{n\to\infty} \frac{a_f(n)}{\log(n)} \geq \frac{1}{\log(p)}.$$

Proof. The modular forms $V^\nu E_2^{(p)}$ are classical modular forms on $\Gamma_0(p^{\nu+1})$ and therefore formula (5.1) exhibits E_2 as having a Katz expansion of the shape of (5.3). Proposition 5.9 then implies the corollary. □

Remark 5.12. Is E_2 *precisely* log-convergent? The minimal c (cf. Definition 5.7) that can be taken in the log-convergence rate for $f = E_2$ is $\limsup_{n\to\infty}(a_f(n)/\log(n))$. Is this minimal c equal to $1/\log(p)$? It is for $p = 5$, as we will show in Section 6. The previous discussion tells us that, as a kind of generalization of the well-known congruence

$$E_2\mathcal{E}_{p-1} \equiv E_{p+1} \pmod{p},$$

we have that for any $\epsilon > 0$, and all but finitely many ν, there are classical modular forms $\mathcal{G}_\nu$ of level 1 and weight $2 + \nu(p-1)$ such that

$$E_2\mathcal{E}_{p-1}^\nu \equiv \mathcal{G}_\nu \pmod{p^{\lfloor(1-\epsilon)\log_p(\nu)\rfloor}}.$$

Let $\theta = qd/dq$ denote the standard shift operator; so that if $f = \sum_{n\geq 0} c_n q^n$, then $\theta(f) = \sum_{n\geq 0} nc_n q^n$. We have $\mathrm{ord}_p(\theta(f)) \geq \mathrm{ord}_p(f)$. The operator θ preserves Katz modular forms, and *almost* preserves classical modular forms in the sense that if f is a classical modular form of weight $k \geq 2$ then so is $F = \theta(f) - kfE_2/12$ (cf. [Kat73]). Note, also, that $\mathrm{ord}_p(F) \geq \mathrm{ord}_p(f)$.

Corollary 5.13. *The operator θ preserves log-convergent Katz modular forms.*

Proof. Let f be a log-convergent Katz p-integral modular form of weight k, of tame conductor N with a Katz expansion,

$$f = \sum_{\nu=0}^{\infty} p^{a(\nu)} f_\nu \mathcal{E}_{p-1}^{-\nu} \tag{5.6}$$

where $a(\nu) \geq c \cdot \log(\nu)$ for some positive c, and the f_ν's are classical p-integral modular forms on $\Gamma_0(N)$. Let $F_\nu = \theta(f_\nu) - (k + \nu(p-1))f_\nu E_2/12$ (which is a classical modular form of weight $k + 2 + \nu(p-1)$ on $\Gamma_0(N)$). Put

$$G = \theta(E_{p-1}) - \frac{p-1}{12}\mathcal{E}_{p-1}E_2.$$

Apply the derivation θ to (5.6) to get

$$\theta(f) = \sum_{\nu=0}^{\infty} p^{a(\nu)} \Big\{ (F_\nu + (k + \nu(p-1))f_\nu E_2/12)\mathcal{E}_{p-1}^{-\nu} - \\ \nu f_\nu \mathcal{E}_{p-1}^{-\nu-1}\left(G + \frac{p-1}{12}\mathcal{E}_{p-1}E_2\right)\Big\}.$$

or:

$$\theta(f) = A + BE_2 - C - DE_2,$$

where

$$A = \sum_{\nu=0}^{\infty} p^{a(\nu)} F_\nu \mathcal{E}_{p-1}^{-\nu},$$
$$B = \sum_{\nu=0}^{\infty} p^{a(\nu)} (k + \nu(p-1))f_\nu/12)\mathcal{E}_{p-1}^{-\nu},$$
$$C = \sum_{\nu=0}^{\infty} p^{a(\nu)} \nu f_\nu G \mathcal{E}_{p-1}^{-\nu-1},$$
$$D = \sum_{\nu=0}^{\infty} p^{a(\nu)} \frac{p-1}{12} \nu f_\nu \mathcal{E}_{p-1}.$$

Now A, B, C, D are all log-convergent, as is E_2 by Corollary 5.11. Therefore so is $\theta(f)$. □

6 Precise log convergence of E_2 for $p = 2, 3, 5$

In this section we assume $p = 2$, 3 or 5 and let P, Q, R denote the Eisenstein series of level 1 of weights $2, 4, 6$, respectively, normalized so that the constant term in its Fourier expansion is 1. Let f be a Katz form of tame level 1 and weight k. Write $k = 4d + 6e$, with d an integer ≥ -1 and $e = 0$ or 1. Then $fQ^{-d}R^{-e}$ is a Katz form of weight 0, that is, a Katz function. Since 0 is the only supersingular value of j for $p = 2, 3, 5$, a Katz function has an expansion in powers of j^{-1} convergent everywhere on the disc $|j^{-1}| \le 1$. Hence, putting $z = j^{-1}$, we can write

$$f = Q^d R^e \sum_{n=0}^{\infty} c_f(n) z^n = \sum_{n=0}^{\infty} R^e \Delta^n Q^{-3n+d}.$$

with $c_f(n) \in \mathbf{Q}_p$ and $c_f(n) \to 0$ as $n \to \infty$. Let

$$C_{f,p}(N) = \min_{n>N}(\mathrm{ord}_p(c_f(n))).$$

Theorem 6.1. *For $p = 5$, we have $C_{f,5}(N) = a_f(3N + 1 - d)$, for all large N.*

Proof. Notice that for $p = 5$, $\mathcal{E}_{p-1} = Q$. Let $\nu = 3N + 1 - d$ for large N. Then

$$Q^{\nu-1} f = \sum_{n=0}^{N} c(n) R^e \Delta^n Q^{3(N-n)} + R^e Q^d \sum_{n>N} c(n) z^n = F + G,$$

say. We have $\mathrm{ord}_5(G) = \min_{n>N}(\mathrm{ord}_5(c(n)) = C_{f,5}(N)$. [5]

Since F is a classical modular form of weight $12N + 6e$ it follows from the definition of a_f that $a_f(\nu) \ge C_{f,5}(N)$. On the other hand, since $\{R^e \Delta^n Q^{3(N-n)} : 0 \le n \le N\}$ is a basis for the space of classical modular forms of weight $12N + 6e$, it is clear that for any such classical form F', the difference $Q^{\nu-1} f - F'$ is a 5-adic Katz form which can be written as $R^e Q^{3N} g$ with g a Katz function whose z-expansion coefficients are $c(n)$ for $n > N$. Thus $\mathrm{ord}_5(Q^{\nu-1} f - F') \le C_{f,5}(N)$. □

We have defined f to be log convergent if

$$\liminf_{n\to\infty} \frac{a_f(n)}{\log(n)} > 0,$$

and to be precisely log convergent if in addition

$$\limsup_{n\to\infty} \frac{a_f(n)}{\log(n)} < \infty.$$

Lemma 6.2. *Suppose $h(n)$ and $H(n)$ are nondecreasing funcions defined for all sufficiently large positive integers n. If for some integers $r > 0$ and s we have $H(N) = h(rN + s)$ for all sufficiently large integers N, then*

$$\liminf_{n\to\infty} \frac{h(n)}{\log(n)} = \liminf_{N\to\infty} \frac{H(N)}{\log(N)},$$

[5] To justify this claim we extend our definition of ord_p from the ring of Katz forms with Fourier coefficients in $\mathbf{Z}$ to the ring $\mathbf{Z}_p[[q]]$ of all formal power series with coefficients in $\mathbf{Z}$. Moreover, since $z \in q + q^2\mathbf{Z}_p[[q]]$, we have $\mathbf{Z}_p[[q]] = \mathbf{Z}_p[[z]]$, and for a formal series $g = \sum a_n q^n = \sum b^n z^n$, we have $\mathrm{ord}_p(g) = \min(\mathrm{ord}_p(a_n)) = \min(\mathrm{ord}_p(b_n))$. Also (Gauss Lemma) the rule $\mathrm{ord}(g_1 g_2) = \mathrm{ord}(g_1) + \mathrm{ord}(g_2)$ holds. Since $\mathrm{ord}_5(R) = \mathrm{ord}_5(Q) = 0$, it follows that $\mathrm{ord}_5(G) = C_{f,5}(N)$ as claimed.

and

$$\limsup_{n\to\infty} \frac{h(n)}{\log(n)} = \limsup_{N\to\infty} \frac{H(N)}{\log(N)}.$$

Proof. We use the fact that $\frac{\log(rx+s)}{\log(x)} \to 1$ as $x \to \infty$. For n and N related by

$$rN + s \leq n \leq r(N+1) + s$$

we have

$$\frac{h(n)}{\log(n)} \leq \frac{h(r(N+1)+s}{\log(rN+s)} = \frac{H(N+1)}{\log(N+1)} \cdot \frac{\log(N+1)}{\log(rN+s)}.$$

Similarly,

$$\frac{h(n)}{\log(n)} \geq \frac{h(rN+s}{\log(r(N+1)+s)} = \frac{H(N)}{\log(N)} \cdot \frac{\log(N)}{\log(r(N+1)+s)}.$$

This proves the lemma, because the second factor of the right hand term in each line approaches 1 as N goes to infinity. □

Theorem 6.1 and Lemma 6.2 show that for $p = 5$ we can replace a_f by C_f in the definition of log convergent and precisely log convergent. Therefore we define log convergent and precisely log convergent for $p = 2$ and $p = 3$ by using $C_{f,p}$ as a replacement for a_f.

Theorem 6.3. *For $p = 2, 3$ or 5, the weight 2 Eisenstein series $P = \mathbf{E}_2$ is precisely log convergent. In fact,*

$$\lim_{n\to\infty} \frac{C_{P,p}(n)}{\log(n)} = \frac{1}{\log(p)}.$$

During the proof of this theorem we write $c(n) = c_P(n)$ and $C_p(n) = C_{P,p}$.

The cases $p = 2, 3$ follow immediately from results of Koblitz (cf. [Kob77]). Koblitz writes $P = \sum a_n j^{-n} \frac{qdj}{jdq}$. Since $dj/j = -dz/z$, and as we will see later in this proof, $qdz/zdq = R/Q$, Koblitz's a_n is the negative of our $c(n)$, hence $\operatorname{ord}_p(c(n)) = \operatorname{ord}_p(a_n)$. Koblitz shows that if we let $l_p(n) = 1 + \lfloor \log(n)/\log(p) \rfloor$, the number of digits in the expression of n in base p, and let $s_p(n)$ denote the sum of those digits, then $\operatorname{ord}_2(c(n)) = l_2(n) + 3s_2(n)$ and $\operatorname{ord}_3(c(n)) = l_3(n) + s_3(n)$. From this it is an easy exercise to show

$$C_2(n) = \lfloor \log(n+1)/\log(2) \rfloor + 4 \quad \text{and} \quad C_3(n) = \lfloor (\log(n+1)/\log(3) \rfloor + 2,$$

formulas from which cases $p = 2$ and $p = 3$ of the theorem are evident.

Investigating the case $p = 5$ we found experimentally with a PARI program that the following conjecture holds for $n < 1029$.

Conjecture 6.4. *We have* $\operatorname{ord}_5(c(n)) \geq l_5(2n)$*, with equality if n written in base 5 contains only the digits 0,1 or 2, but no 3 or 4.*

It is easy to see that Conjecture 6.4 implies that

$$\limsup_{n\to\infty} \frac{C_5(n)}{\log(n)} = \frac{1}{\log(5)}.$$

We already know from Corollary 5.11 that

$$\liminf_{n\to\infty} \frac{a_P(n)}{\log(n)} \geq \frac{1}{\log(5)}.$$

By Lemma 6.2, this is equivalent to

$$\liminf_{n\to\infty} \frac{C_{P,5}(n)}{\log(n)} \geq \frac{1}{\log(5)}.$$

Hence to finish the proof of Theorem 6.3, we need only prove

$$\limsup_{n\to\infty} \frac{C_{P,5}(n)}{\log(n)} \leq \frac{1}{\log(5)}. \tag{6.1}$$

To prove (6.1) it is enough to prove that Conjecture 6.4 holds for $n = 5^m$, that is, $\mathrm{ord}_5(c(n)) = m+1$. Indeed that equality implies that $C_5(n) \leq m+1$ for $n < 5^m$ and, choosing m such that $5^{m-1} \leq n < 5^m$, shows that for every n we have $C_5(n) \leq m+1 \leq \log(n)/\log(5) + 2$.

To prove $\mathrm{ord}_5(c(n)) = m+1$ we use two lemmas.

Lemma 6.5. *We have* $\frac{PQ}{R} - 1 = 3\frac{zdQ}{Qdz}$.

Proof. Let θ denote the classical operator qd/dq. From the formula $\Delta = q\prod_{n\geq 1}(1-q^n)^{24}$ we get by logarithmic differentiation the classical formula

$$\frac{\theta\Delta}{\Delta} = P.$$

From $z = 1/j = \Delta/Q^3$ we get by logarithmic differentiation that

$$\frac{\theta z}{z} = \frac{\theta\Delta}{\Delta} - 3\frac{\theta Q}{Q} = P - 3\frac{\theta Q}{Q}.$$

By a formula of Ramanujan (cf. [Ser73, Thm. 4]) we have

$$3\frac{\theta Q}{Q} = P - \frac{R}{Q}.$$

Substituting gives

$$\frac{\theta z}{z} = \frac{R}{Q},$$

and dividing the next to last equation by the last proves the lemma. □

Lemma 6.6. *Let* $F = \sum_{n\geq 1}\sigma_3(n)q^n$, *so that* $Q = 1 + 240F$. *Then* $F \equiv \sum_{m\geq 0}(z^{5^m} + z^{2\cdot 5^m}) \pmod 5$.

Proof. Guessing this result by computer experiment, we asked Serre for a proof. He immediately supplied two, one of which is the following. During the rest of this proof all congruences are understood to be modulo 5. Since $F = z + 3z^2 + \cdots$, the statement to be proved is equivalent to $F - F^5 \equiv z + 3z^2$. Using the trivial congruence $Q \equiv 1$ and the congruence $P \equiv R$ (the case $p = 5$ of a congruence of Swinnerton-Dyer, (cf. [Ser73, Thm. 5]), we note that

$$z = \Delta/Q^3 \equiv \Delta = (Q^3 - R^2)/1728 \equiv 2 - 2R^2.$$

The case $p = 5, k = 4$ of formula (**) in section 2.2 of [Ser73] reads $F - F^5 \equiv \theta^3 R$. By Ramanujan's formula

$$\theta R = (PR - Q^2)/2 \equiv 3R^2 - 3,$$

one finds that indeed

$$\theta^3 R \equiv 2R^4 - R^2 - 1 \equiv z + 3z^2,$$

which proves Lemma 6.6. □

Let $F = \sum_{n\geq 1} b(n)z^n$. By Lemma 6.6, $b(5^m)$ and $b(2\cdot 5^m)$ are not divisible by 5. Therefore the 5^mth and $2\cdot 5^m$th coefficients of $zdF/dz = \sum_{n\geq 1} nb(n)z^n$ are divisible exactly by 5^m. By Lemma 6.5 we have

$$\sum_{n\geq 1} c(n)z^n = \frac{PQ}{R} - 1 = 3\frac{zdQ}{Qdz} = 3\frac{240zdF}{(1+240F)dz}.$$

This shows that $\mathrm{ord}_5(c(5^m)) = \mathrm{ord}_5(c(2\cdot 5^m)) = m + 1$ thereby completing the proof of Theorem 6.3.

Remark 6.7. For $p = 2$ or 3 a simple analogue of Lemma 6.6 holds, namely $F \equiv \sum_{m\geq 0} z^{p^m} \pmod{p}$. This can be used to obtain Koblitz's result for the very special case $n = p^m$.

7 Discussion

7.1 Log convergence

The running hypothesis in Section 5 is that $p \geq 5$, but in Section 6 we considered only $p = 2, 3, 5$. In dealing with the different primes, our discussion changes strikingly, depending on the three slightly different cases:

(1) $p = 2, 3$

(2) $p = 5$

(3) $p \geq 5$

For (7.1), in Section 6 we used expansions in powers of $z = 1/j$ to give a careful analysis of convergence rates, and in contrast, the general discussion of Section 5 *must* keep away from those cases $p = 2, 3$, in order to maintain the formulation that it currently has. The prime $p = 5$ is in a very fortunate position because it can be covered by the general discussion a la (7.1); but we have also given a precise "power series in $1/j$" treatment of $p = 5$. These issues suggest four questions:

1. Is there any relationship between the convergence rate analysis we give, and computation-time estimates for the actual algorithms?

2. We have produced an algebra of log-convergent modular forms, and it has at least one new element that the overconvergent forms do not have, namely $\mathbf{E}_2$. Moreover, it is closed under the action of θ, i.e., "Tate twist". Are there other interesting Hecke eigenforms in this algebra that we should know about? Going the other way, are there any Hecke eigenforms that are *not* log-convergent? Is there something corresponding to the "eigencurve" (it would have to be, at the very least, a surface) that p-adically interpolates log-convergent eigenforms? Is a limit (in the sense of ord_p's of Fourier coefficients) of log-convergent eigenforms again log-convergent? For this last question to make sense, we probably need to know the following:

3. Is there a rigid-analytic growth type of definition (growth at the rim of the Hasse domain) that characterizes log-convergence, just as there is such a definition characterizing overconvergence?

4. Almost certainly one could treat the case $p = 7$ by expansions in powers of $1/(j - 1728) = \Delta/R^2$ in the same way that we did $p = 5$ with powers of $1/j = \Delta/Q^3$. The case $p = 13$ might be more interesting.

7.2 Uniformity in the algorithms

We are most thankful to Kiran Kedlaya and Alan Lauder for some e-mail communications regarding an early draft of our article. The topic they address is the extent to which the algorithms for the computation of $\mathbf{E}_2$ of an elliptic curve are "uniform" in the elliptic curve, and, in particular, whether one can get fast algorithms for computing $\mathbf{E}_2$ of specific families of elliptic curves. In this section we give a brief synopsis of their comments.

A "reason" why $\mathbf{E}_2$ should turn out not to be overconvergent is that Katz's formula relates it to the direction of the unit-root subspace in one-dimensional de Rham cohomology, and that seems only to make (at least naive) sense in the ordinary case (and not for points in a supersingular disc, not even ones close to the boundary).

Nevertheless, part of the algorithm has good uniformity properties.

1. *Calculating the matrix of Frobenius:* One can calculate the matrix of Frobenius for, say, all elliptic curves in the Legendre family (or any one-parameter family) and the result is overconvergent everywhere, so this should be relatively efficient. This can be done either by the algorithm developed by Kedlaya, or also using the Gauss-Manin connection, as in Lauder's work, which is probably faster. An approach to computing the "full" Frobenius matrix "all at once" for elliptic curves in the Legendre family has been written up and implemented in Magma by Ralf Gerkmann: See [Ger05] for the paper and program. Lauder's paper [Lau03] also discusses Kedlaya's algorithm "all at once" for a one-parameter family of hyperelliptic curves using the Gauss-Manin connection.

2. *Extracting the unit root subspace in de Rham cohomology:* To compute $\mathbf{E}_2$ for an individual elliptic curve, one can specialize the Frobenius matrix and extract the unit root. But extracting only the unit root part over the entire family at once would involve non-overconvergent series, and consequently might be slow. The *unit root zeta function*, which encodes the unit root of Frobenius over a family of ordinary elliptic curves, has been very well studied by Dwork and Wan (cf. [Wan99]).

7.3 Other future projects

1. Explicitly compute anticyclotomic p-adic heights, and apply this to the study of universal norm questions that arise in [RM05].

2. Further investigate Kedlaya's algorithm with a parameter in connection with log convergence and computation of heights.

3. Determine if the equality $\lim_{n\to\infty} a_P(n)/\log(n) = 1/\log(p)$ holds for all primes p, as it does for $p = 5$ by Theorem 6.3.

References

[Bes04] Amnon Besser, *The p-adic height pairings of Coleman-Gross and of Nekovář*, Number theory, CRM Proc. Lecture Notes, vol. 36, Amer. Math. Soc., Providence, RI, 2004, pp. 13–25. MR 2076563 (2005f:11130)

[Blu] Antonia W. Bluher, *A Leisurely Introduction to Formal Groups and Elliptic Curves,*
`http://www.math.uiuc.edu/algebraic-number-theory/0076/`.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. 24 (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[Col91] Robert F. Coleman, *The universal vectorial bi-extension and p-adic heights*, Invent. Math. 103 (1991), no. 3, 631–650. MR 1091621 (92k:14021)

[CGJ95] Robert F. Coleman, Fernando Q. Gouvêa, and Naomi Jochnowitz, *E_2, Θ, and overconvergence*, Internat. Math. Res. Notices (1995), no. 1, 23–41 (electronic). MR 1317641 (96d:11047)

[Ger05] Ralf Gerkmann, `http://www.mathematik.uni-mainz.de/~gerkmann/ellcurves.html`, (2005).

[Gre03] Ralph Greenberg, *Galois theory for the Selmer group of an abelian variety*, Compositio Math. 136 (2003), no. 3, 255–297. MR 1977007 (2004c:11097)

[GJP$^+$05] G. Grigorov, A. Jorza, S. Patrikis, C. Patrascu, and W. Stein, *Verification of the Birch and Swinnerton-Dyer Conjecture for Specific Elliptic Curves*, (Submitted) `http://modular.fas.harvard.edu/papers/bsdalg/` (2005).

[Gou88] F. Q. Gouvêa, *Arithmetic of p-adic modular forms*, Springer-Verlag, Berlin, 1988. MR 91e:11056

[IW03] Adrian Iovita and Annette Werner, *p-adic height pairings on abelian varieties with semistable ordinary reduction*, J. Reine Angew. Math. 564 (2003), 181–203. MR 2021039 (2004j:11066)

[SJ05] William Stein and David Joyner, *Sage: System for algebra and geometry experimentation*, Communications in Computer Algebra (SIGSAM Bulletin) (July 2005), `http://sage.sourceforge.net/`.

[Kat73] Nicholas M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 69–190. Lecture Notes in Mathematics, Vol. 350. MR 0447119 (56 #5434)

[Kat76] ———, *p-adic interpolation of real analytic Eisenstein series*, Ann. of Math. (2) 104 (1976), no. 3, 459–571. MR 0506271 (58 #22071)

[Ked01] Kiran S. Kedlaya, *Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology*, J. Ramanujan Math. Soc. 16 (2001), no. 4, 323–338. MR 1877805 (2002m:14019)

[Ked03] K. S. Kedlaya, *Errata for: "Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology" [J. Ramanujan Math. Soc.* 16 *(2001), no. 4, 323–338*, J. Ramanujan Math. Soc. 18 (2003), no. 4, 417–418, Dedicated to Professor K. S. Padmanabhan. MR 2 043 934

[Kob77] Neil Koblitz, 2*-adic and* 3*-adic ordinals of the (1/j)-expansion coefficients for the weight 2 Eisenstein series, Bull. L.M.S. 9 (1977), 188-192.*

[Lan95] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[Lau03] A. G. B. Lauder, *Rigid cohomology and p-adic point counting*, to appear in a special issue of J. de Thorie des Nombres de Bordeaux, `http://www.maths.ox.ac.uk/~lauder/`.

[LW02] A. G. B. Lauder and D. Wan, *Counting rational points on varieties over finite fields of small characteristic*, to appear in an MSRI Computational Number Theory Proceedings (October, 2002).

[MR04] Barry Mazur and Karl Rubin, *Pairings in the arithmetic of elliptic curves*, Modular curves and abelian varieties, Progr. Math., vol. 224, Birkhäuser, Basel, 2004, pp. 151–163. MR MR2058649 (2005g:11095)

[MT83] B. Mazur and J. Tate, *Canonical height pairings via biextensions*, Arithmetic and geometry, Vol. I, Progr. Math., vol. 35, Birkhäuser Boston, Boston, MA, 1983, pp. 195–237. MR 717595 (85j:14081)

[MT87] ______, *Refined conjectures of the "Birch and Swinnerton-Dyer type"*, Duke Math. J. 54 (1987), no. 2, 711–750. MR 899413 (88k:11039)

[MT91] ______, *The p-adic sigma function*, Duke Math. J. 62 (1991), no. 3, 663–688. MR 93d:11059

[MTT86] B. Mazur, J. Tate, and J. Teitelbaum, *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer*, Invent. Math. 84 (1986), no. 1, 1–48. MR 830037 (87e:11076)

[Nek93] Jan Nekovář, *On p-adic height pairings*, Séminaire de Théorie des Nombres, Paris, 1990–91, Progr. Math., vol. 108, Birkhäuser Boston, Boston, MA, 1993, pp. 127–202. MR 1263527 (95j:11050)

[Nek03] ______, *Selmer Complexes*, 2003, see `http://www.math.jussieu.fr/~nekovar/pu/`.

[Pla94] Andrew Plater, *Supersingular p-adic height pairings on elliptic curves*, Arithmetic geometry (Tempe, AZ, 1993), Contemp. Math., vol. 174, Amer. Math. Soc., Providence, RI, 1994, pp. 95–105. MR 1299736 (95h:11056)

[PR03a] Bernadette Perrin-Riou, *Arithmétique des courbes elliptiques à réduction supersingulière en p*, Experiment. Math. 12 (2003), no. 2, 155–186. MR 2016704 (2005h:11138)

[PR03b] ———, *Arithmétique des courbes elliptiques à réduction supersingulière en p*, Experiment. Math. 12 (2003), no. 2, 155–186. MR 2016704

[RM05] K. Rubin and B. Mazur, *Organizing the arithmetic of elliptic curves*, in preparation.

[Sch82] Peter Schneider, *p-adic height pairings. I*, Invent. Math. 69 (1982), no. 3, 401–409. MR 679765 (84e:14034)

[Sch85] ———, *p-adic height pairings. II*, Invent. Math. 79 (1985), no. 2, 329–374. MR 778132 (86j:11063)

[Ser73] J-P. Serre, *Congruences et formes modulaires [d'après H. P. F. Swinnerton-Dyer]*, Séminaire Bourbaki, 24e année (1971/1972), Exp. No. 416 (Berlin), Springer, 1973, pp. 319–338. Lecture Notes in Math., Vol. 317.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[Wan99] Daqing Wan, *Dwork's conjecture on unit root zeta functions*, Ann. of Math. (2) 150 (1999), no. 3, 867–927. MR MR1740990 (2001a:11108)

[Wut04] Christian Wuthrich, *On p-adic heights in families of elliptic curves*, J. London Math. Soc. (2) 70 (2004), no. 1, 23–40. MR 2064750

[Zar90] Yuri G. Zarhin, *p-adic heights on abelian varieties*, Séminaire de Théorie des Nombres, Paris 1987–88, Progr. Math., vol. 81, Birkhäuser Boston, Boston, MA, 1990, pp. 317–341. MR 1042777 (91f:11043)

Barry Mazur
Department of Mathematics
Harvard University
mazur@math.harvard.edu

John Tate
Department of Mathematics
University of Texas at Austin
tate@math.utexas.edu

William A. Stein
Department of Mathematics
University of California at San Diego
wstein@ucsd.edu

26 Computational Verification Of The Birch And Swinnerton-Dyer Conjecture For Individual Elliptic Curves, with G. Grigorov, A. Jorza, S. Patrikis, and Corina Tarnita

MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

COMPUTATIONAL VERIFICATION OF THE BIRCH AND SWINNERTON-DYER CONJECTURE FOR INDIVIDUAL ELLIPTIC CURVES

GRIGOR GRIGOROV, ANDREI JORZA, STEFAN PATRIKIS, WILLIAM A. STEIN, AND CORINA TARNIŢĂ-PĂTRAŞCU

Abstract. We describe theorems and computational methods for verifying the Birch and Swinnerton-Dyer conjecture for specific elliptic curves over $\mathbb{Q}$. We apply our techniques to show that if E is a non-CM elliptic curve over $\mathbb{Q}$ of conductor ≤ 1000 and rank ≤ 1, then the full Birch and Swinnerton-Dyer conjecture is true for E up to odd primes that divide either a Tamagawa number of E or the degree of some rational cyclic isogeny with domain E.

Contents

THIS MATERIAL IS BASED UPON WORK SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. 0400386.

©1997 American Mathematical Society

1. Introduction

Let E be an elliptic curve over $\mathbb{Q}$. The L-function $L(E,s)$ of E is a holomorphic function on $\mathbb{C}$ that encodes deep arithmetic information about E. This paper is about a connection between the behavior of $L(E,s)$ at $s = 1$ and the arithmetic of E.

We use theorems and computation to attack the following conjecture for many specific elliptic curves of conductor ≤ 1000:

Conjecture 1.1 (Birch and Swinnerton-Dyer)**.** *The order of vanishing* $\mathrm{ord}_{s=1} L(E,s)$ *equals the rank r of E, the group $\text{Ш}(E)$ is finite, and*

$$\frac{L^{(r)}(E,1)}{r!} = \frac{\Omega_E \cdot \mathrm{Reg}_E \cdot \prod_p c_p \cdot \#\text{Ш}(E)}{(\#E(\mathbb{Q})_{\mathrm{tor}})^2}.$$

For more about Conjecture 1.1, see [Lan91, Wil00] and the papers they reference. See also Section 1.2 below for the notation used in the conjecture. Henceforth we call it the BSD conjecture.

Definition 1.2 (Analytic Ш)**.** If E has rank r, let

$$\#\text{Ш}(E)_{\mathrm{an}} = \frac{L^{(r)}(E,1) \cdot (\#E(\mathbb{Q})_{\mathrm{tor}})^2}{r! \cdot \Omega_E \cdot \mathrm{Reg}_E \cdot \prod_p c_p}$$

denote the order of $\text{Ш}(E)$ predicted by Conjecture 1.1. We call this the *analytic order* of $\text{Ш}(E)$.

Conjecture 1.3 (BSD(E,p))**.** *Let (E,p) denote a pair consisting of an elliptic curve E over $\mathbb{Q}$ and a prime p. We also call the assertion that* $\mathrm{ord}_{s=1} L(E,s)$ *equals the rank r, that $\text{Ш}(E)[p^\infty]$ is finite, and*

$$\mathrm{ord}_p(\#\text{Ш}(E)[p^\infty]) = \mathrm{ord}_p(\#\text{Ш}(E)_{\mathrm{an}})$$

the BSD conjecture at *p, and denote it* BSD(E,p).

The BSD conjecture is invariant under isogeny.

Theorem 1.4 (Cassels)**.** *If E and F are $\mathbb{Q}$-isogeneous and p is a prime, then* BSD(E,p) *is true if and only if* BSD(F,p) *is true.*

Proof. See [Cas65, Mil86, Jor05]. □

One way to give evidence for the conjecture is to compute $\#\text{Ш}(E)_{\mathrm{an}}$ and note that it is a perfect square, in accord with the following theorem:

Theorem 1.5 (Cassels)**.** *If E is an elliptic curve over $\mathbb{Q}$ and p is a prime such that $\text{Ш}(E)[p^\infty]$ is finite, then $\#\text{Ш}(E)[p^\infty]$ is a perfect square.*

Proof. See [Cas62, PS99]. □

We use the notation of [Crea] to refer to specific elliptic curves over $\mathbb{Q}$.

Conjecture 1.6 (Birch and Swinnerton-Dyer ≤ 1000)**.** *For* all *optimal curves of conductor* ≤ 1000 *we have $\text{Ш}(E) = 0$, except for the following four rank* 0 *elliptic curves, where $\text{Ш}(E)$ has the indicated order:*

Curve	571A	681B	960D	960N
$\#\text{Ш}(E)_{\mathrm{an}}$	4	9	4	4

Theorem 1.7 (Cremona)**.** *Conjecture 1.1 is true for all elliptic curves of conductor* ≤ 1000 *if and only if Conjecture 1.6 is true.*

Proof. In the book [Cre97], Cremona computed $\#Ш(E)_{\mathrm{an}}$ for every curve of conductor ≤ 1000. By Theorem 1.4 it suffices to consider only the optimal ones, and the four listed are the only ones with nontrivial $\#Ш(E)_{\mathrm{an}}$. □

In view of Theorem 1.7, the main goal of this paper is to obtain results in support of Conjecture 1.6. The results of Section 4.2 below together imply the theorem we claimed in the abstract:

Theorem 1.8. *Suppose that E is a non-CM elliptic curve of rank* ≤ 1*, conductor* ≤ 1000 *and that p is a prime. If p is odd, assume further that the mod p representation $\overline{\rho}_{E,p}$ is irreducible and p does not divide any Tamagawa number of E. Then* $\mathrm{BSD}(E,p)$ *is true.*

Proof. Combine Theorem 3.27, Theorem 3.31, and Theorem 4.4. □

For example, if E is the elliptic curve 37A, then according to [Cre97], all $\overline{\rho}_{E,p}$ are irreducible and the Tamagawa numbers of E are 1. Thus Theorem 1.8 asserts that the full BSD conjecture for E is true.

There are 18 optimal curves of conductor ≤ 1000 of rank 2 (and none of rank > 2). For these E of rank 2, nobody has proved that $Ш(E)$ is finite in even a single case. We exclude CM elliptic curves from most of our computations. The methods for dealing with the BSD conjecture for CM elliptic curves are different than for general curves, and will be the subject of another paper. The same is true for $\mathrm{BSD}(E,p)$ when $\overline{\rho}_{E,p}$ is reducible.

1.1. **Acknowledgement.** We thank Michael Stoll for suggesting this project at an American Institute of Mathematics meeting and for initial feedback and ideas, and Stephen Donnelly and Michael Stoll for key ideas about Section 5. We thank John Cremona for many discussions and his immensely useful computer software. Finally we thank Benedict Gross and Noam Elkies for helpful feedback and encouragement throughout the project.

We thank the Harvard College Research Program for funding Pătraşcu's work on this paper, and the Herchel Smith Harvard Undergraduate Research Fellowship for supporting Patrikis. Jorza and Stein were supported by the National Science Foundation under Grant No. 0400386.

1.2. **Notation and Background.** If G is an abelian group, let G_{tor} denote the torsion subgroup and $G_{/\,\mathrm{tor}}$ denote the quotient G/G_{tor}. For an integer m, let $G[m]$ be the kernel of multiplication by m on G. For a commutative ring R, we let R^* denote the group of units in R.

1.2.1. *Galois Cohomology of Elliptic Curves.* For a number field K, let $G_K = \mathrm{Gal}(\overline{\mathbb{Q}}/K)$. Let E be an elliptic curve defined over a number field K, and consider the first Galois cohomology group $\mathrm{H}^1(K,E) = \mathrm{H}^1(G_K, E(\overline{K}))$, and the local Galois cohomology groups $\mathrm{H}^1(K_v,E) = \mathrm{H}^1(\mathrm{Gal}(\overline{K}_v/K_v), E(\overline{K}_v))$, for each place v of K.

Definition 1.9 (Shafarevich-Tate group)**.** The *Shafarevich-Tate group*

$$Ш(E/K) = \mathrm{Ker}\Big(\mathrm{H}^1(K,E) \to \bigoplus_v \mathrm{H}^1(K_v,E)\Big),$$

of E measures the failure of global cohomology classes to be determined by their localizations at all places.

If E is an elliptic curve over a field F and the field F is clear from context, we write $Ш(E) = Ш(E/F)$. For example, if E is an elliptic curve over $\mathbb{Q}$, then $Ш(E)$ means $Ш(E/\mathbb{Q})$.

Definition 1.10 (Selmer group)**.** For each positive integer m, the *m-Selmer group* is

$$\mathrm{Sel}^{(m)}(E/K) = \mathrm{Ker}\Big(\mathrm{H}^1(K, E[m]) \to \bigoplus_v \mathrm{H}^1(K_v, E)\Big).$$

The Selmer group relates the Mordell-Weil and Shafarevich-Tate groups of E via the exact sequence

$$0 \to E(K)/mE(K) \to \mathrm{Sel}^{(m)}(E/K) \to Ш(E/K)[m] \to 0,$$

where $Ш(E/K)[m]$ denotes the m-torsion subgroup of $Ш(E/K)$. Note that $Ш(E/K)$ is a torsion group since $\mathrm{H}^1(K, E)$ is torsion.

1.2.2. *Elliptic Curves over* $\mathbb{Q}$*.* See [Sil92, pp. 360–361] for the definition of $L(E, s)$ and [Wil95, BCDT01] for why $L(E, s)$ is entire.

Let E be an elliptic curve over $\mathbb{Q}$. We use the notation of [Crea] to refer to certain elliptic curves. Thus, e.g., 37B3 refers to the third elliptic curve in the second isogeny class of elliptic curves of conductor 37, i.e., the curve $y^2+y = x^3+x^2-3x+1$. The ordering of isogeny classes and curves in isogeny classes is as specified in [Cre97]. If the last number is omitted, it is assumed to be 1, so 37B refers to the first curve in the second isogeny class of curves of conductor 37.

Let Reg_E be the absolute value of the discriminant of the canonical height pairing on $E(\mathbb{Q})_{/\,\mathrm{tor}}$. Let $c_p = [E(\mathbb{Q}_p) : E_0(\mathbb{Q}_p)]$ be the Tamagawa number of E at p, where $E_0(\mathbb{Q}_p)$ is the subgroup of points that reduce to a nonsingular point modulo p. Let $\Omega_E = \int_{E(\mathbb{R})} |\omega|$, where

$$\omega = \frac{dx}{2y + a_1x + a_3}$$

is the invariant differential attached to a minimal Weierstrass model for E.

For any prime p, let $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p])$ denote the mod p representation and $\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(T_pE)$ the representation on the p-adic Tate module T_pE of E.

It follows from [BCDT01] that every elliptic curve E over $\mathbb{Q}$ is a factor of the modular curve $X_0(N)$, where N is the conductor of E.

Definition 1.11 (Optimal)**.** An elliptic curve E over $\mathbb{Q}$ is *optimal* if for every elliptic curve F and surjective morphisms $X_0(N) \to F \to E$, we have $E \cong F$. (Optimal curves are also called “strong Weil curves” in the literature.)

We say E is a *complex multiplication* (CM) curve, if $\mathrm{End}(E/\overline{\mathbb{Q}}) \neq \mathbb{Z}$.

2. Elliptic Curve Algorithms

2.1. **Images of Galois Representations.** Let E be an elliptic curve over $\mathbb{Q}$. Many theorems that provide explicit bounds on $\#Ш(E)[p^\infty]$ have as a hypothesis that $\overline{\rho}_{E,p}$ or $\rho_{E,p}$ be either surjective or irreducible. In this section we explain how to prove that $\overline{\rho}_{E,p}$ or $\rho_{E,p}$ is surjective or irreducible, in particular cases.

2.1.1. *Irreducibility.* Regarding irreducibility, note that $\overline{\rho}_{E,p}$ is irreducible if and only if there is no isogeny $E \to F$ over $\mathbb{Q}$ of degree p. The degrees of all such isogenies for curves of conductor ≤ 1000 are recorded in [Cre97], which were computed using Cremona's program `allisog`. This program uses results of Mazur [Maz78] along with computations involving modular curves of genus 0.

2.1.2. *Surjectivity.* We discuss surjectivity of $\rho_{E,p}$ in the rest of this section.

Theorem 2.1 (Mazur)**.** *If E is semistable and $p \geq 11$, then $\overline{\rho}_{E,p}$ is surjective.*

Proof. See [Maz78, Thm. 4]. □

Example 2.2. Mazur's theorem implies that the representations $\overline{\rho}_{E,p}$ attached to the semistable elliptic curve $E = X_0(11)$ are surjective for $p \geq 11$. Note that $\overline{\rho}_{E,5}$ is reducible.

Theorem 2.3 (Cojocaru, Kani, and Serre)**.** *If E is a non-CM elliptic curve of conductor N, and*

$$p \geq 1 + \frac{4\sqrt{6}}{3} \cdot N \cdot \prod_{\textit{prime } \ell | N} \left(1 + \frac{1}{\ell}\right)^{1/2},$$

then $\overline{\rho}_{E,p}$ is surjective.

Proof. See Theorem 2 of [CK], whose proof relies on the results of [Ser72]. □

Example 2.4. When $N = 11$, the bound of Theorem 2.3 is ~ 38.52. When $N = 997$, the bound is ~ 3258.8. For $N = 40000$, the bound is ~ 143109.35.

Proposition 2.5. *Let E be a non-CM elliptic curve over $\mathbb{Q}$ of conductor N and let $p \geq 5$ be a prime. For each prime $\ell \nmid p \cdot N$ with $a_\ell \not\equiv 0 \pmod{p}$, let*

$$s(\ell) = \left(\frac{a_\ell^2 - 4\ell}{p}\right) \in \{0, -1, +1\},$$

where the symbol $\left(\frac{\cdot}{\cdot}\right)$ is the Legendre symbol. If -1 and $+1$ both occur as values of $s(\ell)$, then $\overline{\rho}_{E,p}$ is surjective. If $s(\ell) \in \{0, 1\}$ for all ℓ, then $\mathrm{Im}(\overline{\rho}_{E,p})$ is contained in a Borel subgroup (i.e., reducible), and if $s(\ell) \in \{0, -1\}$ for all ℓ, then $\mathrm{Im}(\overline{\rho}_{E,p})$ is a nonsplit torus.

Proof. This is an application of [Ser72, §4], where we use the quadratic formula to convert the condition that certain polynomials modulo p be reducible or irreducible into a quadratic residue symbol. □

For computational applications we apply Proposition 2.5 as follows. We choose a bound B and compute values $s(\ell)$; if both -1 and $+1$ occur as values of $s(\ell)$, we stop computing $s(\ell)$ and conclude that $\overline{\rho}_{E,p}$ is surjective. If for $\ell \leq B$ we find that $s(\ell) \in \{0, 1\}$, we suspect that $\mathrm{Im}(\overline{\rho}_{E,p})$ is Borel, and attempt to show this (see Section 2.1.1). If for $\ell \leq B$, we have $s(\ell) \in \{0, -1\}$, we suspect that $\mathrm{Im}(\overline{\rho}_{E,p})$ is contained in a nonsplit torus, and try to show this by computing and analyzing the p-division polynomial of E. If this approach is inconclusive, we can alway increase B and eventually the process terminates. In practice we often apply some theorem under the hypothesis that $\overline{\rho}_{E,p}$ is surjective, which is something that in practice we verify for a particular p using Proposition 2.5.

Example 2.4 suggests that the bound of Theorem 2.3 is probably far larger than necessary. Nonetheless, it is small enough that in a reasonable amount of time we can determine whether $\overline{\rho}_{E,p}$ is surjective, using the above process, for all p up to the bound. In this way we determine the exact image of Galois.

Remark 2.6. We can also determine surjectivity of the mod 2 and mod 3 representations directly using the 3-division polynomial of E. For $p \leq 3$ one can show that $\overline{\rho}_{E,p}$ is surjective if and only if the p-division polynomial (of degree n) has Galois group S_n.

Theorem 2.7 (Serre)**.** *If $p \geq 5$ is a prime of good reduction, then $\rho_{E,p}$ is surjective if and only if $\overline{\rho}_{E,p}$ is surjective.*

Proof. This is proved in greater generality as [Ser72, Thm. 4′, pg. 300]. □

Remark 2.8. This result does not extend to $p = 3$ (see [Ser98, Ex. 3, pg IV-28]). In fact, there are infinitely many elliptic curves with $\overline{\rho}_{E,p}$ surjective, but $\rho_{E,p}$ not surjective (see forthcoming work of Noam Elkies).

2.2. Special Values of L-Functions.

Let E be an elliptic curve over $\mathbb{Q}$ of conductor N, and let $f = \sum a_n q^n$ be the corresponding cusp form.

The following lemma will be useful in determining how many terms of the L-series of E are needed to compute the L-series to a given precision. (We could give a strong bound, but for our application this will be enough, and is simplest to apply in practice.)

Lemma 2.9. *For any positive integer n, we have $|a_n| \leq n$.*

Proof. For p prime we know that $a_p = \alpha + \beta$, where α and β are the roots of $x^2 - a_p x + p = 0$. Note that $|\alpha| = |\beta| = \sqrt{p}$.

Since a_n is multiplicative, it is enough to show $|a_n| \leq n$ for prime powers p^r. Let $r > 1$. Then $a_{p^r} = a_p a_{p^{r-1}} - p a_{p^{r-2}}$, and by induction,

$$a_{p^r} = \frac{\alpha^{r+1} - \beta^{r+1}}{\alpha - \beta}.$$

Then

$$|a_{p^r}| \leq \frac{2p^{(r+1)/2}}{|\alpha - \beta|} = \frac{2p^{(r+1)/2}}{\left|\sqrt{4p - a_p^2}\right|}.$$

Note that the sign is changed since we only deal with absolute values. We need to show that this is $\leq p^r$. This happens if

$$\frac{2}{\sqrt{4p - a_p^2}} \leq p^{(r-1)/2}.$$

Since $a_p^2 < 4p$ the difference is at least 1 so it is enough to show that $2 \leq p^{(r-1)/2}$. This is true as long as $p > 3$. For $p = 2$ and $p = 3$ note that a_p is an integer with $|a_p| < 2\sqrt{p}$. For $p = 2$ this integer is at most 2 and so $4p - a_p^2 \geq 4$. Similarly for $p = 3$ this is at most 3 and so $4p - a_p^2 \geq 4$. Therefore it is enough to show that $1 \leq p^{(r-1)/2}$, which is true for all $r > 1$. □

Suppose E has even analytic rank. By [Cre97, §2.13] or [Coh93, Prop. 7.5.8], we have

$$L(E,1) = 2\sum_{n=1}^{\infty} \frac{a_n}{n} e^{-2\pi n/\sqrt{N}}, \tag{2.1}$$

where a_n are the Fourier coefficients of the normalized eigenform associated with E. Using the bound $|a_n| \leq n$ of Lemma 2.9, we see that if we truncate the series (2.1) at the kth term, the error is at most

$$\varepsilon = 2\sum_{n=k}^{\infty} e^{-2\pi n/\sqrt{N}} = \frac{2e^{-2\pi k/\sqrt{N}}}{1 - e^{-2\pi/\sqrt{N}}},$$

and the quantity on the right can easily be evaluated.

Next suppose E has odd analytic rank. In [Cre97, §2.13] or [Coh93, Prop. 7.5.9] we find that

$$L'(E,1) = 2\sum_{n=1}^{\infty} \frac{a_n}{n} G_1(2\pi n/\sqrt{N}).$$

We have

$$G_1(x) = \int_1^{\infty} e^{-xy}\frac{dy}{y} = \int_x^{\infty} e^{-y}\frac{dy}{y} \leq e^{-x},$$

and we obtain the same error bound as for $L(E,1)$. (In fact, $G_1(x) \leq e^{-x}/x$ but we will not need this stronger bound.)

2.3. **Mordell-Weil Groups.**

If E is an elliptic curve over $\mathbb{Q}$ of analytic rank ≤ 1, there are algorithms to compute $E(\mathbb{Q})$ that are guaranteed to succeed. This is because $\#Ш(E)$ is finite, by [Kol91]. Independent implementations of these algorithms are available as part of `mwrank` [Creb] and MAGMA [BCP97]. We did most of our computations of $E(\mathbb{Q})$ using `mwrank`, but use MAGMA in a few cases, since it implements 3-descents, 4-descents and Heegner points methods (thanks to work of Tom Womack, Mark Watkins, and others).

2.4. **Other Algorithms.**

We use many other elliptic curves algorithms, for example, for computing root numbers and the coefficients a_n of the modular form associated to E. For the most part, we used the PARI (see [ABC$^+$]) C-library via SAGE (see [Ste]). For descriptions of these general elliptic curves algorithms, see [Coh93, Cre97].

3. The Kolyvagin Bound

In this section we describe a bound due to Kolyvagin on $\#Ш(E)$, and compute it for many specific elliptic curves over $\mathbb{Q}$. In fact, the bound is on $\#Ш(E/K)$, where K is a quadratic imaginary field; this is not a problem, because the natural map $Ш(E/\mathbb{Q}) \to Ш(E/K)$ has kernel of order a power of 2, so the bound is also a bound on the odd part of $\#Ш(E)$.

Let E be an elliptic curve over $\mathbb{Q}$ of conductor N. For any quadratic imaginary field $K = \mathbb{Q}(\sqrt{-D})$, let E^D denote the twist of E by D. If E is defined by $y^2 = x^3 + ax + b$, then E^D is defined by $y^2 = x^3 + D^2ax + D^3b$, and

$$L(E/K,s) = L(E,s)\cdot L(E^D,s).$$

Definition 3.1 (Heegner Hypothesis)**.** We say that K satisfies the *Heegner hypothesis* for E if $K \neq \mathbb{Q}(\sqrt{-1}), \mathbb{Q}(\sqrt{-3})$, and every prime factor of N splits as a product of two distinct primes in the ring of integers of K. (The condition $K \neq \mathbb{Q}(\sqrt{-1}), \mathbb{Q}(\sqrt{-3})$ is not necessary for some of the results below, but we include it for simplicity.)

If K satisfies the Heegner hypothesis for E, then there is a Heegner point $y_K \in E(K)$, which is the sum of images of certain complex multiplication (CM) points on $X_0(N)$ (see [GZ86, §I.3]). Properties of this point impact the arithmetic of E over K.

3.1. **Bounds on $\#\text{Ш}(E/K)$.** Suppose that K is an imaginary quadratic extension of $\mathbb{Q}$ that satisfies the Heegner hypothesis for E. Kolyvagin proved the following theorem in [Kol90]:

Theorem 3.2 (Kolyvagin)**.** *Let $R = \mathrm{End}(E/\mathbb{C})$ and let $F = \mathrm{Frac}(R)$, so if E is non-CM then $F = \mathbb{Q}$. If p is an odd prime unramified in F such that $\mathrm{Gal}(F(E[p])/F) = \mathrm{Aut}_R(E[p])$, i.e., $\mathrm{Im}(\overline{\rho}_{E,p})$ is as large as possible, then*

$$\mathrm{ord}_p(\#\text{Ш}(E/K)) \leq 2 \cdot \mathrm{ord}_p([E(K) : \mathbb{Z}y_K]).$$

Note that if E does not have complex multiplication, the hypotheses of both these theorems imply that $p \nmid \#E(K)_{\mathrm{tor}}$ (see Lemma 5.7).

Cha [Cha03, Cha05] extended Kolyvagin's method to provide better bounds on $\text{Ш}(E/K)$ in some cases. Let K be a number field, let D_K be the discriminant of K, and let N be the conductor of E.

Theorem 3.3 (Cha)**.** *If $p \nmid D_K$, $p^2 \nmid N$, and $\overline{\rho}_{E,p}$ is irreducible, then*

$$\mathrm{ord}_p(\#\text{Ш}(E/K)) \leq 2 \cdot \mathrm{ord}_p([E(K) : \mathbb{Z}y_K]).$$

As we will see in the proof of Theorem 4.3 below, there is one curve that satisfies the hypotheses of that theorem, but for which we cannot use Theorem 3.2 to prove $\mathrm{BSD}(E, 5)$. The problem is that $\overline{\rho}_{E,5}$ is not surjective. We can use Cha's theorem though:

Lemma 3.4. *Let E be the elliptic curve* 608B*, which has rank* 0*. Then* $\mathrm{BSD}(E, 5)$ *is true for E.*

Proof. Since E admits no 5-isogeny (see [Cre97]), $\overline{\rho}_{E,5}$ is irreducible. Also, $5^2 \nmid 608$, so for any Heegner K of discriminant coprime to 5 we can apply Theorem 3.3. Taking $K = \mathbb{Q}(\sqrt{-79})$, we find that the odd part of $[E(K) : \mathbb{Z}y_K]$ is 1, so $5 \nmid \#\text{Ш}(E/K)$. It follows that $5 \nmid \#\text{Ш}(E)$, so $\mathrm{BSD}(E, 5)$ is true, according to Theorem 1.7. □

Cha's assumption on the reduction of E at p and that $p \nmid D_K$ is problematic when there is a prime $p \geq 5$ of additive reduction or one uses only one K. This situation does occur in several cases, which motivated us to prove the following theorem:

Theorem 3.5. *Suppose E is a non-CM elliptic curve over $\mathbb{Q}$. Suppose K is a quadratic imaginary field that satisfies the Heegner hypothesis and p is an odd prime such that $p \nmid \#E'(K)_{\mathrm{tor}}$ for any curve E' that is $\mathbb{Q}$-isogenous to E. Then*

$$\mathrm{ord}_p(\#\text{Ш}(E)) \leq 2\,\mathrm{ord}_p([E(K) : \mathbb{Z}y_K]),$$

unless $\mathrm{disc}(K)$ *is divisible by exactly one prime* ℓ*, in which case the conclusion is only valid if* $p \neq \ell$*.*

Since the proof of Theorem 3.5 is somewhat long and technical, we defer the proof until Section 5.

Remark 3.6. If in Theorem 3.5, $\overline{\rho}_{E,p}$ is irreducible, then $p \nmid \#E'(K)_{\mathrm{tor}}$ for all E' isogenous to E. This is because the isogeny $E \to E'$ has degree coprime to p, so $E[p] \cong E'[p]$. Also, since $E[p]$ is irreducible, if $E'(K)$ were to contain a p-torsion point, it would have to contain all of them, a contradiction since $\boldsymbol{\mu}_p \not\subset K$ (recall that we exclude $\mathbb{Q}(\sqrt{-3})$ and $\mathbb{Q}(\sqrt{-4})$).

Theorem 3.7 (Bump-Friedberg-Hoffstein, Murty-Murty, Waldspurger)**.** *There are infinitely many quadratic imaginary extensions* $K/\mathbb{Q}$ *such that* K *satisfies the Heegner hypothesis and* $\mathrm{ord}_{s=1} L(E/K) = 1$*.*

Proof. If $\mathrm{ord}_{s=1}L(E,s) = 0$, then the papers [MM91] and [BFH90] both imply the existence of infinitely many K such that y_K has infinite order. If $\mathrm{ord}_{s=1}L(E,s) = 1$, then a result of Waldspurger ([Wal85]) applies, as does [BFH90]. □

Theorem 3.7 is not used in our computations, but ensures that our procedure for bounding $\#Ш(E)$, when E has analytic rank ≤ 1, is an algorithm, i.e., it always terminates with a nontrivial upper bound.

3.2. **The Gross-Zagier Formula.** We use the Gross-Zagier formula to compute the index $[E(K) : \mathbb{Z}y_K]$ without explicitly computing y_K.

The modularity theorem of [BCDT01] asserts that there exists a surjective morphism $\pi : X_0(N) \to E$. Choose π to have minimal degree among all such morphisms. Let $\pi^*(\omega)$ be the pullback of a minimal invariant differential ω on E. Then $\pi^*(\omega) = \alpha \cdot f$, for some constant α and some normalized cusp form f. By [Edi91, Prop. 2], we know that $\alpha \in \mathbb{Z}$.

Definition 3.8 (Manin Constant)**.** The *Manin constant* of E is $c = |\alpha|$.

Manin conjectured in [Man72, §5] that $c = 1$ for the optimal curve in the $\mathbb{Q}$-isogeny class of E.

Theorem 3.9 (Gross-Zagier)**.** *If* K *satisfies the Heegner hypothesis for* E*, then the Néron-Tate canonical height of* y_K *is*

$$h(y_K) = \frac{\sqrt{D}}{c^2 \cdot \int_{E(\mathbb{C})} \omega \wedge \overline{i\omega}} \cdot L'(E/K, 1).$$

Proof. Gross and Zagier proved the following formula in [GZ86] under the hypothesis that D is odd. For the general assertion see [Zha04, Thm. 6.1]. □

3.3. **Remarks on the Index.** Suppose that E is an elliptic curve over $\mathbb{Q}$ of conductor N and that E has analytic rank 1 over a quadratic imaginary field K that satisfies the Heegner hypothesis. In [McC91], McCallum rephrases the analogue of Conjecture 1.1 for E over K using the Gross-Zagier formula as follows:

Conjecture 3.10 (Birch and Swinnerton-Dyer)**.** *Suppose* K *is a quadratic imaginary field that satisfies the Heegner hypothesis, and that* E *has analytic rank* 1 *over* K*. Then*

$$\#Ш(E/K) = \left(\frac{[E(K) : \mathbb{Z}y_K]}{c^2 \cdot \prod_{p \mid N} c_p} \right)^2 .$$

Here the c_p are the Tamagawa numbers of E over $\mathbb{Q}$, c is the Manin constant of E, and $\mathbb{Z}y_K$ is the cyclic group generated by y_K.

Remark 3.11. A serious issue is that Conjecture 3.10 implies that the index $I_K = [E(K) : \mathbb{Z}y_K]$ will be divisible by the Tamagawa numbers c_p. One sees using Tate curves that these Tamagawa numbers can be arbitrarily large. In many cases when E has analytic rank 0, we could instead apply Theorem 4.1 below, but when E has analytic rank 1 a new approach is required, e.g., computation of p-adic regulators and use of results of P. Schneider and others toward p-adic analogues of the BSD conjecture. This will be the subject of a future paper.

Remark 3.12. Conjecture 3.10 has interesting implications in certain special cases. For example, if E is the curve 91B, then $c_7 = c_{13} = 1$. Also $c = 1$, as Cremona has verified, and $\#E(\mathbb{Q})_{\mathrm{tor}} = 3$. Thus for any K, we have $3 \mid [E(K) : \mathbb{Z}y_K]$. If y_K has infinite order, then Conjecture 3.10 implies that $3^2 \mid \#Ш(E/K)$. For $K = \mathbb{Q}(\sqrt{-103})$, the point y_K is torsion, and in this case $E(K)$ has rank 3 and (conjecturally) $Ш(E/K)[3] = 0$. See Remark 3.23 for another example along these lines.

3.4. **Mordell-Weil Groups and Quadratic Imaginary Fields.** Let E be an elliptic curve over $\mathbb{Q}$ and $K = \mathbb{Q}(\sqrt{D})$ a quadratic imaginary field such that $E(K)$ has rank 1. In this section we explain how to understand $E(K)$ in terms of $E(\mathbb{Q})$ and $E^D(\mathbb{Q})$.

Proposition 3.13. *Let $R = \mathbb{Z}[1/2]$ and $K = \mathbb{Q}(\sqrt{D})$. For any squarefree integer $D \neq 1$, we have*

$$E(K) \otimes R = (E(\mathbb{Q}) \otimes R) \oplus (E^D(\mathbb{Q}) \otimes R).$$

Proof. Let τ be the complex conjugation automorphism on $E(K) \otimes R$. The characteristic polynomial of τ is $x^2 - 1$, which is squarefree, so $E(K) \otimes R$ is a direct sum of its $+1$ and -1 eigenspaces for τ. The natural map $E(\mathbb{Q}) \hookrightarrow E(K)$ identifies $E(\mathbb{Q}) \otimes R$ with the $+1$ eigenspace for τ since $E(K)^{G_\mathbb{Q}} = E(\mathbb{Q})$; likewise, $E^D(\mathbb{Q}) \hookrightarrow E(K)$ identifies $E^D(\mathbb{Q}) \otimes R$ with the -1 eigenspace for τ. □

The following slightly more refined proposition will be important for certain explicit Heegner point computations (directly after Equation 3.1).

Proposition 3.14. *Suppose $E(K)$ has rank 1. Then the image of either $E(\mathbb{Q})_{/\,\mathrm{tor}}$ or $E^D(\mathbb{Q})_{/\,\mathrm{tor}}$ has index at most 2 in $E(K)_{/\,\mathrm{tor}}$.*

Proof. Since $E(K)$ has rank 1, Proposition 3.13 implies that exactly one of $E(\mathbb{Q})$ and $E^D(\mathbb{Q})$ has rank 1 and the other has rank 0. We may assume that $E(\mathbb{Q})$ has rank 1 (otherwise, swap E and E^D). Let i be the natural inclusion $E(\mathbb{Q}) \hookrightarrow E(K)$, and let τ denote the automorphism of $E(K)$ induced by complex conjugation. Then $P \mapsto (1+\tau)P$ induces a map $E(K) \to E(K)^+ = E(\mathbb{Q})$ that, upon taking quotients by torsion, induces a map $\psi : E(K)_{/\,\mathrm{tor}} \to E(\mathbb{Q})_{/\,\mathrm{tor}}$. Let P_1 be a generator for $E(\mathbb{Q})_{/\,\mathrm{tor}}$ and P_2 a generator for $E(K)_{/\,\mathrm{tor}}$, and write $i(P_1) = nP_2$, for some nonzero integer n. Then

$$[2]P_1 = \psi(i(P_1)) = \psi(nP_2) = [n]\psi(P_2) = [nm]P_1 \pmod{E(\mathbb{Q})_{\mathrm{tor}}},$$

for some nonzero integer m. Thus $2 = nm$, so $n \leq 2$. □

If D satisfies the Heegner hypothesis, then by computing the residue symbol $\left(\frac{N}{D}\right)$ and understanding how the sign of the functional equation changes under twist, we see that

$$\operatorname{ord}_{s=1} L(E,s) \not\equiv \operatorname{ord}_{s=1} L(E^{(D)},s) \pmod 2.$$

Suppose K satisfies the Heegner hypothesis and $\operatorname{ord}_{s=1} L(E/K,s) = 1$. Then work of Kolyvagin (see [Kol91, Kol88]) implies that $E(K)$ has rank 1.

The root number $\varepsilon_E = \pm 1$ of E is the sign of the functional equation of $L(E,s)$. If $\varepsilon_E = +1$, then the analytic rank $\operatorname{ord}_{s=1} L(E,s)$ is even, and if $\varepsilon_E = -1$, then it is odd.

Proposition 3.15. *Let E be an elliptic curve, let $D = D_K$ be a discriminant that satisfies the Heegner hypothesis such that $\operatorname{ord}_{s=1} L(E/K,s) = 1$, and let $R = \mathbb{Z}[1/2]$. Then*

(1) *If $\varepsilon_E = +1$, then a generator of $E(K) \otimes R$ is the image of a generator of $E^D(\mathbb{Q}) \otimes R$ and $L'(E/K,1) = L(E,1) \cdot L'(E^D,1)$.*

(2) *If $\varepsilon_E = -1$, then a generator of $E(K) \otimes R$ is the image of a generator of $E(\mathbb{Q}) \otimes R$ and $L'(E/K,1) = L'(E,1) \cdot L(E^D,1)$.*

We will use the above proposition to relate computation of $E(K) \otimes R$ to computation of Mordell-Weil groups of elliptic curves defined over $\mathbb{Q}$.

3.5. **Computing the Index of the Heegner Point.** A key input to the theorems of Section 3.1 is computation of the index $[E(K) : \mathbb{Z}y_K]$. We have

$$[E(K)_{/\mathrm{tor}} : \mathbb{Z}y_K]^2 = h(y_K)/h(z), \tag{3.1}$$

where z is a generator of $E(K)_{/\,\mathrm{tor}}$.

In the Gross-Zagier formula we have $h = h_K$, the Néron-Tate canonical height on $E(K) = E^D(K)$ relative to K. Let $h_{\mathbb{Q}}$ denote the height on $E(\mathbb{Q})$ or $E^D(\mathbb{Q})$. Note that if $P \in E(\mathbb{Q})$ or $E^D(\mathbb{Q})$, then

$$h_{\mathbb{Q}}(P) = \frac{1}{[K:\mathbb{Q}]} \cdot h_K(P) = \frac{h_K(P)}{2}. \tag{3.2}$$

Using Proposition 3.14 and algorithms for computing Mordell-Weil groups (see Section 2.3), we can compute z or $2z$, so we can compute $h(z)$ or $2h(z)$. In practice, even for curves of conductor up to 1000, it can take a huge amount of time to compute z. This section about practical methods to either compute the index or at least bound it.

It is not difficult to compute $h(y_K)$, without computing y_K itself, using the Gross-Zagier formula (Section 3.2). We compute $L'(E/K,1)$ by computing L-functions of elliptic curves defined over $\mathbb{Q}$ as explained in Proposition 3.15. It remains to compute

$$\alpha = \frac{\sqrt{|D|}}{c^2 \int_{E(\mathbb{C})} \omega \wedge \overline{i\omega}}. \tag{3.3}$$

3.5.1. *The Manin Constant.* Manin conjectured that the Manin constant c for any optimal elliptic curve factor E of $X_0(N)$ is 1, and there are bounds on the possibilities for c (see, e.g., [Edi91, ARS05]). There is an algorithm to verify in any particular case that $c = 1$, as explained in the proof of the following proposition.

Proposition 3.16 (Cremona)**.** *If E is an optimal elliptic curve of conductor at most 80000, then the Manin constant of E is 1.*

Proof. For each level $N \leq 80000$ we do the following. Using the modular symbols algorithms of [Cre97], we enumerate the rational newforms $f_1, \ldots, f_d$, which correspond (via the modularity theorem) to the optimal elliptic curves $E_1, \ldots, E_d$ of conductor N, respectively. For each f_i we compute approximations to **xxx** decimal digits of the c_4 and c_6 invariants of the lattice Λ_{E_i} attached to the optimal curve in the isogeny class. We then guess integers c_4' and c_6' that are close to the computed approximations, and verify that the elliptic curve E_i' with invariants c_4', c_6' is an elliptic curve of conductor N. We also compute the full isogeny class of E_i' using the program `allisog`. Repeating this procedure for each newform f, we obtain d distinct isogeny classes of elliptic curves of conductor N, and by modularity these must be in bijection with the newforms f_i. However, at this point we have not *proved* that $E_i = E_i'$ or even that E_i' is an optimal quotient. However, we have provably found all elliptic curves over $\mathbb{Q}$ of conductor N.

We next compute the c_4 and c_6 invariants of all curves of conductor N, and observe that the first 12 digits of the c-invariants for these curves are sufficient to distinguish them. (12 digits is enough for every curve up to conductor 80000.) If we had guessed c_4' and c_6' incorrectly above, so that $E_i' \neq E_i$, there would be two curves of conductor N both of whose c-invariants have the same initial **xxx** decimal digits, which is impossible since 12 digits of precision are sufficient to distinguish any two. Thus $E_i' = E_i$, and the c_4', c_6' we computed are the correct invariants of the optimal quotient attached to f_i.

Finally, we observe that c_4' and c_6' are the invariants of a minimal Weierstrass equation, which implies that the Manin constant of E_i is 1. □

3.5.2. *The Integral.* We have the following lemma regarding the integral in (3.3):

Lemma 3.17. *We have* $\int_{E(\mathbb{C})} \omega \wedge \overline{i\omega} = 2 \cdot \mathrm{Vol}(\mathbb{C}/\Lambda)$, *where the volume* $\mathrm{Vol}(\mathbb{C}/\Lambda)$ *is the absolute value of the determinant of a matrix formed from a basis for the lattice in* $\mathbb{C}$ *obtained by integrating the Néron differential* ω_E *against all homology classes in* $\mathrm{H}_1(E, \mathbb{Z})$.

Proof. Fix the Weierstrass equation $y^2 = 4x^3 + g_4 x + g_6$ for E, so $x = \wp(z)$ and $y = \wp'(z)$. First note that

$$\omega = \frac{dx}{y} = \frac{d\wp(z)}{\wp'(z)} = \frac{\wp'(z)dz}{\wp'(z)} = dz.$$

Thus

$$\begin{aligned}\int_{E(\mathbb{C})} \omega \wedge \overline{i\omega} &= \int_{\mathbb{C}/\Lambda} dz \wedge \overline{idz} \\ &= -i \int_{\mathbb{C}/\Lambda} (dx + idy) \wedge (dx - idy) \\ &= -i(2i) \int_{\mathbb{C}/\Lambda} dx \wedge dy = 2 \cdot \mathrm{Vol}(\mathbb{C}/\Lambda).\end{aligned}$$

□

Note that $\mathrm{Vol}(\mathbb{C}/\Lambda)$ can be computed to high precision using the Gauss arithmetic-geometric mean, as described in [Cre97, §3.7].

3.5.3. *Mordell-Weil Groups and Heights.* For the curves that we run our computation on, we use [Creb] (via [Ste]), which computes a basis for $E^D(\mathbb{Q})$, and not just a basis for a subgroup of finite index.

Cremona describes the computation of heights of points on curves defined over $\mathbb{Q}$ in detail in [Cre97, §3.4]. There is an explicit bound on the error in the height computation, which shrinks exponentially in terms of the precision of approximating series, and can be made arbitrarily small. For the L-function computations, see Section 2.2.

3.5.4. *Indexes of Heegner Points on Rank* 1 *Curves.* Suppose E is an elliptic curve over $\mathbb{Q}$ of analytic rank 1, and we wish to compute indexes $i_K = [E(K)_{/\,\mathrm{tor}} : \mathbb{Z}y_K]$ for various K. Assume that $E(\mathbb{Q})$ is known, so we can compute $h(z)$ to high precision, where z generates $E(\mathbb{Q})/_{\mathrm{tor}}$. Then computing the indexes i_K is relatively easy. For each K, compute $h(y_K)$ as described above using the Gross-Zagier formula, so

$$h(y_K) = \alpha \cdot L'(E,1) \cdot L(E^D,1).$$

Then

$$i_K = \sqrt{\frac{h(y_K)}{h(z)}} = \sqrt{\frac{h(y_K)}{2h_{\mathbb{Q}}(z)}}.$$

We emphasize that computation of the Heegner point itself is not necessary. For the results of this index computation for E of conductor ≤ 1000, see Section 3.6.1.

Example 3.18. Let E be the elliptic curve 540B, which has rank 1, and conductor $540 = 2^2 \cdot 3^3 \cdot 5$. The first K that satisfies the Heegner hypothesis is $\mathbb{Q}(\sqrt{-71})$. The group $E(\mathbb{Q})$ is generated by $z = (0,1)$, and we have $h_{\mathbb{Q}}(z) \sim 0.656622630$. We have

$$\alpha \sim \frac{\sqrt{71}}{2 \cdot 3.832955} \sim 1.09917,$$

so

$$h(y_K) \sim 1.09917 \cdot 1.9340458 \cdot 5.559761726 \sim 11.819.$$

Thus

$$i_K = \sqrt{\frac{11.819}{2 \cdot 0.656622630}} \sim \sqrt{8.99999} \sim 3.$$

3.5.5. *Indexes of Heegner Points on Rank* 0 *Curves.* Assume that the analytic rank of E is 0. In practice, computing the indexes of Heegner points in this case is substantially more difficult than the rank 1 case. For a Heegner quadratic imaginary field $K = \mathbb{Q}(\sqrt{D})$, we have

$$i_K = [E(K)_{/\,\mathrm{tor}} : \mathbb{Z}y_K]^2 = \frac{h(y_K)}{h(z)} = \alpha \cdot \frac{L(E,1) \cdot L'(E^D,1)}{h(z)},$$

so one method to find i_K is to find a generator $z \in E^D(\mathbb{Q})$ exactly using descent algorithms, which will terminate since we know that $Ш(E^D)$ is finite, by Kolyvagin's theorem. However, since E^D has potentially large conductor and rank 1, in practice the Mordell-Weil group will sometimes be generated by a point of large height, hence be extremely time consuming to find. One can use 2-descent, 3-descent, 4-descent, and Heegner points methods (i.e., explicitly compute the coordinates of the Heegner point as decimals and try to recognize them using continued fractions.) In some cases these methods produce in a reasonable amount of time an element of

$E^D(\mathbb{Q})$ of infinite order, and one can then saturate the point using [Creb] to find a generator z.

Example 3.19. Let E be the curve 11A. The first field that satisfies the Heegner hypothesis is $K = \mathbb{Q}(\sqrt{-7})$. The conductor of $F = E^{-7}$ is 539, and we find a generator $z \in F(\mathbb{Q})$ for the Mordell-Weil group of this twist. This point has height $h_{\mathbb{Q}}(z) \sim 0.1111361471$. We have

$$\alpha \sim \frac{\sqrt{7}}{2 \cdot 1.8515436234} \sim 0.71447177.$$

The height over K of the Heegner point is thus

$$h(y_K) \sim 0.71447177 \cdot 0.25384186 \cdot 1.225566874 \sim 0.2222722925.$$

Thus by (3.2)

$$i_K = \frac{h(z)}{h(y_K)} = \frac{2h_{\mathbb{Q}}(z)}{h(y_K)} \sim 1.$$

There is a trick to bound the index i_K without computing *any* elements of $E(K)$. This is useful when the algorithms mentioned above for computing a generator of $E^D(\mathbb{Q})$ produce no information in a reasonable amount of time. First compute the height $h(y_K)$ using the Gross-Zagier formula. Next compute the Cremona-Prickett-Siksek [Pri04, Ch. 4] bound B for E^D, which is a number such that if $P \in E^D(\mathbb{Q})$, then the naive logarithmetic height of P is off from the canonical height of P by at most B. Using standard sieving methods implemented in [Creb], we compute all points on E of naive logarithmic height up to some number h_0. If we find any point of infinite order, we saturate, and hence compute $E^D(\mathbb{Q})$, then use the above methods. If we find no point of infinite order, we conclude that there is no point in $E^D(\mathbb{Q})$ of canonical height $\leq h_0 - B$. If $h_0 - B > 0$, we obtain an upper bound on i_K as follows. If z is a generator for $E^D(\mathbb{Q})$, then $h_{\mathbb{Q}}(z) > h_0 - B$, so using (3.2) we have

$$h_{\mathbb{Q}}(z) = \frac{1}{2} \cdot h_K(z) = \frac{h(y_K)}{2 \cdot i_K^2} > h_0 - B.$$

Solving for i_K gives

$$i_K < \sqrt{\frac{h(y_K)}{2(h_0 - B)}}, \tag{3.4}$$

so to bound i_K we consider many K (e.g., the first 30), and for each compute the quantity on the right side of (3.4) for a fixed choice of h_0. We then use a K that minimizes this quantity.

Remark 3.20. Another approach to finding some Heegner point, which we discussed with Noam Elkies, is to search for small points on $E(K)$ over various fields K, until finding a K that satisfies the Heegner hypothesis and is such that $E(K)$ has rank 1. For example, if E is given by $y^2 = x^3 + ax + b$, and x_0 is a small integer, write $y_0^2 \cdot D = x_0^3 + ax_0 + b$, where y_0 and D are integers, and D is square free. Then (x_0, y_0) is a point on the quadratic twist of E by D. We did not use this approach, since it was not necessary in order to prove Theorem 1.8.

Example 3.21. Let E be the elliptic curve 546E. Then $K = \mathbb{Q}(\sqrt{-311})$ satisfies the Heegner hypothesis, since the prime divisors of $546 = 2 \cdot 3 \cdot 7 \cdot 13$ split completely in K. We compute the height of the Heegner point y_K. Let F be the quadratic twist of E by -311. We have

$$\alpha \sim \frac{\sqrt{311}}{2 \cdot 0.0340964942689662168001} \sim 258.60711587$$

Thus

$$\begin{aligned} h(y_K) &\sim \alpha \cdot L(E,1) \cdot L'(F,1) \\ &\sim 258.60711587 \cdot 2.2783578 \cdot 12.41550 \sim 7315.20688, \end{aligned}$$

where in each case we compute the L-series using enough terms to obtain a value correct to $\pm 10^{-5}$. Thus 7320 is a conservative upper bound on $h(y_K)$. The Cremona-Prickett-Siksek bound for F is $B = 13.0825747$. We search for points on F of naive logarithmic height ≤ 18, and find no points. Thus (3.4) implies that

$$i_K < \sqrt{7320/(2 \cdot (18 - 13.0825747))} \sim 27.28171 < 28.$$

It follows that if $p \mid i_K$, then $p \leq 23$. Searching up to height 21 would (presumably) allow us to remove 23, but this might take much longer.

For the results of our computations for all E of conductor ≤ 1000, see Section 3.6.2.

3.6. **Results of Computations.**

3.6.1. *Curves of Rank* 1. First we consider curves of rank 1. Recall from Conjecture 1.6 that we expect Ш to be trivial for all optimal rank 1 curves of conductor at most 1000.

Proposition 3.22. *Suppose (E, p) is a pair with E an optimal elliptic curve of conductor up to* 1000 *of rank* 1*. Let I be the greatest common divisor of $[E(K)_{/\,\mathrm{tor}} : \mathbb{Z}y_K]$ for the first four quadratic imaginary fields $K = \mathbb{Q}(\sqrt{D})$ that satisfy the Heegner hypothesis. If $p \mid I$, then*

$$p \mid 2 \cdot \#E(\mathbb{Q})_{\mathrm{tor}} \cdot \prod_{q \mid N} c_{E,q},$$

except if (E, p) is $(540B, 3)$ or $(756B, 3)$.

Proof. For each rank 1 curve E of conductor up to 1000 we perform the following computation.

1. Let R_E be the regulator of E, correct to precision at least 10^{-10}, which we look up in the `allbsd` table of [Crea].
2. List the first four discriminants $D = D_0, D_1, D_2, D_3$ such that $K = \mathbb{Q}(\sqrt{D})$ satisfies the Heegner hypothesis. For each $D = D_i$ do the following computation:
   - (a) Compute $L'(E,1)$ to some bounded precision ε, using $2\sqrt{N}+10$ terms. The bound ε is determined as explained in Section 2.2.
   - (b) Compute $L(E^D,1)$ to some bounded precision ε' using $2\sqrt{N}+10$ terms.
   - (c) Compute $\alpha = \sqrt{|D|}/(2\,\mathrm{Vol}(\mathbb{C}/\Lambda))$ to precision at least 10^{-10} using PARI.

(d) Using a simple implementation of classical interval arithmetic (in [Ste]) and the bounds above, we compute an interval in which the real number

$$\alpha \cdot L'(E,1) \cdot L'(E^D,1)/(\mathrm{Reg}_E/2)$$

must lie. If there is a unique integer in this interval, by Theorem 3.9 this must be the square of the index $[E(K) : \mathbb{Z}y_K]^2$. If there is no unique integer in this interval, we increase the precision of the computation of L' and L and repeat the above steps. In all cases in the range of our computation, we find a unique integer in the interval; as a double check on our calculations we verify that the integer is a perfect square.

□

Remark 3.23. For the curves 540B and 756B there is no 3-torsion, but there is a rational 3-isogeny. In each case we verified in addition that 3 divides the GCD of the indexes for at least the first 16 fields K that satisfy the Heegner hypothesis. Thus as in Remark 3.12, Conjecture 3.10 asserts that $9 \mid \#\text{Ш}(E/K)$ for the first sixteen K. This illustrates that not only Tamagawa numbers but also isogenies can be an obstruction to applying Kolyvagin's theorem to bound $\#\text{Ш}(E)$, even if the irreducibility hypothesis on $\overline{\rho}_{E,p}$ is removed.

Proposition 3.24. *Suppose E is a non-CM optimal curve of conductor ≤ 1000 and p is an odd prime such that $\overline{\rho}_{E,p}$ is irreducible but not surjective. If E has rank 0 then (E,p) is one of the following:* (245B,3), (338D,3), (352E,3), (608B,5), (675D,5), (675F,5), (704H,3), (722D,3), (726F,3), (800E,5), (800F,5), (864D,3), (864F,3), (864G,3), (864I,3). *If E has rank 1, then (E,p) is one of the following:* (245A,3), (338E,3), (352F,3), (608E,5), (675B,5), (675I,5), (704L,3), (722B,3), (726A,3), (800B,5), (800I,5), (864A,3), (864B,3), (864J,3), (864L,3). *There are no curves of rank ≥ 2 with the above property.*

Proof. Using Proposition 2.5 we make a list of pairs (E,p) such that $\overline{\rho}_{E,p}$ might not be surjective, and such that if (E,p) is not in this list, then $\overline{\rho}_{E,p}$ is surjective. Then using the program `allisog`, we compute for each curve E, a list of all degrees of isogenies emanating from E, and remove those pairs (E,p) for which p divides the degree of one of those isogenies. The curves listed above are the ones that remain. □

Remark 3.25. In Proposition 3.24, the non-surjective irreducible (E,p) come in pairs, one of rank 0 and one of rank 1 having the same conductor. Each pair of curves are related by a quadratic twist. This pattern is common, but does not always occur. For example, (1184F,3) and (1184H,3) are both of rank 0 and have non-surjective irreducible representation, and no curve of conductor 1184 and rank 1 has this property. Note that $1184 = 2^5 \cdot 37$ and 1184F and 1184H are quadratic twists of each other by -1.

Remark 3.26. Proposition 3.24 suggests that it is rare for $\overline{\rho}_{E,p}$ to be non-surjective yet irreducible. When this does occur, frequently $p^2 \mid N$, though not always. Continuing the computation to conductor 10000 we find that $p^2 \mid N$ about half the time in which $\overline{\rho}_{E,p}$ is non-surjective yet irreducible. This gives a sense of the extent to which Theorem 3.3 improves on Theorem 3.2.

Theorem 3.27. *Suppose (E,p) is a pair consisting of a rank 1 non-CM elliptic curve E of conductor ≤ 1000 and a prime p such that $\rho_{E,p}$ is irreducible and p does not divide any Tamagawa number of E. Then $\mathrm{BSD}(E,p)$ is true.*

Proof. By Theorem 3.31 we may assume that p is odd. The pairs that do not satisfy the Heegner point divisibility hypothesis in Proposition 3.22 are those in $S = \{(540B,3),(756B,3)\}$. However, both of these curves admit a rational 3-isogeny, so are excluded by the hypothesis of Theorem 3.27.

Let

$$\begin{aligned} T = \{&(245A,3),(338E,3),(352F,3),(608E,5),(675B,5),(675I,5),\\ &(704L,3),(722B,3),(726A,3),(800B,5),(800I,5),(864A,3),\\ &(864B,3),(864J,3),(864L,3)\}. \end{aligned}$$

Then Proposition 3.24, Theorem 1.7, and Theorem 3.2 together imply $\mathrm{BSD}(E,p)$ for all pairs as in the hypothesis of Theorem 3.27, except the pairs in $S \cup T$. Note that for each $(E,p) \in T$, we have $p^2 \mid N$, so Theorem 3.3 does not apply either. We eliminate the pairs

$$(245A,3),(338E,3),(352F,3),(608E,5),(704L,3),(864J,3),(864L,3)$$

from consideration because in each case $p \mid \prod c_\ell$.

For each $(E,p) \in T$ the representation $\overline{\rho}_{E,p}$ is irreducible and E does not have CM, so the hypothesis of Theorem 3.5 are satisfied. For the pairs

$$\{((245A,3),(338E,3),(352F,3),(608E,5),(704L,3),(864J,3),(864L,3)\}$$

we have $p \mid [E(K) : \mathbb{Z}y_K]$ for the first six Heegner K, but that is not a problem since we eliminated these pairs from consideration. For the remaining pairs, in each case we find a K such that $p \nmid [E(K) : \mathbb{Z}y_K] \cdot \mathrm{disc}(K)$, so Theorem 3.5 implies that $p \nmid \#\text{Ш}(E)$, so $\mathrm{BSD}(E,p)$ is true. □

3.6.2. *Curves of Rank* 0.

Proposition 3.28. *Suppose (E,p) is a pair with E an optimal elliptic curve of conductor ≤ 1000 of rank 0. Let I be the greatest common divisor of $[E(K)_{/\mathrm{tor}} : \mathbb{Z}y_K]$ as K varies over quadratic imaginary fields that satisfy the Heegner hypothesis. If $p \mid I$ and $\overline{\rho}_{E,p}$ is irreducible, then*

$$p \mid 2 \cdot \#E(\mathbb{Q})_{\mathrm{tor}} \cdot \prod_{q \mid N} c_{E,q},$$

except possibly for the curves in the following table:

E	$p \mid I$?	D used	E	$p \mid I$?	D used
258E	3	-983	777B	3	-215
378G	3	-47	780B	3,7	-191
594F	3	-359	819D	3,5	-404
600G	3	-71	850I	3	-151
612D	3	-359	858D	5, 7	-95
626B	3	-39	858K	7	-1031
658A	3	-31	900A	3	-71
676E	5	-23	906E	$p \leq 19$	-23
681B	3	-8	924A	5	-1679
735B	3	-479	978C	3	-431
738B	3	-23	980I	3	-671
742F	3, 5	-199			

In this table, the first column gives an elliptic curve, the second column gives the primes p (with $\overline{\rho}_{E,p}$ irreducible) that might divide the GCD *of indexes, and the third column gives the discriminant used to make this deduction.*

Proof. We use the methods described in Section 3.5.5, and precision bounds as in the proof of Proposition 3.22. In many cases we combined explicit computation of a Heegner point for one prime, with the bounding technique explained in Section 3.5.5, or only computed information using the bound.

For the curve 910E, we used four-descent via MAGMA to compute the point $(3257919871/16641, 133897822473008/2146689)$ on the -159 twist E^D, found using [Creb] that it generates $E^D(\mathbb{Q})$, and obtained an index that is a power of 2 and 3. Since 3 divides a Tamagawa number, we do not include 910E in our table. Likewise, for 930F and $D = -119$, we used MAGMA's four-descent commands to find a point of height ~ 85.3, and deduced that the only odd prime that divides the index is 11; since 11 is a Tamagawa number, we do not include 930F. Similar remarks apply for 966J with $D = -143$. We were unable to use 4-descent to find a generator for a twist of 906E1. (Fortunately, $906 = 2 \cdot 3 \cdot 151$, so Theorem 4.3 implies $\mathrm{BSD}(E, p)$ except for $p = 2, 3, 151$, and for our purposes we will only need that 151 does not divide the Heegner point index.) □

Remark 3.29. We could likely shrink the table in Proposition 3.28 further using MAGMA's four descent command. However, we will not need a smaller table for our ultimate application to the BSD conjecture (Theorem 4.4).

Theorem 3.30. *Suppose (E, p) is a pair with E a rank 0 non-CM curve of conductor ≤ 1000 and p a prime such that $\overline{\rho}_{E,p}$ is irreducible and p does not divide any Tamagawa number of E. Then* $\mathrm{BSD}(E, p)$ *is true except possibly if (E, p) appears in the table in the statement of Proposition 3.22, i.e., E appears in column 1 and p appears in the column directly to the right of p.*

Proof. The argument is similar to the proof of Theorem 3.27. By Theorem 3.31 we may assume that p is odd. Let S be the set of pairs (E, p) in the table in Proposition 3.22. Let

$$\begin{aligned} T = \{&(245B, 3), (338D, 3), (352E, 3), (608B, 5), (675D, 5), (675F, 5), \\ &(704H, 3), (722D, 3), (726F, 3), (800E, 5), (800F, 5), (864D, 3), \\ &(864F, 3), (864G, 3), (864I, 3)\}. \end{aligned}$$

Then Proposition 3.24, Theorem 1.7, and Theorem 3.2 together imply $\mathrm{BSD}(E,p)$ for all pairs as in the hypothesis of Theorem 3.27, except the pairs in $S \cup T$, since the representation $\overline{\rho}_{E,p}$ is surjective and we have verified that $p \nmid [E(K) : \mathbb{Z}y_K]$ for some K. We eliminate the pairs $(722D, 3)$ and $(726F, 3)$ from consideration because in each case $p \mid \prod c_\ell$.

For each $(E,p) \in T$ the representation $\overline{\rho}_{E,p}$ is irreducible and E does not have CM, so the hypotheses of Theorem 3.5 are satisfied. Next for each pair $(E,p) \in T$ except for $(722D, 3)$ and $(726F, 3)$, which we eliminated already, we find a K such that $p \nmid [E(K) : \mathbb{Z}y_K]$ and $\mathrm{disc}(K)$ is not divisible only be p. Theorem 3.5 implies that $p \nmid \#Ш(E)$, hence $\mathrm{BSD}(E,p)$ is true. □

3.6.3. *Two Descent.* In this section, we explain how descent computations imply that $\mathrm{BSD}(E,2)$ is true for curves of conductor $N \leq 1000$.

Theorem 3.31. *If E is an elliptic curve with $N \leq 1000$, then $\mathrm{BSD}(E,2)$ is true.*

Proof. According to Theorem 1.4, it suffices to prove the theorem for the set S of optimal elliptic curves with $N \leq 1000$. By doing an explicit 2-descent, Cremona computed $\mathrm{Sel}^{(2)}(E/\mathbb{Q})$ for every curve $E \in S$, as explained in [Cre97]. This implies that $Ш(E)[2]$ has order the predicted order of $Ш(E)[2^\infty]$ for all $E \in S$. Using MAGMA's `FourDescent` command, we compute $\mathrm{Sel}^{(4)}(E/\mathbb{Q})$ in the three cases in which $Ш(E)[2] \neq 0$, and find that $Ш(E)[4] = Ш(E)[2]$. By Theorem 1.7, it follows that $\mathrm{BSD}(E,2)$ is true for all $E \in S$. □

3.6.4. *Three Descent.* We sharpen Theorem 3.30 using Stoll's 3-descent package (see [Sto05]).

Proposition 3.32. *We have $3 \nmid \#Ш(E)$ for each of the curves listed in the Table in Proposition 3.28 with 3 in the second column and $\overline{\rho}_{E,3}$ irreducible, except for 681B where $\#Ш(E)[3^\infty] = 9$.*

Proof. We use Stoll's package [Sto05] to compute each of the Selmer groups

$$\mathrm{Sel}^{(3)}(E) \cong Ш(E)[3],$$

and obtain the claimed dimensions. When computing class groups in Stoll's package one must take care to not assume any conjectures (by slightly modifying the call to `ClassGroup` in `3descent.m`). Finally, that $Ш(E)[3^\infty] = 9$ follows by applying Theorem 3.2 with $K = \mathbb{Q}(\sqrt{-8})$, and noting that $\overline{\rho}_{E,3}$ is surjective and the index is exactly divisible by 3. □

4. The Kato Bound

Kato proved a theorem that bounds $Ш(E)$ from above when $L(E,1) \neq 0$.

Theorem 4.1 (Kato)**.** *Let E be an optimal elliptic curve over $\mathbb{Q}$ of conductor N, and let p be a prime such that $p \nmid 6N$ and $\rho_{E,p}$ is surjective. If $L(E,1) \neq 0$, then $Ш(E)$ is finite and*

$$\mathrm{ord}_p(\#Ш(E)) \leq \mathrm{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right).$$

This theorem follows from the existence of an "optimal" Kato Euler system (see [Kat04] and [MR04]) combined with a recent result of Matsuno [Mat03] on finite submodules of Selmer groups over $\mathbb{Z}_p$-extensions. For more details, look at the proof of [Rub98, Cor. 8.9] where one replaces an unknown module with the module Matsuno computes. See also [Gri05] for further discussion and recent results on lower bounds on Ш(E) that make use of optimal Kato Euler systems.

4.1. **Computations.** When $L(E,1) \neq 0$ the group Ш(E) is finite, so $\mathrm{ord}_p(\#Ш(E))$ is even. Thus if $\mathrm{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right)$ is odd, we conclude that

$$\mathrm{ord}_p(\#Ш(E)) \leq \mathrm{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right) - 1.$$

Lemma 4.2. *There are no pairs (E,p) that satisfy the conditions of Theorem 4.1 with $N \leq 1000$, such that*

$$\mathrm{ord}_p(\#Ш(E)_{\mathrm{an}}) < \mathrm{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right) - 1.$$

Proof. First we make a table of ratios $L(E,1)/\Omega_E$ for all curves of conductor $\leq$ 1000. For each of these with $L(E,1) \neq 0$, we factor the numerator of the rational number $L(E,1)/\Omega_E$. We then observe that the displayed inequality in the statement of the proposition does not occur. □

Theorem 4.3. *Suppose (E,p) is a pair such that $N \leq 1000$, $p \nmid 3N$, E is a non-CM elliptic curve of rank 0, and $\overline{\rho}_{E,p}$ is irreducible. Then* $\mathrm{BSD}(E,p)$ *is true.*

Proof. The statement for $p = 2$ follows from Theorem 3.31.

Let S be the set of pairs (E,p) as in the statement of Theorem 4.3 for which E is optimal and $p > 2$. By Theorem 1.7 it suffices to prove that $p \nmid \#Ш(E)$ for all $(E,p) \in S$. Using Proposition 2.5 with $A = 1000$, we compute for each rank 0 non-CM elliptic curve of conductor $N \leq 1000$, all primes $p \nmid 6N$ such that $\rho_{E,p}$ might not be surjective. This occurs for 53 pairs (E,p), with the E's all distinct. For these 53 pairs (E,p), we find that the representation $\overline{\rho}_{E,p}$ is reducible (since there is an explicit p isogeny listed in [Cre97]), except for the pair $(608B,5)$, for which $\overline{\rho}_{E,5}$ is irreducible.

Thus Theorem 4.1 implies that for each pair $(E,p) \in S$, except $(608B,5)$, we have the bound

$$\mathrm{ord}_p(\#Ш(E)) \leq \mathrm{ord}_p(L(E,1)/\Omega_E).$$

By Theorem 1.5, $\mathrm{ord}_p(\#Ш(E))$ is even, so Ш$(E)[p^\infty]$ is trivial whenever

$$\mathrm{ord}_p(L(E,1)/\Omega_E) \leq 1.$$

By Theorem 1.7, $\mathrm{ord}_p(\#Ш(E)_{\mathrm{an}}) = 0$ for all $p \geq 5$. Thus by Lemma 4.2, there are no pairs $(E,p) \in S$ with $\mathrm{ord}_p(L(E,1)/\Omega_E) > 1$ (since otherwise some $\mathrm{ord}_p(\#Ш(E)_{\mathrm{an}})$ would be nontrivial).

Finally, note that we dealt with $(608B,5)$ in Lemma 3.4 using Cha's theorem. This completes the proof. □

4.2. **Combining Kato and Kolyvagin.** In this section we bound Ш(E) for rank 0 curves by combining the Kato and Kolyvagin approaches.

Theorem 4.4. *Suppose E is a non-CM elliptic curve of rank 0 with conductor $N \leq 1000$, that $\overline{\rho}_{E,p}$ is irreducible, and that p does not divide any Tamagawa number of E. Then* $\mathrm{BSD}(E,p)$ *is true.*

Proof. Let (E,p) be as in the hypothesis to Theorem 4.4. By Theorem 4.3, $\mathrm{BSD}(E,p)$ is true, except possibly if $p \mid 3N$. Theorem 3.30 implies $\mathrm{BSD}(E,p)$, except if (E,p) appear in the Table of Proposition 3.28. Inspecting the table, we see that whenever a prime $p \geq 5$ is in the second column, then p does not divide the conductor N of E. This proves $\mathrm{BSD}(E,p)$ for $p \geq 5$.

Let E be the curve 681B. Then $\mathrm{BSD}(E,3)$ asserts that $\#\text{Ш}(E)[3^\infty] = 9$. It follows from [CM00] and [AS05, App.], or from the 3-descent of Section 3.6.4 that $\#\text{Ш}(E)[3] = 9$. Also, $\overline{\rho}_{E,3}$ is surjective and for $D = -8$ we have $\mathrm{ord}_3([E(K) : \mathbb{Z}y_K]) = 1$, so $\#\text{Ш}(E)[3^\infty] \mid 9$, which proves $\mathrm{BSD}(E,3)$.

Finally Proposition 3.32 implies $\mathrm{BSD}(E,3)$ for the remaining curves, which proves the theorem. □

5. Proof of Theorem 3.5

In this section we prove Theorem 3.5. Assume that E and K are as in the statement of the theorem, and assume that $\mathrm{ord}_{s=1} L(E/K,1) = 1$. Then the Heegner point y_K has infinite order. Kolyvagin ([Kol90]) shows that in this case the rank of $E(K)$ is 1 and $\text{Ш}(E/K)$ is finite.

5.1. **Gross's Account.**

Gross's account of Kolyvagin's work in [Gro91] contains a proof that if E does not have complex multiplication, then

$$\#\text{Ш}(E/K) \mid t \cdot [E(K) : \mathbb{Z}y_K]^2,$$

where t is an integer divisible only by primes p such that the representation $\overline{\rho}_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[p])$ is not surjective. Gross makes no claim about the powers of primes that divide t (though Kolyvagin does in his papers). Our Theorem 3.5 provides a better bound, which removes the condition that E not have CM, and relaxes the surjectivity hypothesis on $\overline{\rho}_{E,p}$.

Gross uses surjectivity of $\overline{\rho}_{E,p}$ as a hypothesis only to prove the following two propositions. We will prove analogous propositions below, but under weaker hypotheses, which yields our claimed improvement to [Gro91].

Proposition 5.1 (Gross)**.** *Assume that $\overline{\rho}_{E,p}$ is surjective. For any integer n, let K_n be the ring class field of K of conductor n. The restriction map*

$$\mathrm{Res} : \mathrm{H}^1(K, E[p]) \to \mathrm{H}^1(K_n, E[p])^{\mathrm{Gal}(K_n/K)}$$

is an isomorphism.

Proof. That $\overline{\rho}_{E,p}$ is surjective implies that $E(K_n)[p] = 0$. The inflation-restriction-transgression sequence then implies that Res is an isomorphism. □

Gross also uses surjectivity of $\overline{\rho}_{E,p}$ when proving that the pairing

$$\mathrm{H}^1(K, E[p]) \otimes \mathrm{Gal}(K(E[p])/K) \to E[p]$$

is nondegenerate, as follows. Setting $L = K(E[p])$, we have that

$$\mathrm{H}^1(L/K, E(L)[p]) \to \mathrm{H}^1(K, E[p]) \to \mathrm{H}^1(L, E[p])^{\mathrm{Gal}(L/K)} \to \mathrm{H}^2(L/K, E(L)[p]).$$

To see that the pairing is nondegenerate, it suffices to know that the groups $\mathrm{H}^i(L/K, E[p])$ vanish for $i = 1, 2$. This is because we have

$$\mathrm{H}^1(L, E[p])^{\mathrm{Gal}(L/K)} = \mathrm{Hom}(G_L, E[p])^{\mathrm{Gal}(L/K)}$$

since $K(E[p]) \subset L$ and the pairing is $(c, \sigma) = \mathrm{res}_L(c)(\sigma)$. Thus nondegeneracy of the pairing then follows from the following proposition.

Proposition 5.2 (Gross)**.** *Let E be an elliptic curve over a number field K and let p be a prime. Assume that $\overline{\rho}_{E,p}$ is surjective. Then*

$$\mathrm{H}^i(K(E[p])/K, E[p]) = 0 \qquad \textit{for all} \quad i \geq 1.$$

Proof. As above set $L = K(E[p])$. The surjectivity of $\overline{\rho}_{E,p}$ implies that

$$G = \mathrm{Gal}(L/K) \cong \mathrm{Gal}(\mathbb{Q}(E[p])/\mathbb{Q}) \cong \mathrm{GL}_2(\mathbb{F}_p).$$

If $Z \subset G$ is the subgroup corresponding to the scalars in $\mathrm{GL}_2(\mathbb{F}_p)$, then the Hochschild-Serre spectral sequence implies that

$$\mathrm{H}^i(G/Z, \mathrm{H}^j(Z, E(L)[p])) \Longrightarrow \mathrm{H}^{i+j}(L/K, E(L)[p]).$$

Since $\#Z = p-1$, and $E(L)[p]$ is a p-group, and p is odd, we have $\mathrm{H}^j(Z, E(L)[p]) = 0$ for all $j \geq 1$. Also, since p is odd, and non-identity scalars have no nonzero fixed points, $\mathrm{H}^0(Z, E(L)[p]) = 0$. Thus for all i, j we have

$$\mathrm{H}^i(G/Z, \mathrm{H}^j(Z, E(L)[p])) = 0,$$

which implies that the groups $\mathrm{H}^{i+j}(L/K, E(L)[p])$ are all 0. □

Thus our goal is to prove analogues of Propositions 5.1–5.2 under hypotheses that are more amenable to computation.

5.2. **Preliminaries.**

Lemma 5.3. *The determinant of $\overline{\rho}_{E,p}$ is the cyclotomic character, hence $\det(\overline{\rho}_{E,p})$ is surjective.*

Proof. For the convenience of the reader, we give a proof here. The Weil pairing induces an isomorphism of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-modules $E[p] \wedge E[p] \cong \mu_p$. Fix a basis $\{e_1, e_2\}$ of $E[p]$, with respect to which $\rho_p(\sigma)$ has the form $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then

$$\sigma(e_1 \wedge e_2) = (ae_1 + ce_2) \wedge (be_1 + de_2) = \det(\rho_p(\sigma)) \cdot e_1 \wedge e_2.$$

It follows that composition with the determinant gives the cyclotomic character (i.e., the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on μ_p), which is surjective since no nontrivial roots of unity lie in $\mathbb{Q}$. □

We will choose the quadratic field K to be linearly disjoint from $\mathbb{Q}(E[p])$, so $\mathrm{Gal}(K(E[p])/K) \cong \mathrm{Gal}(\mathbb{Q}(E[p])/\mathbb{Q})$. Thus, for our application, it will suffice to show vanishing of $\mathrm{H}^i(\mathbb{Q}(E[p])/\mathbb{Q}, E[p])$, for $i > 0$.

Let $G \subseteq \mathrm{Gal}(\mathbb{Q}(E[p])/\mathbb{Q})$ be the image of $\overline{\rho}_{E,p}$. If $p \nmid \#G$, then for $i > 0$ we have $\mathrm{H}^i(G, E[p]) = 0$ since $E[p]$ is a p-group. Therefore we may assume that $p \mid \#G$. By [Ser72, Prop. 15], the image G either contains $\mathrm{SL}_2(\mathbb{F}_p)$ or is contained in a Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. If G contains $\mathrm{SL}_2(\mathbb{F}_p)$ then properties of the Weil pairing imply that

$$\det : G \to \mathbb{F}_p^*$$

is surjective, so $G = \mathrm{GL}_2(\mathbb{F}_p)$. In this case, we already know Propositions 5.1–5.2.

Lemma 5.4. *Assume that G is contained in a Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Moreover, assume that there is a basis of $E[p]$ so that G acts as $\left(\begin{smallmatrix} \chi & * \\ 0 & \psi \end{smallmatrix}\right)$ where χ and ψ are nontrivial characters. Then $\mathrm{H}^i(G, E[p]) = 0$.*

Proof. Let $W = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ be the unique p-Sylow subgroup of $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \subset \mathrm{GL}_2(\mathbb{F}_p)$. We may assume $W \subset G$, for otherwise G has order prime to p, and the cohomology vanishes.

We begin by explicitly computing $\mathrm{H}^j(W, E[p])$ using the fact that W is cyclic, generated by $w = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Recall that for cyclic groups we can compute cohomology using the projective resolution

$$\cdots \to \mathbb{Z}[W] \to \mathbb{Z}[W] \to \mathbb{Z} \to 0$$

where the boundary maps alternate between multiplication by $w - 1$ and $\mathrm{Norm} = \sum_{i=0}^{p-1} w^i$.

Then we see that

$$\mathrm{H}^j(W, E[p]) = \begin{cases} \mathrm{Ker}(1-w)/\,\mathrm{Im}(\mathrm{Norm}(w)) = \langle \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \rangle\,, & \text{if } j \text{ is even,} \\ \mathrm{Ker}(\mathrm{Norm}(w))/\,\mathrm{Im}(1-w) = \mathbb{F}_p^2/\,\langle \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \rangle\,, & \text{if } j \text{ is odd.} \end{cases}$$

Since χ and ψ are nontrivial by assumption, the G/W-invariants for both of these groups are trivial. Thus $\mathrm{H}^j(W, E[p])^{G/W} = 0$ for $j \geq 0$. Consider the Hochschild-Serre spectral sequence

$$\mathrm{H}^i(G/W, \mathrm{H}^j(W, E[p])) \Rightarrow \mathrm{H}^{i+j}(G, E[p]).$$

For $i > 0$, since $\#(G/W)$ is prime to p, and $\mathrm{H}^j(W, E[p])$ is a p-group for all j, the group $\mathrm{H}^i(G/W, \mathrm{H}^j(W, E[p]))$ is trivial. But when $i = 0$ we have just computed that $\mathrm{H}^i(G/W, \mathrm{H}^j(W, E[p])) = \mathrm{H}^j(W, E[p])^{G/W} = 0$, so the entire spectral sequence is trivial, and we conclude that $\mathrm{H}^n(G, E[p]) = 0$ for all $n \geq 0$. □

5.3. Analogue of Proposition 5.1.

In this section we verify that $\mathrm{H}^i(K_n/K, E(K_n)[p]) = 0$ under a simple condition on p-torsion over K.

Proposition 5.5. *Let E be an elliptic curve over $\mathbb{Q}$ and K be a quadratic imaginary extension of $\mathbb{Q}$. Assume that p is a prime with $p \nmid \#E(K)_{\mathrm{tor}}$ and if $p = 3$ assume that $K \neq \mathbb{Q}(\zeta_3)$. Then for every finite abelian extension L of K we have*

$$\mathrm{H}^i(L/K, E(L)[p]) = 0 \qquad \textit{for all} \quad i \geq 1.$$

Proof. Write the abelian group $\mathrm{Gal}(L/K)$ as a direct sum $P \oplus P'$, where P is its Sylow p-subgroup, so $p \nmid \#P'$. First we show that the subgroup of $E(L)[p]$ invariant under P' is trivial. Let $G = \mathrm{Gal}(L/K)/H$, where H is the subgroup of $\mathrm{Gal}(L/K)$ that acts trivially on $E(L)[p]$. Thus $G \subset \mathrm{Aut}(E(L)[p])$.

Case 1. If $p \nmid \#G$, then $P \subseteq H$, so P' surjects onto G. There is no nonzero element of $E(L)[p]$ invariant under $\mathrm{Gal}(L/K)$ by our assumption that $p \nmid \#E(K)$, so the same holds for P'.

Case 2. If $p \mid \#G$, we cannot have $E(L)[p] = \mathbb{F}_p$, since $\mathbb{F}_p$ has automorphism group isomorphic to $\mathbb{F}_p^*$, of order $p - 1$, but $G \subset \mathrm{Aut}(E(L)[p])$ and $\#G > p - 1$. Thus, $E(L)[p]$ is the full p-torsion subgroup of E, and we identify G with a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ acting on $E(L)[p] = \mathbb{F}_p^2$.

We can choose a basis of $\mathbb{F}_p^2$ so that G contains the subgroup generated by $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Since G is abelian, it must be contained in the normalizer of this subgroup, so $G \subseteq \{\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right) : a \in \mathbb{F}_p^*, b \in \mathbb{F}_p\}$. We claim that G contains an element with $a \neq 1$. Since $E[p] = E(L)[p]$, the representation $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{Aut}(E[p])$ factors through $\mathrm{Gal}(L/K)$. The determinant of $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p])$ is surjective onto $\mathbb{F}_p^*$, and $[K : \mathbb{Q}] = 2$, so the character $\mathrm{Gal}(\overline{K}/K) \to \mathbb{F}_p^*$ has image of index at most 2 in

F_p^*. That is, it contains at least $(p-1)/2$ elements, the squares in $\mathbb{F}_p^*$. Thus, for $p > 3$, the group G contains an element with non-trivial determinant having the form $\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right)$ with $a \neq 1$. Now, $\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right)^p = \left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$ since $a, b \in \mathbb{F}_p$, so $\mathrm{Gal}(L/K)$ contains an element that acts as a nontrivial scalar. Since the group of scalars in $\mathrm{GL}_2(\mathbb{F}_p)$ has $p-1)$ elements, this nontrivial scalar must be in P', so $E(L)[p]^{P'} = 0$.

We have shown in each case that $E(L)[p]^{P'} = 0$. Because $p \nmid \#P'$ we have $\mathrm{H}^i(P', E(L)[p]) = 0$ for all $i \geq 1$, so for each $i \geq 1$ there is an exact inflation-restriction sequence

$$0 \to \mathrm{H}^i(P, E(L)[p]^{P'}) \to \mathrm{H}^i(L/K, E(L)[p]) \to \mathrm{H}^i(P', E(L)[p]).$$

The first group vanishes since $E(L)[p]^{P'} = 0$, and the third group vanishes as mentioned above. We conclude that $\mathrm{H}^i(L/K, E(L)[p]) = 0$, as claimed.

Finally we deal with the case $p = 3$. The only situation in the above argument where $p = 3$ is relevant is in Case 2, when $3 \mid \#G$. Our hypothesis that $K \neq \mathbb{Q}(\zeta_3)$ implies that $\det(\rho_{E,3}) : \mathrm{Gal}(\overline{K}/K) \to \mathbb{F}_3^*$ is surjective, since the fixed field of the kernel of the mod 3 cyclotomic character is $\mathbb{Q}(\zeta_3)$. If we are in Case 2, then the image of $\mathrm{Gal}(\overline{K}/K)$ in $\mathrm{GL}_2(\mathbb{F}_3)$ is contained in $\{\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right) : a \in \mathbb{F}_p^*, b \in \mathbb{F}_p\}$. Since no upper triangular matrix has determinant 2, this contradicts surjectivity of $\det(\rho_{E,3})$. Thus our hypothesis that $K \neq \mathbb{Q}(\zeta_3)$ implies that Case 2 does not occur. □

Corollary 5.6. *Let E be an elliptic curve with $p \nmid \#E(K)_{\mathrm{tor}}$, where $p > 3$ or, if $p = 3$, $K \neq \mathbb{Q}(\zeta_3)$. Let K_n be the ring class field of conductor n of K. Then $\mathrm{H}^i(K_n/K, E(K_n)[p]) = 0$ for all $i \geq 1$.*

Lemma 5.7. *Let E be an elliptic curve over $\mathbb{Q}$, let K be a quadratic imaginary extension, and let $p \mid \#E(K)_{\mathrm{tor}}$ an odd prime. If $p = 3$, assume $K \neq \mathbb{Q}(\zeta_3)$. Then $\overline{\rho}_{E,p}$ is reducible.*

Proof. Let $P \in E(K)[p]$ be nonzero, and let τ be a lift of the generator of $\mathrm{Gal}(K/\mathbb{Q})$ to $G_{\mathbb{Q}}$. If τP is a multiple of P, then the one-dimensional subspace of $E[p]$ generated by P is $G_{\mathbb{Q}}$-stable, so $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p])$ is is reducible. If τP is not a multiple of P, then P and τP generate all of $E[p]$. Since $\tau P \in E(K)$, we have $E(K)[p] = E(\overline{\mathbb{Q}})[p]$. Because the Weil pairing in nondegenerate we have $\mu_p \subset K$. This is a contradiction by our hypothesis on K and p. Since $p > 3$, this is a contradiction. □

5.4. **Analogue of Proposition 5.2.** In this section we show how vanishing of $\mathrm{H}^i(\mathbb{Q}(E[p])/\mathbb{Q}, E[p])$ follows from a statement about torsion and rational isogenies.

Note that E has no $\mathbb{Q}$-rational p-isogeny if and only if $\overline{\rho}_{E,p}$ is irreducible.

Proposition 5.8. *If p is an odd prime and E has no $\mathbb{Q}$-rational p-isogeny, then $\mathrm{H}^i(\mathbb{Q}(E[p])/\mathbb{Q}, E[p]) = 0$ for all $i > 0$.*

Proof. Our hypothesis that E has no $\mathbb{Q}$-rational p-isogeny implies that $\overline{\rho}_{E,p}$ is irreducible. As we already noted, the problem reduces to the case when either G is contained in a Borel subgroup or $G = \mathrm{GL}_2(\mathbb{F}_p)$. The latter case follows from Proposition 5.2. The former case contradicts the hypothesis since the module $E[p]$ is reducible as a module over a Borel subgroup. □

For the above result, we used the irreducibility of the representation to deal with the case when G was contained in a Borel subgroup. The following proposition completes the proof of the general case:

Proposition 5.9. *Suppose p is an odd prime and that $E(\mathbb{Q})[p] = 0$ and for all elliptic curves E' that are p-isogenous to E over $\mathbb{Q}$ we have $E'(\mathbb{Q})[p] = 0$. Then*

$$\mathrm{H}^i(\mathbb{Q}(E[p])/\mathbb{Q}, E[p]) = 0 \qquad \text{for all} \qquad i > 0.$$

Proof. If E admits no p-isogeny, then Proposition 5.8 implies the required vanishing. Thus E admits a rational p-isogeny, so $E[p]$ is reducible, and $G = \mathrm{Im}(\overline{\rho}_{E,p})$ is contained in a Borel subgroup. In particular, for some basis of $E[p]$, the image G acts as $\left(\begin{smallmatrix} \chi & * \\ 0 & \psi \end{smallmatrix}\right)$ for characters χ and ψ. If both χ and ψ are nontrivial, then Lemma 5.4 implies the proposition and we are done. Thus assume that either χ or ψ is trivial.

First suppose that χ is trivial. Then all matrices of the above form fix $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. Therefore there is a point of $E[p]$ fixed by the action of G, which contradicts the assumption that $E(\mathbb{Q})[p] = 0$.

Next suppose that ψ is trivial. Matrices of the above form preserve the line generated by $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, so this line forms a $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-stable subspace of $E[p]$. In particular, there exists an isogeny over $\mathbb{Q}$ to a curve E' having this line as kernel. The image under this isogeny of the line generated by $\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ is a 1-dimensional subspace of $E'[p]$, and since $\psi = 1$, $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts trivially on this subspace (we have an isomorphism of Galois modules $E/\left\langle \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \right\rangle \cong E'$). Thus, $E'(\mathbb{Q})[p]$ is nontrivial, contradicting our assumption.

□

References

[ABC+] B. Allombert, K. Belabas, H. Cohen, X. Roblot, and I. Zakharevitch, `PARI/GP`, `http://pari.math.u-bordeaux.fr/`.

[ARS05] A. Agashe, K. A. Ribet, and W. A. Stein, *The manin constant, congruence primes, and the modular degree*, Preprint, `http://modular.fas.harvard.edu/papers/manin-agashe/` (2005).

[AS05] A. Agashe and W. Stein, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank zero*, Math. Comp. **74** (2005), no. 249, 455–484 (electronic), With an appendix by J. Cremona and B. Mazur. MR 2085902

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[BFH90] Daniel Bump, Solomon Friedberg, and Jeffrey Hoffstein, *Nonvanishing theorems for L-functions of modular forms and their derivatives*, Invent. Math. **102** (1990), no. 3, 543–618. MR MR1074487 (92a:11058)

[Cas62] J. W. S. Cassels, *Arithmetic on curves of genus* 1*. III. The Tate-Šafarevič and Selmer groups*, Proc. London Math. Soc. (3) **12** (1962), 259–296. MR 29 #1212

[Cas65] J. W. S. Cassels, *Arithmetic on curves of genus 1. VIII. On conjectures of Birch and Swinnerton-Dyer*, J. Reine Angew. Math. **217** (1965), 180–199. MR 31 #3420

[Cha03] Byungchul Cha, *Vanishing of Some Cohomology Groups and Bounds for the Shafarevich-Tate Groups of Elliptic Curves*, Johns-Hopkins Ph.D. Thesis (2003).

[Cha05] ———, *Vanishing of some cohomology groups and bounds for the Shafarevich-Tate groups of elliptic curves*, J. Number Theory. **111** (2005), 154–178.

[CK] Alina Carmen Cojocaru and Ernst Kani, *On the surjectivity of the galois representations associated to non-cm elliptic curves.*

[CM00] J.E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. **9** (2000), no. 1, 13–28. MR 1 758 797

[Coh93] H. Cohen, *A course in computational algebraic number theory*, Springer-Verlag, Berlin, 1993. MR 94i:11105

[Crea] J. E. Cremona, *Elliptic curves of conductor* ≤ 25000, `http://www.maths.nott.ac.uk/personal/jec/ftp/data/`.

[Creb] ______, `mwrank` *(computer software)*, `http://www.maths.nott.ac.uk/personal/jec/ftp/progs/`.

[Cre97] ______, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, `http://www.maths.nott.ac.uk/personal/jec/book/`.

[Edi91] B. Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Birkhäuser Boston, Boston, MA, 1991, pp. 25–39. MR 92a:11066

[Gri05] G. Grigorov, *Kato's Euler System and the Main Conjecture*, Harvard Ph.D. Thesis (2005).

[Gro91] B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 235–256.

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[Jor05] A. Jorza, *The Birch and Swinnerton-Dyer Conjecture for Abelian Varieties over Number Fields*, Harvard University Senior Thesis (2005).

[Kat04] Kazuya Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, Astérisque (2004), no. 295, ix, 117–290, Cohomologies p-adiques et applications arithmétiques. III. MR MR2104361

[Kol88] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E,\mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671. MR 89m:11056

[Kol90] ______, *Euler systems*, The Grothendieck Festschrift, Vol. II, Birkhäuser Boston, Boston, MA, 1990, pp. 435–483. MR 92g:11109

[Kol91] V. A. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 429–436. MR 93c:11046

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR 93a:11048

[Man72] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66. MR 47 #3396

[Mat03] Kazuo Matsuno, *Finite Λ-submodules of Selmer groups of abelian varieties over cyclotomic $\mathbb{Z}_p$-extensions*, J. Number Theory **99** (2003), no. 2, 415–443. MR MR1969183 (2004c:11098)

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[McC91] W. G. McCallum, *Kolyvagin's work on Shafarevich-Tate groups*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 295–316. MR 92m:11062

[Mil86] J. S. Milne, *Arithmetic duality theorems*, Academic Press Inc., Boston, Mass., 1986.

[MM91] M. Ram Murty and V. Kumar Murty, *Mean values of derivatives of modular L-series*, Ann. of Math. (2) **133** (1991), no. 3, 447–475. MR MR1109350 (92e:11050)

[MR04] Barry Mazur and Karl Rubin, *Kolyvagin systems*, Mem. Amer. Math. Soc. **168** (2004), no. 799, viii+96. MR MR2031496 (2005b:11179)

[Pri04] M. Prickett, *Saturation of Mordell-Weil Groups of Elliptic Curves over Number Fields*, U. Nottingham, Ph.D. thesis `http://etheses.nottingham.ac.uk/archive/00000052/` (2004).

[PS99] B. Poonen and M. Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2) **150** (1999), no. 3, 1109–1149. MR 2000m:11048

[Rub98] K. Rubin, *Euler systems and modular elliptic curves*, Galois representations in arithmetic algebraic geometry (Durham, 1996), Cambridge Univ. Press, Cambridge, 1998, pp. 351–367. MR 2001a:11106

[Ser72] J-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331.

[Ser98] ———, *Abelian ℓ-adic representations and elliptic curves*, A K Peters Ltd., Wellesley, MA, 1998, With the collaboration of Willem Kuyk and John Labute, Revised reprint of the 1968 original.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[Ste] W. A. Stein, `SAGE`, `http://modular.fas.harvard.edu/SAGE`.

[Sto05] M. Stoll, *Explicit 3-descent in Magma* `http://www.faculty.iu-bremen.de/stoll/magma/explicit-3descent/`.

[Wal85] J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie*, Compositio Math. **54** (1985), no. 2, 173–242. MR MR783511 (87g:11061b)

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Wil00] ———, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[Zha04] Shou-Wu Zhang, *Gross-Zagier formula for* GL(2). *II*, Heegner points and Rankin L-series, Math. Sci. Res. Inst. Publ., vol. 49, Cambridge Univ. Press, Cambridge, 2004, pp. 191–214. MR MR2083213

27 The Manin Constant, with A. Agashe and K. Ribet

Pure and Applied Mathematics Quarterly
Volume 2, Number 2
(*Special Issue: In honor of
John H. Coates, Part 2 of 2*)
617—636, 2006

The Manin Constant

Amod Agashe, Kenneth Ribet and William A. Stein

Abstract: The Manin constant of an elliptic curve is an invariant that arises in connection with the conjecture of Birch and Swinnerton-Dyer. One conjectures that this constant is 1; it is known to be an integer. After surveying what is known about the Manin constant, we establish a new sufficient condition that ensures that the Manin constant is an *odd* integer. Next, we generalize the notion of the Manin constant to certain abelian variety quotients of the Jacobians of modular curves; these quotients are attached to ideals of Hecke algebras. We also generalize many of the results for elliptic curves to quotients of the new part of $J_0(N)$, and conjecture that the generalized Manin constant is 1 for newform quotients. Finally an appendix by John Cremona discusses computation of the Manin constant for all elliptic curves of conductor up to 130000.

1. Introduction

Let E be an elliptic curve over $\mathbf{Q}$, and and let N be the conductor of E. By [BCDT01], we may view E as a quotient of the modular Jacobian $J_0(N)$. After possibly replacing E by an isogenous curve, we may assume that the kernel of the map $J_0(N) \to E$ is connected, i.e., that E is an *optimal* quotient of $J_0(N)$.

Let ω be the unique (up to sign) rational 1-form on a minimal Weierstrass model of E over $\mathbf{Z}$ that restricts to a nowhere-vanishing 1-form on the smooth locus. The pullback of ω is a rational multiple of the differential associated to the normalized new cuspidal eigenform $f_E \in S_2(\Gamma_0(N))$ associated to E. The Manin constant c_E of is E is the absolute value of this rational multiple. The Manin constant plays a role in the conjecture of Birch and Swinnerton-Dyer (see, e.g., [GZ86, p. 310]) and in work on modular parametrizations (see [Ste89, SW04, Vat05]). It is known that the Manin constant is an integer; this fact is important to Cremona's computations of elliptic curves (see [Cre97, pg. 45]), and algorithms

Received January 25, 2006.
Stein was supported by the National Science Foundation by Grant No. 0400386.

for computing special values of elliptic curve L-functions. Manin conjectured that $c_E = 1$. In Section 2, we summarize known results about c_E, and give the new result that $2 \nmid c_E$ if 2 is not a congruence prime and $4 \nmid N$.

In Section 3, we generalize the definition of the Manin constant and many of the results mentioned so far to optimal quotients of $J_0(N)$ and $J_1(N)$ of any dimension associated to ideals of the Hecke algebra. The generalized Manin constant comes up naturally in studying the conjecture of Birch and Swinnerton-Dyer for such quotients (see [AS05, §4]), which is our motivation for studying the generalization. We state what we can prove about the generalized Manin constant, and make a conjecture that the constant is also 1 for quotients associated to newforms. The proofs of the theorems stated in Section 3 are in Section 4. Section 5 is an appendix written by J. Cremona about computational verification that the Manin constant is 1 for many elliptic curves.

Acknowledgments. The authors are grateful to A. Abbes, K. Buzzard, R. Coleman, B. Conrad, B. Edixhoven, A. Joyce, L. Merel, and R. Taylor for discussions and advice regarding this paper. The authors wish to thank the referee for helpful comments and suggestions.

2. Optimal Elliptic Curve Quotients

Let N be a positive integer and let $X_0(N)$ be the modular curve over $\mathbf{Q}$ that classifies isomorphism classes of elliptic curves with a cyclic subgroup of order N. The Hecke algebra $\mathbf{T}$ of level N is the subring of the ring of endomorphisms of $J_0(N) = \mathrm{Jac}(X_0(N))$ generated by the Hecke operators T_n for all $n \geq 1$. Suppose f is a newform of weight 2 for $\Gamma_0(N)$ with integer Fourier coefficients, and let I_f be kernel of the homomorphism $\mathbf{T} \to \mathbf{Z}[\ldots, a_n(f), \ldots]$ that sends T_n to $a_n(f)$. Then the quotient $E = J_0(N)/I_f J_0(N)$ is an elliptic curve over $\mathbf{Q}$. We call E the *optimal quotient* associated to f. Composing the embedding $X_0(N) \hookrightarrow J_0(N)$ that sends x to $(\infty) - (x)$ with the quotient map $J_0(N) \to E$, we obtain a surjective morphism of curves $\phi_E : X_0(N) \to E$. The *modular degree* m_E of E is the degree of ϕ_E.

Let $E_{\mathbf{Z}}$ denote the Néron model of E over $\mathbf{Z}$. A general reference for Néron models is [BLR90]; for the special case of elliptic curves, see, e.g., [Sil92, App. C, §15], and [Sil94]. Let ω be a generator for the rank 1 $\mathbf{Z}$-module of invariant differential 1-forms on $E_{\mathbf{Z}}$. The pullback of ω to $X_0(N)$ is a differential $\phi_E^*\omega$ on $X_0(N)$. The newform f defines another differential $2\pi i f(z)dz = f(q)dq/q$ on $X_0(N)$. Because the action of Hecke operators is compatible with the map $X_0(N) \to E$, the differential $\phi_E^*\omega$ is a $\mathbf{T}$-eigenvector with the same eigenvalues as $f(z)$, so by [AL70] we have $\phi_E^*\omega = c \cdot 2\pi i f(z)dz$ for some $c \in \mathbf{Q}^*$ (see also

[Man72, §5]). The *Manin constant* c_E of E is the absolute value of the rational number c defined above.

The following conjecture is implicit in [Man72, §5].

Conjecture 2.1 (Manin)**.** *We have* $c_E = 1$*.*

Significant progress has been made towards this conjecture. In the following theorems, p denotes a prime and N denotes the conductor of E.

Theorem 2.2 (Edixhoven [Edi91, Prop. 2])**.** *The constant* c_E *is an integer.*

Edixhoven proved this using an integral q-expansion map, whose existence and properties follow from results in [KM85]. We generalize his theorem to quotients of arbitrary dimension in Theorem 3.4.

Theorem 2.3 (Mazur, [Maz78, Cor. 4.1])**.** *If* $p \mid c_E$*, then* $p^2 \mid 4N$*.*

Mazur proved this by applying theorems of Raynaud about exactness of sequences of differentials, then using the "q-expansion principle" in characteristic p and a property of the Atkin-Lehner involution. We generalize Mazur's theorem in Corollary 3.7.

The following two results refine the above results at $p = 2$.

Theorem 2.4 (Raynaud [AU96, Prop. 3.1])**.** *If* $4 \mid c_E$*, then* $4 \mid N$*.*

Theorem 2.5 (Abbes-Ullmo [AU96, Thm. A])**.** *If* $p \mid c_E$*, then* $p \mid N$*.*

We generalize Theorem 2.4 in Theorem 3.10. However, it is not clear if Theorem 2.5 generalizes to dimension greater than 1. It would be fantastic if the theorem could be generalized. It would imply that the Manin constant is 1 for newform quotients A_f of $J_0(N)$, with N odd and square free, which be useful for computations regarding the conjecture of Birch and Swinnerton-Dyer.

B. Edixhoven also has unpublished results (see [Edi89]) which assert that the only primes that can divide c_E are 2, 3, 5, and 7; he also gives bounds that are independent of E on the valuations of c_E at 2, 3, 5, and 7. His arguments rely on the construction of certain stable integral models for $X_0(p^2)$.

See Section 5 for more details about the following computation:

Theorem 2.6 (Cremona)**.** *If E is an optimal elliptic curve over* $\mathbf{Q}$ *with conductor at most* 130000*, then* $c_E = 1$*.*

To the above list of theorems we add the following:

Theorem 2.7. *If* $p \mid c_E$ *then* $p^2 \mid N$ *or* $p \mid m_E$*.*

This theorem is a special case of Theorem 3.11 below. In view of Theorem 2.3, our new contribution is that if m_E is odd and $\mathrm{ord}_2(N) = 1$, then c_E is odd. This hypothesis is *very stringent*—of the optimal elliptic curve quotients of conductor ≤ 120000, only 56 of them satisfy the hypothesis.

3. Quotients of arbitrary dimension

For $N \geq 4$, let Γ a subgroup of $\Gamma_1(N)$ that contains $\Gamma_0(N)$, let X be the modular curve over $\mathbf{Q}$ associated to Γ, and let J be the Jacobian of X. Let I be a *saturated* ideal of the corresponding Hecke algebra $\mathbf{T}$, so $\mathbf{T}/I$ is torsion free. Then $A = A_I = J/IJ$ is an optimal quotient of J.

For a newform $f = \sum a_n(f)q^n \in S_2(\Gamma)$, consider the ring homomorphism $\mathbf{T} \to \mathbf{Z}[\ldots, a_n(f), \ldots]$ that sends T_n to $a_n(f)$. The kernel $I_f \subset \mathbf{T}$ of this homomorphism is a saturated prime ideal of $\mathbf{T}$. The *newform quotient* A_f associated to f is the quotient J/I_fJ. Shimura introduced A_f in [Shi73] where he proved that A_f is an abelian variety over $\mathbf{Q}$ of dimension equal to the degree of the field $\mathbf{Q}(\ldots, a_n(f), \ldots)$. He also observed that there is a natural map $\mathbf{T} \to \mathrm{End}(A_f)$ with kernel I_f.

For the rest of this section, fix a quotient A associated to a saturated ideal I of $\mathbf{T}$; note that A may or may not be attached to a newform.

3.1. Generalization to quotients of arbitrary dimension.

If R is a subring of $\mathbf{C}$, let $S_2(R) = S_2(\Gamma; R)$ denote the $\mathbf{T}$-submodule of $S_2(\Gamma; \mathbf{C})$ of modular forms whose Fourier expansions have all coefficients in R.

Lemma 3.1. *The Hecke operators leave $S_2(R)$ stable.*

Proof. If $\Gamma = \Gamma_0(N)$, then by the explicit description of the Hecke operators on Fourier expansions (e.g., see [DI95, Prop. 3.4.3]), it is clear that the Hecke operators leave $S_2(R)$ stable. For a general Γ, by [DI95, (12.4.1)], one just has to check in addition that the diamond operators also leave $S_2(R)$ stable, which in turn follows from [DI95, Prop. 12.3.11]. □

Lemma 3.2. *We have $S_2(R) \cong S_2(\mathbf{Z}) \otimes R$.*

Proof. This is [DI95, Thm. 12.3.2] when our spaces $S_2(R)$ and $S_2(\mathbf{Z})$ are replaced by their algebraic analogues (see [DI95, pg. 111]). Our spaces and their algebraic analogues are identified by the natural q-expansion maps according to [DI95, Thm. 12.3.7]. □

If B is an abelian variety over $\mathbf{Q}$ and S is a Dedekind domain with field of fractions $\mathbf{Q}$, then we denote by B_S the Néron model of B over S; also, for ease of notation, we will abbreviate $H^0(B_S, \Omega^1_{B_S/S})$ by $H^0(B_S, \Omega^1_{B/S})$.

The inclusion $X \hookrightarrow J$ that sends the cusp ∞ to 0 induces an isomorphism

$$H^0(X, \Omega^1_{X/\mathbf{Q}}) \cong H^0(J, \Omega^1_{J/\mathbf{Q}}).$$

Let ϕ_2 be the optimal quotient map $J \to A$. Then ϕ_2^* induces an inclusion $\psi : H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}) \hookrightarrow H^0(J, \Omega^1_{J/\mathbf{Q}})[I] \cong S_2(\mathbf{Q})[I]$, and we have the following commutative diagram:

$$\begin{tikzcd}
H^0(A, \Omega^1_{A/\mathbf{Q}}) \arrow[r, hook, "\cong"] & H^0(J, \Omega^1_{J/\mathbf{Q}})[I] \arrow[r, "\cong"] & S_2(\mathbf{Q})[I] \\
H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}) \arrow[u, hook] \arrow[urr, hook, "\psi"'] & & S_2(\mathbf{Z})[I] \arrow[u, hook]
\end{tikzcd}$$

Definition 3.3. The *Manin constant* of A is the (lattice) index

$$c_A = [S_2(\mathbf{Z})[I] : \psi(H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}))].$$

Theorem 3.4 below asserts that $c_A \in \mathbf{Z}$, so we may also consider the Manin module of A, which is the quotient $M = S_2(\mathbf{Z})[I]/\psi(H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}))$, and the Manin ideal of A, which is the annihilator of M in $\mathbf{T}$.

If A is an elliptic curve, then c_A is the usual Manin constant. The constant c as defined above was also considered by Gross [Gro82, 2.5, p.222] and Lang [Lan91, III.5, p.95]. The constant c_A was defined for the winding quotient in [Aga99], where it was called the generalized Manin constant. A Manin constant is defined in the context of $\mathbf{Q}$-curves in [GL01].

3.2. **Motivation: connection with the conjecture of Birch and Swinnerton-Dyer.** On a Néron model, the global differentials are the same as the invariant differentials, so $H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}})$ is a free $\mathbf{Z}$-module of rank $d = \dim(A)$. The *real measure* Ω_A of A is the measure of $A(\mathbf{R})$ with respect to the volume given by a generator of $\bigwedge^d H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}) \simeq \mathrm{H}^0(A_{\mathbf{Z}}, \Omega^d_{A_{\mathbf{Z}}/\mathbf{Z}})$. This quantity is of interest because it appears in the conjecture of Birch and Swinnerton-Dyer, which expresses the ratio $L^{(r)}(A,1)/\Omega_A$, in terms of arithmetic invariants of A, where $r = \mathrm{ord}_{s=1} L(A,s)$ (see, e.g., [Lan91, Chap. III, §5] and [AS05, §2.3]).

If we take a $\mathbf{Z}$-basis of $S_2(\mathbf{Z})[I]$ and take the inverse image via the top chain of arrows in the commutative diagram above, we get a $\mathbf{Q}$-basis of $H^0(A, \Omega^1_{A/\mathbf{Q}})$; let Ω'_A denote the volume of $A(\mathbf{R})$ with respect to the wedge product of the elements in the latter basis (this is independent of the choice of the former basis). In doing calculations or proving formulas regarding the ratio in the Birch and Swinnerton-Dyer conjecture mentioned above, it is easier to work with the volume Ω'_A instead of working with Ω_A. If one works with the easier-to-compute volume Ω'_A instead of Ω_A, it is necessary to obtain information about c_A in order to make conclusions regarding the conjecture of Birch and Swinnerton-Dyer, since $\Omega_A = c_A \cdot \Omega'_A$. For example, see [AS05, §4.2] when $r = 0$ and [GZ86, p. 310–311] when $r = 1$; in each case, one gets a formula for computing the Birch and Swinnerton-Dyer conjectural order of the Shafarevich-Tate group, and the formula contains the Manin constant (see, e.g., [Mc91]).

The method of Section 5 for verifying that $c_A = 1$ for specific elliptic curves is of little use when applied to general abelian varieties, since there is no simple analogue of the minimal Weierstrass model (but see [GL01] for **Q**-curves). This emphasizes the need for general theorems regarding the Manin constant of quotients of dimension bigger than one.

3.3. **Results and a conjecture.** We start by giving several results regarding the Manin constant for quotients of arbitrary dimension. The proofs of most of the theorems are given in Section 4.

Let Γ be a subgroup of $\Gamma_0(N)$ that contains $\Gamma_1(N)$. We have the following generalization of Edixhoven's Theorem 2.2.

Theorem 3.4. *The Manin constant c_A is an integer. (In the notation of Section 3.1 we even have that $\psi(H^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}})) \subseteq S_2(\mathbf{Z})[I]$.)*

Proof. Let $J = \mathrm{Jac}(X_\Gamma)$ and $J' = J_1(N)$. Suppose A is an optimal quotient of J. We have natural maps $\mathrm{H}^0(J'_{\mathbf{Z}}, \Omega^1_{J'/\mathbf{Z}}) \hookrightarrow \mathrm{H}^0(J', \Omega^1_{J'/\mathbf{Q}}) \overset{\cong}{\to} S_2(\Gamma_1(N); \mathbf{Q})$; from the proof of Lemma 6.1.6 of [CES03], the image of the composite is contained in $S_2(\Gamma_1(N); \mathbf{Z})$. The maps $J' \to J \to A$ induce a chain of inclusions

$$\mathrm{H}^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}}) \hookrightarrow \mathrm{H}^0(J_{\mathbf{Z}}, \Omega^1_{J/\mathbf{Z}}) \hookrightarrow \mathrm{H}^0(J'_{\mathbf{Z}}, \Omega^1_{J'/\mathbf{Z}}) \hookrightarrow S_2(\Gamma_1(N); \mathbf{Z}) \hookrightarrow \mathbf{Z}[[q]].$$

Combining this chain of inclusions with commutativity of the diagram

$$\begin{tikzcd} & S_2(\Gamma_1(N)) \arrow[dr, "F\text{-exp}"] & \\ S_2(\Gamma) \arrow[ur, "f(q)\mapsto f(q)"] \arrow[rr, "F\text{-exp}"] & & \mathbf{C}[[q]], \end{tikzcd}$$

where F-exp is the Fourier expansion map, we see that the image of $\mathrm{H}^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}})$ lies in $S_2(\mathbf{Z})[I]$, as claimed. □

For the rest of the paper, we take $\Gamma = \Gamma_0(N)$. For each prime $\ell \mid N$ with $\mathrm{ord}_\ell(N) = 1$, let W_ℓ be the ℓth Atkin-Lehner operator. Let $J = J_0(N)$ and $A = A_I = J/IJ$ be an optimal quotient of J attached to a saturated ideal I. If ℓ is a prime, then as usual, $\mathbf{Z}_{(\ell)}$ will denote the localization of $\mathbf{Z}$ at ℓ.

Theorem 3.5. *Suppose that ℓ is an odd prime such that $\ell^2 \nmid N$, and that if $\ell \mid N$, then $A^\vee \subset J$ is stable under W_ℓ. Then $\ell \mid c_A$ if and only if $\ell \mid N$ and $S_2(\mathbf{Z}_{(\ell)})[I]$ is not stable under the action of W_ℓ.*

We will prove this theorem in Section 4.2.

Remark 3.6. The condition that $S_2(\mathbf{Z}_{(\ell)})[I]$ is stable under W_ℓ can be verified using standard algorithms. Thus in light of Theorem 3.5, if A is stable under all Atkin-Lehner operators and N is square free, then one can compute the set of odd primes that divide c_A. It would be interesting to refine the arguments of this paper to find an algorithm to compute c_A exactly.

Let J_{old} denote the abelian subvariety of J generated by the images of the degeneracy maps from levels that properly divide N (see, e.g., [Maz78, §2(b)]) and let J^{new} denote the quotient of J by J_{old}. A *new quotient* is a quotient $J \to A$ that factors through the map $J \to J^{\text{new}}$. The following corollary generalizes Mazur's Theorem 2.3:

Corollary 3.7. *If $A = A_f$ is an optimal newform quotient of $J_0(N)$ and $\ell \mid c_A$ is a prime, then $\ell = 2$ or $\ell^2 \mid N$.*

Proof. Since f is a newform, W_ℓ acts as either 1 or -1 on A hence on $S_2(\mathbf{Z}_{(\ell)})[I]$. Thus $S_2(\mathbf{Z}_{(\ell)})[I]$ is W_ℓ-stable. □

Corollary 3.8. *If $A = J_0(N)_{\text{new}}$ is the new subvariety of $J_0(N)$ and $\ell \mid c_A$ is a prime, then $\ell = 2$ or $\ell^2 \mid N$. (In particular, if N is prime then the Manin constant of $J_0(N)$ is a power of 2, since $A = J_0(N)[I]$ for $I = 0$.)*

Proof. We have $W_\ell = -T_\ell$ on A (e.g., see the end of [DI95, §6.3]). Also the new subspace $S_2(\mathbf{Z})_{\text{new}}$ of $S_2(\Gamma_0(N))$ is T_ℓ-stable. □

Remark 3.9. If $A = J_0(33)$, then

$$W_3 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & -\frac{4}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix}$$

with respect to the basis

$$\begin{aligned} f_1 &= q - q^5 - 2q^6 + 2q^7 + \cdots, \\ f_2 &= q^2 - q^4 - q^5 - q^6 + 2q^7 + \cdots, \\ f_3 &= q^3 - 2q^6 + \cdots \end{aligned}$$

for $S_2(\mathbf{Z})$. Thus W_3 does not preserve $S_2(\mathbf{Z}_{(3)})$. In fact, the Manin constant of $J_0(33)$ is not 1 in this case (see Section 3.4).

The hypothesis of Theorem 3.5 sometimes holds for non-new A. For example, take $J = J_0(33)$ and $\ell = 3$. Then W_3 acts as an endomorphism of J, and a computation shows that the characteristic polynomial of W_3 on $S_2(33)_{\mathrm{new}}$ is $x-1$ and on $S_2(33)_{\mathrm{old}}$ is $(x-1)(x+1)$, where $S_2(33)_{\mathrm{old}}$ is the old subspace of $S_2(33)$. Consider the optimal elliptic curve quotient $A = J/(W_3+1)J$, which is isogenous to $J_0(11)$. Then A is an optimal old quotient of J, and W_3 acts as -1 on A, so W_3 preserves the corresponding spaces of modular forms. Thus Theorem 3.5 implies that $3 \nmid c_A$.

The following theorem generalizes Raynaud's Theorem 2.4 (see also [GL01] for generalizations to **Q**-curves).

Theorem 3.10. *If $f \in S_2(\Gamma_0(N))$ is a newform and ℓ is a prime such that $\ell^2 \nmid N$, then* $\mathrm{ord}_\ell(c_{A_f}) \le \dim A_f$.

Note that in light of Theorem 3.5, this theorem gives new information only at $\ell = 2$, when $2 \parallel N$. We prove the theorem in Section 4.4

Let π denote the natural quotient map $J \to A$. When we compose π with its dual $A^\vee \to J^\vee$ (identifying $J^\vee$ with J using the inverse of the principal polarization of J), we get an isogeny $\phi : A^\vee \to A$. The *modular exponent* m_A of A is the exponent of the group $\ker(\phi)$. When A is an elliptic curve, the modular exponent is just the modular degree of A (see, e.g., [AU96, p. 278]).

Theorem 3.11. *If $f \in S_2(\Gamma_0(N))$ is a newform and $\ell \mid c_{A_f}$ is a prime, then $\ell^2 \mid N$ or $\ell \mid m_A$.*

Again, in view of Corollary 3.7, this theorem gives new information only at $\ell = 2$, when $\mathrm{ord}_2(N) \le 1$. We prove the theorem in Section 4.3.

The theorems above suggest that the Manin constant is 1 for quotients associated to newforms of square-free level. In the case when the level is not square free, computations of [FpS$^+$01] involving Jacobians of genus 2 curves that are quotients of $J_0(N)^{\mathrm{new}}$ show that $c_A = 1$ for 28 two-dimensional newform quotients. These include quotients having the following non-square-free levels:

$$3^2\cdot 7, \quad 3^2\cdot 13, \quad 5^3, \quad 3^3\cdot 5, \quad 3\cdot 7^2, \quad 5^2\cdot 7, \quad 2^2\cdot 47, \quad 3^3\cdot 7.$$

The above observations suggest the following conjecture, which generalizes Conjecture 2.1:

Conjecture 3.12. *If f is a newform on $\Gamma_0(N)$ then $c_{A_f} = 1$.*

It is plausible that $c_{A_f} = 1$ for any newform on any congruence subgroup between $\Gamma_0(N)$ and $\Gamma_1(N)$. However, we do not have enough data to justify making a conjecture in this context.

3.4. **Examples of nontrivial Manin constants.** We present two sets of examples in which the Manin constant is not 1.

Using results of [Kil02], Adam Joyce [Joy05] proves that there is a new optimal quotient of $J_0(431)$ with Manin constant 2. Joyce's methods also produce examples with Manin constant 2 at levels 503 and 2089. For the convenience of the reader, we breifly discuss his example at level 431. There are exactly two elliptic curves E_1 and E_2 of prime conductor 431, and $E_1 \cap E_2 = 0$ as subvarieties of $J_0(431)$, so $A = E_1 \times E_2$ is an optimal quotient of $J_0(431)$ attached to a saturated ideal I. If f_i is the newform corresponding to E_i, then one finds that $f_1 \equiv f_2 \pmod 2$, and so $g = (f_1 - f_2)/2 \in S_2(\mathbf{Z})[I]$. However g is not in the image of $\mathrm{H}^0(A_{\mathbf{Z}}, \Omega^1_{A/\mathbf{Z}})$. Thus the Manin constant of A is divisible by 2.

As another class of examples, one finds by computation for each prime $\ell \leq 100$ that W_ℓ does not leave $S_2(\Gamma_0(11\ell); \mathbf{Z}_{(\ell)})$ stable. Theorem 3.5 (with $I = 0$) then implies that the Manin constant of $J_0(11\ell)$ is divisible by ℓ for these values of ℓ.

4. Proofs of some of the Theorems

In Sections 4.2, 4.3, and 4.4, we prove Theorems 3.5, 3.11, and 3.10 respectively. In Section 4.1, we state two lemmas that will be used in these proofs. The proofs of the theorems may be read independently of each other, after reading Section 4.1.

4.1. **Two lemmas.** The following lemma is a standard fact; we state it as a lemma merely because it is used several times.

Lemma 4.1. *Suppose $i : A \hookrightarrow B$ is an injective homomorphism of torsion-free abelian groups. If p is a prime, then $B/i(A)$ has no nonzero p-torsion if and only if the induced map $A \otimes \mathbf{F}_p \to B \otimes \mathbf{F}_p$ is injective.*

Proof. Let Q denote the quotient $B/i(A)$. Tensor the exact sequence $0 \to A \to B \to Q \to 0$ with $\mathbf{F}_p$. The associated long exact sequences reveal that $\ker(A \otimes \mathbf{F}_p \to B \otimes \mathbf{F}_p) \cong Q_{\mathrm{tor}}[p]$. □

Suppose ℓ is a prime such that $\ell^2 \nmid N$. In what follows, we will be stating some standard facts taken from [Maz78, §2(e)] (which in turn relies on [DR73]). Let $\mathcal{X}_{\mathbf{Z}_{(\ell)}}$ be the minimal regular resolution of the coarse moduli scheme associated to $\Gamma_0(N)$ (as in [DR73, § VI.6.9]) over $\mathbf{Z}_{(\ell)}$, and let $\Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}$ denote the relative dualizing sheaf of $\mathcal{X}_{\mathbf{Z}_{(\ell)}}$ over $\mathbf{Z}_{(\ell)}$. The Tate curve over $\mathbf{Z}_{(\ell)}[[q]]$ gives rise to a morphism from Spec $\mathbf{Z}_{(\ell)}[[q]]$ to the smooth locus of $\mathcal{X}_{\mathbf{Z}_{(\ell)}} \to \operatorname{Spec} \mathbf{Z}_{(\ell)}$. Since the module of completed Kahler differentials for $\mathbf{Z}_{(\ell)}[[q]]$ over $\mathbf{Z}_{(\ell)}$ is free of rank 1 on the basis dq, we obtain a map q-exp $: H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}) \to \mathbf{Z}_{(\ell)}[[q]]$.

The natural morphism $\mathrm{Pic}^0_{\mathcal{X}/\mathbf{Z}_{(\ell)}} \to J_{\mathbf{Z}_{(\ell)}}$ identifies $\mathrm{Pic}^0_{\mathcal{X}/\mathbf{Z}_{(\ell)}}$ with the identity component of $J_{\mathbf{Z}_{(\ell)}}$ (see, e.g., [BLR90, §9.4–9.5]). Passing to tangent spaces along the identity section over $\mathbf{Z}_{(\ell)}$, we obtain an isomorphism $H^1(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \mathcal{O}_{\mathcal{X}_{\mathbf{Z}_{(\ell)}}}) \cong \mathrm{Tan}(J_{\mathbf{Z}_{(\ell)}})$. Using Grothendieck duality, one gets an isomorphism $\mathrm{Cot}(J_{\mathbf{Z}_{(\ell)}}) \overset{\cong}{\to} H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}})$, where $\mathrm{Cot}(J_{\mathbf{Z}_{(\ell)}})$ is the cotangent space at the identity section. On the Néron model $J_{\mathbf{Z}_{(\ell)}}$, the group of global differentials is the same as the group of invariant differentials, which in turn is naturally isomorphic to $\mathrm{Cot}(J_{\mathbf{Z}_{(\ell)}})$. Thus we obtain an isomorphism $H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \cong H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}})$.

Let G be a $\mathbf{T}$-module equipped with an injection $G \hookrightarrow H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}})$ of $\mathbf{T}$-modules such that G is annihilated by I. If $\ell \mid N$, assume moreover that G is a $\mathbf{T}[W_\ell]$-module and that the inclusion in the previous sentence is a homomorphism of $\mathbf{T}[W_\ell]$-modules. As a typical example, $G = H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$, with the injection $\pi^* : H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \hookrightarrow H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}})$. Let Φ be the composition of the inclusions

$$G \hookrightarrow H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \cong H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}) \xrightarrow{q\text{-exp}} \mathbf{Z}_{(\ell)}[[q]], \tag{1}$$

and let ψ' be the composition of

$$G \hookrightarrow H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}})[I] \hookrightarrow S_2(\mathbf{Z}_{(\ell)})[I],$$

where the last inclusion follows from a "local" version of Theorem 3.4. The maps Φ and ψ' are related by the commutative diagram

$$\begin{tikzcd}
 & S_2(\mathbf{Z}_{(\ell)})[I] \arrow[dr, "F\text{-exp}"] & \\
G \arrow[ur, "\psi'"] \arrow[rr, "\Phi"] & & \mathbf{Z}_{(\ell)}[[q]],
\end{tikzcd} \tag{2}$$

where F-exp is the Fourier expansion map (at infinity), as before.

We say that a subgroup B of an abelian group C is *saturated* (in C) if the quotient C/B is torsion free.

Lemma 4.2. *Recall that ℓ is a prime such that $\ell^2 \nmid N$. If ℓ divides N, suppose that $S_2(\mathbf{Z}_{(\ell)})[I]$ is stable under the action of W_ℓ; if $\ell = 2$ assume moreover that W_ℓ acts as a scalar on A. Consider the map*

$$G \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell,$$

which is obtained by tensoring the inclusion $G \hookrightarrow H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}})$ with $\mathbf{F}_\ell$. If this map is injective, then the image of G under the map Φ of (2) is saturated in $\mathbf{Z}_{(\ell)}[[q]]$.

Proof. By Lemma 4.1, it suffices to prove that the map

$$\Phi_\ell : G \otimes \mathbf{F}_\ell \to \mathbf{Z}_{(\ell)}[[q]] \otimes \mathbf{F}_\ell = \mathbf{F}_\ell[[q]]$$

obtained by tensoring (1) with $\mathbf{F}_\ell$ is injective. Let $\mathcal{X}_{\mathbf{F}_\ell}$ denote the special fiber of $\mathcal{X}_{\mathbf{Z}_{(\ell)}}$ and let $\Omega_{\mathcal{X}/\mathbf{F}_\ell}$ denote the relative dualizing sheaf of $\mathcal{X}_{\mathbf{F}_\ell}$ over $\mathbf{F}_\ell$.

First suppose that ℓ does not divide N. Then $\mathcal{X}_{\mathbf{Z}_{(\ell)}}$ is smooth and proper over $\mathbf{Z}_{(\ell)}$. Thus the formation of $\mathrm{H}^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}_{\mathbf{Z}_{(\ell)}}})$ is compatible with any base change on $\mathbf{Z}_{(\ell)}$ (such as reduction modulo ℓ). The injectivity of Φ_ℓ now follows since by hypothesis the induced map $G \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$ is injective, and

$$H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{X}_{\mathbf{F}_\ell}, \Omega_{\mathcal{X}/\mathbf{F}_\ell}) \to \mathbf{F}_\ell[[q]]$$

is injective by the q-expansion principle (which is easy in this setting, since $\mathcal{X}_{\mathbf{F}_\ell}$ is a smooth and geometrically connected curve).

Next suppose that ℓ divides N. First we verify that $\ker(\Phi_\ell)$ is stable under W_ℓ. Suppose $\omega \in \ker(\Phi_\ell)$. Choose $\omega' \in G$ such that the image of ω' in $G \otimes \mathbf{F}_\ell$ is ω, and let $f = \psi'(\omega')$. Because $\Phi_\ell(\omega) = 0$ in $\mathbf{F}_\ell[[q]]$, there exists $h \in \mathbf{Z}_{(\ell)}[[q]]$ such that $\ell h = F\text{-exp}(f)$. Let $f' = f/\ell \in S_2(\mathbf{Q})$; then f' is actually in $S_2(\mathbf{Z}_{(\ell)})$ (since $F\text{-exp}(f/\ell) = h \in \mathbf{Z}_{(\ell)}[[q]]$). Now $\ell f' = f$ is annihilated by every element of I, hence so is f'; thus $f' \in S_2(\mathbf{Z}_{(\ell)})[I]$. By hypothesis, $W_\ell(f') \in S_2(\mathbf{Z}_{(\ell)})[I]$. Then

$$\Phi(W_\ell \omega') = F\text{-exp}(W_\ell f) = \ell \cdot F\text{-exp}(W_\ell f') \in \ell \mathbf{Z}_{(\ell)}[[q]].$$

Reducing modulo ℓ, we get $\Phi_\ell(W_\ell \omega) = 0$ in $\mathbf{F}_\ell[[q]]$. Thus $W_\ell \omega \in \ker(\Phi_\ell)$, which proves that $\ker(\Phi_\ell)$ is stable under W_ℓ.

Since W_ℓ is an involution, and by hypothesis either ℓ is odd or W_ℓ is a scalar, the space $\ker(\Phi_\ell)$ breaks up into a direct sum of eigenspaces under W_ℓ with eigenvalues ± 1. It suffices to show that if $\omega \in \ker(\Phi_\ell)$ is an element of either eigenspace, then $\omega = 0$. For this, we use a standard argument that goes back to Mazur (see, e.g., the proof of Prop. 22 in [MR91]); we give some details to clarify the argument in our situation.

Following the proof of Prop. 3.3 on p. 68 of [Maz77], we have

$$H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{X}_{\mathbf{F}_\ell}, \Omega_{\mathcal{X}/\mathbf{F}_\ell}).$$

In the following, we shall think of $G \otimes \mathbf{F}_\ell$ as a subgroup of $H^0(\mathcal{X}_{\mathbf{F}_\ell}, \Omega_{\mathcal{X}/\mathbf{F}_\ell})$, which we can do by the hypothesis that the induced map $G \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$ is injective and that

$$H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{X}_{\mathbf{Z}_{(\ell)}}, \Omega_{\mathcal{X}/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \cong H^0(\mathcal{X}_{\mathbf{F}_\ell}, \Omega_{\mathcal{X}/\mathbf{F}_\ell}).$$

Suppose $\omega \in \ker(\Phi_\ell)$ is in the ± 1 eigenspace (we will treat the cases of $+1$ and -1 eigenspaces together). We will show that ω is trivial over $\mathcal{X}_{\overline{\mathbf{F}}_\ell}$, the base change of $\mathcal{X}_{\mathbf{F}_\ell}$ to an algebraic closure $\overline{\mathbf{F}}_\ell$, which suffices for our purposes. Since $\ell^2 \nmid N$, we have $\ell \mid\mid N$, and so the special fiber $\mathcal{X}_{\overline{\mathbf{F}}_\ell}$ is as depicted on p. 177 of [Maz77]: it consists of the union of two copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$ identified transversely at the supersingular points, and some copies of $\mathbf{P}^1$, each of which intersects exactly one of the two copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$ and perhaps another $\mathbf{P}^1$, all of them transversally. All the singular points are ordinary double points, and the cusp ∞ lies on one of the two copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$.

In particular, $\mathcal{X}_{\overline{\mathbf{F}}_\ell} \to \operatorname{Spec} \overline{\mathbf{F}}_\ell$ is locally a complete intersection, hence Gorenstein, and so by [DR73, § I.2.2, p. 162], the sheaf $\Omega_{\mathcal{X}/\overline{\mathbf{F}}_\ell} = \Omega_{\mathcal{X}/\mathbf{F}_\ell} \otimes \overline{\mathbf{F}}_\ell$ is invertible. Since $\omega \in \ker(\Phi_\ell)$, the differential ω vanishes on the copy of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$ containing the cusp ∞ by the q-expansion principle (which is easy in this case, since all that is being invoked here is that on an integral curve, the natural map from the group of global sections of an invertible sheaf into the completion of the sheaf's stalk at a point is injective). The two copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$ are swapped under the action of the Atkin-Lehner involution W_ℓ, and thus $W_\ell(\omega)$ vanishes on the other copy that does not contain the cusp ∞. Since $W_\ell(\omega) = \pm\omega$, we see that ω is zero on both copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$. Also, by the description of the relative dualizing sheaf in [DR73, § I.2.3, p. 162], if $\pi : \widetilde{\mathcal{X}}_{\overline{\mathbf{F}}_\ell} \to \mathcal{X}_{\overline{\mathbf{F}}_\ell}$ is a normalization, then ω correponds to a meromorphic differential $\widetilde{\omega}$ on $\widetilde{\mathcal{X}}_{\overline{\mathbf{F}}_\ell}$ which is regular outside the inverse images (under π) of the double points on $\mathcal{X}_{\overline{\mathbf{F}}_\ell}$ and has at worst a simple pole at any point that lies over a double point on $\mathcal{X}_{\overline{\mathbf{F}}_\ell}$. Moreover, on the inverse image of any double point on $\mathcal{X}_{\overline{\mathbf{F}}_\ell}$, the residues of $\widetilde{\omega}$ add to zero. For any of the $\mathbf{P}^1$'s, above a point of intersection of the $\mathbf{P}^1$ with a copy of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$, the residue of $\widetilde{\omega}$ on the inverse image of the copy of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$ is zero (since ω is trivial on both copies of $X_0(N/\ell)_{\overline{\mathbf{F}}_\ell}$), and thus the residue of $\widetilde{\omega}$ on the inverse image of $\mathbf{P}^1$ is zero. Thus $\widetilde{\omega}$ restricted to the inverse image of $\mathbf{P}^1$ is regular away from the inverse image of any point where the $\mathbf{P}^1$ meets another $\mathbf{P}^1$ (recall that there can be at most one such point). Hence the restriction of $\widetilde{\omega}$ to the inverse image of the $\mathbf{P}^1$ is either regular everywhere or is regular away from one point where it has at most a simple pole; in the latter case, the residue is zero by the residue theorem. Thus in either case, $\widetilde{\omega}$ restricted to the inverse image of the $\mathbf{P}^1$ is regular, and therefore is zero. Thus ω is trivial on all the copies of $\mathbf{P}^1$ as well. Hence $\omega = 0$, as was to be shown. □

4.2. **Proof of Theorem 3.5.** We continue to use the notation of Section 4.1.

First suppose that $\ell \mid N$ and $S_2(\mathbf{Z}_{(\ell)})[I]$ is not stable under the action of W_ℓ. Relative differentials and Néron models are functorial, so $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$ is W_ℓ-stable. Thus the map $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \to S_2(\mathbf{Z}_{(\ell)})[I]$ is not surjective. But c_A is the order of the cokernel, so $\ell \mid c_A$.

Next we prove the other implication, namely that if $\ell \mid c_A$, then $\ell \mid N$ and $S_2(\mathbf{Z}_{(\ell)})[I]$ is not stable under W_ℓ. We will prove this by proving the contrapositive, i.e., that if either $\ell \nmid N$ or $S_2(\mathbf{Z}_{(\ell)})[I]$ is stable under W_ℓ, then $\ell \nmid c_A$.

We now follow the discussion preceding Lemma 4.2, taking $G = H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$. To show that $\ell \nmid c_A$, we have to show that c_A is a unit in $\mathbf{Z}_{(\ell)}$. For this, it suffices to check that in diagram (2), the image of $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$ in $\mathbf{Z}_{(\ell)}[[q]]$ under Φ is saturated, since the image of $S_2(\Gamma_0(N); \mathbf{Z}_{(\ell)})[I]$ under F-exp is saturated in $\mathbf{Z}_{(\ell)}[[q]]$. In view of Lemma 4.2, it suffices to show that the map

$$H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$$

is injective.

Since A is an optimal quotient, $\ell \neq 2$, and J has good or semistable reduction at ℓ, [Maz78, Cor 1.1] yields an exact sequence

$$0 \to H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \to H^0(B_{\mathbf{Z}_{(\ell)}}, \Omega^1_{B/\mathbf{Z}_{(\ell)}}) \to 0$$

where $B = \ker(J \to A)$. Since $H^0(B_{\mathbf{Z}_{(\ell)}}, \Omega^1_{B/\mathbf{Z}_{(\ell)}})$ is torsion free, by Lemma 4.1 the map $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$ is injective, as was to be shown.

4.3. **Proof of Theorem 3.11.** We continue to use the notation and hypotheses of Section 4.1 (so $\ell^2 \nmid N$) and assume in addition that A is a newform quotient, and that $\ell \nmid m_A$. We have to show that then $\ell \nmid c_A$. Just as in the previous proof, it suffices to check that the image of $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$ in $\mathbf{Z}_{(\ell)}[[q]]$ is saturated. Since A is a newform quotient, if $\ell \mid N$, then W_ℓ acts as a scalar on A and on $S_2(\Gamma_0(N); \mathbf{Z}_{(\ell)})[I]$. So again, using Lemma 4.2, it suffices to show that the map $H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell \to H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$ is injective.

The composition of pullback and pushforward in the following diagram is multiplication by the modular exponent of A:

$$\begin{tikzcd} & H^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \arrow[dr, "\pi_*"] & \\ H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \arrow[ur, "\pi^*"] \arrow[rr, "m_A"] & & H^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \end{tikzcd}$$

Since $m_A \in {\mathbf{Z}_{(\ell)}}^\times$, the map π^* is a section to the map π_* up to a unit and hence its reduction modulo ℓ is injective, which is what was left to be shown.

4.4. **Proof of Theorem 3.10.** Theorem 3.10 asserts that if $A = A_f$ is a quotient of $J = J_0(N)$ attached to a newform f, and ℓ is a prime such that $\ell^2 \nmid N$, then $\mathrm{ord}_\ell(c_{\scriptscriptstyle A}) \leq \dim(A)$. Our proof follows [AU96], except at the end we argue using lattice indices instead of multiples.

Let B denote the kernel of the quotient map $J \to A$. Consider the exact sequence $0 \to B \to J \to A \to 0$, and the corresponding complex $B_{\mathbf{Z}_{(\ell)}} \to J_{\mathbf{Z}_{(\ell)}} \to A_{J_{\mathbf{Z}_{(\ell)}}}$ of Néron models. Because $J_{\mathbf{Z}_{(\ell)}}$ has semiabelian reduction (since $\ell^2 \nmid N$), Theorem A.1 of the appendix of [AU96, pg. 279–280], due to Raynaud, implies that there is an integer r and an exact sequence

$$0 \to \mathrm{Tan}(B_{\mathbf{Z}_{(\ell)}}) \to \mathrm{Tan}(J_{\mathbf{Z}_{(\ell)}}) \to \mathrm{Tan}(A_{\mathbf{Z}_{(\ell)}}) \to (\mathbf{Z}/\ell\mathbf{Z})^r \to 0.$$

Here Tan is the tangent space at the 0 section; it is a finite free $\mathbf{Z}_{(\ell)}$-module of rank equal to the dimension. In particular, we have $r \leq d = \dim(A)$. Note that Tan is $\mathbf{Z}_{(\ell)}$-dual to the cotangent space, and the cotangent space is isomorphic to the space of global differential 1-forms. The theorem of Raynaud mentioned above is the generalization to $e = \ell - 1$ of [Maz78, Cor. 1.1], which we used above in the proof of Theorem 3.5.

Let C be the cokernel of $\mathrm{Tan}(B_{\mathbf{Z}_{(\ell)}}) \to \mathrm{Tan}(J_{\mathbf{Z}_{(\ell)}})$. We have a diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Tan}(B_{\mathbf{Z}_{(\ell)}}) \arrow[r] & \mathrm{Tan}(J_{\mathbf{Z}_{(\ell)}}) \arrow[rr] \arrow[dr] & & \mathrm{Tan}(A_{\mathbf{Z}_{(\ell)}}) \arrow[r] & (\mathbf{Z}/\ell\mathbf{Z})^r \arrow[r] & 0. \\
& & & C \arrow[ur, hook] & & &
\end{tikzcd} \tag{3}$$

Since $C \subset \mathrm{Tan}(A_{\mathbf{Z}_{(\ell)}})$, so C is torsion free, we see that C is a free $\mathbf{Z}_{(\ell)}$-module of rank d. Let $C^* = \mathrm{Hom}_{\mathbf{Z}_{(\ell)}}(C, \mathbf{Z}_{(\ell)})$ be the $\mathbf{Z}_{(\ell)}$-linear dual of C. Applying the $\mathrm{Hom}_{\mathbf{Z}_{(\ell)}}(-, \mathbf{Z}_{(\ell)})$ functor to the two short exact sequences in (3), we obtain exact sequences

$$0 \to C^* \to \mathrm{H}^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \to \mathrm{H}^0(B_{\mathbf{Z}_{(\ell)}}, \Omega^1_{B/\mathbf{Z}_{(\ell)}}) \to 0,$$

and

$$0 \to \mathrm{H}^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}}) \to C^* \to (\mathbf{Z}/\ell\mathbf{Z})^r \to 0. \tag{4}$$

The $(\mathbf{Z}/\ell\mathbf{Z})^r$ on the right in (4) occurs as $\mathrm{Ext}^1_{\mathbf{Z}_{(\ell)}}((\mathbf{Z}/\ell\mathbf{Z})^r, \mathbf{Z}_{(\ell)})$.

Since $\mathrm{H}^0(B_{\mathbf{Z}_{(\ell)}}, \Omega^1_{B/\mathbf{Z}_{(\ell)}})$ is torsion free, by Lemma 4.1, the induced map

$$C^* \otimes \mathbf{F}_\ell \to \mathrm{H}^0(J_{\mathbf{Z}_{(\ell)}}, \Omega^1_{J/\mathbf{Z}_{(\ell)}}) \otimes \mathbf{F}_\ell$$

is injective. Since A is a newform quotient, if $\ell \mid N$ then W_ℓ acts as a scalar on C^* and on $S_2(\Gamma_0(N); \mathbf{Z}_{(\ell)})[I]$. Using Lemma 4.2, with $G = C^*$, we see that the image of C^* in $\mathbf{Z}_{(\ell)}[[q]]$ under the composite of the maps in (1) is saturated. The Manin constant for A at ℓ is the index of the image via q-expansion of $\mathrm{H}^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$ in $\mathbf{Z}_{(\ell)}[[q]]$ in its saturation. Since the image of C^* in $\mathbf{Z}_{(\ell)}[[q]]$ is saturated, the image of C^* is the saturation of the image of $\mathrm{H}^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$, so the Manin constant at ℓ is the index of $\mathrm{H}^0(A_{\mathbf{Z}_{(\ell)}}, \Omega^1_{A/\mathbf{Z}_{(\ell)}})$ in C^*, which is ℓ^r by (4), hence is at most ℓ^d.

5. Appendix by J. Cremona: Verifying that $c = 1$

Let f be a normalised rational newform for $\Gamma_0(N)$. Let Λ_f be its period lattice; that is, the lattice of periods of $2\pi i f(z)dz$ over $H_1(X_0(N), \mathbf{Z})$.

We know that $E_f = \mathbf{C}/\Lambda_f$ is an elliptic curve E_f defined over $\mathbf{Q}$ and of conductor N. This is the optimal quotient of $J_0(N)$ associated to f. Our goal is two-fold: to identify E_f (by giving an explicit Weierstrass model for it with integer coeffients); and to show that the associated Manin constant for E_f is 1. In this section we will give an algorithm for this; our algorithm applies equally to optimal quotients of $J_1(N)$.

As input to our algorithm, we have the following data:

(1) a $\mathbf{Z}$-basis for Λ_f, known to a specific precision;
(2) the type of the lattice Λ_f (defined below); and
(3) a complete isogeny class of elliptic curves $\{E_1, \ldots, E_m\}$ of conductor N, given by minimal models, all with $L(E_j, s) = L(f, s)$.

So E_f is isomorphic over $\mathbf{Q}$ to E_{j_0} for a unique $j_0 \in \{1, \ldots, m\}$.

The justification for this uses the full force of the modularity of elliptic curves defined over $\mathbf{Q}$: we have computed a full set of newforms f at level N, and the same number of isogeny classes of elliptic curves, and the theory tells us that there is a bijection between these sets. Checking the first few terms of the L-series (i.e., comparing the Hecke eigenforms of the newforms with the traces of Frobenius for the curves) allows us to pair up each isogeny class with a newform.

We will assume that one of the E_j, which we always label E_1, is such that Λ_f and Λ_1 (the period lattice of E_1) are approximately equal. This is true in practice, because our method of finding the curves in the isogeny class is to compute the coefficients of a curve from numerical approximations to the c_4 and c_6 invariants of $\mathbf{C}/\Lambda_f$; in all cases these are very close to integers which are the invariants of the minimal model of an elliptic curve of conductor N, which we call E_1. The other curves in the isogeny class are then computed from E_1. For the algorithm described here, however, it is irrelevant how the curves E_j were obtained, provided that Λ_1 and Λ_f are close (in a precise sense defined below).

Normalisation of lattices: every lattice Λ in $\mathbf{C}$ which defined over $\mathbf{R}$ has a unique $\mathbf{Z}$-basis ω_1, ω_2 satisfying one of the following:

- **Type 1:** ω_1 and $(2\omega_2 - \omega_1)/i$ are real and positive; or

- **Type 2:** ω_1 and ω_2/i are real and positive.

For Λ_f we know the type from modular symbol calculations, and we know ω_1, ω_2 to a certain precision by numerical integration; modular symbols provide us with cycles $\gamma_1, \gamma_2 \in H_1(X_0(N), \mathbf{Z})$ such that the integral of $2\pi i f(z)dz$ over γ_1, γ_2 give ω_1, ω_2.

For each curve E_j we compute (to a specific precision) a $\mathbf{Z}$-basis for its period lattice Λ_j using the standard AGM method. Here, Λ_j is the lattice of periods of the Néron differential on E_j. The type of Λ_j is determined by the sign of the discriminant of E_j: type 1 for negative discriminant, and type 2 for positive discriminant.

For our algorithm we will need to know that Λ_1 and Λ_f are approximately equal. To be precise, we know that they have the same type, and also we verify, for a specific positive ε, that

$$(*) \qquad \left|\frac{\omega_{1,1}}{\omega_{1,f}} - 1\right| < \varepsilon \qquad \text{and} \qquad \left|\frac{\mathrm{im}(\omega_{2,1})}{\mathrm{im}(\omega_{2,f})} - 1\right| < \varepsilon.$$

Here $\omega_{1,j}$, $\omega_{2,j}$ denote the normalised generators of Λ_j, and $\omega_{1,f}$, $\omega_{2,f}$ those of Λ_f.

Pulling back the Néron differential on E_{j_0} to $X_0(N)$ gives $c \cdot 2\pi i f(z)dz$ where $c \in \mathbf{Z}$ is the Manin constant for f. Hence

$$c\Lambda_f = \Lambda_{j_0}.$$

Our task is now to

(1) identify j_0, to find which of the E_j is (isomorphic to) the "optimal" curve E_f; and
(2) determine the value of c.

Our main result is that $j_0 = 1$ and $c = 1$, provided that the precision bound ε in (*) is sufficiently small (in most cases, $\varepsilon < 1$ suffices). In order to state this precisely, we need some further definitions.

A result of Stevens says that in the isogeny class there is a curve, say E_{j_1}, whose period lattice Λ_{j_1} is contained in every Λ_j; this is the unique curve in the class with minimal Faltings height. (It is conjectured that E_{j_1} is the $\Gamma_1(N)$-optimal curve, but we do not need or use this fact. In many cases, the $\Gamma_0(N)$- and $\Gamma_1(N)$-optimal curves are the same, so we expect that $j_0 = j_1$ often; indeed, this holds for the vast majority of cases.)

For each j, we know therefore that $a_j = \omega_{1,j_1}/\omega_{1,j} \in \mathbf{N}$ and also $b_j = \mathrm{im}(\omega_{2,j_1})/\mathrm{im}(\omega_{2,j}) \in \mathbf{N}$. Let B be the maximum of a_1 and b_1.

Proposition 5.1. *Suppose that (*) holds with $\epsilon = B^{-1}$; then $j_0 = 1$ and $c = 1$. That is, the curve E_1 is the optimal quotient and its Manin constant is* 1.

Proof. Let $\varepsilon = B^{-1}$ and $\lambda = \frac{\omega_{1,1}}{\omega_{1,f}}$, so $|\lambda - 1| < \varepsilon$. For some j we have $c\Lambda_f = \Lambda_j$. The idea is that $\operatorname{lcm}(a_1, b_1)\Lambda_1 \subseteq \Lambda_{j_1} \subseteq \Lambda_j = c\Lambda_f$; if $a_1 = b_1 = 1$, then the closeness of Λ_1 and Λ_f forces $c = 1$ and equality throughout. To cover the general case it is simpler to work with the real and imaginary periods separately.

Firstly,

$$\frac{\omega_{1,j}}{\omega_{1,f}} = c \in \mathbf{Z}.$$

Then

$$c = \frac{\omega_{1,1}}{\omega_{1,f}} \frac{\omega_{1,j}}{\omega_{1,1}} = \frac{a_1}{a_j}\lambda.$$

Hence

$$0 \le |\lambda - 1| = \frac{|a_j c - a_1|}{a_1} < \varepsilon.$$

If $\lambda \neq 1$, then $\varepsilon > |\lambda - 1| \ge a_1^{-1} \ge B^{-1} = \varepsilon$, contradiction. Hence $\lambda = 1$, so $\omega_{1,1} = \omega_{1,f}$. Similarly, we have

$$\frac{\operatorname{im}(\omega_{2,j})}{\operatorname{im}(\omega_{2,f})} = c \in \mathbf{Z}$$

and again we can conclude that $\operatorname{im}(\omega_{2,1}) = \operatorname{im}(\omega_{2,f})$, and hence $\omega_{2,1} = \omega_{2,f}$.

Thus $\Lambda_1 = \Lambda_f$, from which the result follows. □

Theorem 5.2. *For all $N < 60000$, every optimal elliptic quotient of $J_0(N)$ has Manin constant equal to* 1. *Moreover, the optimal curve in each class is the one whose identifying number on the tables* [Cre] *is* 1 *(except for class* 990h *where the optimal curve is* 990h3*).*

Proof. For all $N < 60000$ we used modular symbols to find all newforms f and their period lattices, and also the corresponding isogeny classes of curves. In all cases we verified that (*) held with the appropriate value of ε. (The case of 990h is only exceptional on account of an error in labelling the curves several years ago, and is not significant.) □

Remark 5.3. In the vast majority of cases, the value of B is 1, so the precision needed for the computation of the periods is very low. For $N < 60000$, out of 258502 isogeny classes, only 136 have $B > 1$: we found $a_1 = 2$ in 13 cases, $a_1 = 3$ in 29 cases, and $a_1 = 4$ and $a_1 = 5$ once each (for $N = 15$ and $N = 11$ respectively); $b_1 = 2$ in 93 cases; and no larger values. Class 17a is the only one for which both a_1 and b_1 are greater than 1 (both are 2).

Finally, we give a slightly weaker result for $60000 < N < 130000$; in this range we do not know Λ_f precisely, but only its projection onto the real line. (The reason for this is that we can find the newforms using modular symbols for $H_1^+(X_0(N), \mathbf{Z})$, which has half the dimension of $H_1(X_0(N), \mathbf{Z})$; but to find the exact period lattice requires working in $H_1(X_0(N), \mathbf{Z})$.) The argument is similar to the one given above, using $B = a_1$.

Theorem 5.4. *For all N in the range $60000 < N < 130000$, every optimal elliptic quotient of $J_0(N)$ has Manin constant equal to 1.*

Proof. We continue to use the notation above. We do not know the lattice Λ_f but only (to a certain precision) a positive real number $\omega_{1,f}^{+}$ such that either Λ_f has type 1 and $\omega_{1,f} = 2\omega_{1,f}^{+}$, or Λ_f has type 2 and $\omega_{1,f} = \omega_{1,f}^{+}$. Curve E_1 has lattice Λ_1, and the ratio $\lambda = \omega_{1,1}^{+}/\omega_{1,f}^{+}$ satisfies $|\lambda - 1| < \varepsilon$. In all cases this holds with $\varepsilon = \frac{1}{3}$, which will suffice.

First assume that $a_1 = 1$.

If the type of Λ_f is the same as that of Λ_1 (for example, this must be the case if all the Λ_j have the same type, which will hold whenever all the isogenies between the E_j have odd degree) then from $c\Lambda_f = \Lambda_j$ we deduce as before that $\lambda = 1$ exactly, and $c = a_1/a_j = 1/a_j$, hence $c = a_j = 1$. So in this case we have that $c = 1$, though there might be some ambiguity in which curve is optimal if $a_j = 1$ for more than one value of j.

Assume next that Λ_1 has type 1 but Λ_f has type 2. Now $\lambda = \omega_{1,1}/2\omega_{1,f}$. The usual argument now gives $ca_j = 2$. Hence either $c = 1$ and $a_j = 2$, or $c = 2$ and $a_j = 1$. To see if the latter case could occur, we look for classes in which $a_1 = 1$ and Λ_1 has type 1, while for some $j > 1$ we also have $a_j = 1$ and Λ_j of type 2. This occurs 28 times for $60000 < N < 130000$, but for 15 of these the level N is odd, so we know that c must be odd. The remaining 13 cases are

$$62516a, 67664a, 71888e, 72916a, 75092a, 85328d, 86452a, 96116a,$$
$$106292b, 111572a, 115664a, 121168e, 125332a;$$

we have been able to eliminate these by carrying out the extra computations necessary as in the proof of Theorem 5.2. We note that in all of these 13 cases, the isogeny class consists of two curves, E_1 of type 1 and E_2 of type 2, with $[\Lambda_1 : \Lambda_2] = 2$, so that E_2 rather than E_1 has minimal Faltings height.

Next suppose that Λ_1 has type 2 but Λ_f has type 1. Now $\lambda = 2\omega_{1,1}/\omega_{1,f}$. The usual argument now gives $2ca_j = 1$, which is impossible; so this case cannot occur.

Finally we consider the cases where $a_1 > 1$. There are only three of these for $60000 < N < 130000$: namely, $91270a$, $117622a$ and $124973b$, where $a_1 = 3$. In each case the Λ_j all have the same type (they are linked via 3-isogenies) and the usual argument shows that $ca_j = 3$. But none of these levels is divisible by 3, so $c = 1$ in each case. □

References

[AU96] A. Abbes and E. Ullmo, *À propos de la conjecture de Manin pour les courbes elliptiques modulaires*, Compositio Math. **103** (1996), no. 3, 269–286.

[Aga99] A. Agashe, *On invisible elements of the Tate–Shafarevich group*, C. R. Acad. Sci. Paris Sér. I Math. **328** (1999), no. 5, 369–374.

[AL70] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[ARS] A. Agashe, K. Ribet, and W. A. Stein, *The modular degree, congruence primes, and multiplicity one* (2005).

[AS05] A. Agashe and W. A. Stein, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank zero*, Math. Comp. **74** (2005), no. 249, 455–484, with an appendix by J. Cremona and B. Mazur.

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

[CES03] B. Conrad, S. Edixhoven, and W. A. Stein, $J_1(p)$ *Has Connected Fibers*, Documenta Mathematica **8** (2003), 331–408.

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997,
`http://www.maths.nott.ac.uk/personal/jec/book/`.

[Cre] J. E. Cremona, *Tables of Elliptic Curves,*
`http://www.maths.nott.ac.uk/personal/jec/ftp/data/`.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993–1994), Amer. Math. Soc., Providence, RI, 1995, pp. 39–133.

[DR73] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349.

[Edi89] B. Edixhoven, *Stable models of modular curves and applications*, Thèse de doctorat à l'université d'Utrecht (1989),
`http://www.maths.univ-rennes1.fr/~edix/publications/ prschr.html/`.

[Edi91] B. Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Birkhäuser Boston, Boston, MA, 1991, pp. 25–39.

[FpS$^+$01] E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. Comp. **70** (2001), no. 236, 1675–1697 (electronic).

[GL01] Josep González and Joan-C. Lario, $\mathbf{Q}$*-curves and their Manin ideals*, Amer. J. Math. **123** (2001), no. 3, 475–503.

[Gro82] B. H. Gross, *On the conjecture of Birch and Swinnerton-Dyer for elliptic curves with complex multiplication*, Number theory related to Fermat's last theorem (Cambridge, Mass., 1981), Birkhäuser Boston, Mass., 1982, pp. 219–236.

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

[Joy05] A. Joyce, *The Manin constant of an optimal quotient of* $J_0(431)$, J. Number Theory **110** (2005), no. 2, 325–330.

[Kil02] L. J. P. Kilford, *Some non-Gorenstein Hecke algebras attached to spaces of modular forms*, J. Number Theory **97** (2002), no. 1, 157–164.

[KM85] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry.

[Man72] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[MR91] B. Mazur and K. A. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque (1991), no. 196–197, 6, 215–255 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[Mc91] W. G. McCallum, *Kolyvagin's work on Shafarevich–Tate groups*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 295–316.

[Shi73] G. Shimura, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544.

[Shi94] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[Sil94] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Springer-Verlag, New York, 1994.

[SW04] W. Stein and M. Watkins, *Modular parametrizations of Neumann–Setzer elliptic curves*, Int. Math. Res. Not. (2004), no. 27, 1395–1405.

[Ste89] G. Stevens, *Stickelberger elements and modular parametrizations of elliptic curves*, Invent. Math. **98** (1989), no. 1, 75–106.

[Vat05] V. Vatsal, *Multiplicative subgroups of $J_0(N)$ and applications to elliptic curves*, J. Inst. Math. Jussieu **4** (2005), no. 2, 281–316.

Amod Agashe
Department of Mathematics
Florida State University
208 Love Building Tallahassee
FL 32306 U.S.A.
E-mail: agashe@mail.math.fsu.edu

Kenneth A. Ribet
Department of Mathematics, M/C 3840
University of California
Berkeley, CA 94720-3840, USA.
E-mail: ribet@math.berkeley.edu

William A. Stein
Department of Mathematics
Harvard University
Cambridge, MA 02138
E-mail: was@math.harvard.edu

28 Average Ranks Of Elliptic Curves: Tension Between Data And Conjecture, with B. Bektemirov, B. Mazur, and M. Watkins

AVERAGE RANKS OF ELLIPTIC CURVES: TENSION BETWEEN DATA AND CONJECTURE

BAUR BEKTEMIROV, BARRY MAZUR, WILLIAM STEIN AND MARK WATKINS

Abstract. Rational points on elliptic curves are the gems of arithmetic: they are, to diophantine geometry, what units in rings of integers are to algebraic number theory, what algebraic cycles are to algebraic geometry. A rational point in just the right context, at one place in the theory, can inhibit and control—thanks to ideas of Kolyvagin [Kol88]—the existence of rational points and other mathematical structures elsewhere. Despite all that we know about these objects, the initial mystery and excitement that drew mathematicians to this arena in the first place remains in full force today.

We have a network of heuristics and conjectures regarding rational points, and we have massive data accumulated to exhibit instances of the phenomena. Generally, we would expect that our data support our conjectures; and if not, we lose faith in our conjectures. But here there is a somewhat more surprising interrelation between data and conjecture: they are not exactly in open conflict one with the other, but they are no great comfort to each other either. We discuss various aspects of this story, including recent heuristics and data that attempt to resolve this mystery. We shall try to convince the reader that, despite seeming discrepancy, data and conjecture are, in fact, in harmony.

1. Introduction

Suppose you are given an algebraic curve C defined, let us say, as the locus of zeroes of a polynomial $f(x, y)$ in two variables with rational coefficients. Suppose you are told that C has at least one rational point, i.e., there is a pair of rational numbers (a, b) such that $f(a, b) = 0$. How likely is it that C will have infinitely many rational points?

Such a question, on the one hand, clearly touches on a fundamental issue in diophantine geometry, and on the other, is somewhat meaningless until it is made more precise and appropriately organized. The question we have just asked has distinctly different features when considered for each of the three basic "types" of algebraic curves: curves of (geometric) genus 0, 1, and > 1. Curves of genus 0 possessing a rational point *always* have infinitely many rational points (an easy fact; indeed, even known to the ancient Greeks, since our curve can be written as a conic in this case); curves of genus > 1 never do (a hard fact; indeed a theorem of Faltings [Fal86], for which he received the Fields Medal).

This leaves curves of genus 1 as the unresolved, and thus most interesting, case of the problem we posed, since some elliptic curves, like

$$x^3 + y^3 = 1,$$

only have finitely many rational points (two, in this instance) and others, like

$$y^2 + y = x^3 - x,$$

have infinitely many, starting with $(0,0)$, $(1,0)$, $(-1,-1)$, $(2,-3)$, $(1/4,-5/8)$, $(6,14)$, $(-5/9,8/27)$, $(21/25,-69/125)$, $(-20/49,-435/343)$, ...

If we are to try to extract an actual number between 0 and 1 that will describe "the" probability that a curve of genus 1 possessing at least one rational point has infinitely many, we have to be precise about exactly which curves we want to count, and how we propose to "sort" them. Let us agree, then (with details later):

- to deal only with the smooth projective models of the curves of genus 1 possessing a rational point (these being precisely the *elliptic curves* defined over $\mathbb{Q}$),
- to count their isomorphism classes over $\mathbb{Q}$, and
- to list them in order of increasing conductor, banking on the theorem that tells us that there are only finitely many isomorphism classes of elliptic curves over $\mathbb{Q}$ with any given conductor.

We can now pose our question. Does

$$P(X) = \frac{\#\{\text{elliptic curves of conductor } \leq X \text{ with infinitely many rational points}\}}{\#\{\text{elliptic curves of conductor} \leq X\}}$$

converge as X tends to infinity, and if so, what is the limit

$$P = \lim_{X\to\infty} P(X)?$$

In this way we have made our initial question precise:

What is the probability P that an elliptic curve has infinitely many rational points?

It is extraordinary how much vacillation there has been in the past three decades, in the various guesses about the answer to this—clearly basic—question. The subject of this paper is to discuss aspects of this drama. Its see-saw history, involving a network of heuristics and conjectures and massive data that seemed not to offer much comfort to the conjecturers, comes in four parts.

(1) **The minimalist conjecture.** The "classical" Birch and Swinnerton-Dyer conjecture (see Section 2) suggests that the probability P described by our question is at least $1/2$. The reason for this is the (heuristic) phenomenon of *parity*: elliptic curves can be sorted into two classes, those of **even parity**, where the "sign in the functional equation of the L-function" is $+1$, and those of **odd parity**, where the "sign" is -1. The (conjectural) probability that an elliptic curve is of even parity is $1/2$, and the same—of course—for odd parity. A consequence of the Birch and Swinnerton-Dyer conjecture is that *all* elliptic curves of odd parity have infinitely many rational points. This is why no one doubts that the probability P described above is $\geq 1/2$.

It has long been a folk conjecture that P is *exactly* $1/2$—let us call this the **minimalist conjecture**. Given the Birch and Swinnerton-Dyer conjecture and the Parity Principle, an equivalent, and cleaner, way of stating it is as follows:

Conjecture 1.1. *An elliptic curve of even parity has probability zero of having infinitely many rational points.*

This minimalist conjecture might seem appealing purely on the grounds that rational points of elliptic curves are accidental gems of mathematics, and it is hard to imagine that there could be bulk occurrence of these precious accidents—or at least substantially more bulk than is already predicted.

It seems that one cannot find such a minimalist conjecture explicitly in the literature until very recently (see [Wat06] and Conjecture 3.4). Nevertheless, for some particular families of elliptic curves (the "quadratic twist" families—see Section 3.3 below) the conjecture is much older. Over a quarter of a century ago, Dorian Goldfeld conjectured that for any elliptic curve E, the probability

$$G(D) = \frac{\#\{\text{quadratic twists up to } D \text{ of } E \text{ with infinitely many rational points}\}}{\#\{\text{quadratic twists up to } D \text{ of } E\}}$$

has $G = 1/2$ as its limit as $D \to \infty$.

(2) **Contrary numerical data.** The next phase of our story involves the accumulation of numerical data regarding this probability P taken over the entirety of elliptic curves, and also over various selected families of elliptic curves. The short description of this data (but see the detailed discussion in the body of our article) is the following. Over every data set accumulated so far, about 2/3 (or sometimes more) of the curves in the families being considered have had infinitely many rational points, and rather flatly so over the range of conductors involved in the computations; these now include a large set (over 100 million curves) of elliptic curves of conductor $< 10^8$.

(3) **A gross heuristic, for special families.** To get the most precise results we change the data set, and restrict attention to the probability that a member of even parity of a *quadratic twist family* of elliptic curves has infinitely many rational points. As a refinement to Goldfeld's conjecture, Peter Sarnak gave a heuristic that predicts that among the first D members of such a quadratic twist family (essentially arranged in order of increasing conductor) the number of those with even parity and infinitely many rational points is caught between $D^{3/4-\epsilon}$ and $D^{3/4+\epsilon}$ for any positive ϵ and D sufficiently large. This guess, based on consideration of the size of Fourier coefficients of modular forms of half-integral weight, revived the minimalist conjecture: if Sarnak's estimate is correct, we would indeed have $G = 1/2$ in Goldfeld's conjecture, and even-parity members of a quadratic twist family would have probability zero of having infinitely many rational points.

At this point in our story, there is decided friction between accumulated data which suggests something like 2/3 as the probability for the general member to have infinitely many rational points, and a reasoned theoretical expectation, which suggests exactly 1/2 for that probability. Generally, the least we would expect of our data is that they either support our conjectures, or overthrow them. Here there was a somewhat more surprising interrelation between data and conjecture: a kind of truce between them: we believed our guesses, we believed the data, and acknowledged the apparent gap between them.

(4) **A refined heuristic, for special families.** More recently, another twist to this story has developed. The work of Katz and Sarnak [KaSa99] regarding symmetry groups of the analogous families of curves over function

fields[1] gave impetus to the random matrix theory calculations of Keating and Snaith [KeSn00] regarding moments of L-functions and their value distribution. This was then combined with a discretization process by Conrey, Keating, Rubinstein, and Snaith in [CKRS02] to give a more precise guess for the (asymptotic) number of even parity curves with infinitely many rational points in a given quadratic twist family. For example, for the quadratic twist family $y^2 = x^3 - d^2x$, the prediction is that among the first D members of this family, the number of those with even parity and infinitely many rational points is asymptotic to

$$F(D) = c \cdot D^{3/4} \log(D)^{11/8} \tag{1}$$

for some (positive) constant c.

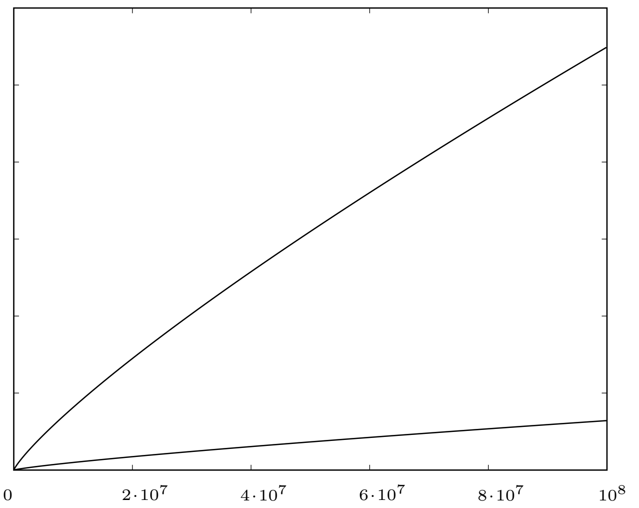


FIGURE 1. Plots of $D^{3/4}\log(D)^{11/8}$ (upper) and $\Delta^{19/24}(\log\Delta)^{3/8}$ (lower) up to 10^8

On the one hand, this is a sharpening of the prior heuristic, for $F(D)$ is comfortably sandwiched between $D^{3/4\pm\epsilon}$. On the other hand, we may be in for a surprise when we actually plot the graph of the function $F(D)$. See Figure 1. The striking aspect of the graph in Figure 1 is how "linear" it looks. Indeed, if $F(D)$ were replaced by a linear function with roughly the slope that appears in Figure 1, it would predict something closer to 2/3 than 1/2 for the proportion of curves in the family with infinitely many rational points.

Similarly—cf. Section 3.5 below—if we order all elliptic curves by discriminant, one of us (see [Wat06]) has conjectured that the number of even

[1] More recently, Kowalski [KowB] has used monodromy results of Katz to prove upper bounds for average ranks in the function field analogues.

parity elliptic curves with infinitely many rational points and absolute discriminant less than X is asymptotically given by

$$\Phi(X) \;=\; cX^{19/24}(\log X)^{3/8}.$$

See Figure 1.

Here, roughly speaking, is where the story is at present, as we will explain in detail in the body of this article. The curious last phase of it, focussing on *special families*, makes it seem now that data for these families is (a) more closely adhering to the refined guess than one might expect, even for relatively small values of the conductor, and (b) a refined guess predicts an asymptotic behavior that is far from linear, but within the currently attainable range is so close to linear, that the numerical evidence elucidating these phenomena (even the very large data sets that computers have amassed) seem indecisive when it comes to distinguishing convincingly between such gross questions as: is the probability closer to 1/2 or to 2/3?

It may very well be that until we actually prove our conjectures, no data that we can accumulate, however massive it may appear, will give even lukewarm comfort to the conjecturers.[2] This conflict raises the question of whether we as mathematicians may, at times, face a situation where the substance we study has one shape asymptotically, and yet all computational evidence elucidating this substance, even up to the very large numbers that computers today, or in our lifetime, can compute, seems consistent with the possibility that the data have a different asymptotic shape.

But, of course, our story will continue. We would hope for

- a refined heuristic that covers the full set of elliptic curves, and not just quadratic twist families,
- an extension of the numerical computation to conductors $< 10^{10}$, which is a range where we may begin to see some significant differences between the graph of $F(D)$ and a linear function,
- a conceptual understanding of how to obtain—by more unified means—this impressive bulk of rational points that we see occurring for even parity elliptic curves, at least for curves of "small" conductor.

We find it useful to compare our question *what is the probability that an elliptic curve has infinitely many points* with some of the other counting problems of current interest. Specifically, consider the problem of *counting quartic fields* and sorting them into classes corresponding to the isomorphy type of the Galois group of their Galois closure. We have to be exceedingly careful when choosing the coefficients of a degree 4 polynomial if we want a root of that polynomial to generate anything other than a field whose Galois group is S_4. Hilbert's irreducibility theorem provides corroboration of this with a proof that if you rank algebraic numbers of degree 4 by the size of the coefficients of their minimal polynomial (monic, over $\mathbb{Q}$) then 100% of them have Galois group S_4. But consider the problem of counting quartic

[2] We are reminded of the challenge of Shanks [Sha85, §69] regarding Carmichael numbers; with respect to the conjecture of Erdős that, for every $\epsilon > 0$, there are, for sufficiently large X, at least $X^{1-\epsilon}$ Carmichael numbers up to X, Shanks (essentially) noted that the data for small X did not remotely conform to this, and proposed giving an explicit X for which there were at least (say) $\sqrt{X}$ Carmichael numbers up to X, suspecting that exhibiting such an X would be much beyond the capabilities of computers.

fields (rather than the algebraic numbers that generate them) listed by the size (absolute value) of their discriminant. Counting field extensions of a given field whose Galois closure has its Galois group of a particular isomorphy type has been the subject of a number of precise conjectures (initially [CDO], and then successively refined in [Mal02, Mal04]). Bhargava's remarkable paper [Bha05], which is further evidence for these conjectures, proves that when we count quartic fields, nested by absolute discriminant, we do *not* get that 100% of them have Galois group S_4.

Bhargava regards the problem of counting quartic fields as a problem purely in the Geometry of Numbers, and *proves* the following theorem:

Theorem 1.2. *[Bhargava]. When ordered by absolute discriminant,* a positive proportion *(approximately 0.17111) of quartic fields have associated Galois group D_4 (the dihedral group). The remaining approximately 0.82889 of quartic fields have Galois group S_4, and the other three transitive subgroups occur with probability 0 asymptotically.*

It should be noted that these are Bhargava's percentages when counting fields up to isomorphism; when working in a fixed algebraic closure of the rationals, the percentages are not the same.

We would be more than delighted to see unconditional results of this precision established for questions such as the one motivating this survey article.

Acknowledgment. We thank Armand Brumer, Frank Calegari, Noam Elkies and Oisin McGuinness for stimulating conversations. We used PARI [ABCRZ] and SAGE [SJ05] to compute and analyze the data, and matplotlib [Mpl05] to draw the graphs. We thank Bob Guralnik for references to work about distributions of Galois groups. Similar surveys to this are those of Rubin and Silverberg [RS02] and Kowalski [KowA], the latter of which also relates random matrix theory to the theory of elliptic curves, and then discusses questions related to the inability of reconciling experimental data with theoretical asymptotics, particularly with respect to the work of Heath-Brown [HB93] on the 2-Selmer rank of the congruent number curves.

2. Elliptic Curves

An *elliptic curve* E over $\mathbb{Q}$ is a projective nonsingular curve defined as the projective closure of the zero locus of an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{2}$$

with the a_i in $\mathbb{Q}$. The set $E(\mathbb{Q})$ of rational points on E is equipped with an abelian group structure (see [Sil92]).

Via completing the square in the y-variable, translating to eliminate the quadratic x-term, and then re-scaling, we find that the equation (2) is rationally equivalent to exactly one of the form

$$y^2 = x^3 - 27c_4x - 54c_6, \tag{3}$$

with $c_4, c_6, \Delta = (c_4^3 - c_6^2)/1728 \in \mathbb{Z}$ and for which there is no prime p with $p^4 \mid c_4$ and $p^{12} \mid \Delta$. We call Δ the *minimal discriminant* of E. (For example, the minimal discriminant of the curve $y^2 + y = x^3 - x$ mentioned in Section 1 is $\Delta = 37$; also $c_4 = 48$ and $c_6 = -216$ for this curve.)

The *conductor* of an elliptic curve E over $\mathbb{Q}$ is a positive integer $N = N_E$ that is a measure of the nature of the reduction of the elliptic curve modulo the prime divisors of Δ. For example, a prime $p \geq 5$ divides the conductor N only if there is no way of finding another defining equation (2) of E so that when reduced modulo p we obtain an equation over the field $\mathbb{F}_p$ without multiple roots; the maximal power of such a prime p dividing N is 2 and whether it is 1 or 2 is determined by the nature of the *best* reduction of E modulo p, i.e., whether its defining cubic polynomial has a double or a triple root modulo p. There is a slightly more involved, but elementary, recipe to give the power of the primes 2 and 3 dividing the conductor (see [Tat75]).

Mordell proved in 1922 (see [Mor22]) that the *Mordell-Weil group* $E(\mathbb{Q})$ of rational points on E is a finitely generated abelian group, so $E(\mathbb{Q}) \approx \mathbb{Z}^r \oplus E(\mathbb{Q})_{\mathrm{tor}}$. The integer r is the *rank* of E, and is the main statistic that we will discuss below; in contrast, the torsion group is rather well-understood, and is thus of less interest.

Let Δ be the minimal discriminant of E. The *L-function* $L(E,s)$ of E is a Dirichlet series given by a simple recipe in terms of the number of points N_p of the reduction of E over $\mathbb{F}_p$ for all primes p. Specifically,

(4)
$$L(E,s) = \prod_{p \nmid \Delta} \frac{1}{1-(1+p-N_p)p^{-s}+p^{1-2s}} \cdot \prod_{p \mid \Delta} \frac{1}{1-(1+p-N_p)p^{-s}} = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

The integers a_n are defined by expanding the Euler product; e.g., $a_p = p+1-N_p$ and $a_{p^2} = a_p^2 - p$ when $p \nmid \Delta$, etc. As an example, if E is $y^2 + y = x^3 - x$ then

$$L(E,s) = 1 - \frac{2}{2^s} - \frac{3}{3^s} + \frac{2}{4^s} - \frac{2}{5^s} + \frac{6}{6^s} - \frac{1}{7^s} + \frac{6}{9^s} + \frac{4}{10^s} - \frac{5}{11^s} + \cdots .$$

For any elliptic curve E, the celebrated papers of Wiles [Wil95] and others [BCDT01] imply that $L(E,s)$ extends to an entire analytic function on the complex plane. Moreover these results imply that the completed L-function $\Lambda(E,s) = N^{s/2} \cdot (2\pi)^{-s} \cdot \Gamma(s) \cdot L(E,s)$ satisfies the functional equation

$$\Lambda(E,s) = u_E \cdot \Lambda(E, 2-s),$$

where u_E is either -1 or 1, and is called the *sign in the functional equation for E*. Note that $u_E = 1$ if and only if $L(E,s)$ vanishes to even order at $s=1$.

The classical Birch and Swinnerton-Dyer Conjecture [BSD] asserts that the order of vanishing of $\Lambda(E,s)$ at the point $s=1$ is equal to the rank of the Mordell-Weil group $E(\mathbf{Q})$. In the data regarding *rank* that we will be reporting below, at times the Mordell-Weil rank r has been computed directly by finding r rational points of E that are linearly independent and span a subgroup of finite index in $E(\mathbf{Q})$ and we will refer to this r as the *arithmetic rank* of E. At times, however, what is computed is the apparent order of vanishing of $L(E,s)$ at $s=1$; we refer to this order of vanishing as the *analytic rank* of E. The BSD conjecture asserts that the ranks are in fact equal. We say a curve has *even parity* if the analytic rank is even, and *odd parity* if it is odd.

We now state the refined BSD conjecture for curves of rank 0. When E is given by (3), the real period Ω_{re} is, up to easily determined factors of 2 and 3, equal to the integral $\int_{E(\mathbb{R})} dx/y$. For a prime p, the Tamagawa number Ω_p is the index in $E(\mathbb{Q}_p)$ of the subgroup of p-adic points that reduce to a nonsingular point in $E(\mathbb{F}_p)$.

Conjecture 2.1. *[Birch and Swinnerton-Dyer]. If $L(E,1) \neq 0$, then*

$$L(E,1) = \frac{\Omega(E) \cdot \#\text{Ш}(E)}{\#E(\mathbb{Q})^2_{\text{tor}}}, \tag{5}$$

where Ш(E)*, the* Shafarevich-Tate group of E*, is a certain (mysterious) group associated to* E *(it measures the failure of a local-global principle), and*

$$\Omega(E) = \Omega_{\text{re}} \cdot \prod_p \Omega_p.$$

Since $L(E,1) \neq 0$, the group Ш(E) is known to be finite [Kol88] of order a perfect square [Cas62].

For any $r = 0, 1, 2, \ldots$ the question we now may ask is: what percentage of elliptic curves (nested according to size of conductor) have rank r? More correctly, we should ask: do these percentages exist, and if so what are they?

3. Conjectures

One fairly firm anchor in the study of elliptic curves is a principle that goes under the heading of parity. This principle is still only conjectural, but is amply confirmed numerically in our accumulated data, and we also have theoretical reasons to believe it.[3] The parity principle is that 50% of the members of any of the sets of elliptic curves we will be considering have even parity, and 50% have odd parity (under reasonable orderings). So by BSD, 50% should have even rank, and 50% should have odd rank.

In general terms, the minimalist principle proclaims that from the rough viewpoint of percentages, there are as few rational points on elliptic curves as is possible, given the constraint of the parity principle. That is, 50% of the members of any of the families of elliptic curves we will be considering have rank $r = 0$, and 50% have rank $r = 1$, and the remaining ranks $r \geq 2$ account for 0% of the family.

As one thing or another things comes to light in the subject, the minimalist position is sometimes favored, and sometimes not. For certain special families of elliptic curves this minimalist conjecture has long been in print, and has had a wild ride in terms of its being believed, and doubted.

3.1. The form of the conjectures.

There are two types of asymptotic conjectures that we encounter in discussions regarding rank statistics. The first we might call a *rough conjecture* where it is asserted, or conjectured, for a certain collection $\mathcal{F}(x)$ of items indexed by a variable x that there is an *exponent* a and a function $x(\epsilon)$ such that the cardinality of $\mathcal{F}(x)$ is bounded above by $x^{a+\epsilon}$ and below by $x^{a-\epsilon}$ for any positive ϵ and any $x \geq x(\epsilon)$.

We also will be discussing *fine conjectures* where such collections $\mathcal{F}(x)$ will be conjectured to have asymptotic estimates of the form

$$\#\mathcal{F}(x) \sim x^a \cdot (\log x)^b \cdot c,$$

for constants a, b, c; the delicacy, of course, of these constants is inversely related to their alphabetical order.

What seems to be a pattern is that the exponent a, appearing in both rough and fine versions, in specific contexts under discussion, can usually be guessed by more

[3] In particular, the sign of the functional equation is a product of local signs for primes $p|\Delta$, each of which is ±1 with equal proportion. See the work of Helfgott [Hel04] for the latest results.

old-fashioned heuristics. But—at present, at least—the on-going work regarding random matrix eigenvalues is the only source of heuristics that lead us to formulate specific "fine conjectures" regarding ranks, and specifically for guesses regarding the exponent b of the $\log x$ term. The fact that the graphs of some of the specific concoctions of the form $x^a \cdot (\log x)^b \cdot c$ predicted by random matrix statistics can look deceivingly like x^1 even though $a < 1$ (and for a significant range of the variable x) is one of the curiosities of our story.

3.2. **Random matrix statistics.** Originally developed in mathematical physics, random matrix theory [Meh04] has now found many applications in number theory, the first being the oft-told story [Dy] of Dyson's remark to Montgomery regarding the pair-correlation of zeros of the Riemann ζ-function. Based on substantial numerical evidence, random matrix theory appears to give reasonable models for the distribution of L-values in families, though the issue of what constitutes a proper "family" can be delicate. The work of Katz and Sarnak [KaSa99] regarding families of curves over function fields implies that for quadratic twists of even parity, we should expect orthogonal symmetry with even parity. Though we have no function field analogue when considering all curves of even parity, we still brazenly assume (largely from looking at the sign in the functional equation) that the symmetry is orthogonal with even parity. What this means is that we want to model properties of the L-function via random matrices taken from $\mathrm{SO}(2M)$ with respect to Haar measure, for an appropriate value of M.[4] We suspect that the L-value distribution is approximately given by the distribution of the evaluations at 1 of the characteristic polynomials of our random matrices. In the large, this distribution is determined entirely by the symmetry type, while finer considerations are distinguished via arithmetic considerations.

Via the moment conjectures [KeSn00] of random matrix theory and then using Mellin inversion, we expect that (for some constant $c > 0$)

$$\mathrm{Prob}[L(E,1) < t] \sim ct^{1/2}(\log N)^{3/8} \qquad \text{as} \qquad t \to 0, \tag{6}$$

when the curves E are taken from a suitable family.

3.3. **Conjectures about twist families.** Let E be an elliptic curve over $\mathbb{Q}$ defined by an equation $y^2 = x^3 + ax + b$. The *quadratic twist* E_d of E by a nonzero integer d is the elliptic curve defined by $y^2 = x^3 + ad^2x + bd^3$. The twist E_d is isomorphic to E over the field $\mathbb{Q}(\sqrt{d})$, and (when d is a fundamental discriminant relatively prime to N_E) the conductor of E_d is $d^2 \cdot N_E$.

Conjecture 3.1. *[Goldfeld,* [Gol79]*]. The average rank of the curves E_d is $\frac{1}{2}$, in the sense that*

$$\lim_{D\to\infty} \frac{\sum_{|d|<D} \mathrm{rank}(E_d)}{\#\{d : |d| < D\}} = \frac{1}{2}.$$

(Here the integers d are fundamental discriminants.)

There are many conditional and unconditional results regarding Goldfeld's conjecture. For a survey, see the papers of Rubin and Silverberg [RS02, Sil01].

The values $L(E_d, 1)$ of quadratic twists E_d of a given curve E essentially appear in a single object, as the coefficients (weighted by the real period and Tamagawa

[4] Here we wish the mean density of zeros of the L-functions to match the mean density of eigenvalues of our matrices, and so, as in [KeSn00], we should take $2M \approx 2 \log N$.

numbers) of an integral modular form g_E of weight $3/2$ (this follows from work of Waldspurger, see [Wal81]). In particular, for many d, we have that $L(E_d, 1) = 0$ precisely when the dth (or $-d$th, depending on the case) coefficient of g_E is zero. This object g_E does not give us values of $L(E_d, 1)$ for all d, but does provide a large proportion of them. The Ramunajan conjecture for modular forms implies that the coefficients of g_E should be bounded by about $|d|^{1/4}$, and so if we assume a coefficient distribution that is somewhat uniform, we approximate the count $F(D)$ of quadratic twists up to D with even parity that have $L(E_d, 1) = 0$ by $\sum_{|d|<D} 1/|d|^{1/4}$. Sarnak's rough heuristic asserts that this count lies between $D^{3/4-\varepsilon}$ and $D^{3/4+\varepsilon}$. Using random matrix theory, the paper [CKRS06] gets the refined heuristic that

$$F(D) \sim D^{3/4} \cdot (\log D)^b \cdot c,$$

where there are four possibilities for b (depending on the Galois group of the cubic polynomial $x^3 - 27c_4x - 54c_6$), and c is still mysterious.

In [CKRS06], Rubinstein used weight 3/2 forms to give data about $L(E_d, 1)$ for over 2000 elliptic curves E. For each of these he computed $L(E_d, 1)$ for a substantial subset of the quadratic twists by fundamental discriminants d with $|d| < 10^8$. (For example, for the curve E given by $y^2 + y = x^3 - x^2$ of conductor 11, the only twist E_{-d} of even parity with $L(E_{-d}, 1) = 0$ for $3 < d < 91$ is $d = 47$.) The data of Rubinstein agree fairly well with predictions such as (1).

It is possible, however, to ameliorate the effects of b and c (and the 3/4-exponent for that matter) via the ratio conjecture of [CKRS02]. Fix an elliptic curve E and a modulus q, prime for simplicity. Consider the d with $\gcd(q, d) = 1$ for which E_d has even parity and $L(E_d, 1) = 0$, and divide these into two classes depending on whether d is a square modulo q. The ratio conjecture asserts that the (asymptotic) ratio of the sizes of these two classes is $\left(\frac{q+1+a_q}{q+1-a_q}\right)^{-1/2}$, where the exponent $-1/2$ comes from the arguments leading to (6). In essence, the d's that are squares should give $c_S X^{3/4}(\log X)^{b_E}$ while those that are not should yield $c_N X^{3/4}(\log X)^{b_E}$, and [CKRS02] predicts c_S/c_N via a clever methodology. The data match this prediction fairly well,[5] especially for $a_q = 0$, when the convergence is quite rapid.

We can also consider other twist families. For example, Kramarz and Zagier [ZK87] considered cubic twists $x^3 + y^3 = m$ of the Fermat cubic[6] $x^3 + y^3 = 1$ and found in their data that 23.3% of the curves with even parity have rank at least 2, and 2.2% of those with odd parity have rank at least 3. One of the authors of the present article [Wat04] and independently Fermigier (unpublished) have followed up on these computations. Also, Patricia Quattrini (Universidad de Buenos Aires) as part of her thesis work (to appear in Experimental Mathematics) did some extensive calculations of the analytic rank for the curves $y^2 = x^3 - nx$. As in the Kramarz-Zagier case, the percentage of curves with analytic rank ≥ 2 was in the 20% range but did seem to be going down. Similar computations [DFK04] have also been undertaken for twists by other (complex) Dirichlet characters, which are related to ranks over number fields. Finally, Fermigier [Fer96] investigated specializations of various (about 100) elliptic curves defined over $\mathbb{Q}(t)$, and found that

[5] In [CPRW] a secondary term is computed, and the fit to the data becomes even better. The paper [Wat04] notes similar data for cubic twists, while [CRSW] analyses the data of Elkies for the congruent number curve in the odd parity case.

[6] Note that this is rationally isomorphic to the elliptic curve in the form (3) given by the equation $Y^2 = X^3 - 54 \cdot 5832$ via the map $(X, Y) = \big(108/(x + y), 972(y - x)/(y + x)\big)$.

typically 10-20% of the specializations had excess rank that could not be explained simply from parity.

3.4. **Conjectures when counting all elliptic curves.** Before we can count curves with even parity and infinitely many points, we might first take a step back, and just try to count curves. Though before we ordered curves by conductor, when deriving heuristics it is often easier to sort by discriminant. Indeed, Brumer and McGuinness [BM90, §5] state a heuristic estimate for the number of minimal discriminants of elliptic curves up to a given bound:

Conjecture 3.2. *[Brumer-McGuinness]. We have the following estimates for the number of positive or negative minimal discriminants of elliptic curves of absolute value at most X (respectively):*

$$A_{\pm}(X) \sim \frac{\alpha_{\pm}}{\zeta(10)} X^{5/6}$$

where $\alpha_+ = 0.4206\ldots$ and $\alpha_- = \sqrt{3}\alpha_+ = 0.7285\ldots$ are given by

$$\alpha_{\pm} = \frac{\sqrt{3}}{10} \int_{\pm 1}^{\infty} \frac{du}{\sqrt{u^3 \mp 1}}.$$

Brumer and McGuinness say little about their derivation of this heuristic, but remark that it suggests a heuristic for prime discriminants that matches very well with their data. We can derive their heuristic by counting lattice points in the (c_4, c_6)-plane, restricting to congruence classes modulo powers of 2 and 3 to ensure that Δ is integral. Because $\Delta = (c_4^3 - c_6^2)/1728$, we heuristically have that $A_+(X)$ is proportional to the area of the region $0 < c_4^3 - c_6^2 < 1728X$, and similarly with $A_-(X)$. This gives $\alpha_{\pm} X^{5/6}$; the extra factor of $\zeta(10)$ comes about since we need (for $p \geq 5$, and similarly for $p = 2, 3$) to eliminate (c_4, c_6) pairs with $p^4 | c_4$ and $p^6 | c_6$. For a more complete derivation of the value of $\alpha_{\pm}$ see [Wat06].

We expect that half of these curves have even parity. Now we wish to estimate how many of the curves with even parity have $L(E, 1) = 0$.

3.5. **Rank conjectures for all curves.** To make use of the heuristic (6), we introduce a discretization process. We want to connect $L(E, 1)$ with the *winding number* $W = W(E) = |L(E, 1)/\Omega_{\mathrm{re}}|$ (see [MSD74, §2.2]),[7] and measure the likelihood that W is 0. Ignoring torsion (so that W is an integer) we are trying to estimate the probability that $L(E, 1) < \Omega(E)$. If we consider *only* elliptic curves for which $\Omega(E)$ lies in a fixed interval $c_1 < \Omega(E) < c_2$ then we get a neat estimate of this probability. So this line of reasoning leads one to try to deal with the statistics of the invariant $\Omega(E)$ for varying E.

Next, we simplify matters by restricting to curves with prime positive discriminant (and even parity). Three nice things about these curves are that (except for a sparse subset): all have trivial torsion; all have $\Omega_p = 1$ for all (finite) primes p; and all have that $N = \Delta$. The idea of our discretization is that W can only take on integral values (note that when $W \neq 0$, Conjecture 2.1 implies that $W = \#Ш(E)$, which is a perfect square, but we will not use this). Thus, in terms of our probability distribution of L-values, we get that $L(E, 1) < \Omega_{\mathrm{re}}$ if and only if $L(E, 1) = 0$;

[7] We could relate W to the Birch and Swinnerton-Dyer conjecture, but the (topological) winding number interpretation is rigorous and sufficient for our needs.

this is because

$$0 \leq W = \left| \frac{L(E,1)}{\Omega_{\mathrm{re}}} \right| < 1$$

and W is an integer.

Putting $t = \Omega_{\mathrm{re}}$ and $N = \Delta$ in (6) we get:

Heuristic 3.3. *A curve with positive prime discriminant and even parity has infinitely many points with probability* $c\Omega_{\mathrm{re}}^{1/2}(\log \Delta)^{3/8}$.

Using the above, the number $B(X)$ of such curves up to X with even parity and infinitely many points is estimated by integrating $\int\int c\Omega_{\mathrm{re}}^{1/2}(\log\Delta)^{3/8-1}du_4du_6$ over the region $|u_4^3 - u_6^2| < 1728X$, where the "-1" in the exponent of $\log\Delta$ comes about from the prime number theorem. Also, the integral makes sense because Ω_{re} and Δ are smooth functions of c_4 and c_6, that is, we can define Ω_{re} and Δ for c_4 and c_6 that are not necessarily integral (or even rational). So, similar to the above discussion of the Brumer-McGuinness heuristic, we have replaced a (weighted) lattice-point problem with the area of a region in a plane, weighted by a factor depending on the real period and the discriminant (and congruence restrictions as before).

We expect that the typical size[8] of the real period Ω_{re} is $1/|\Delta|^{1/12}$, and so, from the above heuristic,[9] we thus get a crude estimate that $B(X)$ is of size $X^{19/24}$.

The preprint [Wat06] handles more of the details, and considers all curves, not only those with prime discriminant. Indeed, if $F(X)$ is the number of elliptic curves E with even parity and $L(E,1) = 0$ and $|\Delta| \leq X$ then [Wat06] predicts

$$F(X) \sim c_1 X^{19/24}(\log X)^{3/8},$$

with a computable[10] positive constant c_1.

In any case, since we expect $cX^{5/6}$ curves with $|\Delta| \leq X$, this heuristic says that 100% of the even parity curves have rank 0.

The best known results (conditional on a Generalized Riemann Hypothesis and a Parity Principle analogue) on nonvanishing of even parity L-functions appear in the work of Young [YouB], and results about average (analytic) ranks and their relation to random matrix theory appear in [YouA].

3.6. **Ordering by conductor.** The predictions become more difficult to derive when we order by conductor instead of discriminant, as this introduces arithmetic considerations related to the ABC-conjecture (see [GT02]) in the accounting. Even giving a heuristic for the number $C(X)$ of curves of conductor less than X is nontrivial. The preprint [Wat06] asserts heuristic asymptotics of $c_2 X^{5/6}$ for $C(X)$ and similarly $c_3 X^{19/24}(\log X)^{3/8}$ for the number of rank 2 curves with conductor less than X. However, Cremona's data (see below) might suggest linear growth

[8] This is also an upper bound; the ABC conjecture says Ω_{re} is never much smaller than $1/|\Delta|^{1/2}$.

[9] Note that random matrix theory is largely used to determine the power of logarithm in this heuristic. The cruder estimate of $X^{19/24}$ can alternatively be obtained by assuming that the winding number is a random square integer of size up to $1/\Omega_{\mathrm{re}}$ (this is similar to Sarnak's heuristic); indeed this was probably known to Brumer and McGuinness, as they conclude their paper with

> *While our data may seem massive,* $N = 10^8$ *is not sufficient to distinguish growth laws of* $\log\log N$, $N^{1/12}$ *or* $N^{1/24}$ *from constants. So we have to be cautious in formulating conjectures based on the numerical evidence.*

[10] Our imprecise discretization might make the computed value of c_1 not too relevant.

for $C(X)$. In any event, in all cases we expect that 100% of the even parity curves have rank 0. Despite our lack of numerical confirmation, we label these guesses as "conjectures":

Conjecture 3.4. *The number of even parity elliptic curves with infinitely many rational points and absolute discriminant less than X is asymptotically given by $c_1 X^{19/24}(\log X)^{3/8}$ for some positive computable constant c_1 as $X \to \infty$. If we replace absolute discriminant by conductor, we get an asymptotic of $c_3 X^{19/24}(\log X)^{3/8}$. In particular, asymptotically almost all elliptic curves with even parity have finitely many rational points.*

See [Wat06] for more details.

4. Data

> The opinion had been expressed that, in general, an elliptic curve might tend to have the smallest possible rank, namely 0 or 1, compatible with the rank parity predictions of Birch and Swinnerton-Dyer. We present evidence that this may not be the case. [...] This proportion of rank 2 curves seemed too large to conform to the conventional wisdom. – Brumer and McGuinness [BM90]

In [BM90], Brumer and McGuinness considered over 310000 curves of prime conductor $\leq 10^8$. In this section we discuss extensions of their data, and answer in the affirmative that there is a similar large proportion of rank 2 curves for composite conductor $\leq 10^8$, and for prime conductor $\leq 10^{10}$. More precisely, we consider 136832795 curves of all conductors $\leq 10^8$, and 11378911 curves of prime conductor $\leq 10^{10}$. The results of the rank computation we describe are similar to those of Brumer and McGuinness, which appear to suggest that if one orders all elliptic curves over $\mathbb{Q}$ by conductor, then the average rank is bigger than 0.5. However, as discussed above, we conjecture that the average rank is 0.5.

4.1. Brumer-McGuinness.

In [BM90], Brumer and McGuinness found, by thousands of hours of computer search, 311219 curves of prime conductor $\leq 10^8$. For 310716 of these curves they computed the probable rank by a combination of point searches and computation of apparent order of vanishing of L-functions. Table 1 (expanded from [BM90]) summarizes the rank distribution that they found.[11]

Table 1. Brumer-McGuinness Rank Distribution

Rank	0	1	2	3	4	5
$\Delta > 0$	31748	51871	24706	5267	377	0
$\Delta < 0$	61589	91321	36811	6594	427	5
Total # Curves	93337	143192	61517	11861	804	5
Proportion	0.300	0.461	0.198	0.038	0.0026	0.00002
Proportion $\Delta > 0$	0.279	0.455	0.217	0.046	0.0033	0.00000
Proportion $\Delta < 0$	0.313	0.464	0.187	0.034	0.0022	0.00003

[11] Some of their counts were computed incorrectly (for instance, they only used 4000 terms of the L-series, and thus mis-identified 11 curves of rank 0 as having rank 2), but this has little influence on the overall statistics.

In Table 1, note that curves with $\Delta > 0$ are more likely to have large rank. Let $r_\varepsilon(X)$ be the average rank of elliptic curves in [BM90] with conductor at most X and discriminant sign ε. They observe that in their data, r_+ climbs to 1.04 and r_- climbs to 0.94, and they remark that *"An interesting phenomenon was the systematic influence of the discriminant sign on all aspects of the arithmetic of the curve."* The more extensive computations do *not* always find this to be the case; see, in particular, Figure 3 below, where the graphs split by discriminant cross.

4.2. **The Stein-Watkins Database.** Brumer and McGuinness fixed the a_1, a_2, a_3 invariants (12 total possibilities, as (2) can be modified first to be integral, and then to ensure that $a_1, a_3 \in \{0, 1\}$ and $|a_2| \leq 1$) and then searched for a_4 and a_6 that made $|\Delta|$ small. Stein and Watkins [SW02] broke the c_4 and c_6 invariants into congruence classes, and then found small solutions to $c_4^3 - c_6^2 = 1728\Delta$, with c_4, c_6 minimal in the sense of (3). There is little theoretical advantage in this approach; more computing power and disk space were the main advances in [SW02]. Stein and Watkins searched for curves with prime conductor up to 10^{10}, and for composite conductor chose $|\Delta| < 10^{12}$ and $N \leq 10^8$ as search bounds, and then included isogenous curves and twists (with $N \leq 10^8$) of the curves they found.

4.3. **Completeness of the Databases.** Note that neither the method of Brumer-McGuinness nor Stein-Watkins is guaranteed to find all curves of prime (absolute) discriminant up to a given bound (indeed, it is more likely that they miss a few curves), but we think that their data sets are reasonable surrogates, and should exhibit validity when compared to the predictions of the theoretical model.

For curves of composite conductor, the Stein-Watkins database is much more likely to miss curves. Here the comparison is to the data set of Cremona [Cre], who used the algorithms of [Cre97] and the modularity theorem of [BCDT01] to find *every* elliptic curve of conductor up to 120000. Cremona found 782493 curves up to conductor 120000. In the Stein-Watkins computation, they found 614442 curves of conductor up to 120000, so they found over 78.5% of the curves. The first case in which Cremona has a curve and Stein-Watkins do not is the curve $y^2 + xy + y = x^3 - 7705x + 1226492$ of conductor 174, which has discriminant $-621261297432576 = -2^{11} \cdot 3^{21} \cdot 29$, whose absolute value is substantially larger than 10^{12}. The conductors up to 500 where they miss curves are

$$174, 222, 273, 291, 330, 354, 357, 390, 420, 442, 462, 493.$$

Figure 2 shows the proportion of the number of curves in the Stein-Watkins database to the number of curves in Cremona's database, as a function of the conductor.

The rank distribution of Cremona's curves is given in Table 2. The average rank for Cremona's curves is about 0.688. This is smaller than the average rank in other data sets we consider (and is probably explainable via the real period considerations of the last section), but we prefer to highlight the results from other data sets.

Table 2. Rank Distribution of All Curves of Conductor $\leq$ 120000

Rank	0	1	2	3
Proportion	0.404	0.505	0.090	0.001
Proportion $\Delta > 0$	0.408	0.503	0.087	0.001
Proportion $\Delta < 0$	0.401	0.506	0.092	0.001

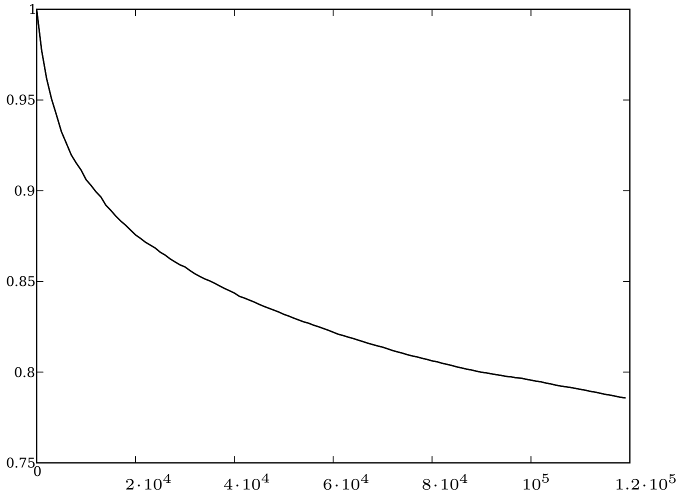


Figure 2. Proportion of Cremona's Curves obtained by Stein and Watkins for $N \leq 120000$

As noted above, when ordering by conductor there is presently no consensus guesstimate for the number of curves up to X. Cremona has commented that there is approximately linear growth in the number of curves of conductor less than 120000, and extrapolating this gives a prediction close to 650 million for the number of curves with $N \leq 10^8$.

D. J. Bernstein suggested that we try to quantify the completeness of the Stein-Watkins database by considering what percentage of Cremona's curves would be obtained by using their search methods with a smaller discriminant bound. That is, for a parameter B, if we find all curves with $N \leq B^2$, $|\Delta| \leq B^3$, and $c_4 \leq 100 \cdot (12B)^2$, and then take all isogenous curves and twists of these with conductor less than B^2, what percentage of Cremona's curves do we obtain? With $B = 300$, we get 246532 curves, while Cremona has 592519 curves of conductor $\leq 90000 = 300^2$, so we get about 42%. Applying this percentage to the Stein-Watkins database with $B = 10^4$, this would suggest that there are about 325 million elliptic curves with conductor less than 10^8. So the two guesses differ by a factor of two, exemplifying our ignorance on so basic an issue.

5. Average Ranks: Graphs of Data

This section contains graphs that at a glance suggest that the minimalist principle is contradicted by the data for curves of conductor $\leq 10^8$; indeed, particularly in Figure 3, we see that *the average rank is increasing!* However, for prime conductor $\leq 10^{10}$ the average rank drops, though only slightly from 0.978 to 0.964. With some imagination, the distribution of rank for prime conductor might appear to support the minimalist conjecture that the average rank is 0.5. Table 3 gives the average rank for various collections of curves that are described in more detail elsewhere in this paper and section.

TABLE 3. Average Ranks

Cremona's curves of conductor ≤ 120000	0.688
All Stein-Watkins curves of conductor $\leq 10^8$	0.865
Brumer-McGuinness curves of prime conductor $\leq 10^8$	0.982
Stein-Watkins curves of prime conductor $\leq 10^{10}$	0.964
Selected curves of prime conductor near 10^{14} with $\Delta < 0$	0.869
Selected curves of prime conductor near 10^{14} with $\Delta > 0$	0.938

In this section when we write elliptic curves with property P, we mean elliptic curves in the Stein-Watkins database with property P.

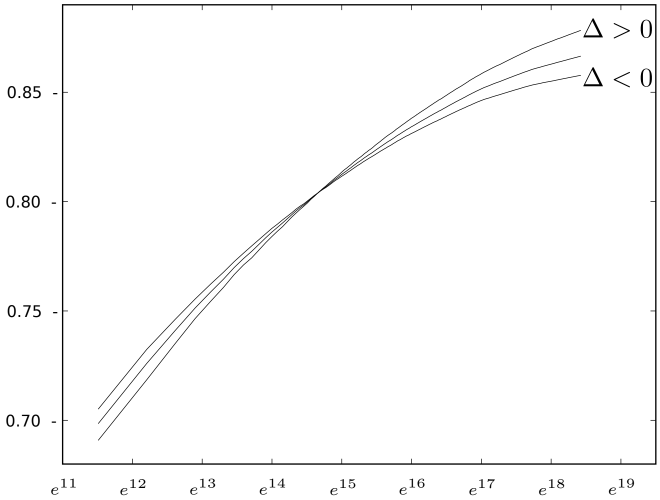


FIGURE 3. Average Rank of Stein-Watkins Curves of Conductor $\leq 10^8$

5.1. **Curves Ordered By Conductor.** The average rank of all Stein-Watkins curves with conductor $\leq 10^8$ is about 0.87. Figure 3 gives the average rank as a function of log of the conductor, and also the average rank for curves of positive and negative discriminant. We created this graph by computing the average rank of curves of conductor up to $n \cdot 10^5$ for $1 \leq n \leq 1000$. Figure 4 graphs the proportion of curves with each rank 0, 1, 2, and 3, as a function of log of the conductor, all on a single graph. The overall rank proportions are in Table 4.

TABLE 4. Rank Distribution for Stein-Watkins Curves with $N \leq 10^8$

Rank	0	1	2	3	≥ 4
Proportion	0.336	0.482	0.163	0.019	0.000
Proportion $\Delta > 0$	0.331	0.480	0.168	0.020	0.000
Proportion $\Delta < 0$	0.339	0.482	0.160	0.018	0.000

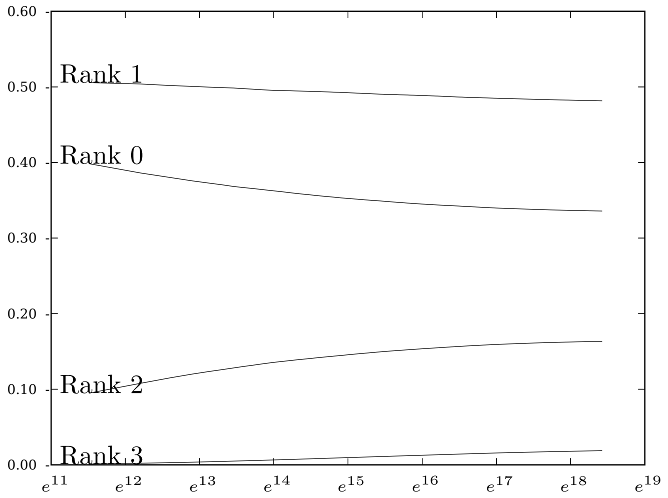


FIGURE 4. Rank Distribution of Stein-Watkins Curves with $N \leq 10^8$

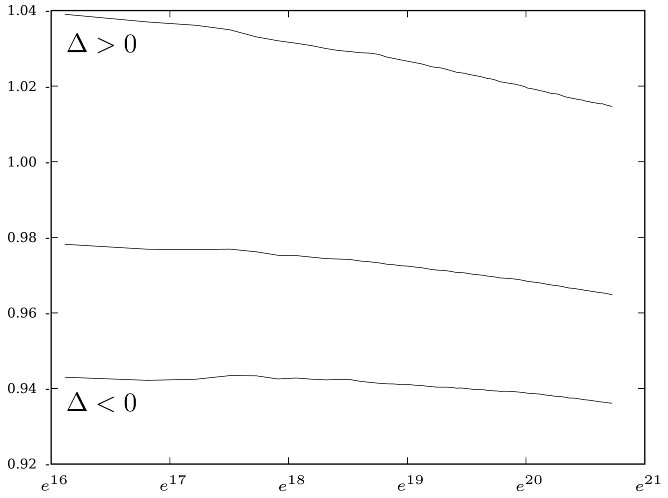


FIGURE 5. Average Rank of Curves with prime $N \leq 10^{10}$

5.2. **Prime Conductor Curves.** The average rank for the curves of prime conductor $\leq 10^{10}$ is about 0.964; see Table 5 for the rank distribution. Figure 5 plots the average rank of curves of prime conductor $\leq 10^{10}$ as a function of log of the conductor. Note that here the average ranks are decreasing, unlike in Figure 3.

5.2.1. *An experiment.* The data of [BM90] and [SW02] for curves of prime conductor up to 10^8 and 10^{10} show very little drop in the observed average rank.

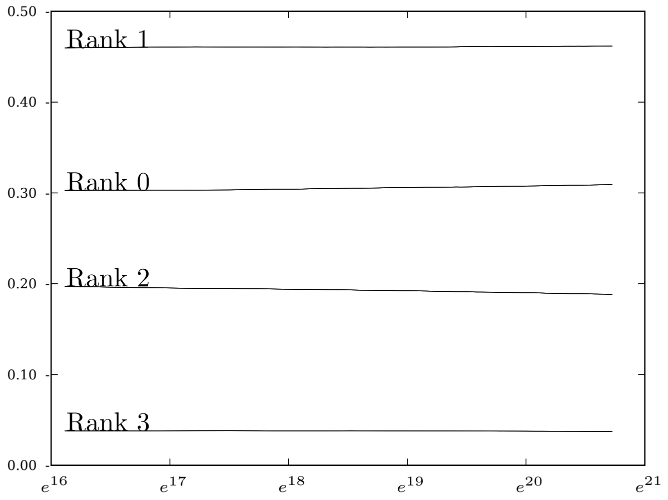


FIGURE 6. Rank Distribution of Curves with prime $N \leq 10^{10}$

TABLE 5. Rank Distribution for Prime Conductor $\leq 10^{10}$

Rank	0	1	2	3	≥ 4
Proportion	0.309	0.462	0.188	0.037	0.004
Proportion $\Delta > 0$	0.291	0.457	0.204	0.044	0.004
Proportion $\Delta < 0$	0.320	0.465	0.179	0.033	0.003

To investigate the possibility that the average rank might not decrease much below 0.964 we chose a selection of curves with prime conductor of size 10^{14}. It is non-trivial to get a good data set, since we must take congruence conditions on the elliptic curve coefficients and the variation of the size of the real period into account; see [Wat06] for more details on how to account for this.

Our data sets contained 89913 curves of positive prime discriminant, and 89749 similar curves with negative discriminant, with $|\Delta|$ near 10^{14} for all the curves. It then took a few months to compute the analytic rank for these curves. We found that for positive discriminant the average analytic rank is approximately 0.937 and for negative discriminant it is approximately 0.869 (see Table 6 for more details). Note that this is significantly less than the average rank found in [BM90] and [SW02]. It could be said that this is the strongest numerical evidence yet for the Minimalist Conjecture, though, it is still very weak. Incidentally, the largest rank found in any of these data sets is 6.

Let $f(\Delta)$ be the "probability" that $L(E, 1) = 0$ for an even parity curve of discriminant near Δ for Δ positive. For example, Tables 5 and 6 suggests that

$$f(10^{10}) \sim \frac{0.204 + 0.004}{0.291 + 0.204 + 0.004} = 0.417\ldots$$

$$f(10^{14}) \sim \frac{0.176 + 0.004}{0.319 + 0.176 + 0.004} = 0.361\ldots$$

Table 6. Rank Distribution For a Selection of Curves With Prime Conductor Near 10^{14}

Rank	0	1	2	3	≥ 4
Proportion $\Delta > 0$	0.319	0.467	0.176	0.034	0.004
Proportion $\Delta < 0$	0.343	0.475	0.154	0.025	0.002

(Note that we approximated $f(10^{10})$ using data for all $|\Delta| < 10^{10}$.) Motivated by the discussion in Section 3, we might heuristically approximate this probability function by $\hat{f}(\Delta) = c \cdot (\log \Delta)^{3/8}/\Delta^{1/24}$, where $\Delta^{1/24}$ comes about as the square root of the "typical" real period. The value of $\hat{f}(10^{10})/\hat{f}(10^{14})$ is about 1.29, which is not ridiculously far from the observed ratio of

$$\frac{f(10^{10})}{f(10^{14})} \sim \frac{0.417}{0.361} \sim 1.16.$$

5.3. **Variants.** We also carried out computations similar to the ones described above when counting isogeny classes instead of isomorphism classes of curves (isogeny is a coarser equivalence relation than isomorphism, grouping together curves between which there is a finite degree morphism). In our data the average size of isogeny classes for all curves of conductor up to X converges reasonably quickly to 1 (Duke has shown [Duk97] that this is indeed the case under a different ordering). Thus the data and graphs look almost identical to those presented above. Table 7 gives rank data for other subsets of the Stein-Watkins database of curves of conductor $\leq 10^8$. In the table, "has CM" refers to curves that have complex multiplication, i.e., whose endomorphism ring (over $\mathbb{C}$) is bigger than $\mathbb{Z}$.

Table 7. Distribution of Rank in Various Subsets of the Stein-Watkins Database with Conductor $N < 10^8$

Description	**Number**	**Rank 0**	**Rank 1**	**Rank 2**	**Rank** ≥ 3
All Curves	136832795	0.336	0.482	0.163	0.019
All Isogeny Classes	115821258	0.328	0.480	0.171	0.021
Has Isogeny	38599162	0.375	0.492	0.125	0.008
Has nontrivial torsion	35249448	0.373	0.492	0.127	0.008
N squarefree	21841534	0.296	0.467	0.202	0.034
Has Full 2-torsion	1674285	0.392	0.496	0.107	0.005
N is square	538558	0.416	0.496	0.084	0.004
N is prime	312435	0.303	0.460	0.197	0.041
Has 3-torsion	184590	0.422	0.498	0.078	0.002
Has CM	135226	0.411	0.498	0.087	0.005
N is prime squared	517	0.439	0.480	0.072	0.010

6. How can we *systematically* account for the Mordell-Weil rank we have already computed?

Forget all questions of asymptotics. Consider only the curves of prime conductor up to 10^{10} in our data. Is there an argument other than just computing ranks for each of the elliptic curves in the databases—is there a pure thought heuristic—that explains why we are witnessing so much Mordell-Weil rank? In a sense, these rational points are both analogous, and not analogous, to the physicist's dark matter.[12] This large mass of rational points for elliptic curves of prime conductor $\leq 10^{10}$ is palpably there. We aren't in the dark about that. We are merely in the dark about how to give a satisfactory account of it being there, other than computing instances, one after another.

We are, in a word, just at the very beginning of this story.

References

[ABCRZ] B. Allombert, K. Belabas, H. Cohen, X. Roblot, and I. Zakharevitch, `PARI/GP`, computer software, `pari.math.u-bordeaux.fr`

[Bek04] J. D. Bekenstein, *An alternative to the dark matter paradigm: relativistic MOND gravitation.* Invited talk at the 28th Johns Hopkins Workshop on Current Problems in Particle Theory, June 2004, Johns Hopkins University, Baltimore. Published online in JHEP Proceedings of Science. Online at `arxiv.org/astro-ph/0412652`

[Bha05] M. Bhargava, *The density of discriminants of quartic rings and fields*, Annals of Mathematics **162** (2005), no. 2, 1031–1063. Can be obtained online: Project Euclid Identifier is `euclid.annm/1134163091`

[BSD] B. J. Birch, H. P. F. Swinnerton-Dyer, *Notes on elliptic curves. I, II,* J. Reine Angew. **212**, 7–25 (1963), **218**, 79–108 (1965).

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). Online at `www.ams.org/journal-getitem?pii=S0894034701003708`

[BM90] A. Brumer and O. McGuinness, *The behavior of the Mordell-Weil group of elliptic curves*, Bull. Amer. Math. Soc. (N.S.) **23** (1990), no. 2, 375–382, data available from `oisinmc.com/math/310716`

[Cas62] J. W. S. Cassels, *Arithmetic on curves of genus* 1*. IV. Proof of the Hauptvermutung*, J. Reine Angew. Math. **211** (1962), 95–112.

[CDO] H. Cohen, F. Diaz y Diaz, and M. Olivier, *Counting discriminants of number fields of degree up to four*, Algorithmic number theory (Leiden, 2000), Lecture Notes in Comput. Sci., vol. 1838, Springer, Berlin, 2000, pp. 269–283.

[CKRS02] J. B. Conrey, J. P. Keating, M. O. Rubinstein, N. C. Snaith, *On the frequency of vanishing of quadratic twists of modular L-functions.* In *Number theory for the millennium, I* (Urbana, IL, 2000), edited by M. A. Bennett, B. C. Berndt, N. Boston, H. G. Diamond, A. J. Hildebrand and W. Philipp, A K Peters, Natick, MA (2002), 301–315. Available online at `arxiv.org/math.NT/0012043`

[CKRS06] ———, *Random Matrix Theory and the Fourier Coefficients of Half-Integral Weight Forms*, Experimental Math. **15** (2006), no. 1. Online at `arxiv.org/math/0412083`

[CPRW] J. B. Conrey, A. Pokharel, M. O. Rubinstein, and M. Watkins, *Secondary terms in the number of vanishings of quadratic twists of elliptic curve L-functions* (2005), `arxiv.org/math.NT/0509059`

[CRSW] J. B. Conrey, M. O. Rubinstein, N. C. Snaith, and M. Watkins, *Discretisation for odd quadratic twists* (2006), `arxiv.org/math.NT/0509428`

[Cre] J. E. Cremona, *Elliptic Curve Tables*, `www.maths.nott.ac.uk/personal/jec/ftp/data`

[Cre97] ———, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997. Online at `www.maths.nott.ac.uk/personal/jec/book`

[DFK04] C. David, J. Fearnley, H. Kisilevsky, *On the Vanishing of Twisted L-Functions of Elliptic Curves.* Experiment. Math. **13** (2004), no. 2, 185–198. Available online at `arxiv.org/math.NT/0406012`

[Duk97] W. Duke, *Elliptic curves with no exceptional primes.* C. R. Acad. Sci. Paris Sér. I Math. **325** (1997), no. 8, 813–818.

[12] The original idea is due to Zwicky [Zwi33]; recent SDSS and WMAP data [Teg04] seem to confirm its existence, though there are still some doubters (such as [Bek04]).

[Dy] See, for instance, a book review of S. W. Graham in the MAA Online book review column at www.maa.org/reviews/stalkingrh.html where discrepancies in versions of the story are discussed.

[Fal86] G. Faltings, *Finiteness theorems for abelian varieties over number fields*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, Translated from the German original [Invent. Math. **73** (1983), no. 3, 349–366; ibid. **75** (1984), no. 2, 381] by Edward Shipz, pp. 9–27.

[Fer96] S. Fermigier, *Etude experimentale du rang de familles de courbes elliptiques sur Q.* (French) [Experimental study of the rank of families of elliptic curves over Q]. Experiment. Math. **5** (1996), no. 2, 119–130. Available online at www.expmath.org/restricted/5/5.2/fermigier.ps

[Gol79] D. Goldfeld, *Conjectures on elliptic curves over quadratic fields*, Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979), Lecture Notes in Math., vol. 751, Springer, Berlin, 1979, pp. 108–118.

[GT02] A. Granville and T. J. Tucker, *It's as easy as abc*, Notices Amer. Math. Soc. **49** (2002), no. 10, 1224–1231. Online at www.ams.org/notices/200210/fea-granville.pdf

[HB93] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem. I. II.* Invent. Math. **111** (1993), no. 1, 171–195, **118** (1994), no. 2, 331–370.

[Hel04] H. A. Helfgott, *On the behaviour of root numbers in families of elliptic curves* (2004), submitted, arxiv.org/math.NT/0408141

[KaSa99] N. M. Katz and P.Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, vol. 45, American Mathematical Society, Providence, RI, 1999.

[KeSn00] J. P. Keating, N. C. Snaith, *Random matrix theory and $\zeta(1/2+it)$*, *Random matrix theory and L-functions at $s = 1/2$*. Comm. Math. Phys. **214** (2000), no. 1, 57–89, 91–110. Online from www.maths.bris.ac.uk/~mancs/publications.html

[Kol88] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E, \mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671.

[KowA] E. Kowalski, *Elliptic curves, rank in families and random matrices.* To appear in the Proceedings of the Isaac Newton Institute workshop on random matrices and L-functions (July 2004). Also related to the author's lecture at the AIM/Princeton workshop on the Birch and Swinnerton-Dyer conjecture (November 2003). See www.math.u-bordeaux1.fr/~kowalski/elliptic-curves-families.pdf

[KowB] E. Kowalski, *On the rank of quadratic twists of elliptic curves over function fields.* To appear in International J. Number Theory, online at arxiv.org/math.NT/0503732

[Mal02] G. Malle, *On the distribution of Galois groups*, J. Number Theory **92** (2002), no. 2, 315–329. Online from www.mathematik.uni-kl.de/~malle/en/publications.html

[Mal04] ———, *On the distribution of Galois groups. II*, Experiment. Math. **13** (2004), no. 2, 129–135. Online from www.expmath.org/expmath/volumes/13/13.2/Malle.pdf

[Mpl05] *Matplotlib*, computer software, matplotlib.sourceforge.net

[MSD74] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61.

[Meh04] M. L. Mehta, *Random matrices.* Third edition. Pure and Applied Mathematics (Amsterdam), 142. Elsevier/Academic Press, Amsterdam, 2004. xviii+688 pp.

[Mor22] L. J. Mordell, *On the Rational Solutions of the Indeterminate Equations of the Third and Fourth Degrees*, Proc. Cambridge Philos. Soc. 21, (1922-23), 179–192.

[RS02] K. Rubin and A. Silverberg, *Ranks of elliptic curves*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), no. 4, 455–474, available online from the AMS website at www.ams.org/bull/2002-39-04/S0273-0979-02-00952-7/home.html

[Sha85] D. Shanks, *Solved and unsolved problems in number theory.* Third edition, Chelsea, New York, 1985.

[Sil01] A. Silverberg, *Open questions in arithmetic algebraic geometry*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 83–142. Can also be obtained online from www.math.uci.edu/~asilverb/bibliography/pcmibook.ps

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[SJ05] W. Stein and D. Joyner, *Sage: System for algebra and geometry experimentation*, Communications in Computer Algebra (SIGSAM Bulletin) (July 2005). Online at `sage.sourceforge.net`

[SW02] W. A. Stein and M. Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, pp. 267–275. Online from `modular.ucsd.edu/papers/stein-watkins`

[Tat75] J. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1975, pp. 33–52. Lecture Notes in Math., Vol. 476.

[Teg04] M. Tegmark, *et al.*, *Cosmological parameters from SDSS and WMAP,* Phys. Rev. D69 (2004) 103501. Online at `arxiv.org/astro-ph/0310723`.

[Wal81] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier.* (French) [On the Fourier coefficients of modular forms of half-integral weight]. J. Math. Pures Appl. (9) **60** (1981), no. 4, 375–484.

[Wat04] M. Watkins, *Rank distribution in a family of cubic twists.* To appear in Proceedings of the Issac Newton Institute Workshop on Elliptic Curves and Random Matrix Theory. Online at `arxiv.org/math.NT/0412427`

[Wat06] ———, *Some heuristics about elliptic curves*, Preprint (2006). To be submitted to Experimental Mathematics; currently online at `www.maths.bris.ac.uk/~mamjw/heur.ps`

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551. Online: `www.jstor.org/view/0003486x/di976377/97p0063l/0`

[YouA] M. P. Young, *Low-lying zeros of families of elliptic curves*, J. Amer. Math. Soc. **19** (2006), no. 1, 205–250. Online from `dx.doi.org/10.1090/S0894-0347-05-00503-5`

[YouB] M. P. Young, *On the nonvanishing of elliptic curve L-functions at the central point* (2005), to appear in Proc. London Math. Soc., `arxiv.org/math.NT/0508185`, AIM preprint 2004-30.

[ZK87] D. Zagier and G. Kramarz, *Numerical investigations related to the L-series of certain elliptic curves*, J. Indian Math. Soc. (N.S.) **52**, 51–69 (1988).

[Zwi33] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln.* (German) [The red shift of extragalactic nebulae]. Helvetica Phys. Acta, vol. 6 (1933), 110–127.

29 Book – Modular Forms: A Computational Approach, with an appendix by Paul Gunnels

This book is published by the American Mathematical Society:

`http://www.ams.org/bookstore-getitem/item=gsm-79`

You can show your support by buying a copy. See also the Amazon.com page:

`http://www.amazon.com/gp/product/0821839608/ref=pd_lpo_k2_dp_sr_1?pf_rd_p=486539851&pf_rd_s=lpo-top-stripe-1&pf_rd_t=201&pf_rd_i=1848162138&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=1H4GENWMXXAJH2F1FQ6H`

Modular Forms:
A Computational Approach

William A. Stein
(with an appendix by Paul E. Gunnells)

Department of Mathematics, University of Washington
E-mail address: wstein@math.washington.edu

Department of Mathematics and Statistics, University of Massachusetts
E-mail address: gunnells@math.umass.edu

1991 *Mathematics Subject Classification.* Primary 11;
Secondary 11-04

Key words and phrases. abelian varieties, cohomology of arithmetic groups, computation, elliptic curves, Hecke operators, modular curves, modular forms, modular symbols, Manin symbols, number theory

Abstract. This is a textbook about algorithms for computing with modular forms. It is nontraditional in that the primary focus is not on underlying theory; instead, it answers the question *"how do you explicitly compute spaces of modular forms?"*

To my grandmother, Annette Maurer.

Contents

Preface

This is a graduate-level textbook about algorithms for computing with modular forms. It is nontraditional in that the primary focus is not on underlying theory; instead, it answers the question *"how do you use a computer to explicitly compute spaces of modular forms?"*

This book emerged from notes for a course the author taught at Harvard University in 2004, a course at UC San Diego in 2005, and a course at the University of Washington in 2006.

The author has spent years trying to find good practical ways to compute with classical modular forms for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ and has implemented most of these algorithms several times, first in C++ [**Ste99b**], then in MAGMA [**BCP97**], and as part of the free open source computer algebra system SAGE (see [**Ste06**]). Much of this work has involved turning formulas and constructions buried in obscure research papers into precise computational recipes then testing these and eliminating inaccuracies.

The author is aware of no other textbooks on computing with modular forms, the closest work being Cremona's book [**Cre97a**], which is about computing with elliptic curves, and Cohen's book [**Coh93**] about algebraic number theory.

In this book we focus on how to compute *in practice* the spaces $M_k(N, \varepsilon)$ of modular forms, where $k \geq 2$ is an integer and ε is a Dirichlet character of modulus N (the appendix treats modular forms for higher rank groups). We spend the most effort explaining the general algorithms that appear so far to be the best (in practice!) for such computations. We will not discuss in any detail computing with quaternion algebras, half-integral weight forms, weight 1 forms, forms for noncongruence subgroups or groups other than GL_2, Hilbert and Siegel modular forms, trace formulas, p-adic modular forms, and modular abelian varieties, all of which are topics for additional books. We also rarely analyze the complexity of the algorithms, but instead settle for occasional remarks about their practical efficiency.

For most of this book we assume the reader has some prior exposure to modular forms (e.g., [**DS05**]), though we recall many of the basic definitions. We cite standard books for proofs of the fundamental results about modular forms that we will use. The reader should also be familiar with basic algebraic number theory, linear algebra, complex analysis (at the level of [**Ahl78**]), and algorithms (e.g., know what an algorithm is and what big oh notation means). In some of the examples and applications we assume that the reader knows about elliptic curves at the level of [**Sil92**].

Chapter 1 is foundational for the rest of this book. It introduces congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ and modular forms as functions on the complex upper half plane. We discuss q-expansions, which provide an important computational handle on modular forms. We also study an algorithm for computing with congruence subgroups. The chapter ends with a list of applications of modular forms throughout mathematics.

In Chapter 2 we discuss level 1 modular forms in much more detail. In particular, we introduce Eisenstein series and the cusp form Δ and describe their q-expansions and basic properties. Then we prove a structure theorem for level 1 modular forms and use it to deduce dimension formulas and give an algorithm for explicitly computing a basis. We next introduce Hecke operators on level 1 modular forms, prove several results about them, and deduce multiplicativity of the Ramanujan τ function as an application. We also discuss explicit computation of Hecke operators. In Section 2.6 we make some brief remarks on recent work on asymptotically fast computation of values of τ. Finally, we describe computation of constant terms of Eisenstein series using an analytic algorithm. We generalize many of the constructions in this chapter to higher level in subsequent chapters.

In Chapter 3 we turn to modular forms of higher level but restrict for simplicity to weight 2 since much is clearer in this case. (We remove the weight restriction later in Chapter 8.) We describe a geometric way of viewing cuspidal modular forms as differentials on modular curves, which leads to modular symbols, which are an explicit way to present a certain homology group. This chapter closes with methods for explicitly computing cusp forms of weight 2 using modular symbols, which we generalize in Chapter 9.

In Chapter 4 we introduce Dirichlet characters, which are important both in explicit construction of Eisenstein series (in Chapter 5) and in decomposing spaces of modular forms as direct sums of simpler spaces. The main focus of this chapter is a detailed study of how to explicitly represent and compute with Dirichlet characters.

Chapter 5 is about how to explicitly construct the Eisenstein subspace of modular forms. First we define generalized Bernoulli numbers attached to a Dirichlet character and an integer then explain a new analytic algorithm for computing them (which generalizes the algorithm in Chapter 2). Finally we give without proof an explicit description of a basis of Eisenstein series, explain how to compute it, and give some examples.

Chapter 6 records a wide range of dimension formulas for spaces of modular forms, along with a few remarks about where they come from and how to compute them.

Chapter 7 is about linear algebra over exact fields, mainly the rational numbers. This chapter can be read independently of the others and does not require any background in modular forms. Nonetheless, this chapter occupies a central position in this book, because the algorithms in this chapter are of crucial importance to any actual implementation of algorithms for computing with modular forms.

Chapter 8 is the most important chapter in this book; it generalizes Chapter 3 to higher weight and general level. The modular symbols formulation described here is central to general algorithms for computing with modular forms.

Chapter 9 applies the algorithms from Chapter 8 to the problem of computing with modular forms. First we discuss decomposing spaces of modular forms using Dirichlet characters, and then explain how to compute a basis of Hecke eigenforms for each subspace using several approaches. We also discuss congruences between modular forms and bounds needed to provably generate the Hecke algebra.

Chapter 10 is about computing analytic invariants of modular forms. It discusses tricks for speeding convergence of certain infinite series and sketches how to compute every elliptic curve over $\mathbb{Q}$ with given conductor.

Chapter 11 contains detailed solutions to most of the exercises in this book. (Many of these were written by students in a course taught at the University of Washington.)

Appendix A deals with computational techniques for working with generalizations of modular forms to more general groups than $\mathrm{SL}_2(\mathbb{Z})$, such as $\mathrm{SL}_n(\mathbb{Z})$ for $n \geq 3$. Some of this material requires more prerequisites than the rest of the book. Nonetheless, seeing a natural generalization of the material in the rest of this book helps to clarify the key ideas. The topics in the appendix are directly related to the main themes of this book: modular symbols, Manin symbols, cohomology of subgroups of $\mathrm{SL}_2(\mathbb{Z})$ with various coefficients, explicit computation of modular forms, etc.

Software. We use SAGE, Software for Algebra and Geometry Experimentation (see [**Ste06**]), to illustrate how to do many of the examples. SAGE is completely free and packages together a wide range of open source mathematics software for doing much more than just computing with modular forms. SAGE can be downloaded and run on your computer or can be used via a web browser over the Internet. The reader is encouraged to experiment with many of the objects in this book using SAGE. We do not describe the basics of using SAGE in this book; the reader should read the SAGE tutorial (and other documentation) available at the SAGE website [**Ste06**]. All examples in this book have been automatically tested and should work exactly as indicated in SAGE version at least 1.5.

Acknowledgements. David Joyner and Gabor Wiese carefully read the book and provided a huge number of helpful comments.

John Cremona and Kevin Buzzard both made many helpful remarks that were important in the development of the algorithms in this book. Much of the mathematics (and some of the writing) in Chapter 10 is joint work with Helena Verrill.

Noam Elkies made remarks about Chapters 1 and 2. Sándor Kovács provided interesting comments on Chapter 1. Allan Steel provided helpful feedback on Chapter 7. Jordi Quer made useful remarks about Chapter 4 and Chapter 6.

The students in the courses that I taught on this material at Harvard, San Diego, and Washington provided substantial feedback: in particular, Abhinav Kumar made numerous observations about computing widths of cusps (see Section 1.4.1) and Thomas James Barnet-Lamb made helpful remarks about how to represent Dirichlet characters. James Merryfield made helpful remarks about complex analytic issues and about convergence in Stirling's formula. Robert Bradshaw, Andrew Crites (who wrote Exercise 7.5), Michael Goff, Dustin Moody, and Koopa Koo wrote most of the solutions included in Chapter 11 and found numerous typos throughout the book. Dustin Moody also carefully read through the book and provided feedback.

H. Stark suggested using Stirling's formula in Section 2.7.1, and Mark Watkins and Lynn Walling made comments on Chapter 3.

Justin Walker found typos in the first published version of the book.

Parts of Chapter 1 follow Serre's beautiful introduction to modular forms [**Ser73**, Ch. VII] closely, though we adjust the notation, definitions, and order of presentation to be consistent with the rest of this book.

I would like to acknowledge the partial support of NSF Grant DMS 05-55776. Gunnells was supported in part by NSF Grants DMS 02-45580 and DMS 04-01525.

Notation and Conventions. We denote canonical isomorphisms by $\cong$ and noncanonical isomorphisms by $\approx$. If V is a vector space and s denotes some sort of construction involving V, we let V_s denote the corresponding subspace and V^s the quotient space. E.g., if ι is an involution of V, then V_+ is $\mathrm{Ker}(\iota - 1)$ and $V^+ = V/\mathrm{Im}(\iota - 1)$. If A is a finite abelian group, then A_{tor} denotes the torsion subgroup and A/tor denotes the quotient A/A_{tor}. We denote right group actions using exponential notation. Everywhere in this book, N is a positive integer and k is an integer.

If N is an integer, a *divisor* t of N is a *positive* integer such that N/t is an integer.

Chapter 1

Modular Forms

This chapter introduces modular forms and congruence subgroups, which are central objects in this book. We first introduce the upper half plane and the group $\mathrm{SL}_2(\mathbb{Z})$ then recall some definitions from complex analysis. Next we define modular forms of level 1 followed by modular forms of general level. In Section 1.4 we discuss congruence subgroups and explain a simple way to compute generators for them and determine element membership. Section 1.5 lists applications of modular forms.

We assume familiarity with basic number theory, group theory, and complex analysis. For a deeper understanding of modular forms, the reader is urged to consult the standard books in the field, e.g., [**Lan95, Ser73, DI95, Miy89, Shi94, Kob84**]. See also [**DS05**], which is an excellent first introduction to the theoretical foundations of modular forms.

1.1. Basic Definitions

The group

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1 \text{ and } a, b, c, d \in \mathbb{R} \right\}$$

acts on the *complex upper half plane*

$$\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$$

by *linear fractional transformations*, as follows. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$, then for any $z \in \mathfrak{h}$ we let

$$\gamma(z) = \frac{az + b}{cz + d} \in \mathfrak{h}. \tag{1.1.1}$$

Since the determinant of γ is 1, we have

$$\left(\frac{d}{dz}\gamma\right)(z) = \frac{1}{(cz+d)^2}.$$

Definition 1.1 (Modular Group)**.** The *modular group* is the group of all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$.

For example, the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{1.1.2}$$

are both elements of $\mathrm{SL}_2(\mathbb{Z})$; the matrix S induces the function $z \mapsto -1/z$ on $\mathfrak{h}$, and T induces the function $z \mapsto z + 1$.

Theorem 1.2. *The group* $\mathrm{SL}_2(\mathbb{Z})$ *is generated by* S *and* T*.*

Proof. See e.g. [**Ser73**, §VII.1]. □

In SAGE we compute the group $\mathrm{SL}_2(\mathbb{Z})$ and its generators as follows:

```
sage: G = SL(2,ZZ); G
Modular Group SL(2,Z)
sage: S, T = G.gens()
sage: S
[ 0 -1]
[ 1  0]
sage: T
[1 1]
[0 1]
```

Definition 1.3 (Holomorphic and Meromorphic)**.** Let R be an open subset of $\mathbb{C}$. A function $f : R \to \mathbb{C}$ is *holomorphic* if f is complex differentiable at every point $z \in R$, i.e., for each $z \in R$ the limit

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists, where h may approach 0 along any path. A function $f : R \to \mathbb{C} \cup \{\infty\}$ is *meromorphic* if it is holomorphic except (possibly) at a discrete set S of points in R, and at each $\alpha \in S$ there is a positive integer n such that $(z - \alpha)^n f(z)$ is holomorphic at α.

The function $f(z) = e^z$ is a holomorphic function on $\mathbb{C}$; in contrast, $1/(z - i)$ is meromorphic on $\mathbb{C}$ but not holomorphic since it has a pole at i. The function $e^{-1/z}$ is not even meromorphic on $\mathbb{C}$.

Modular forms are holomorphic functions on $\mathfrak{h}$ that transform in a particular way under a certain subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Before defining general modular forms, we define modular forms of level 1.

1.2. Modular Forms of Level 1

Definition 1.4 (Weakly Modular Function)**.** A *weakly modular function* of *weight* $k \in \mathbb{Z}$ is a meromorphic function f on $\mathfrak{h}$ such that for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and all $z \in \mathfrak{h}$ we have

$$f(z) = (cz+d)^{-k} f(\gamma(z)). \tag{1.2.1}$$

The constant functions are weakly modular of weight 0. There are no nonzero weakly modular functions of odd weight (see Exercise 1.4), and it is not obvious that there are any weakly modular functions of even weight $k \geq 2$ (but there are, as we will see!). The product of two weakly modular functions of weights k_1 and k_2 is a weakly modular function of weight $k_1 + k_2$ (see Exercise 1.3).

When k is even, (1.2.1) has a possibly more conceptual interpretation; namely (1.2.1) is the same as

$$f(\gamma(z))(d(\gamma(z)))^{k/2} = f(z)(dz)^{k/2}.$$

Thus (1.2.1) simply says that the weight k "differential form" $f(z)(dz)^{k/2}$ is fixed under the action of every element of $\mathrm{SL}_2(\mathbb{Z})$.

By Theorem 1.2, the group $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices S and T of (1.1.2), so to show that a meromorphic function f on $\mathfrak{h}$ is a weakly modular function, all we have to do is show that for all $z \in \mathfrak{h}$ we have

$$f(z+1) = f(z) \qquad \text{and} \qquad f(-1/z) = z^k f(z). \tag{1.2.2}$$

Suppose f is a weakly modular function of weight k. A *Fourier expansion* of f, if it exists, is a representation of f as $f(z) = \sum_{n=m}^{\infty} a_n e^{2\pi i n z}$, for all $z \in \mathfrak{h}$. Let $q = q(z) = e^{2\pi i z}$, which we view as a holomorphic function on $\mathbb{C}$. Let D' be the open unit disk with the origin removed, and note that q defines a map $\mathfrak{h} \to D'$. By (1.2.2) we have $f(z+1) = f(z)$, so there is a function $F : D' \to \mathbb{C}$ such that $F(q(z)) = f(z)$. This function F is a complex-valued function on D', but it may or may not be well behaved at 0.

Suppose that F is well behaved at 0, in the sense that for some $m \in \mathbb{Z}$ and all q in a neighborhood of 0 we have the equality

$$F(q) = \sum_{n=m}^{\infty} a_n q^n. \tag{1.2.3}$$

If this is the case, we say that f is *meromorphic at* ∞. If, moreover, $m \geq 0$, we say that f is *holomorphic at* ∞. We also call (1.2.3) the *q-expansion* of f about ∞.

Definition 1.5 (Modular Function)**.** A *modular function* of *weight k* is a weakly modular function of weight k that is meromorphic at ∞.

Definition 1.6 (Modular Form)**.** A *modular form* of *weight k* (and *level* 1) is a modular function of weight k that is holomorphic on $\mathfrak{h}$ and at ∞.

If f is a modular form, then there are numbers a_n such that for all $z \in \mathfrak{h}$,

$$f(z) = \sum_{n=0}^{\infty} a_n q^n. \tag{1.2.4}$$

Proposition 1.7. *The above series converges for all $z \in \mathfrak{h}$.*

Proof. The function $f(q)$ is holomorphic on D, so its Taylor series converges absolutely in D. □

Since $e^{2\pi i z} \to 0$ as $z \to i\infty$, we set $f(\infty) = a_0$.

Definition 1.8 (Cusp Form)**.** A *cusp form* of *weight k* (and *level* 1) is a modular form of weight k such that $f(\infty) = 0$, i.e., $a_0 = 0$.

Let $\mathbb{C}[[q]]$ be the ring of all *formal power series* in q. If $k = 2$, then $dq = 2\pi i q dz$, so $dz = \frac{1}{2\pi i}\frac{dq}{q}$. If $f(q)$ is a cusp form of weight 2, then

$$2\pi i f(z) dz = f(q)\frac{dq}{q} = \frac{f(q)}{q} dq \in \mathbb{C}[[q]]dq.$$

Thus the differential $2\pi i f(z)dz$ is holomorphic at ∞, since q is a local parameter at ∞.

1.3. Modular Forms of Any Level

In this section we define spaces of modular forms of arbitrary level.

Definition 1.9 (Congruence Subgroup)**.** A *congruence subgroup* of $\mathrm{SL}_2(\mathbb{Z})$ is any subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains

$$\Gamma(N) = \mathrm{Ker}(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$$

for some positive integer N. The smallest such N is the *level of* Γ.

The most important congruence subgroups in this book are

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}$$

and

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\},$$

where $*$ means any element. Both groups have level N (see Exercise 1.6).

Let k be an integer. Define the *weight k right action* of $\mathrm{GL}_2(\mathbb{Q})$ on the set of all functions $f : \mathfrak{h} \to \mathbb{C}$ as follows. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$, let

$$(f^{[\gamma]_k})(z) = \det(\gamma)^{k-1}(cz+d)^{-k} f(\gamma(z)). \tag{1.3.1}$$

Proposition 1.10. *Formula* (1.3.1) *defines a right action of* $\mathrm{GL}_2(\mathbb{Z})$ *on the set of all functions* $f : \mathfrak{h} \to \mathbb{C}$*; in particular,*

$$f^{[\gamma_1\gamma_2]_k} = (f^{[\gamma_1]_k})^{[\gamma_2]_k}.$$

Proof. See Exercise 1.7. □

Definition 1.11 (Weakly Modular Function)**.** A *weakly modular function* of weight k for a congruence subgroup Γ is a meromorphic function $f : \mathfrak{h} \to \mathbb{C}$ such that $f^{[\gamma]_k} = f$ for all $\gamma \in \Gamma$.

A central object in the theory of modular forms is the *set of cusps*

$$\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}.$$

An element $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ acts on $\mathbb{P}^1(\mathbb{Q})$ by

$$\gamma(z) = \begin{cases} \frac{az+b}{cz+d} & \text{if } z \neq \infty, \\ \frac{a}{c} & \text{if } z = \infty. \end{cases}$$

Also, note that if the denominator c or $cz+d$ is 0 above, then

$$\gamma(z) = \infty \in \mathbb{P}^1(\mathbb{Q}).$$

The set of *cusps for a congruence subgroup* Γ is the set $C(\Gamma)$ of Γ-orbits of $\mathbb{P}^1(\mathbb{Q})$. (We will often identify elements of $C(\Gamma)$ with a representative element from the orbit.) For example, the lemma below asserts that if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, then there is exactly one orbit, so $C(\mathrm{SL}_2(\mathbb{Z})) = \{[\infty]\}$.

Lemma 1.12. *For any cusps* $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$ *there exists* $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ *such that* $\gamma(\alpha) = \beta$.

Proof. This is Exercise 1.8. □

Proposition 1.13. *For any congruence subgroup* Γ*, the set* $C(\Gamma)$ *of cusps is finite.*

Proof. This is Exercise 1.9. □

See [**DS05**, §3.8] and Algorithm 8.12 below for more discussion of cusps and results relevant to their enumeration.

In order to define modular forms for general congruence subgroups, we next explain what it means for a function to be holomorphic on the *extended upper half plane*

$$\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q}).$$

See [**Shi94**, §1.3–1.5] for a detailed description of the correct topology to consider on $\mathfrak{h}^*$. In particular, a basis of neighborhoods for $\alpha \in \mathbb{Q}$ is given by the sets $\{\alpha\} \cup D$, where D is an open disc in $\mathfrak{h}$ that is tangent to the real line at α.

Recall from Section 1.2 that a weakly modular function f on $\mathrm{SL}_2(\mathbb{Z})$ is holomorphic at ∞ if its q-expansion is of the form $\sum_{n=0}^{\infty} a_n q^n$.

In order to make sense of holomorphicity of a weakly modular function f for an arbitrary congruence subgroup Γ at any $\alpha \in \mathbb{Q}$, we first prove a lemma.

Lemma 1.14. *If $f : \mathfrak{h} \to \mathbb{C}$ is a weakly modular function of weight k for a congruence subgroup Γ and if $\delta \in \mathrm{SL}_2(\mathbb{Z})$, then $f^{[\delta]_k}$ is a weakly modular function for $\delta^{-1}\Gamma\delta$.*

Proof. If $s = \delta^{-1}\gamma\delta \in \delta^{-1}\Gamma\delta$, then

$$(f^{[\delta]_k})^{[s]_k} = f^{[\delta s]_k} = f^{[\delta\delta^{-1}\gamma\delta]_k} = f^{[\gamma\delta]_k} = f^{[\delta]_k}.$$

□

Fix a weakly modular function f of weight k for a congruence subgroup Γ, and suppose $\alpha \in \mathbb{Q}$. In Section 1.2 we constructed the q-expansion of f by using that $f(z) = f(z+1)$, which held since $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. There are congruence subgroups Γ such that $T \not\in \Gamma$. Moreover, even if we are interested only in modular forms for $\Gamma_1(N)$, where we have $T \in \Gamma_1(N)$ for all N, we will still have to consider q-expansions at infinity for modular forms on groups $\delta^{-1}\Gamma_1(N)\delta$, and these need not contain T. Fortunately, $T^N = \left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma(N)$, so a congruence subgroup of level N contains T^N. Thus we have $f(z+H) = f(H)$ for some positive integer H, e.g., $H = N$ always works, but there may be a smaller choice of H. The minimal choice of $H > 0$ such that $\left(\begin{smallmatrix} 1 & H \\ 0 & 1 \end{smallmatrix}\right) \in \delta^{-1}\Gamma\delta$, where $\delta(\infty) = \alpha$, is called the *width of the cusp* α relative to the group Γ (see Section 1.4.1). When f is meromorphic at infinity, we obtain a Fourier expansion

$$f(z) = \sum_{n=m}^{\infty} a_n q^{n/H} \tag{1.3.2}$$

in powers of the function $q^{1/H} = e^{2\pi i z/H}$. We say that f is holomorphic at ∞ if in (1.3.2) we have $m \geq 0$.

What about the other cusps $\alpha \in \mathbb{P}^1(\mathbb{Q})$? By Lemma 1.12 there is a $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\infty) = \alpha$. We declare f to be *holomorphic at the cusp* α if the weakly modular function $f^{[\gamma]_k}$ is holomorphic at ∞.

Definition 1.15 (Modular Form)**.** A *modular form* of integer *weight* k for a congruence subgroup Γ is a weakly modular function $f : \mathfrak{h} \to \mathbb{C}$ that is holomorphic on $\mathfrak{h}^*$. We let $M_k(\Gamma)$ denote the space of weight k modular forms of weight k for Γ.

Proposition 1.16. *If a weakly modular function f is holomorphic at a set of representative elements for $C(\Gamma)$, then it is holomorphic at every element of $\mathbb{P}^1(\mathbb{Q})$.*

Proof. Let $c_1, \ldots, c_n \in \mathbb{P}^1(\mathbb{Q})$ be representatives for the set of cusps for Γ. If $\alpha \in \mathbb{P}^1(\mathbb{Q})$, then there is $\gamma \in \Gamma$ such that $\alpha = \gamma(c_i)$ for some i. By hypothesis f is holomorphic at c_i, so if $\delta \in \mathrm{SL}_2(\mathbb{Z})$ is such that $\delta(\infty) = c_i$, then $f^{[\delta]_k}$ is holomorphic at ∞. Since f is a weakly modular function for Γ,

$$f^{[\delta]_k} = (f^{[\gamma]_k})^{[\delta]_k} = f^{[\gamma\delta]_k}. \tag{1.3.3}$$

But $\gamma(\delta(\infty)) = \gamma(c_i) = \alpha$, so (1.3.3) implies that f is holomorphic at α. $\square$

1.4. Remarks on Congruence Subgroups

Recall that a congruence subgroup is a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma(N)$ for some N. Any congruence subgroup has finite index in $\mathrm{SL}_2(\mathbb{Z})$, since $\Gamma(N)$ does. What about the converse: is every finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ a congruence subgroup? This is the *congruence subgroup problem.* One can ask about the congruence subgroup problem with $\mathrm{SL}_2(\mathbb{Z})$ replaced by many similar groups. If p is a prime, then one can prove that every finite index subgroup of $\mathrm{SL}_2(\mathbb{Z}[1/p])$ is a congruence subgroup (i.e., contains the kernel of reduction modulo some integer coprime to p), and for any $n > 2$, all finite index subgroups of $\mathrm{SL}_n(\mathbb{Z})$ are congruence subgroups (see [**Hum80**]). However, there are numerous finite index subgroups of $\mathrm{SL}_2(\mathbb{Z})$ that are not congruence subgroups. The paper [**Hsu96**] contains an *algorithm* to decide if certain finite index subgroups are congruence subgroups and gives an example of a subgroup of index 12 that is not a congruence subgroup.

One can consider modular forms even for noncongruence subgroups. See, e.g., [**Tho89**] and the papers it references for work on this topic. We will not consider such modular forms further in this book. Note that modular symbols (which we define later in this book) *are* computable for noncongruence subgroups.

Finding coset representatives for $\Gamma_0(N)$, $\Gamma_1(N)$ and $\Gamma(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ is straightforward and will be discussed at length later in this book. To make the problem more explicit, note that you can quotient out by $\Gamma(N)$ first. Then the question amounts to finding coset representatives for a subgroup of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ (and lifting), which is reasonably straightforward.

Given coset representatives for a finite index subgroup G of $\mathrm{SL}_2(\mathbb{Z})$, we can compute generators for G as follows. Let R be a set of coset representatives for G. Let $\sigma, \tau \in \mathrm{SL}_2(\mathbb{Z})$ be the matrices denoted by S and T in (1.1.2). Define maps $s, t : R \to G$ as follows. If $r \in R$, then there exists a unique $\alpha_r \in R$ such that $Gr\sigma = G\alpha_r$. Let $s(r) = r\sigma\alpha_r^{-1}$. Likewise, there is a unique β_r such that $Gr\tau = G\beta_r$ and we let $t(r) = r\tau\beta_r^{-1}$. Note that $s(r)$ and $t(r)$ are in G for all r. Then G is generated by $s(R) \cup t(R)$.

Proposition 1.17. *The above procedure computes generators for G.*

Proof. Without loss of generality, assume that $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ represents the coset of G. Let g be an element of G. Since σ and τ generate $\mathrm{SL}_2(\mathbb{Z})$, it is possible to write g as a product of powers of σ and τ. There is a procedure, which we explain below with an example in order to avoid cumbersome notation, which writes g as a product of elements of $s(R) \cup t(R)$ times a right coset representative $r \in R$. For example, if

$$g = \sigma\tau^2\sigma\tau,$$

then $g = I\sigma\tau^2\sigma\tau = s(I)y\tau^2\sigma\tau$ for some $y \in R$. Continuing,

$$s(I)y\tau^2\sigma\tau = s(I)(y\tau)\tau\sigma\tau = s(I)(t(y)z)\tau\sigma\tau$$

for some $z \in R$. Again,

$$s(I)(t(y)z)\tau\sigma\tau = s(I)t(y)(z\tau)\sigma\tau = \cdots .$$

The procedure illustrated above (with an example) makes sense for arbitrary g and, after carrying it out, writes g as a product of elements of $s(R) \cup t(R)$ times a right coset representative $r \in R$. But $g \in G$ and I is the right coset representative for G, so this right coset representative must be I. □

Remark 1.18. We could also apply the proof of Proposition 1.17 to write any element of G in terms of the given generators. Moreover, we could use it to write any element $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ in the form gr, where $g \in G$ and $r \in R$, so we can decide whether or not $\gamma \in G$.

1.4.1. Computing Widths of Cusps. Let Γ be a congruence subgroup of level N. Suppose $\alpha \in C(\Gamma)$ is a cusp, and choose $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\infty) = \alpha$. Recall that the minimal h such that $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \gamma^{-1}\Gamma\gamma$ is called the *width of the cusp* α for the group Γ. In this section we discuss how to compute h.

Algorithm 1.19 (Width of Cusp)**.** *Given a congruence subgroup Γ of level N and a cusp α for Γ, this algorithm computes the width h of α. We assume that Γ is given by congruence conditions, e.g., $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$.*

(1) [Find γ] Use the extended Euclidean algorithm to find $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\infty) = \alpha$, as follows. If $\alpha = \infty$, set $\gamma = 1$; otherwise, write $\alpha = a/b$, find c, d such that $ad - bc = 1$, and set $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

(2) [Compute Conjugate Matrix] Compute the following element of $\mathrm{Mat}_2(\mathbb{Z}[x])$:

$$\delta(x) = \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1}.$$

Note that the entries of $\delta(x)$ are constant or linear in x.

(3) [Solve] The congruence conditions that define Γ give rise to four linear congruence conditions on x. Use techniques from elementary number theory (or enumeration) to find the smallest simultaneous positive solution h to these four equations.

Example 1.20. (1) Suppose $\alpha = 0$ and $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$. Then $\gamma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ has the property that $\gamma(\infty) = \alpha$. Next, the congruence condition is

$$\delta(x) = \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1} = \begin{pmatrix} 1 & 0 \\ -x & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N}.$$

Thus the smallest positive solution is $h = N$, so the width of 0 is N.

(2) Suppose $N = pq$ where p, q are distinct primes, and let $\alpha = 1/p$. Then $\gamma = \left(\begin{smallmatrix} 1 & 0 \\ p & 1 \end{smallmatrix}\right)$ sends ∞ to α. The congruence condition for $\Gamma_0(pq)$ is

$$\delta(x) = \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1} = \begin{pmatrix} 1 - px & x \\ -p^2 x & px + 1 \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{pq}.$$

Since $p^2 x \equiv 0 \pmod{pq}$, we see that $x = q$ is the smallest solution. Thus $1/p$ has width q, and symmetrically $1/q$ has width p.

Remark 1.21. For $\Gamma_0(N)$, once we enforce that the bottom left entry is 0 (mod N) and use that the determinant is 1, the coprimality from the other two congruences is automatic. So there is one congruence to solve in the $\Gamma_0(N)$ case. There are two congruences in the $\Gamma_1(N)$ case.

1.5. Applications of Modular Forms

The above definition of modular forms might leave the impression that modular forms occupy an obscure corner of complex analysis. This is *not* the case! Modular forms are highly geometric, arithmetic, and topological objects that are of extreme interest all over mathematics:

(1) **Fermat's last theorem:** Wiles' proof [**Wil95**] of Fermat's last theorem uses modular forms extensively. The work of Wiles et al. on modularity also massively extends computational methods for elliptic curves over $\mathbb{Q}$, because many elliptic curve algorithms, e.g., for computing L-functions, modular degrees, Heegner points, etc., require that the elliptic curve be modular.

(2) **Diophantine equations:** Wiles' proof of Fermat's last theorem has made available a wide array of new techniques for solving certain diophantine equations. Such work relies crucially on having access to tables or software for computing modular forms. See, e.g., [**Dar97, Mer99, Che05, SC03**]. (Wiles did not need a computer, because the relevant spaces of modular forms that arise in his proof have dimension 0!) Also, according to Siksek (personal communication) the paper [**BMS06**] would "have been entirely impossible to write without [the algorithms described in this book]."

(3) **Congruent number problem:** This ancient open problem is to determine which integers are the area of a right triangle with rational side lengths. There is a potential solution that uses modular forms (of weight 3/2) extensively (the solution is conditional on truth of the Birch and Swinnerton-Dyer conjecture, which is not yet known). See [**Kob84**].

(4) **Topology:** Topological modular forms are a major area of current research.

(5) **Construction of Ramanujan graphs:** Modular forms can be used to construct almost optimal expander graphs, which play a role in communications network theory.

(6) **Cryptography and Coding Theory:** Point counting on elliptic curves over finite fields is crucial to the construction of elliptic curve cryptosystems, and modular forms are relevant to efficient algorithms for point counting (see [**Elk98**]). Algebraic curves that are associated to modular forms are useful in constructing and studying certain error-correcting codes (see [**Ebe02**]).

(7) **The Birch and Swinnerton-Dyer conjecture:** This central open problem in arithmetic geometry relates arithmetic properties of elliptic curves (and abelian varieties) to special values of L-functions. Most deep results toward this conjecture use modular forms extensively (e.g., work of Kolyvagin, Gross-Zagier, and Kato). Also, modular forms are used to compute and prove results about special values of these L-functions. See [**Wil00**].

(8) **Serre's Conjecture on modularity of Galois representation:** Let $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be the Galois group of an algebraic closure of $\mathbb{Q}$. Serre conjectured and many people have (nearly!) proved that every continuous homomorphism $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_q)$, where $\mathbb{F}_q$ is a finite field and $\det(\rho(\text{complex conjugation})) = -1$, "arises" from a modular form. More precisely, for almost all primes p the coefficients a_p of a modular (eigen-)form $\sum a_n q^n$ are congruent to the traces of elements $\rho(\mathrm{Frob}_p)$, where Frob_p are certain special elements of $G_{\mathbb{Q}}$ called Frobenius elements. See [**RS01**] and [**DS05**, Ch. 9].

(9) **Generating functions for partitions:** The generating functions for various kinds of partitions of an integer can often be related to modular forms. Deep theorems about modular forms then translate into results about partitions. See work of Ramanujan, Gordon, Andrews, and Ahlgren and Ono (e.g., [**AO01**]).

(10) **Lattices:** If $L \subset \mathbb{R}^n$ is an even unimodular lattice (the basis matrix has determinant ± 1 and $\lambda \cdot \lambda \in 2\mathbb{Z}$ for all $\lambda \in L$), then the theta series

$$\theta_L(q) = \sum_{\lambda \in L} q^{\lambda\cdot\lambda}$$

is a modular form of weight $n/2$. The coefficient of q^m is the number of lattice vectors with squared length m. Theorems and computational methods for modular forms translate into theorems and computational methods for lattices. For example, the 290 theorem of M. Bharghava and J. Hanke is a theorem about lattices, which asserts that an integer-valued quadratic form represents all positive integers if and only if it represents the integers up to 290; it is proved by doing many calculations with modular forms (both theoretical and with a computer).

1.6. Exercises

1.1 Suppose $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ has positive determinant. Prove that if $z \in \mathbb{C}$ is a complex number with positive imaginary part, then the imaginary part of $\gamma(z) = (az+b)/(cz+d)$ is also positive.

1.2 Prove that every rational function (quotient of two polynomials) is a meromorphic function on $\mathbb{C}$.

1.3 Suppose f and g are weakly modular functions for a congruence subgroup Γ with $f \neq 0$.

(a) Prove that the product fg is a weakly modular function for Γ.

(b) Prove that $1/f$ is a weakly modular function for Γ.

(c) If f and g are modular functions, show that fg is a modular function for Γ.

(d) If f and g are modular forms, show that fg is a modular form for Γ.

1.4 Suppose f is a weakly modular function of odd weight k and level $\Gamma_0(N)$ for some N. Show that $f = 0$.

1.5 Prove that $\mathrm{SL}_2(\mathbb{Z}) = \Gamma_0(1) = \Gamma_1(1) = \Gamma(1)$.

1.6 (a) Prove that $\Gamma_1(N)$ is a group.

(b) Prove that $\Gamma_1(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$ (Hint: It contains the kernel of the homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$.)

(c) Prove that $\Gamma_0(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$.

(d) Prove that $\Gamma_0(N)$ and $\Gamma_1(N)$ have level N.

1.7 Let k be an integer, and for any function $f : \mathfrak{h}^* \to \mathbb{C}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$, set $f^{[\gamma]_k}(z) = \det(\gamma)^{k-1} \cdot (cz+d)^{-k} \cdot f(\gamma(z))$. Prove that if $\gamma_1, \gamma_2 \in \mathrm{GL}_2(\mathbb{Z})$, then for all $z \in \mathfrak{h}^*$ we have

$$f^{[\gamma_1\gamma_2]_k}(z) = ((f^{[\gamma_1]_k})^{[\gamma_2]_k})(z).$$

1.8 Prove that for any $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$, there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\alpha) = \beta$.

1.9 Prove Proposition 1.13, which asserts that the set of cusps $C(\Gamma)$, for any congruence subgroup Γ, is finite.

1.10 Use Algorithm 1.19 to give an example of a group Γ and cusp α with width 2.

Chapter 2

Modular Forms of Level 1

In this chapter we study in detail the structure of level 1 modular forms, i.e., modular forms on $\mathrm{SL}_2(\mathbb{Z}) = \Gamma_0(1) = \Gamma_1(1)$. We assume some complex analysis (e.g., the residue theorem), linear algebra, and that the reader has read Chapter 1.

2.1. Examples of Modular Forms of Level 1

In this section we will finally see some examples of modular forms of level 1! We first introduce the Eisenstein series and then define Δ, which is a cusp form of weight 12. In Section 2.2 we prove the structure theorem, which says that all modular forms of level 1 are polynomials in Eisenstein series.

For an even integer $k \geq 4$, the *nonnormalized weight k Eisenstein series* is the function on the extended upper half plane $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q})$ given by

$$G_k(z) = \sideset{}{^*}\sum_{m,n \in \mathbb{Z}} \frac{1}{(mz+n)^k}. \tag{2.1.1}$$

The star on top of the sum symbol means that for each z the sum is over all $m, n \in \mathbb{Z}$ such that $mz + n \neq 0$.

Proposition 2.1. *The function $G_k(z)$ is a modular form of weight k, i.e., $G_k \in M_k(\mathrm{SL}_2(\mathbb{Z}))$.*

Proof. See [**Ser73**, § VII.2.3] for a proof that $G_k(z)$ defines a holomorphic function on $\mathfrak{h}^*$. To see that G_k is modular, observe that

$$G_k(z+1) = \sum^{*} \frac{1}{(m(z+1)+n)^k} = \sum^{*} \frac{1}{(mz+(n+m))^k} = \sum^{*} \frac{1}{(mz+n)^k},$$

where for the last equality we use that the map $(m, n+m) \mapsto (m,n)$ on $\mathbb{Z} \times \mathbb{Z}$ is invertible. Also,

$$\begin{aligned} G_k(-1/z) &= \sum^{*} \frac{1}{(-m/z+n)^k} \\ &= \sum^{*} \frac{z^k}{(-m+nz)^k} \\ &= z^k \sum^{*} \frac{1}{(mz+n)^k} = z^k G_k(z), \end{aligned}$$

where we use that $(n,-m) \mapsto (m,n)$ is invertible. □

Proposition 2.2. $G_k(\infty) = 2\zeta(k)$, *where* ζ *is the Riemann zeta function.*

Proof. As $z \to \infty$ (along the imaginary axis) in (2.1.1), the terms that involve z with $m \neq 0$ go to 0. Thus

$$G_k(\infty) = \sum_{n \in \mathbb{Z}}^{*} \frac{1}{n^k}.$$

This sum is twice $\zeta(k) = \sum_{n \geq 1} \frac{1}{n^k}$, as claimed. □

2.1.1. The Cusp Form Δ. Suppose $E = \mathbb{C}/\Lambda$ is an elliptic curve over $\mathbb{C}$, viewed as a quotient of $\mathbb{C}$ by a lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, with $\omega_1/\omega_2 \in \mathfrak{h}$ (see [**DS05**, §1.4]). The *Weierstrass $\wp$-function* of the lattice Λ is

$$\wp = \wp_\Lambda(u) = \frac{1}{u^2} + \sum_{k=4,6,8,\ldots} (k-1)G_k(\omega_1/\omega_2)u^{k-2},$$

where the sum is over even integers $k \geq 4$. It satisfies the differential equation

$$(\wp')^2 = 4\wp^3 - 60G_4(\omega_1/\omega_2)\wp - 140G_6(\omega_1/\omega_2).$$

If we set $x = \wp$ and $y = \wp'$, the above is an (affine) equation of the form $y^2 = ax^3+bx+c$ for an elliptic curve that is complex analytically isomorphic to $\mathbb{C}/\Lambda$ (see [**Ahl78**, pg. 277] for why the cubic has distinct roots).

The discriminant of the cubic

$$4x^3 - 60G_4(\omega_1/\omega_2)x - 140G_6(\omega_1/\omega_2)$$

is $16D(\omega_1/\omega_2)$, where

$$D(z) = (60G_4(z))^3 - 27(140G_6(z))^2.$$

Since $D(z)$ is the difference of two modular forms of weight 12 it has weight 12. Moreover,

$$\begin{aligned} D(\infty) &= (60G_4(\infty))^3 - 27\,(140G_6(\infty))^2 \\ &= \left(\frac{60}{3^2\cdot 5}\pi^4\right)^3 - 27\left(\frac{140\cdot 2}{3^3\cdot 5\cdot 7}\pi^6\right)^2 \\ &= 0, \end{aligned}$$

so D is a cusp form of weight 12. Let

$$\Delta = \frac{D}{(2\pi)^{12}}.$$

Lemma 2.3. *If $z \in \mathfrak{h}$, then $\Delta(z) \neq 0$.*

Proof. Let $\omega_1 = z$ and $\omega_2 = 1$. Since $E = \mathbb{C}/(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$ is an elliptic curve, it has nonzero discriminant $\Delta(z) = \Delta(\omega_1/\omega_2) \neq 0$. □

Proposition 2.4. *We have $\Delta = q \cdot \prod_{n=1}^{\infty}(1-q^n)^{24}$.*

Proof. See [**Ser73**, Thm. 6, pg. 95]. □

Remark 2.5. SAGE computes the q-expansion of Δ efficiently to high precision using the command `delta_qexp`:

```
sage: delta_qexp(6)
q - 24*q^2 + 252*q^3 - 1472*q^4 + 4830*q^5 + O(q^6)
```

2.1.2. Fourier Expansions of Eisenstein Series. Recall from (1.2.4) that elements f of $M_k(\mathrm{SL}_2(\mathbb{Z}))$ can be expressed as formal power series in terms of $q(z) = e^{2\pi i z}$ and that this expansion is called the Fourier expansion of f. The following proposition gives the Fourier expansion of the Eisenstein series $G_k(z)$.

Definition 2.6 (Sigma)**.** For any integer $t \geq 0$ and any positive integer n, the *sigma function*

$$\sigma_t(n) = \sum_{1 \leq d | n} d^t$$

is the sum of the tth powers of the positive divisors of n. Also, let $d(n) = \sigma_0(n)$, which is the number of divisors of n, and let $\sigma(n) = \sigma_1(n)$. For example, if p is prime, then $\sigma_t(p) = 1 + p^t$.

Proposition 2.7. *For every even integer $k \geq 4$, we have*

$$G_k(z) = 2\zeta(k) + 2\cdot\frac{(2\pi i)^k}{(k-1)!}\cdot\sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n.$$

Proof. See [**Ser73**, Section VII.4], which uses clever manipulations of series, starting with the identity

$$\pi\cot(\pi z) = \frac{1}{z} + \sum_{m=1}^{\infty}\left(\frac{1}{z+m} + \frac{1}{z-m}\right).$$

□

From a computational point of view, the q-expansion of Proposition 2.7 is unsatisfactory because it involves transcendental numbers. To understand these numbers, we introduce the *Bernoulli numbers* B_n for $n \geq 0$ *defined* by the following equality of formal power series:

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}. \tag{2.1.2}$$

Expanding the power series, we have

$$\frac{x}{e^x-1} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} - \frac{x^8}{1209600} + \cdots .$$

As this expansion suggests, the Bernoulli numbers B_n with $n > 1$ odd are 0 (see Exercise 1.2). Expanding the series further, we obtain the following table:

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30},$$

$$B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}, \quad B_{14} = \frac{7}{6}, \quad B_{16} = -\frac{3617}{510}, \quad B_{18} = \frac{43867}{798},$$

$$B_{20} = -\frac{174611}{330}, \quad B_{22} = \frac{854513}{138}, \quad B_{24} = -\frac{236364091}{2730}, \quad B_{26} = \frac{8553103}{6}.$$

See Section 2.7 for a discussion of fast (analytic) methods for computing Bernoulli numbers.

We compute some Bernoulli numbers in SAGE:

```
sage: bernoulli(12)
-691/2730
sage: bernoulli(50)
495057205241079648212477525/66
sage: len(str(bernoulli(10000)))
27706
```

A key fact is that Bernoulli numbers are rational numbers and they are connected to values of ζ at positive even integers.

Proposition 2.8. *If $k \geq 2$ is an even integer, then*

$$\zeta(k) = -\frac{(2\pi i)^k}{2 \cdot k!} \cdot B_k.$$

Proof. This is proved by manipulating a series expansion of $z\cot(z)$ (see [**Ser73**, Section VII.4]). □

Definition 2.9 (Normalized Eisenstein Series)**.** The *normalized Eisenstein series* of even weight $k \geq 4$ is

$$E_k = \frac{(k-1)!}{2 \cdot (2\pi i)^k} \cdot G_k.$$

Combining Propositions 2.7 and 2.8, we see that

$$E_k = -\frac{B_k}{2k} + q + \sum_{n=2}^{\infty} \sigma_{k-1}(n) q^n. \tag{2.1.3}$$

Warning 2.10. Our series E_k is normalized so that the coefficient of q is 1, but often in the literature E_k is normalized so that the constant coefficient is 1. We use the normalization with the coefficient of q equal to 1, because then the eigenvalue of the nth Hecke operator (see Section 2.4) is the coefficient of q^n. Our normalization is also convenient when considering congruences between cusp forms and Eisenstein series.

2.2. Structure Theorem for Level 1 Modular Forms

In this section we describe a structure theorem for modular forms of level 1. If f is a nonzero meromorphic function on $\mathfrak{h}$ and $w \in \mathfrak{h}$, let $\mathrm{ord}_w(f)$ be the largest integer n such that $f(z)/(w-z)^n$ is holomorphic at w. If $f = \sum_{n=m}^{\infty} a_n q^n$ with $a_m \neq 0$, we set $\mathrm{ord}_\infty(f) = m$. We will use the following theorem to give a presentation for the vector space of modular forms of weight k; this presentation yields an algorithm to compute this space.

Let $M_k = M_k(\mathrm{SL}_2(\mathbb{Z}))$ denote the complex vector space of modular forms of weight k for $\mathrm{SL}_2(\mathbb{Z})$. The *standard fundamental domain* $\mathcal{F}$ for $\mathrm{SL}_2(\mathbb{Z})$ is the set of $z \in \mathfrak{h}$ with $|z| \geq 1$ and $|\mathrm{Re}(z)| \leq 1/2$. Let $\rho = e^{2\pi i/3}$.

Theorem 2.11 (Valence Formula)**.** *Let k be any integer and suppose $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ is nonzero. Then*

$$\mathrm{ord}_\infty(f) + \frac{1}{2}\mathrm{ord}_i(f) + \frac{1}{3}\mathrm{ord}_\rho(f) + \sum_{w \in \mathcal{F}}^{*} \mathrm{ord}_w(f) = \frac{k}{12},$$

where $\displaystyle\sum_{w \in \mathcal{F}}^{*}$ is the sum over elements of $\mathcal{F}$ other than i and ρ.

Proof. The proof in [**Ser73**, §VII.3] uses the residue theorem. □

Let $S_k = S_k(\mathrm{SL}_2(\mathbb{Z}))$ denote the subspace of weight k cusp forms for $\mathrm{SL}_2(\mathbb{Z})$. We have an exact sequence

$$0 \to S_k \to M_k \xrightarrow{\iota_\infty} \mathbb{C}$$

that sends $f \in M_k$ to $f(\infty)$. When $k \geq 4$ is even, the space M_k contains the Eisenstein series G_k, and $G_k(\infty) = 2\zeta(k) \neq 0$, so the map $M_k \to \mathbb{C}$ is surjective. This proves the following lemma.

Lemma 2.12. *If $k \geq 4$ is even, then $M_k = S_k \oplus \mathbb{C}G_k$ and the following sequence is exact:*

$$0 \to S_k \to M_k \xrightarrow{\iota_\infty} \mathbb{C} \to 0.$$

Proposition 2.13. *For $k < 0$ and $k = 2$, we have $M_k = 0$.*

Proof. Suppose $f \in M_k$ is nonzero yet $k = 2$ or $k < 0$. By Theorem 2.11,

$$\mathrm{ord}_\infty(f) + \frac{1}{2}\mathrm{ord}_i(f) + \frac{1}{3}\mathrm{ord}_\rho(f) + \sum_{w\in D}^{*} \mathrm{ord}_w(f) = \frac{k}{12} \leq \frac{1}{6}.$$

This is not possible because each quantity on the left is nonnegative so whatever the sum is, it is too big (or 0, in which case $k = 0$). □

Theorem 2.14. *Multiplication by Δ defines an isomorphism $M_{k-12} \to S_k$.*

Proof. By Lemma 2.3, Δ is not identically 0, so because Δ is holomorphic, multiplication by Δ defines an injective map $M_{k-12} \hookrightarrow S_k$. To see that this map is surjective, we show that if $f \in S_k$, then $f/\Delta \in M_{k-12}$. Since Δ has weight 12 and $\mathrm{ord}_\infty(\Delta) \geq 1$, Theorem 2.11 implies that Δ has a simple zero at ∞ and does not vanish on $\mathfrak{h}$. Thus if $f \in S_k$ and if we let $g = f/\Delta$, then g is holomorphic and satisfies the appropriate transformation formula, so $g \in M_{k-12}$. □

Corollary 2.15. *For $k = 0, 4, 6, 8, 10, 14$, the space M_k has dimension 1, with basis 1, G_4, G_6, G_8, G_{10}, and G_{14}, respectively, and $S_k = 0$.*

Proof. Combining Proposition 2.13 with Theorem 2.14, we see that the spaces M_k for $k \leq 10$ cannot have dimension greater than 1, since otherwise $M_{k'} \neq 0$ for some $k' < 0$. Also M_{14} has dimension at most 1, since M_2 has dimension 0. Each of the indicated spaces of weight ≥ 4 contains the indicated Eisenstein series and so has dimension 1, as claimed. □

Corollary 2.16. $\dim M_k = \begin{cases} 0 & \textit{if } k \textit{ is odd or negative,} \\ \lfloor k/12 \rfloor & \textit{if } k \equiv 2 \pmod{12}, \\ \lfloor k/12 \rfloor + 1 & \textit{if } k \not\equiv 2 \pmod{12}. \end{cases}$

Here $\lfloor x \rfloor$ is the biggest integer $\leq x$.

Proof. As we have already seen above, the formula is true when $k \leq 12$. By Theorem 2.14, the dimension increases by 1 when k is replaced by $k+12$. □

Theorem 2.17. *The space M_k has as basis the modular forms $G_4^a G_6^b$, where a, b run over all pairs of nonnegative integers such that $4a + 6b = k$.*

Proof. Fix an even integer k. We first prove by induction that the modular forms $G_4^a G_6^b$ generate M_k; the cases $k \leq 10$ and $k = 14$ follow from the above arguments (e.g., when $k = 0$, we have $a = b = 0$ and basis 1). Choose some pair of nonnegative integers a, b such that $4a + 6b = k$. The form $g = G_4^a G_6^b$ is not a cusp form, since it is nonzero at ∞. Now suppose $f \in M_k$ is arbitrary. Since $g(\infty) \neq 0$, there exists $\alpha \in \mathbb{C}$ such that $f - \alpha g \in S_k$. Then by Theorem 2.14, there is $h \in M_{k-12}$ such that $f - \alpha g = \Delta \cdot h$. By induction, h is a polynomial in G_4 and G_6 of the required type, and so is Δ, so f is as well. Thus

$$\{G_4^a G_6^b \mid a \geq 0,\ b \geq 0,\ 4a + 6b = k\}$$

spans M_k.

Suppose there is a nontrivial linear relation between the $G_4^a G_6^b$ for a given k. By multiplying the linear relation by a suitable power of G_4 and G_6, we may assume that we have such a nontrivial relation with $k \equiv 0 \pmod{12}$. Now divide the linear relation by the weight k form $G_6^{k/6}$ to see that G_4^3/G_6^2 satisfies a polynomial with coefficients in $\mathbb{C}$ (see Exercise 2.4). Hence G_4^3/G_6^2 is a root of a polynomial, hence a constant, which is a contradiction since the q-expansion of G_4^3/G_6^2 is not constant. □

Algorithm 2.18 (Basis for M_k)**.** *Given integers n and k, this algorithm computes a basis of q-expansions for the complex vector space M_k mod q^n. The q-expansions output by this algorithm have coefficients in $\mathbb{Q}$.*

(1) [Simple Case] If $k = 0$, output the basis with just 1 in it and terminate; otherwise if $k < 4$ or k is odd, output the empty basis and terminate.
(2) [Power Series] Compute E_4 and E_6 mod q^n using the formula from (2.1.3) and Section 2.7.
(3) [Initialize] Set $b = 0$.
(4) [Enumerate Basis] For each integer b between 0 and $\lfloor k/6 \rfloor$, compute $a = (k - 6b)/4$. If a is an integer, compute and output the basis element $E_4^a E_6^b \mod q^n$. When computing E_4^a, find $E_4^m \pmod{q^n}$ for each $m \leq a$, and save these intermediate powers, so they can be reused later, and likewise for powers of E_6.

Proof. This is simply a translation of Theorem 2.17 into an algorithm, since E_k is a nonzero scalar multiple of G_k. That the q-expansions have coefficients in $\mathbb{Q}$ follows from (2.1.3). □

Example 2.19. We compute a basis for M_{24}, which is the space with smallest weight whose dimension is greater than 1. It has as basis E_4^6, $E_4^3E_6^2$, and E_6^4, whose explicit expansions are

$$\begin{aligned} E_4^6 &= \frac{1}{191102976000000} + \frac{1}{132710400000}q + \frac{203}{44236800000}q^2 + \cdots, \\ E_4^3E_6^2 &= \frac{1}{3511517184000} - \frac{1}{12192768000}q - \frac{377}{4064256000}q^2 + \cdots, \\ E_6^4 &= \frac{1}{64524128256} - \frac{1}{32006016}q + \frac{241}{10668672}q^2 + \cdots. \end{aligned}$$

We compute this basis in SAGE as follows:

```
sage: E4 = eisenstein_series_qexp(4, 3)
sage: E6 = eisenstein_series_qexp(6, 3)
sage: E4^6
1/191102976000000 + 1/132710400000*q
                  + 203/44236800000*q^2 + O(q^3)
sage: E4^3*E6^2
1/3511517184000 - 1/12192768000*q
                - 377/4064256000*q^2 + O(q^3)
sage: E6^4
1/64524128256 - 1/32006016*q + 241/10668672*q^2 + O(q^3)
```

In Section 2.3, we will discuss the reduced echelon form basis for M_k.

2.3. The Miller Basis

Lemma 2.20 (V. Miller). *The space S_k has a basis $f_1, \ldots, f_d$ such that if $a_i(f_j)$ is the ith coefficient of f_j, then $a_i(f_j) = \delta_{i,j}$ for $i = 1, \ldots, d$. Moreover the f_j all lie in $\mathbb{Z}[[q]]$. We call this basis the* Miller basis *for S_k.*

This is a straightforward construction involving E_4, E_6 and Δ. The following proof very closely follows [**Lan95**, Ch. X, Thm. 4.4], which in turn follows the first lemma of V. Miller's thesis.

Proof. Let $d = \dim S_k$. Since $B_4 = -1/30$ and $B_6 = 1/42$, we note that

$$F_4 = -\frac{8}{B_4} \cdot E_4 = 1 + 240q + 2160q^2 + 6720q^3 + 17520q^4 + \cdots$$

and

$$F_6 = -\frac{12}{B_6} \cdot E_6 = 1 - 504q - 16632q^2 - 122976q^3 - 532728q^4 + \cdots$$

have q-expansions in $\mathbb{Z}[[q]]$ with leading coefficient 1. Choose integers $a, b \geq 0$ such that

$$4a + 6b \leq 14 \qquad \text{and} \qquad 4a + 6b \equiv k \pmod{12},$$

with $a = b = 0$ when $k \equiv 0 \pmod{12}$, and let

$$g_j = \Delta^j F_6^{2(d-j)+b} F_4^a = \left(\frac{\Delta}{F_6^2}\right)^j F_6^{2d+b} F_4^a, \qquad\qquad \text{for } j = 1, \ldots, d.$$

Then it is elementary to check that g_j has weight k

$$a_j(g_j) = 1 \qquad \text{and} \qquad a_i(g_j) = 0 \qquad \text{when} \qquad i < j.$$

Hence the g_j are linearly independent over $\mathbb{C}$, so form a basis for S_k. Since F_4, F_6, and Δ are all in $\mathbb{Z}[[q]]$, so are the g_j. The f_i may then be constructed from the g_j by Gauss elimination. The coefficients of the resulting power series lie in $\mathbb{Z}$ because each time we clear a column we use the power series g_j whose leading coefficient is 1 (so no denominators are introduced). □

Remark 2.21. The basis coming from Miller's lemma is "canonical", since it is just the reduced row echelon form of any basis. Also the set of all *integral* linear combinations of the elements of the Miller basis are precisely the modular forms of level 1 with integral q-expansion.

We extend the Miller basis to all M_k by taking a multiple of G_k with constant term 1 and subtracting off the f_i from the Miller basis so that the coefficients of $q, q^2, \ldots q^d$ of the resulting expansion are 0. We call the extra basis element f_0.

Example 2.22. If $k = 24$, then $d = 2$. Choose $a = b = 0$, since $k \equiv 0 \pmod{12}$. Then

$$g_1 = \Delta F_6^2 = q - 1032q^2 + 245196q^3 + 10965568q^4 + 60177390q^5 - \cdots$$

and

$$g_2 = \Delta^2 = q^2 - 48q^3 + 1080q^4 - 15040q^5 + \cdots.$$

We let $f_2 = g_2$ and

$$f_1 = g_1 + 1032g_2 = q + 195660q^3 + 12080128q^4 + 44656110q^5 - \cdots.$$

Example 2.23. When $k = 36$, the Miller basis including f_0 is

$$\begin{aligned}
f_0 &= 1 + \quad 6218175600q^4 + 15281788354560q^5 + \cdots ,\\
f_1 &= \quad q + \quad 57093088q^4 + \quad 37927345230q^5 + \cdots ,\\
f_2 &= \qquad q^2 + \quad 194184q^4 + \qquad 7442432q^5 + \cdots ,\\
f_3 &= \qquad\quad q^3 - \quad 72q^4 + \qquad\quad 2484q^5 + \cdots .
\end{aligned}$$

Example 2.24. The SAGE command `victor_miller_basis` computes the Miller basis to any desired precision for a given k.

```
sage: victor_miller_basis(28,5)
[
1 + 15590400*q^3 + 36957286800*q^4 + O(q^5),
q + 151740*q^3 + 61032448*q^4 + O(q^5),
q^2 + 192*q^3 - 8280*q^4 + O(q^5)
]
```

Remark 2.25. To write $f \in M_k$ as a polynomial in E_4 and E_6, it is wasteful to compute the Miller basis. Instead, use the upper triangular (but not echelon!) basis $\Delta^j F_6^{2(d-j)+a} F_4^b$, and match coefficients from q^0 to q^d.

2.4. Hecke Operators

In this section we define Hecke operators on level 1 modular forms and derive their basic properties. We will not give proofs of the analogous properties for Hecke operators on higher level modular forms, since the proofs are clearest in the level 1 case, and the general case is similar (see, e.g., [**Lan95**]).

For any positive integer n, let

$$X_n = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}) \;:\; a \geq 1,\ ad = n,\ \text{and } 0 \leq b < d \right\}.$$

Note that the set X_n is in bijection with the set of subgroups of $\mathbb{Z}^2$ of index n, where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ corresponds to $L = \mathbb{Z} \cdot (a, b) + \mathbb{Z} \cdot (0, d)$, as one can see using Hermite normal form, which is the analogue over $\mathbb{Z}$ of echelon form (see Exercise 7.5).

Recall from (1.3.1) that if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$, then

$$f^{[\gamma]_k} = \det(\gamma)^{k-1}(cz + d)^{-k} f(\gamma(z)).$$

Definition 2.26 (Hecke Operator $T_{n,k}$)**.** The nth Hecke operator $T_{n,k}$ of weight k is the operator on the set of functions on $\mathfrak{h}$ defined by

$$T_{n,k}(f) = \sum_{\gamma \in X_n} f^{[\gamma]_k}.$$

Remark 2.27. It would make more sense to write $T_{n,k}$ on the right, e.g., $f|T_{n,k}$, since $T_{n,k}$ is defined using a right group action. However, if n, m are integers, then the action of $T_{n,k}$ and $T_{m,k}$ on weakly modular functions commutes (by Proposition 2.29 below), so it makes no difference whether we view the Hecke operators of given weight k as acting on the right or left.

Proposition 2.28. *If f is a weakly modular function of weight k, then so is $T_{n,k}(f)$; if f is a modular function, then so is $T_{n,k}(f)$.*

Proof. Suppose $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Since γ induces an automorphism of $\mathbb{Z}^2$,

$$X_n \cdot \gamma = \{\delta\gamma : \delta \in X_n\}$$

is also in bijection with the subgroups of $\mathbb{Z}^2$ of index n. For each element $\delta\gamma \in X_n \cdot \gamma$, there is $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\sigma\delta\gamma \in X_n$ (the element σ transforms $\delta\gamma$ to Hermite normal form), and the set of elements $\sigma\delta\gamma$ is thus equal to X_n. Thus

$$T_{n,k}(f) = \sum_{\sigma\delta\gamma \in X_n} f^{[\sigma\delta\gamma]_k} = \sum_{\delta \in X_n} f^{[\delta\gamma]_k} = T_{n,k}(f)^{[\gamma]_k}.$$

A finite sum of meromorphic function is meromorphic, so $T_{n,k}(f)$ is weakly modular. If f is holomorphic on $\mathfrak{h}$, then each $f^{[\delta]_k}$ is holomorphic on $\mathfrak{h}$ for $\delta \in X_n$. A finite sum of holomorphic functions is holomorphic, so $T_{n,k}(f)$ is holomorphic.

□

We will frequently drop k from the notation in $T_{n,k}$, since the weight k is implicit in the modular function to which we apply the Hecke operator. Henceforth we make the convention that if we write $T_n(f)$ and if f is modular, then we mean $T_{n,k}(f)$, where k is the weight of f.

Proposition 2.29. *On weight k modular functions we have*

$$T_{mn} = T_m T_n \qquad \text{if } (m,n) = 1, \tag{2.4.1}$$

and

$$T_{p^n} = T_{p^{n-1}} T_p - p^{k-1} T_{p^{n-2}} \qquad \text{if } p \text{ is prime.} \tag{2.4.2}$$

Proof. Let L be a subgroup of index mn. The quotient $\mathbb{Z}^2/L$ is an abelian group of order mn, and $(m,n) = 1$, so $\mathbb{Z}^2/L$ decomposes uniquely as a direct sum of a subgroup of order m with a subgroup of order n. Thus there exists a unique subgroup L' such that $L \subset L' \subset \mathbb{Z}^2$, and L' has index m in $\mathbb{Z}^2$. The subgroup L' corresponds to an element of X_m, and the index n subgroup $L \subset L'$ corresponds to multiplying that element on the right by some uniquely determined element of X_n. We thus have

$$\mathrm{SL}_2(\mathbb{Z}) \cdot X_m \cdot X_n = \mathrm{SL}_2(\mathbb{Z}) \cdot X_{mn},$$

i.e., the set products of elements in X_m with elements of X_n equal the elements of X_{mn}, up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence. Thus for any f, we have $T_{mn}(f) = T_n(T_m(f))$. Applying this formula with m and n swapped yields the equality $T_{mn} = T_m T_n$.

We will show that $T_{p^n} + p^{k-1}T_{p^{n-2}} = T_p T_{p^{n-1}}$. Suppose f is a weight k weakly modular function. Using that $f^{[\left(\begin{smallmatrix} p & 0 \\ 0 & p \end{smallmatrix}\right)]_k} = (p^2)^{k-1}p^{-k}f = p^{k-2}f$, we have

$$\sum_{x \in X_{p^n}} f^{[x]_k} \; + \; p^{k-1} \sum_{x \in X_{p^{n-2}}} f^{[x]_k} = \sum_{x \in X_{p^n}} f^{[x]_k} \; + p \sum_{x \in pX_{p^{n-2}}} f^{[x]_k}.$$

Also

$$T_p T_{p^{n-1}}(f) = \sum_{y \in X_p} \sum_{x \in X_{p^{n-1}}} (f^{[x]_k})^{[y]_k} = \sum_{x \in X_{p^{n-1}} \cdot X_p} f^{[x]_k}.$$

Thus it suffices to show that X_{p^n} disjoint union p copies of $pX_{p^{n-2}}$ is equal to $X_{p^{n-1}} \cdot X_p$, where we consider elements with multiplicities and up to left $\mathrm{SL}_2(\mathbb{Z})$-equivalence (i.e., the left action of $\mathrm{SL}_2(\mathbb{Z})$).

Suppose L is a subgroup of $\mathbb{Z}^2$ of index p^n, so L corresponds to an element of X_{p^n}. First suppose L is not contained in $p\mathbb{Z}^2$. Then the image of L in $\mathbb{Z}^2/p\mathbb{Z}^2 = (\mathbb{Z}/p\mathbb{Z})^2$ is of order p, so if $L' = p\mathbb{Z}^2 + L$, then $[\mathbb{Z}^2 : L'] = p$ and $[L : L'] = p^{n-1}$, and L' is the only subgroup with this property. Second, suppose that $L \subset p\mathbb{Z}^2$ if of index p^n and that $x \in X_{p^n}$ corresponds to L. Then every one of the $p+1$ subgroups $L' \subset \mathbb{Z}^2$ of index p contains L. Thus there are $p+1$ chains $L \subset L' \subset \mathbb{Z}^2$ with $[\mathbb{Z}^2 : L'] = p$.

The chains $L \subset L' \subset \mathbb{Z}^2$ with $[\mathbb{Z}^2 : L'] = p$ and $[\mathbb{Z}^2 : L] = p^{n-1}$ are in bijection with the elements of $X_{p^{n-1}} \cdot X_p$. On the other hand the union of X_{p^n} with p copies of $pX_{p^{n-2}}$ corresponds to the subgroups L of index p^n, but with those that contain $p\mathbb{Z}^2$ counted $p+1$ times. The structure of the set of chains $L \subset L' \subset \mathbb{Z}^2$ that we derived in the previous paragraph gives the result. □

Corollary 2.30. *The Hecke operator T_{p^n}, for prime p, is a polynomial in T_p with integer coefficients, i.e., $T_{p^n} \in \mathbb{Z}[T_p]$. If n, m are any integers, then $T_n T_m = T_m T_n$.*

Proof. The first statement follows from (2.4.2) of Proposition 2.29. It then follows that $T_n T_m = T_m T_n$ when m and n are both powers of a single prime p. Combining this with (2.4.1) gives the second statement in general. □

Proposition 2.31. *Let $f = \sum_{n \in \mathbb{Z}} a_n q^n$ be a modular function of weight k. Then*

$$T_n(f) = \sum_{m \in \mathbb{Z}} \left(\sum_{1 \leq d \,|\, \gcd(n,m)} d^{k-1} a_{mn/d^2} \right) q^m.$$

In particular, if $n = p$ is prime, then

$$T_p(f) = \sum_{m\in\mathbb{Z}} \left(a_{mp} + p^{k-1}a_{m/p}\right) q^m,$$

where $a_{m/p} = 0$ if $m/p \notin \mathbb{Z}$.

Proof. This is proved in [**Ser73**, §VII.5.3] by writing out $T_n(f)$ explicitly and using that $\sum_{0\leq b<d} e^{2\pi ibm/d}$ is d if $d \mid m$ and 0 otherwise. □

Corollary 2.32. *The Hecke operators preserve M_k and S_k.*

Remark 2.33. Alternatively, for M_k the above corollary is Proposition 2.28, and for S_k we see from the definitions that if $f(\infty) = 0$, then $T_n f$ also vanishes at ∞.

Example 2.34. Recall from (2.1.3) that

$$E_4 = \frac{1}{240} + q + 9q^2 + 28q^3 + 73q^4 + 126q^5 + 252q^6 + 344q^7 + \cdots .$$

Using the formula of Proposition 2.31, we see that

$$T_2(E_4) = (1/240 + 2^3 \cdot (1/240)) + 9q + (73 + 2^3 \cdot 1)q^2 + \cdots .$$

Since M_4 has dimension 1 and since we have proved that T_2 preserves M_4, we know that T_2 acts as a scalar. Thus we know just from the constant coefficient of $T_2(E_4)$ that

$$T_2(E_4) = 9E_4.$$

More generally, for p prime we see by inspection of the constant coefficient of $T_p(E_4)$ that

$$T_p(E_4) = (1 + p^3)E_4.$$

In fact $T_n(E_k) = \sigma_{k-1}(n)E_k$, for any integer $n \geq 1$ and even weight $k \geq 4$.

Example 2.35. By Corollary 2.32, the Hecke operators T_n also preserve the subspace S_k of M_k. Since S_{12} has dimension 1 (spanned by Δ), we see that Δ is an eigenvector for every T_n. Since the coefficient of q in the q-expansion of Δ is 1, the eigenvalue of T_n on Δ is the nth coefficient of Δ. Since $T_{nm} = T_n T_m$ for $\gcd(n, m) = 1$, we have proved the nonobvious fact that the *Ramanujan function* $\tau(n)$ that gives the nth coefficient of Δ is a multiplicative function, i.e., if $\gcd(n, m) = 1$, then $\tau(nm) = \tau(n)\tau(m)$.

Remark 2.36. The Hecke operators respect the decomposition $M_k = S_k \oplus \mathbb{C}E_k$, i.e., for all k the series E_k are eigenvectors for all T_n.

2.5. Computing Hecke Operators

This section is about how to compute matrices of Hecke operators on M_k.

Algorithm 2.37 (Hecke Operator)**.** *This algorithm computes the matrix of the Hecke operator T_n on the Miller basis for M_k.*

(1) [Dimension] Compute $d = \dim(M_k) - 1$ using Corollary 2.16.
(2) [Basis] Using Lemma 2.20, compute the echelon basis $f_0, \dots, f_d$ for $M_k \pmod{q^{dn+1}}$.
(3) [Hecke operator] Using Proposition 2.31, compute for each i the image $T_n(f_i) \pmod{q^{d+1}}$.
(4) [Write in terms of basis] The elements $T_n(f_i) \pmod{q^{d+1}}$ determine linear combinations of
$$f_0, f_1, \dots, f_d \pmod{q^d}.$$
These linear combinations are easy to find once we compute $T_n(f_i)$ (mod q^{d+1}), since our basis of f_i is in echelon form. The linear combinations are just the coefficients of the power series $T_n(f_i)$ up to and including q^d.
(5) [Write down matrix] The matrix of T_n acting from the right relative to the basis $f_0, \dots, f_d$ is the matrix whose rows are the linear combinations found in the previous step, i.e., whose rows are the coefficients of $T_n(f_i)$.

Proof. By Proposition 2.31, the dth coefficient of $T_n(f)$ involves only a_{dn} and smaller-indexed coefficients of f. We need only compute a modular form f modulo q^{dn+1} in order to compute $T_n(f)$ modulo q^{d+1}. Uniqueness in step (4) follows from Lemma 2.20 above. □

Example 2.38. We compute the Hecke operator T_2 on M_{12} using the above algorithm.

(1) [Compute dimension] We have $d = 2 - 1 = 1$.
(2) [Compute basis] Compute up to (but not including) the coefficient of $q^{dn+1} = q^{1\cdot 2+1} = q^3$. As given in the proof of Lemma 2.20, we have
$$F_4 = 1 + 240q + 2160q^2 + \cdots \quad \text{and} \quad F_6 = 1 - 504q - 16632q^2 + \cdots .$$
Thus M_{12} has basis
$$F_4^3 = 1+720q+179280q^2+\cdots \quad \text{and} \quad \Delta = (F_4^3 - F_6^2)/1728 = q - 24q^2 + \cdots .$$
Subtracting 720Δ from F_4^3 yields the echelon basis, which is
$$f_0 = 1 + 196560q^2 + \cdots \quad \text{and} \quad f_1 = q - 24q^2 + \cdots .$$
SAGE does the arithmetic in the above calculation as follows:

```
sage: R.<q> = QQ[['q']]
sage: F4 =  240 * eisenstein_series_qexp(4,3)
sage: F6 = -504 * eisenstein_series_qexp(6,3)
sage: F4^3
1 + 720*q + 179280*q^2 + O(q^3)
sage: Delta = (F4^3 - F6^2)/1728; Delta
q - 24*q^2 + O(q^3)
sage: F4^3 - 720*Delta
1 + 196560*q^2 + O(q^3)
```

(3) [Compute Hecke operator] In each case letting a_n denote the nth coefficient of f_0 or f_1, respectively, we have

$$\begin{aligned} T_2(f_0) &= T_2(1 + 196560q^2 + \cdots) \\ &= (a_0 + 2^{11}a_0)q^0 + (a_2 + 2^{11}a_{1/2})q^1 + \cdots \\ &= 2049 + 196560q + \cdots, \end{aligned}$$

and

$$\begin{aligned} T_2(f_1) &= T_2(q - 24q^2 + \cdots) \\ &= (a_0 + 2^{11}a_0)q^0 + (a_2 + 2^{11}a_{1/2})q^1 + \cdots \\ &= 0 - 24q + \cdots. \end{aligned}$$

(Note that $a_{1/2} = 0$.)

(4) [Write in terms of basis] We read off at once that

$$T_2(f_0) = 2049 f_0 + 196560 f_1 \quad \text{and} \quad T_2(f_1) = 0 f_0 + (-24) f_1.$$

(5) [Write down matrix] Thus the matrix of T_2, acting from the right on the basis f_0, f_1, is

$$T_2 = \begin{pmatrix} 2049 & 196560 \\ 0 & -24 \end{pmatrix}.$$

As a check note that the characteristic polynomial of T_2 is $(x-2049)(x+24)$ and that $2049 = 1 + 2^{11}$ is the sum of the 11th powers of the divisors of 2.

Example 2.39. The Hecke operator T_2 on M_{36} with respect to the echelon basis is

$$\begin{pmatrix} 34359738369 & 0 & 6218175600 & 9026867482214400 \\ 0 & 0 & 34416831456 & 5681332472832 \\ 0 & 1 & 194184 & -197264484 \\ 0 & 0 & -72 & -54528 \end{pmatrix}.$$

It has characteristic polynomial

$$(x - 34359738369) \cdot (x^3 - 139656x^2 - 59208339456x - 1467625047588864),$$

where the cubic factor is irreducible.

The `echelon_form()` command creates the space of modular forms but with basis in echelon form (which is not the default).

```
sage: M = ModularForms(1,36, prec=6).echelon_form()
sage: M.basis()
[
1 + 6218175600*q^4 + 15281788354560*q^5 + O(q^6),
q + 57093088*q^4 + 37927345230*q^5 + O(q^6),
q^2 + 194184*q^4 + 7442432*q^5 + O(q^6),
q^3 - 72*q^4 + 2484*q^5 + O(q^6)
]
```

Next we compute the matrix of the Hecke operator T_2.

```
sage: T2 = M.hecke_matrix(2); T2
[34359738369   0       6218175600 9026867482214400]
[          0   0      34416831456    5681332472832]
[          0   1           194184       -197264484]
[          0   0              -72           -54528]
```

Finally we compute and factor its characteristic polynomial.

```
sage: T2.charpoly().factor()
(x - 34359738369) *
   (x^3 - 139656*x^2 - 59208339456*x - 1467625047588864)
```

The following is a famous open problem about Hecke operators on modular forms of level 1. It generalizes our above observation that the characteristic polynomial of T_2 on M_k, for $k = 12, 36$, factors as a product of a linear factor and an irreducible factor.

Conjecture 2.40 (Maeda)**.** *The characteristic polynomial of T_2 on S_k is irreducible for any k.*

Kevin Buzzard observed that in several specific cases the Galois group of the characteristic polynomial of T_2 is the full symmetric group (see [**Buz96**]). See also [**FJ02**] for more evidence for the following conjecture:

Conjecture 2.41. *For all primes p and all even $k \geq 2$ the characteristic polynomial of $T_{p,k}$ acting on S_k is irreducible.*

2.6. Fast Computation of Fourier Coefficients

How difficult is it to compute prime-indexed coefficients of

$$\Delta = \sum_{n=1}^{\infty} \tau(n) q^n = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 + \cdots ?$$

Theorem 2.42 (Bosman, Couveignes, Edixhoven, de Jong, Merkl)**.** *Let p be a prime. There is a probabilistic algorithm to compute $\tau(p)$, for prime p, that has expected running time polynomial in* $\log(p)$*.*

Proof. See [**ECdJ$^+$06**]. □

More generally, if $f = \sum a_n q^n$ is an eigenform in some space $M_k(\Gamma_1(N))$, where $k \geq 2$, then one expects that there is an algorithm to compute a_p in time polynomial in $\log(p)$. Bas Edixhoven, Jean-Marc Couveignes and Robin de Jong have proved that $\tau(p)$ can be computed in polynomial time; their approach involves sophisticated techniques from arithmetic geometry (e.g., étale cohomology, motives, Arakelov theory). The ideas they use are inspired by the ones introduced by Schoof, Elkies and Atkin for quickly counting points on elliptic curves over finite fields (see [**Sch95**]).

Edixhoven describes (in an email to the author) the strategy as follows:

(1) We compute the mod ℓ Galois representation ρ associated to Δ. In particular, we produce a polynomial f such that $\mathbb{Q}[x]/(f)$ is the fixed field of $\ker(\rho)$. This is then used to obtain $\tau(p) \pmod{\ell}$ and to do a Schoof-like algorithm for computing $\tau(p)$.

(2) We compute the field of definition of suitable points of order ℓ on the modular Jacobian $J_1(\ell)$ to do part (1) (see [**DS05**, Ch. 6] for the definition of $J_1(\ell)$).

(3) The method is to approximate the polynomial f in some sense (e.g., over the complex numbers or modulo many small primes r) and to use an estimate from Arakelov theory to determine a precision that will suffice.

2.7. Fast Computation of Bernoulli Numbers

This section, which was written jointly with Kevin McGown, is about computing the Bernoulli numbers B_n, for $n \geq 0$, defined in Section 2.1.2 by

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}. \tag{2.7.1}$$

One way to compute B_n is to multiply both sides of (2.7.1) by $e^x - 1$ and equate coefficients of x^{n+1} to obtain the recurrence

$$B_0 = 1, \qquad B_n = -\frac{1}{n+1} \cdot \sum_{k=0}^{n-1} \binom{n+1}{k} B_k.$$

This recurrence provides a straightforward and easy-to-implement method for calculating B_n if one is interested in computing B_n for all n up to some bound. For example,

$$B_1 = -\frac{1}{2} \cdot \left(\binom{2}{0} B_0 \right) = -\frac{1}{2}$$

and

$$B_2 = -\frac{1}{3} \cdot \left(\binom{3}{0} B_0 + \binom{3}{1} B_1 \right) = -\frac{1}{3} \cdot \left(1 - \frac{3}{2} \right) = \frac{1}{6}.$$

However, computing B_n via the recurrence is slow; it requires summing over many large terms, it requires storing the numbers $B_0, \ldots, B_{n-1}$, and it takes only limited advantage of asymptotically fast arithmetic algorithms. There is also an inductive procedure to compute Bernoulli numbers that resembles Pascal's triangle called the Akiyama-Tanigawa algorithm (see [**Kan00**]).

Another approach to computing B_n is to use Newton iteration and asymptotically fast polynomial arithmetic to approximate $1/(e^x - 1)$. This method yields a very fast algorithm to compute $B_0, B_2, \ldots, B_{p-3}$ modulo p. See [**BCS92**] for an application of this method modulo a prime p to the verification of Fermat's last theorem for irregular primes up to one million.

Example 2.43. David Harvey implemented the algorithm of [**BCS92**] in SAGE as the command `bernoulli_mod_p`:

```
sage: bernoulli_mod_p(23)
[1, 4, 13, 17, 13, 6, 10, 5, 10, 9, 15]
```

A third way to compute B_n uses an algorithm based on Proposition 2.8, which we explain below (Algorithm 2.45). This algorithm appears to have been independently invented by several people: by Bernd C. Kellner (see [**Kel06**]); by Bill Dayl; and by H. Cohen and K. Belabas.

We compute B_n as an exact rational number by approximating $\zeta(n)$ to very high precision using Proposition 2.8, the Euler product

$$\zeta(s) = \sum_{m=1}^{\infty} m^{-s} = \prod_{p \text{ prime}} (1 - p^{-s})^{-1},$$

and the following theorem:

Theorem 2.44 (Clausen, von Staudt)**.** *For even $n \geq 2$,*

$$\mathrm{denom}(B_n) = \prod_{p-1\,|\,n} p.$$

Proof. See [**Lan95**, Ch. X, Thm. 2.1]. □

2.7.1. The Number of Digits of B_n. The following is a new quick way to compute the number of digits of the numerator of B_n. For example, using it we can compute the number of digits of $B_{10^{50}}$ in less than a second.

By Theorem 2.44 we have $d_n = \mathrm{denom}(B_n) = \prod_{p-1|n} p$. The number of digits of the numerator is thus

$$\lceil \log_{10}(d_n \cdot |B_n|) \rceil.$$

But

$$\begin{aligned}\log(|B_n|) &= \log\left(\frac{2\cdot n!}{(2\pi)^n}\,\zeta(n)\right)\\ &= \log(2) + \log(n!) - n\log(2) - n\log(\pi) + \log(\zeta(n)),\end{aligned}$$

and $\zeta(n) \sim 1$ so $\log(\zeta(n)) \sim 0$. Finally, Stirling's formula (see [**Ahl78**, pg. 198–206]) gives a fast way to compute $\log(n!) = \log(\Gamma(n+1))$:
(2.7.2)

$$\log(\Gamma(z))\text{``}=\text{''}\,\frac{\log(2\pi)}{2} + \left(z - \frac{1}{2}\right)\log(z) - z + \sum_{m=1}^{\infty} \frac{B_{2m}}{2m(2m-1)z^{2m-1}}.$$

We put quotes around the equality sign because $\log(\Gamma(z))$ does not converge to its Laurent series. Indeed, note that for any fixed value of z the summands on the right side go to ∞ as $m \to \infty$! Nonetheless, we can use this formula to very efficiently compute $\log(\Gamma(z))$, since if we truncate the sum, then the error is smaller than the next term in the infinite sum.

2.7.2. Computing B_n Exactly. We return to the problem of computing B_n. Let

$$K = \frac{2\cdot n!}{(2\pi)^n}$$

so that $|B_n| = K\zeta(n)$. Write

$$B_n = \frac{a}{d},$$

with $a, d \in \mathbb{Z}$, $d \geq 1$, and $\gcd(a, d) = 1$. It is elementary to show that $a = (-1)^{n/2+1}\,|a|$ for even $n \geq 2$. Suppose that using the Euler product we approximate $\zeta(n)$ from below by a number z such that

$$0 \leq \zeta(m) - z < \frac{1}{Kd}.$$

Then $0 \leq |B_n| - zK < d^{-1}$; hence $0 \leq |a| - zKd < 1$. It follows that $|a| = \lceil zKd \rceil$ and hence $a = (-1)^{n/2+1} \lceil zKd \rceil$.

It remains to compute z. Consider the following problem: given $s > 1$ and $\varepsilon > 0$, find $M \in \mathbb{Z}_+$ so that

$$z = \prod_{p \leq M} (1 - p^{-s})^{-1}$$

satisfies $0 \leq \zeta(s) - z < \varepsilon$. We always have $0 \leq \zeta(s) - z$. Also,

$$\sum_{n \leq M} n^{-s} \leq \prod_{p \leq M} (1 - p^{-s})^{-1},$$

so

$$\zeta(s) - z \leq \sum_{n=M+1}^{\infty} n^{-s} \leq \int_M^{\infty} x^{-s}\, dx = \frac{1}{(s-1)M^{s-1}}.$$

Thus if $M > \varepsilon^{-1/(s-1)}$, then

$$\frac{1}{(s-1)M^{s-1}} \leq \frac{1}{M^{s-1}} < \varepsilon\,,$$

so $\zeta(s) - z < \varepsilon$, as required. For our purposes, we have $s = n$ and $\varepsilon = (Kd)^{-1}$, so it suffices to take $M > (Kd)^{1/(n-1)}$.

Algorithm 2.45 (Bernoulli Number B_n)**.** *Given an integer $n \geq 0$, this algorithm computes the Bernoulli number B_n as an exact rational number.*

```
(1) [Special cases] If $n = 0$, return 1; if $n = 1$, return $-1/2$; if $n \geq 3$ is odd, return 0.
(2) [Factorial factor] Compute $K = \frac{2 \cdot n!}{(2\pi)^n}$ to sufficiently many digits of precision so the ceiling in step (6) is uniquely determined (this precision can be determined using Section 2.7.1).
(3) [Denominator] Compute $d = \prod_{p-1|n} p$.
(4) [Bound] Compute $M = \left\lceil (Kd)^{1/(n-1)} \right\rceil$.
(5) [Approximate $\zeta(n)$] Compute $z = \prod_{p \leq M} (1 - p^{-n})^{-1}$.
(6) [Numerator] Compute $a = (-1)^{n/2+1} \lceil dKz \rceil$.
(7) [Output $B_n$] Return $\frac{a}{d}$.
```

In step (5) use a sieve to compute all primes $p \leq M$ efficiently (which is fast, since M is so small). In step (4) we may replace M by any integer greater than the one specified by the formula, so we do not have to compute $(Kd)^{1/(n-1)}$ to very high precision.

In Section 5.2.2 below we will generalize the above algorithm.

Example 2.46. We illustrate Algorithm 2.45 by computing B_{50}. Using 135 binary digits of precision, we compute

$$K = 7500866746076957704747736.71552473164563479.$$

The divisors of n are $1, 2, 5, 10, 25, 50$, so

$$d = 2 \cdot 3 \cdot 11 = 66.$$

We find $M = 4$ and compute

$$z = 1.000000000000000888178421093081590298350 12.$$

Finally we compute

$$dKz = 4950572052410796482124775 24.999999994425778,$$

so

$$B_{50} = \frac{495057205241079648212477525}{66}.$$

2.8. Exercises

2.1 Using Proposition 2.8 and the table on page 16, compute $\sum_{n=1}^{\infty} \frac{1}{n^{26}}$ explicitly.

2.2 Prove that if $n > 1$ is odd, then the Bernoulli number B_n is 0.

2.3 Use (2.1.3) to write down the coefficients of 1, q, q^2, and q^3 of the Eisenstein series E_8.

2.4 Suppose k is a positive integer with $k \equiv 0 \pmod{12}$. Suppose $a, b \geq 0$ are integers with $4a + 6b = k$.
 (a) Prove $3 \mid a$.
 (b) Show that $G_4^a \cdot G_6^b \,/\, G_6^{\frac{k}{6}} = \left(G_4^3/G_6^2\right)^{\frac{a}{3}}$.

2.5 Compute the Miller basis for $M_{28}(\mathrm{SL}_2(\mathbb{Z}))$ with precision $O(q^8)$. Your answer will look like Example 2.23.

2.6 Consider the cusp form $f = q^2+192q^3-8280q^4+\cdots$ in $S_{28}(\mathrm{SL}_2(\mathbb{Z}))$. Write f as a polynomial in E_4 and E_6 (see Remark 2.25).

2.7 Let G_k be the weight k Eisenstein series from equation (2.1.1). Let c be the complex number so that the constant coefficient of the q-expansion of $g = c \cdot G_k$ is 1. Is it always the case that the q-expansion of g lies in $\mathbb{Z}[[q]]$?

2.8 Compute the matrix of the Hecke operator T_2 on the Miller basis for $M_{32}(\mathrm{SL}_2(\mathbb{Z}))$. Then compute its characteristic polynomial and verify it factors as a product of two irreducible polynomials.

What Next? Much of the rest of this book is about methods for computing subspaces of $M_k(\Gamma_1(N))$ for general N and k. These general methods are more complicated than the methods presented in this chapter, since there are many more modular forms of small weight and it can be difficult to obtain them. Forms of level $N > 1$ have subtle connections with elliptic curves, abelian varieties, and motives. Read on for more!

Chapter 3

Modular Forms of Weight 2

We saw in Chapter 2 (especially Section 2.2) that we can compute each space $M_k(\mathrm{SL}_2(\mathbb{Z}))$ explicitly. This involves computing Eisenstein series E_4 and E_6 to some precision, then forming the basis $\{E_4^a E_6^b : 4a + 6b = k, 0 \leq a, b \in \mathbb{Z}\}$ for $M_k(\mathrm{SL}_2(\mathbb{Z}))$. In this chapter we consider the more general problem of computing $S_2(\Gamma_0(N))$, for any positive integer N. Again we have a decomposition

$$M_2(\Gamma_0(N)) = S_2(\Gamma_0(N)) \oplus E_2(\Gamma_0(N)),$$

where $E_2(\Gamma_0(N))$ is spanned by generalized Eisenstein series and $S_2(\Gamma_0(N))$ is the space of cusp forms, i.e., elements of $M_2(\Gamma_0(N))$ that vanish at *all* cusps.

In Chapter 5 we compute the space $E_2(\Gamma_0(N))$ in a similar way to how we computed $M_k(\mathrm{SL}_2(\mathbb{Z}))$. On the other hand, elements of $S_2(\Gamma_0(N))$ often cannot be written as sums or products of generalized Eisenstein series. In fact, the structure of $M_2(\Gamma_0(N))$ is, in general, much more complicated than that of $M_k(\mathrm{SL}_2(\mathbb{Z}))$. For example, when p is a prime, $E_2(\Gamma_0(p))$ has dimension 1, whereas $S_2(\Gamma_0(p))$ has dimension about $p/12$.

Fortunately an idea of Birch, which he called modular symbols, provides a method for computing $S_2(\Gamma_0(N))$ and indeed for much more that is relevant to understanding special values of L-functions. Modular symbols are also a powerful theoretical tool. In this chapter, we explain how $S_2(\Gamma_0(N))$ is related to modular symbols and how to use this relationship to explicitly compute a basis for $S_2(\Gamma_0(N))$. In Chapter 8 we will introduce more general modular symbols and explain how to use them to compute $S_k(\Gamma_0(N))$, $S_k(\Gamma_1(N))$ and $S_k(N, \varepsilon)$ for any integers $k \geq 2$ and N and character ε.

Section 3.1 contains a very brief summary of basic facts about modular forms of weight 2, modular curves, Hecke operators, and integral homology. Section 3.2 introduces modular symbols and describes how to compute with them. In Section 3.5 we talk about how to cut out the subspace of modular symbols corresponding to cusp forms using the boundary map. Section 3.6 is about a straightforward method to compute a basis for $S_2(\Gamma_0(N))$ using modular symbols, and Section 3.7 outlines a more sophisticated algorithm for computing newforms that uses Atkin-Lehner theory.

Before reading this chapter, you should have read Chapter 1 and Chapter 2. We also assume familiarity with algebraic curves, Riemann surfaces, and homology groups of compact Riemann surfaces.

3.1. Hecke Operators

Recall from Chapter 1 that the group $\Gamma_0(N)$ acts on $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q})$ by linear fractional transformations. The quotient $\Gamma_0(N)\backslash\mathfrak{h}^*$ is a Riemann surface, which we denote by $X_0(N)$. See [**DS05**, Ch. 2] for a detailed description of the topology on $X_0(N)$. The Riemann surface $X_0(N)$ also has a canonical structure of algebraic curve over $\mathbb{Q}$, as is explained in [**DS05**, Ch. 7] (see also [**Shi94**, §6.7]).

Recall from Section 1.3 that a cusp form of weight 2 for $\Gamma_0(N)$ is a function f on $\mathfrak{h}$ such that $f(z)dz$ defines a holomorphic differential on $X_0(N)$. Equivalently, a cusp form is a holomorphic function f on $\mathfrak{h}$ such that

(a) the expression $f(z)dz$ is invariant under replacing z by $\gamma(z)$ for each $\gamma \in \Gamma_0(N)$ and

(b) $f(z)$ vanishes at every cusp for $\Gamma_0(N)$.

The space $S_2(\Gamma_0(N))$ of weight 2 cusp forms on $\Gamma_0(N)$ is a finite-dimensional complex vector space, of dimension equal to the genus g of $X_0(N)$. The space $X_0(N)(\mathbb{C})$ is a compact oriented Riemann surface, so it is a 2-dimensional oriented real manifold, i.e., $X_0(N)(\mathbb{C})$ is a g-holed torus (see Figure 3.1.1 on page 38).

Condition (b) in the definition of f means that f has a Fourier expansion about each element of $\mathbb{P}^1(\mathbb{Q})$. Thus, at ∞ we have

$$\begin{aligned} f(z) &= a_1 e^{2\pi i z} + a_2 e^{2\pi i 2z} + a_3 e^{2\pi i 3z} + \cdots \\ &= a_1 q + a_2 q^2 + a_3 q^3 + \cdots, \end{aligned}$$

where, for brevity, we write $q = q(z) = e^{2\pi i z}$.

Example 3.1. Let E be the elliptic curve defined by the equation $y^2+xy = x^3+x^2-4x-5$. Let $a_p = p+1-\#\tilde{E}(\mathbb{F}_p)$, where $\tilde{E}$ is the reduction of E mod p (note that for the primes that divide the conductor of E we have $a_3 = -1$, $a_{13} = 1$). For n composite, define a_n using the relations at the end of Section 3.7. Then the Shimura-Taniyama conjecture asserts that

$$\begin{aligned} f &= q + a_2q^2 + a_3q^3 + a_4q^4 + a_5q^5 + \cdots \\ &= q + q^2 - q^3 - q^4 + 2q^5 + \cdots \end{aligned}$$

is the q-expansion of an element of $S_2(\Gamma_0(39))$. This conjecture, which is now a theorem (see [**BCDT01**]), asserts that any q-expansion constructed as above from an elliptic curve over $\mathbb{Q}$ is a modular form. This conjecture was mostly proved first by Wiles [**Wil95**] as a key step in the proof of Fermat's last theorem.

Just as is the case for level 1 modular forms (see Section 2.4) there are commuting Hecke operators $T_1, T_2, T_3, \ldots$ that act on $S_2(\Gamma_0(N))$. To define them conceptually, we introduce an interpretation of the modular curve $X_0(N)$ as an object whose points *parameterize* elliptic curves with extra structure.

Proposition 3.2. *The complex points of* $Y_0(N) = \Gamma_0(N)\backslash\mathfrak{h}$ *are in natural bijection with isomorphism classes of pairs* (E, C)*, where* E *is an elliptic curve over* $\mathbb{C}$ *and* C *is a cyclic subgroup of* $E(\mathbb{C})$ *of order* N*. The class of the point* $\lambda \in \mathfrak{h}$ *corresponds to the pair*

$$\left(\mathbb{C}/(\mathbb{Z}+\mathbb{Z}\lambda), \quad \left(\frac{1}{N}\mathbb{Z}+\mathbb{Z}\lambda\right)/(\mathbb{Z}+\mathbb{Z}\lambda)\right).$$

Proof. See Exercise 3.1. □

Suppose n and N are coprime positive integers. There are two natural maps π_1 and π_2 from $Y_0(n\cdot N)$ to $Y_0(N)$; the first, π_1, sends $(E, C) \in Y_0(n\cdot N)(\mathbb{C})$ to (E, C'), where C' is the unique cyclic subgroup of C of order N, and the second, π_2, sends (E, C) to $(E/D, C/D)$, where D is the unique cyclic subgroup of C of order n. These maps extend in a unique way to algebraic maps from $X_0(n\cdot N)$ to $X_0(N)$:

(3.1.1)

$$\begin{tikzcd} & X_0(n\cdot N) \arrow[dl, "\pi_2"'] \arrow[dr, "\pi_1"] & \\ X_0(N) & & X_0(N). \end{tikzcd}$$

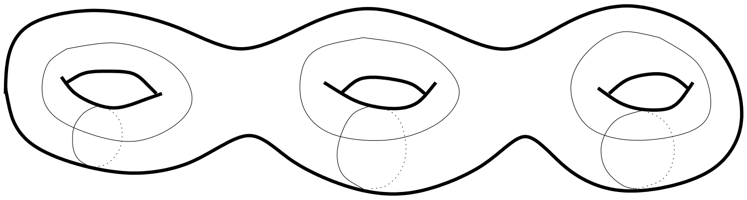
The nth *Hecke operator* T_n is $\pi_{1*}\circ\pi_2^*$, where π_2^* and π_{1*} denote pullback and pushforward of differentials, respectively. (There is a similar definition of T_n when $\gcd(n, N) \neq 1$.) Using our interpretation of $S_2(\Gamma_0(N))$ as differentials on $X_0(N)$, this gives an action of Hecke operators on $S_2(\Gamma_0(N))$. One can show that these induce the maps of Proposition 2.31 on q-expansions.

Example 3.3. There is a basis of $S_2(39)$ so that

$$T_2 = \begin{pmatrix} 1 & 1 & 0 \\ -2 & -3 & -2 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad T_5 = \begin{pmatrix} -4 & -2 & -6 \\ 4 & 4 & 4 \\ 0 & 0 & 2 \end{pmatrix}.$$

Notice that these matrices commute. Also, the characteristic polynomial of T_2 is $(x-1)\cdot(x^2+2x-1)$.

3.1.1. Homology. The first homology group $\mathrm{H}_1(X_0(N),\mathbb{Z})$ is the group of closed 1-cycles modulo boundaries of 2-cycles (formal sums of images of 2-simplexes). Topologically $X_0(N)$ is a g-holed torus, where g is the genus of $X_0(N)$. Thus $\mathrm{H}_1(X_0(N),\mathbb{Z})$ is a free abelian group of rank $2g$ (see, e.g., [**GH81**, Ex. 19.30] and [**DS05**, §6.1]), with two generators corresponding to each hole, as illustrated in the case $N = 39$ in Figure 3.1.1.



$$\mathrm{H}_1(X_0(39),\mathbb{Z}) \cong \mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}$$

Figure 3.1.1. The homology of $X_0(39)$.

The homology of $X_0(N)$ is closely related to modular forms, since the Hecke operators T_n also act on $\mathrm{H}_1(X_0(N),\mathbb{Z})$. The action is by pullback of homology classes by π_2 followed by taking the image under π_1, where π_1 and π_2 are as in (3.1.1).

Integration defines a pairing

$$\langle\, ,\, \rangle : S_2(\Gamma_0(N)) \times \mathrm{H}_1(X_0(N),\mathbb{Z}) \to \mathbb{C}. \tag{3.1.2}$$

Explicitly, for a path x,

$$\langle f, x\rangle = 2\pi i \cdot \int_x f(z)dz.$$

Theorem 3.4. *The pairing* (3.1.2) *is nondegenerate and Hecke equivariant in the sense that for every Hecke operator T_n, we have $\langle fT_n, x\rangle = \langle f, T_n x\rangle$. Moreover, it induces a perfect pairing*

$$\langle\, ,\, \rangle : S_2(\Gamma_0(N)) \times \mathrm{H}_1(X_0(N),\mathbb{R}) \to \mathbb{C}. \tag{3.1.3}$$

This is a special case of the results in Section 8.5.

As we will see, modular symbols allow us to make explicit the action of the Hecke operators on $\mathrm{H}_1(X_0(N), \mathbb{Z})$; the above pairing then translates this into a wealth of information about cusp forms.

We will also consider the relative homology group $\mathrm{H}_1(X_0(N), \mathbb{Z}; \{\text{cusps}\})$ of $X_0(N)$ *relative to the cusps*; it is the same as usual homology, but in addition we allow paths with endpoints in the cusps instead of restricting to closed loops. Modular symbols provide a "combinatorial" presentation of $\mathrm{H}_1(X_0(N), \mathbb{Z})$ in terms of paths between elements of $\mathbb{P}^1(\mathbb{Q})$.

3.2. Modular Symbols

Let $\mathbb{M}_2$ be the free abelian group with basis the set of symbols $\{\alpha, \beta\}$ with $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$ modulo the 3-term relations

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$$

above and modulo any torsion. Since $\mathbb{M}_2$ is torsion-free, we have

$$\{\alpha, \alpha\} = 0 \qquad \text{and} \qquad \{\alpha, \beta\} = -\{\beta, \alpha\}.$$

Remark 3.5 (Warning)**.** The symbols $\{\alpha, \beta\}$ satisfy the relations $\{\alpha, \beta\} = -\{\beta, \alpha\}$, so order matters. The notation $\{\alpha, \beta\}$ looks like the set containing two elements, which strongly (and incorrectly) suggests that the order does not matter. This is the standard notation in the literature.

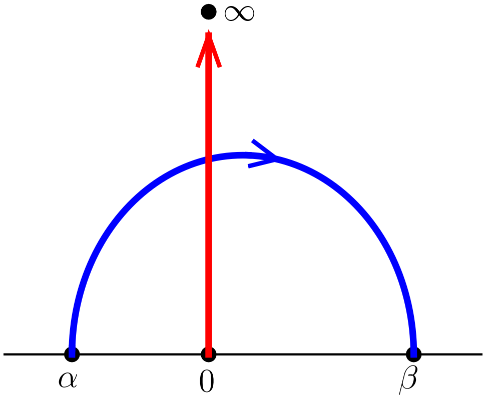


Figure 3.2.1. The modular symbols $\{\alpha, \beta\}$ and $\{0, \infty\}$.

As illustrated in Figure 3.2.1, we "think of" this modular symbol as the homology class, relative to the cusps, of a path from α to β in $\mathfrak{h}^*$.

Define a *left action of* $\mathrm{GL}_2(\mathbb{Q})$ on $\mathbb{M}_2$ by letting $g \in \mathrm{GL}_2(\mathbb{Q})$ act by

$$g\{\alpha, \beta\} = \{g(\alpha), g(\beta)\},$$

and g acts on α and β via the corresponding linear fractional transformation. The space $\mathbb{M}_2(\Gamma_0(N))$ of *modular symbols for* $\Gamma_0(N)$ is the quotient of $\mathbb{M}_2$ by the submodule generated by the infinitely many elements of the form $x - g(x)$, for x in $\mathbb{M}_2$ and g in $\Gamma_0(N)$, and modulo any torsion. A modular symbol for $\Gamma_0(N)$ is an element of this space. We frequently denote the equivalence class of a modular symbol by giving a representative element.

Example 3.6. Some modular symbols are 0 no matter what the level N is! For example, since $\gamma = \left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$\{\infty, 0\} = \{\gamma(\infty), \gamma(0)\} = \{\infty, 1\},$$

so

$$0 = \{\infty, 1\} - \{\infty, 0\} = \{\infty, 1\} + \{0, \infty\} = \{0, \infty\} + \{\infty, 1\} = \{0, 1\}.$$

See Exercise 3.2 for a generalization of this observation.

There is a natural homomorphism

$$\varphi : \mathbb{M}_2(\Gamma_0(N)) \to \mathrm{H}_1(X_0(N), \{\text{cusps}\}, \mathbb{Z}) \tag{3.2.1}$$

that sends a formal linear combination of geodesic paths in the upper half plane to their image as paths on $X_0(N)$. In [**Man72**] Manin proved that (3.2.1) is an isomorphism (this is a fairly involved topological argument).

Manin identified the subspace of $\mathbb{M}_2(\Gamma_0(N))$ that is sent isomorphically onto $\mathrm{H}_1(X_0(N), \mathbb{Z})$. Let $\mathbb{B}_2(\Gamma_0(N))$ denote the free abelian group whose basis is the finite set $C(\Gamma_0(N)) = \Gamma_0(N)\backslash\mathbb{P}^1(\mathbb{Q})$ of cusps for $\Gamma_0(N)$. The *boundary map*

$$\delta : \mathbb{M}_2(\Gamma_0(N)) \to \mathbb{B}_2(\Gamma_0(N))$$

sends $\{\alpha, \beta\}$ to $\{\beta\} - \{\alpha\}$, where $\{\beta\}$ denotes the basis element of $\mathbb{B}_2(\Gamma_0(N))$ corresponding to $\beta \in \mathbb{P}^1(\mathbb{Q})$. The kernel $\mathbb{S}_2(\Gamma_0(N))$ of δ is the subspace of *cuspidal modular symbols*. Thus an element of $\mathbb{S}_2(\Gamma_0(N))$ can be thought of as a linear combination of paths in $\mathfrak{h}^*$ whose endpoints are cusps and whose images in $X_0(N)$ are homologous to a $\mathbb{Z}$-linear combination of closed paths.

Theorem 3.7 (Manin)**.** *The map φ above induces a canonical isomorphism*

$$\mathbb{S}_2(\Gamma_0(N)) \cong \mathrm{H}_1(X_0(N), \mathbb{Z}).$$

Proof. This is [**Man72**, Thm. 1.9]. □

For any (commutative) ring R let

$$\mathbb{M}_2(\Gamma_0(N), R) = \mathbb{M}_2(\Gamma_0(N)) \otimes_{\mathbb{Z}} R$$

and

$$\mathbb{S}_2(\Gamma_0(N), R) = \mathbb{S}_2(\Gamma_0(N)) \otimes_{\mathbb{Z}} R.$$

Proposition 3.8. *We have*

$$\dim_{\mathbb{C}} \mathbb{S}_2(\Gamma_0(N), \mathbb{C}) = 2 \dim_{\mathbb{C}} S_2(\Gamma_0(N)).$$

Proof. We have

$$\dim_{\mathbb{C}} \mathbb{S}_2(\Gamma_0(N), \mathbb{C}) = \operatorname{rank}_{\mathbb{Z}} \mathbb{S}_2(\Gamma_0(N)) = \operatorname{rank}_{\mathbb{Z}} \mathrm{H}_1(X_0(N), \mathbb{Z}) = 2g.$$

□

Example 3.9. We illustrate modular symbols in the case when $N = 11$. Using SAGE (below), which implements the algorithm that we describe below over $\mathbb{Q}$, we find that $\mathbb{M}_2(\Gamma_0(11); \mathbb{Q})$ has basis $\{\infty, 0\}$, $\{-1/8, 0\}$, $\{-1/9, 0\}$. A basis for the integral homology $\mathrm{H}_1(X_0(11), \mathbb{Z})$ is the subgroup generated by $\{-1/8, 0\}$ and $\{-1/9, 0\}$.

```
sage: set_modsym_print_mode ('modular')
sage: M = ModularSymbols(11, 2)
sage: M.basis()
({Infinity,0}, {-1/8,0}, {-1/9,0})
sage: S = M.cuspidal_submodule()
sage: S.integral_basis()     # basis over ZZ.
({-1/8,0}, {-1/9,0})
sage: set_modsym_print_mode ('manin')    # set it back
```

3.3. Computing with Modular Symbols

3.3.1. Manin's Trick. In this section, we describe a trick of Manin that we will use to prove that spaces of modular symbols are computable.

By Exercise 1.6 the group $\Gamma_0(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$. Fix right coset representatives $r_0, r_1, \ldots, r_m$ for $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbb{Z})$, so that

$$\mathrm{SL}_2(\mathbb{Z}) = \Gamma_0(N) r_0 \cup \Gamma_0(N) r_1 \cup \cdots \cup \Gamma_0(N) r_m,$$

where the union is disjoint. For example, when N is prime, a list of coset representatives is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & 0 \\ N-1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Let

$$\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) = \{(a : b) \, : \, a, b \in \mathbb{Z}/N\mathbb{Z}, \ \gcd(a, b, N) = 1\}/\sim \tag{3.3.1}$$

where $(a : b) \sim (a' : b')$ if there is $u \in (\mathbb{Z}/N\mathbb{Z})^*$ such that $a = ua', b = ub'$.

Proposition 3.10. *There is a bijection between* $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ *and the right cosets of* $\Gamma_0(N)$ *in* $\mathrm{SL}_2(\mathbb{Z})$*, which sends a coset representative* $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ *to the class of* $(c : d)$ *in* $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$*.*

Proof. See Exercise 3.3. □

See Proposition 8.6 for the analogous statement for $\Gamma_1(N)$.

We now describe an observation of Manin (see [**Man72**, §1.5]) that is crucial to making $\mathbb{M}_2(\Gamma_0(N))$ computable. It allows us to write any modular symbol $\{\alpha, \beta\}$ as a $\mathbb{Z}$-linear combination of symbols of the form $r_i\{0, \infty\}$, where the $r_i \in \mathrm{SL}_2(\mathbb{Z})$ are coset representatives as above. In particular, the finitely many symbols $r_0\{0,\infty\}, \ldots, r_m\{0,\infty\}$ generate $\mathbb{M}_2(\Gamma_0(N))$.

Proposition 3.11 (Manin)**.** *Let* N *be a positive integer and* $r_0, \ldots, r_m$ *a set of right coset representatives for* $\Gamma_0(N)$ *in* $\mathrm{SL}_2(\mathbb{Z})$*. Every* $\{\alpha, \beta\} \in \mathbb{M}_2(\Gamma_0(N))$ *is a* $\mathbb{Z}$*-linear combination of* $r_0\{0,\infty\}, \ldots, r_m\{0,\infty\}$*.*

We give two proofs of the proposition. The first is useful for computation (see [**Cre97a**, §2.1.6]); the second (see [**MTT86**, §2]) is easier to understand conceptually since it does not require any knowledge of continued fractions.

Continued Fractions Proof of Proposition 3.11. Since

$$\{\alpha, \beta\} = \{0, \beta\} - \{0, \alpha\},$$

it suffices to consider modular symbols of the form $\{0, b/a\}$, where the rational number b/a is in lowest terms. Expand b/a as a continued fraction and consider the successive convergents in lowest terms:

$$\frac{b_{-2}}{a_{-2}} = \frac{0}{1}, \quad \frac{b_{-1}}{a_{-1}} = \frac{1}{0}, \quad \frac{b_0}{a_0} = \frac{b_0}{1}, \ldots, \quad \frac{b_{n-1}}{a_{n-1}}, \quad \frac{b_n}{a_n} = \frac{b}{a}$$

where the first two are included formally. Then

$$b_k a_{k-1} - b_{k-1} a_k = (-1)^{k-1},$$

so that

$$g_k = \begin{pmatrix} b_k & (-1)^{k-1} b_{k-1} \\ a_k & (-1)^{k-1} a_{k-1} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Hence

$$\left\{ \frac{b_{k-1}}{a_{k-1}}, \frac{b_k}{a_k} \right\} = g_k\{0, \infty\} = r_i\{0, \infty\},$$

for some i, is of the required special form. Since

$$\{0, b/a\} = \{0, \infty\} + \{\infty, b_0\} + \left\{ \frac{b_0}{1}, \frac{b_1}{a_1} \right\} + \cdots + \left\{ \frac{b_{n-1}}{a_{n-1}}, \frac{b_n}{a_n} \right\},$$

this completes the proof. □

Inductive Proof of Proposition 3.11. As in the first proof it suffices to prove the proposition for any symbol $\{0, b/a\}$, where b/a is in lowest terms. We will induct on $a \in \mathbb{Z}_{\geq 0}$. If $a = 0$, then the symbol is $\{0, \infty\}$, which corresponds to the identity coset, so assume that $a > 0$. Find $a' \in \mathbb{Z}$ such that

$$ba' \equiv 1 \pmod{a};$$

then $b' = (ba' - 1)/a \in \mathbb{Z}$ so the matrix

$$\delta = \begin{pmatrix} b & b' \\ a & a' \end{pmatrix}$$

is an element of $\mathrm{SL}_2(\mathbb{Z})$. Thus $\delta = \gamma \cdot r_j$ for some right coset representative r_j and $\gamma \in \Gamma_0(N)$. Then

$$\{0, b/a\} - \{0, b'/a'\} = \{b'/a', b/a\} = \begin{pmatrix} b & b' \\ a & a' \end{pmatrix} \cdot \{0, \infty\} = r_j\{0, \infty\},$$

as elements of $\mathbb{M}_2(\Gamma_0(N))$. By induction, $\{0, b'/a'\}$ is a linear combination of symbols of the form $r_k\{0, \infty\}$, which completes the proof. □

Example 3.12. Let $N = 11$, and consider the modular symbol $\{0, 4/7\}$. We have

$$\frac{4}{7} = 0 + \frac{1}{1 + \frac{1}{1+\frac{1}{3}}},$$

so the partial convergents are

$$\frac{b_{-2}}{a_{-2}} = \frac{0}{1}, \quad \frac{b_{-1}}{a_{-1}} = \frac{1}{0}, \quad \frac{b_0}{a_0} = \frac{0}{1}, \quad \frac{b_1}{a_1} = \frac{1}{1}, \quad \frac{b_2}{a_2} = \frac{1}{2}, \quad \frac{b_3}{a_3} = \frac{4}{7}.$$

Thus, noting as in Example 3.6 that $\{0, 1\} = 0$, we have

$$\begin{aligned} \{0, 4/7\} &= \{0, \infty\} + \{\infty, 0\} + \{0, 1\} + \{1, 1/2\} + \{1/2, 4/7\} \\ &= \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix} \{0, \infty\} + \begin{pmatrix} 4 & 1 \\ 7 & 2 \end{pmatrix} \{0, \infty\} \\ &= \begin{pmatrix} 1 & 0 \\ 9 & 1 \end{pmatrix} \{0, \infty\} + \begin{pmatrix} 1 & 0 \\ 9 & 1 \end{pmatrix} \{0, \infty\} \\ &= 2 \cdot \left[\begin{pmatrix} 1 & 0 \\ 9 & 1 \end{pmatrix} \{0, \infty\} \right]. \end{aligned}$$

We compute the convergents of $4/7$ in SAGE as follows (note that 0 and ∞ are excluded):

```
sage: convergents(4/7)
[0, 1, 1/2, 4/7]
```

3.3.2. Manin Symbols. As above, fix coset representatives $r_0, \ldots, r_m$ for $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbb{Z})$. Consider formal symbols $[r_i]'$ for $i = 0, \ldots, m$. Let $[r_i]$ be the modular symbol $r_i\{0, \infty\} = \{r_i(0), r_i(\infty)\}$. We equip the symbols $[r_0]', \ldots, [r_m]'$ with a *right action of* $\mathrm{SL}_2(\mathbb{Z})$, which is given by $[r_i]'.g = [r_j]'$, where $\Gamma_0(N)r_j = \Gamma_0(N)r_i g$. We extend the notation by writing $[\gamma]' = [\Gamma_0(N)\gamma]' = [r_i]'$, where $\gamma \in \Gamma_0(N)r_i$. Then the right action of $\Gamma_0(N)$ is simply $[\gamma]'.g = [\gamma g]'$.

Theorem 1.2 implies that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the two matrices $\sigma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\tau = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Note that $\sigma = S$ from Theorem 1.2 and $\tau = TS$, so $T = \tau\sigma \in \langle \sigma, \tau \rangle$.

The following theorem provides us with a finite presentation for the space $\mathbb{M}_2(\Gamma_0(N))$ of modular symbols.

Theorem 3.13 (Manin)**.** *Consider the quotient M of the free abelian group on Manin symbols $[r_0]', \ldots, [r_m]'$ by the subgroup generated by the elements (for all i):*

$$[r_i]' + [r_i]'\sigma \qquad \textit{and} \qquad [r_i]' + [r_i]'\tau + [r_i]'\tau^2,$$

and modulo any torsion. Then there is an isomorphism

$$\Psi : M \xrightarrow{\sim} \mathbb{M}_2(\Gamma_0(N))$$

given by $[r_i]' \mapsto [r_i] = r_i\{0, \infty\}$.

Proof. We will only prove that Ψ is surjective; the proof that Ψ is injective requires much more work and will be omitted from this book (see [**Man72**, §1.7] for a complete proof).

Proposition 3.11 implies that Ψ is surjective, assuming that Ψ is well defined. We next verify that Ψ is well defined, i.e., that the listed 2-term and 3-term relations hold in the image. To see that the first relation holds, note that

$$\begin{aligned}
[r_i] + [r_i]\sigma &= \{r_i(0), r_i(\infty)\} + \{r_i\sigma(0), r_i\sigma(\infty)\} \\
&= \{r_i(0), r_i(\infty)\} + \{r_i(\infty), r_i(0)\} \\
&= 0.
\end{aligned}$$

For the second relation we have

$$\begin{aligned}
[r_i] + [r_i]\tau + [r_i]\tau^2 &= \{r_i(0), r_i(\infty)\} + \{r_i\tau(0), r_i\tau(\infty)\} + \{r_i\tau^2(0), r_i\tau^2(\infty)\} \\
&= \{r_i(0), r_i(\infty)\} + \{r_i(\infty), r_i(1)\} + \{r_i(1), r_i(0)\} \\
&= 0.
\end{aligned}$$

□

Example 3.14. By default SAGE computes modular symbols spaces over $\mathbb{Q}$, i.e., $\mathbb{M}_2(\Gamma_0(N);\mathbb{Q}) \cong \mathbb{M}_2(\Gamma_0(N))\otimes\mathbb{Q}$. SAGE represents (weight 2) Manin symbols as pairs (c,d). Here c, d are integers that satisfy $0 \leq c, d < N$; they define a point $(c:d) \in \mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, hence a right coset of $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ (see Proposition 3.10).

Create $\mathbb{M}_2(\Gamma_0(N);\mathbb{Q})$ in SAGE by typing `ModularSymbols(N, 2)`. We then use the SAGE command `manin_generators` to enumerate a list of generators $[r_0], \ldots, [r_n]$ as in Theorem 3.13 for several spaces of modular symbols.

```
sage: M = ModularSymbols(2,2)
sage: M
Modular Symbols space of dimension 1 for Gamma_0(2)
of weight 2 with sign 0 over Rational Field
sage: M.manin_generators()
[(0,1), (1,0), (1,1)]

sage: M = ModularSymbols(3,2)
sage: M.manin_generators()
[(0,1), (1,0), (1,1), (1,2)]

sage: M = ModularSymbols(6,2)
sage: M.manin_generators()
[(0,1), (1,0), (1,1), (1,2), (1,3), (1,4), (1,5), (2,1),
 (2,3), (2,5), (3,1), (3,2)]
```

Given `x=(c,d)`, the command `x.lift_to_sl2z(N)` computes an element of $\mathrm{SL}_2(\mathbb{Z})$ whose lower two entries are congruent to (c,d) modulo N.

```
sage: M = ModularSymbols(2,2)
sage: [x.lift_to_sl2z(2) for x in M.manin_generators()]
[[1, 0, 0, 1], [0, -1, 1, 0], [0, -1, 1, 1]]
sage: M = ModularSymbols(6,2)
sage: x = M.manin_generators()[9]
sage: x
(2,5)
sage: x.lift_to_sl2z(6)
[1, 2, 2, 5]
```

The `manin_basis` command returns a list of indices into the Manin generator list such that the corresponding symbols form a basis for the quotient of the $\mathbb{Q}$-vector space spanned by Manin symbols modulo the 2-term and 3-term relations of Theorem 3.13.

```
sage: M = ModularSymbols(2,2)
sage: M.manin_basis()
[1]
sage: [M.manin_generators()[i] for i in M.manin_basis()]
[(1,0)]
sage: M = ModularSymbols(6,2)
sage: M.manin_basis()
[1, 10, 11]
sage: [M.manin_generators()[i] for i in M.manin_basis()]
[(1,0), (3,1), (3,2)]
```

Thus, e.g., every element of $\mathbb{M}_2(\Gamma_0(6))$ is a $\mathbb{Q}$-linear combination of the three symbols $[(1,0)]$, $[(3,1)]$, and $[(3,2)]$. We can write each of these as a modular symbol using the `modular_symbol_rep` function.

```
sage: M.basis()
((1,0), (3,1), (3,2))
sage: [x.modular_symbol_rep() for x in M.basis()]
[{Infinity,0}, {0,1/3}, {-1/2,-1/3}]
```

The `manin_gens_to_basis` function returns a matrix whose rows express each Manin symbol generator in terms of the subset of Manin symbols that forms a basis (as returned by `manin_basis`).

```
sage: M = ModularSymbols(2,2)
sage: M.manin_gens_to_basis()
[-1]
[ 1]
[ 0]
```

Since the basis is $(1,0)$, this means that in $\mathbb{M}_2(\Gamma_0(2);\mathbb{Q})$, we have $[(0,1)] = -[(1,0)]$ and $[(1,1)] = 0$. (Since no denominators are involved, we have in fact computed a presentation of $\mathbb{M}_2(\Gamma_0(2);\mathbb{Z})$.)

To convert a Manin symbol $x = (c,d)$ to an element of a modular symbols space M, use `M(x)`:

```
sage: M = ModularSymbols(2,2)
sage: x = (1,0); M(x)
(1,0)
```

Next consider $\mathbb{M}_2(\Gamma_0(6);\mathbb{Q})$:

```
sage: M = ModularSymbols(6,2)
sage: M.manin_gens_to_basis()
[-1  0  0]
[ 1  0  0]
[ 0  0  0]
[ 0 -1  1]
[ 0 -1  0]
[ 0 -1  1]
[ 0  0  0]
[ 0  1 -1]
[ 0  0 -1]
[ 0  1 -1]
[ 0  1  0]
[ 0  0  1]
```

Recall that our choice of basis for $\mathbb{M}_2(\Gamma_0(6);\mathbb{Q})$ is $[(1,0)], [(3,1)], [(3,2)]$. Thus, e.g., the first row of this matrix says that $[(0,1)] = -[(1,0)]$, and the fourth row asserts that $[(1,2)] = -[(3,1)] + [(3,2)]$.

```
sage: M = ModularSymbols(6,2)
sage: M((0,1))
-(1,0)
sage: M((1,2))
-(3,1) + (3,2)
```

3.4. Hecke Operators

3.4.1. Hecke Operators on Modular Symbols. When p is a prime not dividing N, define

$$T_p(\{\alpha,\beta\}) = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \{\alpha,\beta\} + \sum_{r \bmod p} \begin{pmatrix} 1 & r \\ 0 & p \end{pmatrix} \{\alpha,\beta\}.$$

The Hecke operators are compatible with the integration pairing $\langle\, ,\, \rangle$ of Section 3.1, in the sense that $\langle fT_p, x\rangle = \langle f, T_p x\rangle$. When $p \mid N$, the definition is the same, except that the matrix $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is not included in the sum (see Theorem 8.23). There is a similar definition of T_n for n composite (see Section 8.3.1).

Example 3.15. For example, when $N = 11$, we have

$$\begin{aligned} T_2\{0, 1/5\} &= \{0, 2/5\} + \{0, 1/10\} + \{1/2, 3/5\} \\ &= -2\{0, 1/5\}. \end{aligned}$$

See Figure 3.4.1.

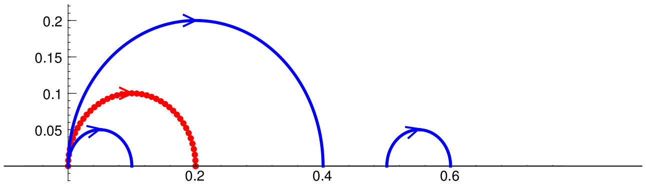


Figure 3.4.1. Image of $\{0, 1/5\}$ under T_2

3.4.2. Hecke Operators on Manin Symbols. In [**Mer94**], L. Merel gives a description of the action of T_p directly on Manin symbols $[r_i]$ (see Section 8.3.2 for details). For example, when $p = 2$ and N is odd, we have

$$T_2([r_i]) = [r_i]\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + [r_i]\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + [r_i]\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} + [r_i]\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}. \tag{3.4.1}$$

For any prime, let C_p be the set of matrices constructed using the following algorithm (see [**Cre97a**, §2.4]):

Algorithm 3.16 (Cremona's Heilbronn Matrices)**.** *Given a prime p, this algorithm outputs a list of 2×2 matrices of determinant p that can be used to compute the Hecke operator T_p.*

```
(1) Output $\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$.
(2) For $r = \left\lceil -\frac{p}{2} \right\rceil, \ldots, \left\lfloor \frac{p}{2} \right\rfloor$:
    (a) Let $x_1 = p$, $x_2 = -r$, $y_1 = 0$, $y_2 = 1$, $a = -p$, $b = r$.
    (b) Output $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$.
    (c) As long as $b \neq 0$, do the following:
        (i) Let $q$ be the integer closest to $a/b$ (if $a/b$ is a half integer,
            round away from 0).
        (ii) Let $c = a - bq$, $a = -b$, $b = c$.
```

(iii) Set $x_3 = qx_2 - x_1$, $x_1 = x_2$, $x_2 = x_3$, and
$y_3 = qy_2 - y_1$, $y_1 = y_2$, $y_2 = y_3$.

(iv) Output $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$.

Proposition 3.17 (Cremona, Merel). *Let C_p be as above. Then for $p \nmid N$ and $[x] \in \mathbb{M}_2(\Gamma_0(N))$ a Manin symbol, we have*

$$T_p([x]) = \sum_{g \in C_p} [xg].$$

Proof. See Proposition 2.4.1 of [**Cre97a**]. □

There are other lists of matrices, due to Merel, that work even when $p \mid N$ (see Section 8.3.2).

The command `HeilbronnCremonaList(p)`, for p prime, outputs the list of matrices from Algorithm 3.16.

```
sage: HeilbronnCremonaList(2)
[[1, 0, 0, 2], [2, 0, 0, 1], [2, 1, 0, 1], [1, 0, 1, 2]]
sage: HeilbronnCremonaList(3)
[[1, 0, 0, 3], [3, 1, 0, 1], [1, 0, 1, 3], [3, 0, 0, 1],
 [3, -1, 0, 1], [-1, 0, 1, -3]]
sage: HeilbronnCremonaList(5)
[[1, 0, 0, 5], [5, 2, 0, 1], [2, 1, 1, 3], [1, 0, 3, 5],
 [5, 1, 0, 1], [1, 0, 1, 5], [5, 0, 0, 1], [5, -1, 0, 1],
 [-1, 0, 1, -5], [5, -2, 0, 1], [-2, 1, 1, -3],
 [1, 0, -3, 5]]
sage: len(HeilbronnCremonaList(37))
128
sage: len(HeilbronnCremonaList(389))
1892
sage: len(HeilbronnCremonaList(2003))
11662
```

Example 3.18. We compute the matrix of T_2 on $\mathbb{M}_2(\Gamma_0(2))$:

```
sage: M = ModularSymbols(2,2)
sage: M.T(2).matrix()
[1]
```

Example 3.19. We compute some Hecke operators on $\mathbb{M}_2(\Gamma_0(6))$:

```
sage: M = ModularSymbols(6, 2)
sage: M.T(2).matrix()
[ 2  1 -1]
[-1  0  1]
[-1 -1  2]
sage: M.T(3).matrix()
[3 2 0]
[0 1 0]
[2 2 1]
sage: M.T(3).fcp()  # factored characteristic polynomial
(x - 3) * (x - 1)^2
```

For $p \geq 5$ we have $T_p = p + 1$, since $M_2(\Gamma_0(6))$ is spanned by generalized Eisenstein series (see Chapter 5).

Example 3.20. We compute the Hecke operators on $\mathbb{M}_2(\Gamma_0(39))$:

```
sage: M = ModularSymbols(39, 2)
sage: T2 = M.T(2)
sage: T2.matrix()
[ 3  0 -1  0  0  1  1 -1  0]
[ 0  0  2  0 -1  1  0  1 -1]
[ 0  1  0 -1  1  1  0  1 -1]
[ 0  0  1  0  0  1  0  1 -1]
[ 0 -1  2  0  0  1  0  1 -1]
[ 0  0  1  1  0  1  1 -1  0]
[ 0  0  0 -1  0  1  1  2  0]
[ 0  0  0  1  0  0  2  0  1]
[ 0  0 -1  0  0  0  1  0  2]
sage: T2.fcp()     # factored characteristic polynomial
(x - 3)^3 * (x - 1)^2 * (x^2 + 2*x - 1)^2
```

The Hecke operators commute, so their eigenspace structures are related.

```
sage: T2 = M.T(2).matrix()
sage: T5 = M.T(5).matrix()
sage: T2*T5 - T5*T2 == 0
True
sage: T5.charpoly().factor()
(x^2 - 8)^2 * (x - 6)^3 * (x - 2)^2
```

The decomposition of T_2 is a list of the kernels of $(f^e)(T_2)$, where f runs through the irreducible factors of the characteristic polynomial of T_2 and f^e exactly divides this characteristic polynomial. Using SAGE, we find them:

```
sage: M = ModularSymbols(39, 2)
sage: M.T(2).decomposition()
[
Modular Symbols subspace of dimension 3 of Modular
Symbols space of dimension 9 for Gamma_0(39) of weight
2 with sign 0 over Rational Field,
Modular Symbols subspace of dimension 2 of Modular
Symbols space of dimension 9 for Gamma_0(39) of weight
2 with sign 0 over Rational Field,
Modular Symbols subspace of dimension 4 of Modular
Symbols space of dimension 9 for Gamma_0(39) of weight
2 with sign 0 over Rational Field
]
```

3.5. Computing the Boundary Map

In Section 3.2 we defined a map $\delta : \mathbb{M}_2(\Gamma_0(N)) \to \mathbb{B}_2(\Gamma_0(N))$. The kernel of this map is the space $\mathbb{S}_2(\Gamma_0(N))$ of cuspidal modular symbols. This kernel will be important in computing cusp forms in Section 3.7 below.

To compute the boundary map on $[\gamma]$, note that $[\gamma] = \{\gamma(0), \gamma(\infty)\}$, so if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$\delta([\gamma]) = \{\gamma(\infty)\} - \{\gamma(0)\} = \{a/c\} - \{b/d\}.$$

Computing this boundary map would appear to first require an algorithm to compute the set $C(\Gamma_0(N)) = \Gamma_0(N)\backslash\mathbb{P}^1(\mathbb{Q})$ of cusps for $\Gamma_0(N)$. In fact, there is a trick that computes the set of cusps in the course of running the algorithm. First, give an algorithm for deciding whether or not two elements of $\mathbb{P}^1(\mathbb{Q})$ are equivalent modulo the action of $\Gamma_0(N)$. Then simply construct $C(\Gamma_0(N))$ in the course of computing the boundary map, i.e., keep a list of cusps found so far, and whenever a new cusp class is discovered, add it to the list. The following proposition, which is proved in [**Cre97a**, Prop. 2.2.3], explains how to determine whether two cusps are equivalent.

Proposition 3.21 (Cremona)**.** *Let (c_i, d_i), $i = 1, 2$, be pairs of integers with $\gcd(c_i, d_i) = 1$ and possibly $d_i = 0$. There is $g \in \Gamma_0(N)$ such that $g(c_1/d_1) = c_2/d_2$ in $\mathbb{P}^1(\mathbb{Q})$ if and only if*

$$s_1 d_2 \equiv s_2 d_1 \pmod{\gcd(d_1 d_2, N)}$$

where s_j satisfies $c_j s_j \equiv 1 \pmod{d_j}$.

In SAGE the command `boundary_map()` computes the boundary map from $\mathbb{M}_2(\Gamma_0(N))$ to $\mathbb{B}_2(\Gamma_0(N))$, and the `cuspidal_submodule()` command computes its kernel. For example, for level 2 the boundary map is given by the matrix $[1\ \ -1]$, and its kernel is the 0 space:

```
sage: M = ModularSymbols(2, 2)
sage: M.boundary_map()
Hecke module morphism boundary map defined by the matrix
[ 1 -1]
Domain: Modular Symbols space of dimension 1 for
Gamma_0(2) of weight ...
Codomain: Space of Boundary Modular Symbols for
Congruence Subgroup Gamma0(2) ...
sage: M.cuspidal_submodule()
Modular Symbols subspace of dimension 0 of Modular
Symbols space of dimension 1 for Gamma_0(2) of weight
2 with sign 0 over Rational Field
```

The smallest level for which the boundary map has nontrivial kernel, i.e., for which $\mathbb{S}_2(\Gamma_0(N)) \neq 0$, is $N = 11$.

```
sage: M = ModularSymbols(11, 2)
sage: M.boundary_map().matrix()
[ 1 -1]
[ 0  0]
[ 0  0]
sage: M.cuspidal_submodule()
Modular Symbols subspace of dimension 2 of Modular
Symbols space of dimension 3 for Gamma_0(11) of weight
2 with sign 0 over Rational Field
sage: S = M.cuspidal_submodule(); S
Modular Symbols subspace of dimension 2 of Modular
Symbols space of dimension 3 for Gamma_0(11) of weight
2 with sign 0 over Rational Field
sage: S.basis()
((1,8), (1,9))
```

The following illustrates that the Hecke operators preserve $\mathbb{S}_2(\Gamma_0(N))$:

```
sage: S.T(2).matrix()
[-2  0]
[ 0 -2]
sage: S.T(3).matrix()
[-1  0]
[ 0 -1]
sage: S.T(5).matrix()
[1 0]
[0 1]
```

A nontrivial fact is that for p prime the eigenvalue of each of these matrices is $p+1-\#E(\mathbb{F}_p)$, where E is the elliptic curve $X_0(11)$ defined by the (affine) equation $y^2+y=x^3-x^2-10x-20$. For example, we have

```
sage: E = EllipticCurve([0,-1,1,-10,-20])
sage: 2 + 1 - E.Np(2)
-2
sage: 3 + 1 - E.Np(3)
-1
sage: 5 + 1 - E.Np(5)
1
sage: 7 + 1 - E.Np(7)
-2
```

The same numbers appear as the eigenvalues of Hecke operators:

```
sage: [S.T(p).matrix()[0,0] for p in [2,3,5,7]]
[-2, -1, 1, -2]
```

In fact, something similar happens for every elliptic curve over $\mathbb{Q}$. The book [**DS05**] (especially Chapter 8) is about this striking numerical relationship between the number of points on elliptic curves modulo p and coefficients of modular forms.

3.6. Computing a Basis for $S_2(\Gamma_0(N))$

This section is about a method for using modular symbols to compute a basis for $S_2(\Gamma_0(N))$. It is not the most efficient for certain applications, but it is easy to explain and understand. See Section 3.7 for a method that takes advantage of additional structure of $S_2(\Gamma_0(N))$.

Let $\mathbb{M}_2(\Gamma_0(N);\mathbb{Q})$ and $\mathbb{S}_2(\Gamma_0(N);\mathbb{Q})$ be the spaces of modular symbols and cuspidal modular symbols over $\mathbb{Q}$. Before we begin, we describe a simple but crucial fact about the relation between cusp forms and Hecke operators.

If $f = \sum b_n q^n \in \mathbb{C}[[q]]$ is a power series, let $a_n(f) = b_n$ be the n coefficient of f. Notice that a_n is a $\mathbb{C}$-linear map $\mathbb{C}[[q]] \to \mathbb{C}$.

As explained in [**DS05**, Prop. 5.3.1] and [**Lan95**, §VII.3] (recall also Proposition 2.31), the Hecke operators T_n act on elements of $M_2(\Gamma_0(N))$ as follows (where $k = 2$ below):

$$T_n\left(\sum_{m=0}^{\infty} a_m q^m\right) = \sum_{m=0}^{\infty}\left(\sum_{1\leq d\,|\,\gcd(n,m)} \varepsilon(d)\cdot d^{k-1}\cdot a_{mn/d^2}\right) q^m, \tag{3.6.1}$$

where $\varepsilon(d) = 1$ if $\gcd(d, N) = 1$ and $\varepsilon(d) = 0$ if $\gcd(d, N) \neq 1$. (Note: More generally, if $f \in M_k(\Gamma_1(N))$ is a modular form with Dirichlet character ε, then the above formula holds; above we are considering this formula in the special case when ε is the trivial character and $k = 2$.)

Lemma 3.22. *Suppose $f \in \mathbb{C}[[q]]$ and n is a positive integer. Let T_n be the operator on q-expansions (formal power series) defined by* (3.6.1). *Then*

$$a_1(T_n(f)) = a_n(f).$$

Proof. The coefficient of q in (3.6.1) is $\varepsilon(1)\cdot 1\cdot a_{1\cdot n/1^2} = a_n$. □

The *Hecke algebra* $\mathbb{T}$ is the ring generated by all Hecke operators T_n acting on $M_k(\Gamma_1(N))$. Let $\mathbb{T}'$ denote the image of the Hecke algebra in $\mathrm{End}(S_2(\Gamma_0(N)))$, and let $\mathbb{T}'_{\mathbb{C}} = \mathbb{T}'\otimes_{\mathbb{Z}}\mathbb{C}$ be the $\mathbb{C}$-span of the Hecke operators. Let $\tilde{\mathbb{T}}_{\mathbb{C}}$ denote the subring of $\mathrm{End}(\mathbb{C}[[q]])$ generated over $\mathbb{C}$ by all Hecke operators acting on formal power series via definition (3.6.1).

Proposition 3.23. *There is a bilinear pairing of complex vector spaces*

$$\mathbb{C}[[q]] \times \tilde{\mathbb{T}}_{\mathbb{C}} \to \mathbb{C}$$

given by

$$\langle f, t\rangle = a_1(t(f)).$$

If f is such that $\langle f, t\rangle = 0$ for all $t \in \tilde{\mathbb{T}}_{\mathbb{C}}$, then $f = 0$.

Proof. The pairing is bilinear since both t and a_1 are linear.

Suppose $f \in \mathbb{C}[[q]]$ is such that $\langle f, t\rangle = 0$ for all $t \in \tilde{\mathbb{T}}_{\mathbb{C}}$. Then $\langle f, T_n\rangle = 0$ for each positive integer n. But by Lemma 3.22 we have

$$a_n(f) = a_1(T_n(f)) = 0$$

for all n; thus $f = 0$. □

Proposition 3.24. *There is a perfect bilinear pairing of complex vector spaces*

$$S_2(\Gamma_0(N)) \times \mathbb{T}'_{\mathbb{C}} \to \mathbb{C}$$

given by

$$\langle f, t\rangle = a_1(t(f)).$$

Proof. The pairing has 0 kernel on the left by Proposition 3.23. Suppose that $t \in \mathbb{T}'_{\mathbb{C}}$ is such that $\langle f, t\rangle = 0$ for all $f \in S_2(\Gamma_0(N))$. Then $a_1(t(f)) = 0$ for all f. For any n, the image $T_n(f)$ is also a cusp form, so $a_1(t(T_n(f))) = 0$ for all n and f. Finally the fact that $\mathbb{T}'$ is commutative and Lemma 3.22 together imply that for all n and f,

$$0 = a_1(t(T_n(f))) = a_1(T_n(t(f))) = a_n(t(f)),$$

so $t(f) = 0$ for all f. Thus t is the 0 operator.

Since $S_2(\Gamma_0(N))$ has finite dimension and the kernel on each side of the pairing is 0, it follows that the pairing is perfect, i.e., defines an *isomorphism*

$$\mathbb{T}'_{\mathbb{C}} \cong \mathrm{Hom}_{\mathbb{C}}(S_2(\Gamma_0(N)); \mathbb{C}).$$

□

By Proposition 3.24 there is an isomorphism of vector spaces

$$\Psi : S_2(\Gamma_0(N)) \xrightarrow{\cong} \mathrm{Hom}(\mathbb{T}'_{\mathbb{C}}, \mathbb{C}) \tag{3.6.2}$$

that sends $f \in S_2(\Gamma_0(N))$ to the homomorphism

$$t \mapsto a_1(t(f)).$$

For any $\mathbb{C}$-linear map $\varphi : \mathbb{T}'_{\mathbb{C}} \to \mathbb{C}$, let

$$f_\varphi = \sum_{n=1}^{\infty} \varphi(T_n) q^n \in \mathbb{C}[[q]].$$

Lemma 3.25. *The series f_φ is the q-expansion of $\Psi^{-1}(\varphi) \in S_2(\Gamma_0(N))$.*

Proof. Note that it is not even *a priori* obvious that f_φ is the q-expansion of a modular form. Let $g = \Psi^{-1}(\varphi)$, which is by definition the unique element of $S_2(\Gamma_0(N))$ such that $\langle g, T_n\rangle = \varphi(T_n)$ for all n. By Lemma 3.22, we have

$$\langle f_\varphi, T_n\rangle = a_1(T_n(f_\varphi)) = a_n(f_\varphi) = \varphi(T_n),$$

so $\langle f_\varphi - g, T_n\rangle = 0$ for all n. Proposition 3.23 implies that $f_\varphi - g = 0$, so $f_\varphi = g = \Psi^{-1}(\varphi)$, as claimed. □

Conclusion: The cusp forms f_φ, as φ varies through a basis of $\mathrm{Hom}(\mathbb{T}'_{\mathbb{C}}, \mathbb{C})$, form a basis for $S_2(\Gamma_0(N))$. In particular, we can compute $S_2(\Gamma_0(N))$ by computing $\mathrm{Hom}(\mathbb{T}'_{\mathbb{C}}, \mathbb{C})$, where we compute $\mathbb{T}'$ in any way we want, e.g., using a space that contains an isomorphic copy of $S_2(\Gamma_0(N))$.

Algorithm 3.26 (Basis of Cusp Forms)**.** *Given positive integers N and B, this algorithm computes a basis for $S_2(\Gamma_0(N))$ to precision $O(q^B)$.*

(1) Compute $\mathbb{M}_2(\Gamma_0(N);\mathbb{Q})$ via the presentation of Section 3.3.2.
(2) Compute the subspace $\mathbb{S}_2(\Gamma_0(N);\mathbb{Q})$ of cuspidal modular symbols as in Section 3.5.
(3) Let $d = \frac{1}{2}\cdot\dim \mathbb{S}_2(\Gamma_0(N);\mathbb{Q})$. By Proposition 3.8, d is the dimension of $S_2(\Gamma_0(N))$.
(4) Let $[T_n]$ denote the matrix of T_n acting on a basis of $\mathbb{S}_2(\Gamma_0(N);\mathbb{Q})$. For a matrix A, let $a_{ij}(A)$ denote the ijth entry of A. For various integers i, j with $0 \leq i, j \leq d-1$, compute formal q-expansions

$$f_{ij}(q) = \sum_{n=1}^{B-1} a_{ij}([T_n])q^n + O(q^B) \in \mathbb{Q}[[q]]$$

until we find enough to span a space of dimension d (or exhaust all of them). These f_{ij} are a basis for $S_2(\Gamma_0(N))$ to precision $O(q^B)$.

3.6.1. Examples.

We use SAGE to demonstrate Algorithm 3.26.

Example 3.27. The smallest N with $S_2(\Gamma_0(N)) \neq 0$ is $N = 11$.

```
sage: M = ModularSymbols(11); M.basis()
((1,0), (1,8), (1,9))
sage: S = M.cuspidal_submodule(); S
Modular Symbols subspace of dimension 2 of Modular
Symbols space of dimension 3 for Gamma_0(11) of weight
2 with sign 0 over Rational Field
```

We compute a few Hecke operators, and then read off a nonzero cusp form, which forms a basis for $S_2(\Gamma_0(11))$:

```
sage: S.T(2).matrix()
[-2  0]
[ 0 -2]
sage: S.T(3).matrix()
[-1  0]
[ 0 -1]
```

Thus

$$f_{0,0} = q - 2q^2 - q^3 + \cdots \in S_2(\Gamma_0(11))$$

forms a basis for $S_2(\Gamma_0(11))$.

Example 3.28. We compute a basis for $S_2(\Gamma_0(33))$ to precision $O(q^6)$.

```
sage: M = ModularSymbols(33)
sage: S = M.cuspidal_submodule(); S
Modular Symbols subspace of dimension 6 of Modular
Symbols space of dimension 9 for Gamma_0(33) of weight
2 with sign 0 over Rational Field
```

Thus $\dim S_2(\Gamma_0(33)) = 3$.

```
sage: R.<q> = PowerSeriesRing(QQ)
sage: v = [S.T(n).matrix()[0,0] for n in range(1,6)]
sage: f00 = sum(v[n-1]*q^n for n in range(1,6)) + O(q^6)
sage: f00
q - q^2 - q^3 + q^4 + O(q^6)
```

This gives us one basis element of $S_2(\Gamma_0(33))$. It remains to find two others. We find

```
sage: v = [S.T(n).matrix()[0,1] for n in range(1,6)]
sage: f01 = sum(v[n-1]*q^n for n in range(1,6)) + O(q^6)
sage: f01
-2*q^3 + O(q^6)
```

and

```
sage: v = [S.T(n).matrix()[1,0] for n in range(1,6)]
sage: f10 = sum(v[n-1]*q^n for n in range(1,6)) + O(q^6)
sage: f10
q^3 + O(q^6)
```

This third one is (to our precision) a scalar multiple of the second, so we look further.

```
sage: v = [S.T(n).matrix()[1,1] for n in range(1,6)]
sage: f11 = sum(v[n-1]*q^n for n in range(1,6)) + O(q^6)
sage: f11
q - 2*q^2 + 2*q^4 + q^5 + O(q^6)
```

This latter form is clearly not in the span of the first two. Thus we have the following basis for $S_2(\Gamma_0(33))$ (to precision $O(q^6)$):

$$\begin{aligned} f_{00} &= q - q^2 - q^3 + q^4 + \cdots, \\ f_{11} &= q - 2q^2 + 2q^4 + q^5 + \cdots, \\ f_{10} &= q^3 + \cdots. \end{aligned}$$

Example 3.29. Next consider $N = 23$, where we have

$$d = \dim S_2(\Gamma_0(23)) = 2.$$

The command q_expansion_cuspforms computes matrices T_n and returns a function f such that $f(i,j)$ is the q-expansion of $f_{i,j}$ to some precision. (For efficiency reasons, $f(i,j)$ in SAGE actually computes matrices of T_n acting on a basis for the linear dual of $\mathbb{S}_2(\Gamma_0(N))$.)

```
sage: M = ModularSymbols(23)
sage: S = M.cuspidal_submodule()
sage: S
Modular Symbols subspace of dimension 4 of Modular
Symbols space of dimension 5 for Gamma_0(23) of weight
2 with sign 0 over Rational Field
sage: f = S.q_expansion_cuspforms(6)
sage: f(0,0)
q - 2/3*q^2 + 1/3*q^3 - 1/3*q^4 - 4/3*q^5 + O(q^6)
sage: f(0,1)
O(q^6)
sage: f(1,0)
-1/3*q^2 + 2/3*q^3 + 1/3*q^4 - 2/3*q^5 + O(q^6)
```

Thus a basis for $S_2(\Gamma_0(23))$ is

$$f_{0,0} = q - \frac{2}{3}q^2 + \frac{1}{3}q^3 - \frac{1}{3}q^4 - \frac{4}{3}q^5 + \cdots,$$
$$f_{1,0} = -\frac{1}{3}q^2 + \frac{2}{3}q^3 + \frac{1}{3}q^4 - \frac{2}{3}q^5 + \cdots.$$

Or, in echelon form,

$$q - q^3 - q^4 + \cdots$$
$$q^2 - 2q^3 - q^4 + 2q^5 + \cdots$$

which we computed using

```
sage: S.q_expansion_basis(6)
[
q - q^3 - q^4 + O(q^6),
q^2 - 2*q^3 - q^4 + 2*q^5 + O(q^6)
]
```

3.7. Computing $S_2(\Gamma_0(N))$ Using Eigenvectors

In this section we describe how to use modular symbols to construct a basis of $S_2(\Gamma_0(N))$ consisting of modular forms that are eigenvectors for every element of the ring $\mathbb{T}^{(N)}$ generated by the Hecke operator T_p, with $p \nmid N$. Such eigenvectors are called *eigenforms*.

Suppose M is a positive integer that divides N. As explained in [**Lan95**, VIII.1–2], for each divisor d of N/M there is a natural *degeneracy map* $\alpha_{M,d}$: $S_2(\Gamma_0(M)) \to S_2(\Gamma_0(N))$ given by $\alpha_{M,d}(f(q)) = f(q^d)$. The *new subspace* of $S_2(\Gamma_0(N))$, denoted $S_2(\Gamma_0(N))_{\text{new}}$, is the complementary $\mathbb{T}$-submodule of the $\mathbb{T}$-module generated by the images of all maps $\alpha_{M,d}$, with M and d as above. It is a nontrivial fact that this complement is well defined; one possible proof uses the Petersson inner product (see [**Lan95**, §VII.5]).

The theory of Atkin and Lehner [**AL70**] (see Theorem 9.4 below) asserts that, as a $\mathbb{T}^{(N)}$-module, $S_2(\Gamma_0(N))$ decomposes as follows:

$$S_2(\Gamma_0(N)) \quad = \bigoplus_{M|N,\ d|N/M} \beta_{M,d}(S_2(\Gamma_0(M))_{\text{new}}).$$

To compute $S_2(\Gamma_0(N))$ it suffices to compute $S_2(\Gamma_0(M))_{\text{new}}$ for each $M \mid N$.

We now turn to the problem of computing $S_2(\Gamma_0(N))_{\text{new}}$. Atkin and Lehner [**AL70**] proved that $S_2(\Gamma_0(N))_{\text{new}}$ is spanned by eigenforms for all T_p with $p \nmid N$ and that the common eigenspaces of all the T_p with $p \nmid N$ each have dimension 1. Moreover, if $f \in S_2(\Gamma_0(N))_{\text{new}}$ is an eigenform then the coefficient of q in the q-expansion of f is nonzero, so it is possible to normalize f so the coefficient of q is 1 (such a *normalized* eigenform in the new subspace is called a *newform*). With f so normalized, if $T_p(f) = a_p f$, then the pth Fourier coefficient of f is a_p. If $f = \sum_{n=1}^{\infty} a_n q^n$ is a normalized eigenvector for all T_p, then the a_n, with n composite, are determined by the a_p, with p prime, by the following formulas: $a_{nm} = a_n a_m$ when n and m are relatively prime and $a_{p^r} = a_{p^{r-1}} a_p - p a_{p^{r-2}}$ for $p \nmid N$ prime. When $p \mid N$, $a_{p^r} = a_p^r$. We conclude that in order to compute $S_2(\Gamma_0(N))_{\text{new}}$, it suffices to compute all systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$ of the prime-indexed Hecke operators $T_2, T_3, T_5, \ldots$ acting on $S_2(\Gamma_0(N))_{\text{new}}$. Given a system of eigenvalues, the corresponding eigenform is $f = \sum_{n=1}^{\infty} a_n q^n$, where the a_n, for n composite, are determined by the recurrence given above.

In light of the pairing $\langle\ ,\ \rangle$ introduced in Section 3.1, computing the above systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$ amounts to computing the systems of eigenvalues of the Hecke operators T_p on the subspace V of $\mathbb{S}_2(\Gamma_0(N))$ that corresponds to the new subspace of $S_2(\Gamma_0(N))$. For each proper divisor M of N and each divisor d of N/M, let $\phi_{M,d} : \mathbb{S}_2(\Gamma_0(N)) \to \mathbb{S}_2(\Gamma_0(M))$ be the map sending x to $\left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right) x$. Then V is the intersection of the kernels of all maps $\phi_{M,d}$.

Computing the systems of eigenvalues of a collection of commuting diagonalizable endomorphisms is a problem in linear algebra (see Chapter 7).

Example 3.30. All forms in $S_2(\Gamma_0(39))$ are new. Up to Galois conjugacy, the eigenvalues of the Hecke operators T_2, T_3, T_5, and T_7 on $\mathbb{S}_2(\Gamma_0(39))$ are $\{1, -1, 2, -4\}$ and $\{a, 1, -2a-2, 2a+2\}$, where $a^2+2a-1=0$. Each of these eigenvalues occur in $\mathbb{S}_2(\Gamma_0(39))$ with multiplicity two; for example, the characteristic polynomial of T_2 on $\mathbb{S}_2(\Gamma_0(39))$ is $(x-1)^2 \cdot (x^2+2x-1)^2$. Thus $S_2(\Gamma_0(39))$ is spanned by

$$\begin{aligned}
f_1 &= q + q^2 - q^3 - q^4 + 2q^5 - q^6 - 4q^7 + \cdots,\\
f_2 &= q + aq^2 + q^3 + (-2a-1)q^4 + (-2a-2)q^5 + aq^6 + (2a+2)q^7 + \cdots,\\
f_3 &= q + \sigma(a)q^2 + q^3 + (-2\sigma(a)-1)q^4 + (-2\sigma(a)-2)q^5 + \sigma(a)q^6 + \cdots,
\end{aligned}$$

where $\sigma(a)$ is the other $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugate of a.

3.7.1. Summary. To compute the q-expansion of a basis for $S_2(\Gamma_0(N))$, we use the degeneracy maps so that we only have to solve the problem for $S_2(\Gamma_0(M))_{\mathrm{new}}$, for all integers $M \mid N$. Using modular symbols, we compute all systems of eigenvalues $\{a_2, a_3, a_5, \ldots\}$, and then write down the corresponding eigenforms $\sum a_n q^n$.

3.8. Exercises

3.1 Suppose that $\lambda, \lambda' \in \mathfrak{h}$ are in the same orbit for the action of $\Gamma_0(N)$, i.e., that there exists $g \in \Gamma_0(N)$ such that $g(\lambda) = \lambda'$. Let $\Lambda = \mathbb{Z} + \mathbb{Z}\lambda$ and $\Lambda' = \mathbb{Z} + \mathbb{Z}\lambda'$. Prove that the pairs $(\mathbb{C}/\Lambda, (\frac{1}{N}\mathbb{Z} + \Lambda)/\Lambda)$ and $(\mathbb{C}/\Lambda', (\frac{1}{N}\mathbb{Z} + \Lambda')/\Lambda')$ are isomorphic. (By an isomorphism $(E, C) \to (F, D)$ of pairs, we mean an isomorphism $\phi : E \to F$ of elliptic curves that sends C to D. You may use the fact that an isomorphism of elliptic curves over $\mathbb{C}$ is a $\mathbb{C}$-linear map $\mathbb{C} \to \mathbb{C}$ that sends the lattice corresponding to one curve onto the lattice corresponding to the other.)

3.2 Let n, m be integers and N a positive integer. Prove that the modular symbol $\{n, m\}$ is 0 as an element of $\mathbb{M}_2(\Gamma_0(N))$. [Hint: See Example 3.6.]

3.3 Let p be a prime.
(a) List representative elements of $\mathbb{P}^1(\mathbb{Z}/p\mathbb{Z})$.
(b) What is the cardinality of $\mathbb{P}^1(\mathbb{Z}/p\mathbb{Z})$ as a function of p?
(c) Prove that there is a bijection between the right cosets of $\Gamma_0(p)$ in $\mathrm{SL}_2(\mathbb{Z})$ and the elements of $\mathbb{P}^1(\mathbb{Z}/p\mathbb{Z})$ that sends $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to $(c : d)$. (As mentioned in this chapter, the analogous statement is also true when the level is composite; see [**Cre97a**, §2.2] for complete details.)

3.4 Use the inductive proof of Proposition 3.11 to write $\{0, 4/7\}$ in terms of Manin symbols for $\Gamma_0(7)$.

3.5 Show that the Hecke operator T_2 acts as multiplication by 3 on the space $\mathbb{M}_2(\Gamma_0(3))$ as follows:

(a) Write down right coset representatives for $\Gamma_0(3)$ in $\mathrm{SL}_2(\mathbb{Z})$.

(b) List all eight relations coming from Theorem 3.13.

(c) Find a single Manin symbols $[r_i]$ so that the three other Manin symbols are a nonzero multiple of $[r_i]$ modulo the relations found in the previous step.

(d) Use formula (3.4.1) to compute $T_2([r_i])$. You will obtain a sum of four symbols. Using the relations above, write this sum as a multiple of $[r_i]$. (The multiple must be 3 or you made a mistake.)

Chapter 4

Dirichlet Characters

In this chapter we develop a theory for computing with Dirichlet characters, which are extremely important to computations with modular forms for (at least) two reasons:

(1) To compute the Eisenstein subspace $E_k(\Gamma_1(N))$ of $M_k(\Gamma_1(N))$, we write down Eisenstein series attached to pairs of Dirichlet characters (the space $E_k(\Gamma_1(N))$ will be defined in Chapter 5).

(2) To compute $S_k(\Gamma_1(N))$, we instead compute a decomposition

$$M_k(\Gamma_1(N)) = \bigoplus M_k(\Gamma_1(N), \varepsilon)$$

and then compute each factor (see Section 9.1). Here the sum is over all Dirichlet characters ε of modulus N.

Dirichlet characters appear frequently in many other areas of number theory. For example, by the Kronecker-Weber theorem, Dirichlet characters correspond to the 1-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

After defining Dirichlet characters in Section 4.1, in Section 4.2 we describe a good way to represent Dirichlet characters using a computer. Section 4.3 is about how to evaluate Dirichlet characters and leads naturally to a discussion of the baby-step giant-step algorithm for solving the discrete log problem and methods for efficiently computing the Kronecker symbol. In Section 4.4 we explain how to factor Dirichlet characters into their prime power constituents and apply this to the computations of conductors. We describe how to carry out a number of standard operations with Dirichlet characters in Section 4.6 and discuss alternative ways to represent them in Section 4.7. Finally, in Section 4.8 we give a very short tutorial about how to compute with Dirichlet characters using SAGE.

4.1. The Definition

Fix an integral domain R and a root ζ of unity in R.

Definition 4.1 (Dirichlet Character)**.** A *Dirichlet character* of modulus N over R is a map $\varepsilon : \mathbb{Z} \to R$ such that there is a homomorphism $f : (\mathbb{Z}/N\mathbb{Z})^* \to \langle\zeta\rangle$ for which

$$\varepsilon(a) = \begin{cases} 0 & \text{if } \gcd(a, N) > 1, \\ f\,(a \bmod N) & \text{if } \gcd(a, N) = 1. \end{cases}$$

We denote the group of such Dirichlet characters by $D(N, R)$. Note that elements of $D(N, R)$ are in bijection with homomorphisms $(\mathbb{Z}/N\mathbb{Z})^* \to \langle\zeta\rangle$.

A familiar Dirichlet character is the Legendre symbol $\left(\frac{a}{p}\right)$, with p an odd prime, that appears in quadratic reciprocity theory. It is a Dirichlet character of modulus p that takes the value 1 on integers that are congruent to a nonzero square modulo p, the value -1 on integers that are congruent to a nonzero nonsquare modulo p, and 0 on integers divisible by p.

4.2. Representing Dirichlet Characters

Lemma 4.2. *The groups $(\mathbb{Z}/N\mathbb{Z})^*$ and $D(N, \mathbb{C})$ are isomorphic.*

Proof. We prove the more general fact that for any finite abelian group G, we have that $G \approx \mathrm{Hom}(G, \mathbb{C}^*)$. To deduce this latter isomorphism, first reduce to the case when G is cyclic by writing G as a product of cyclic groups. The cyclic case follows because if G is cyclic of order n, then $\mathbb{C}^*$ contains an nth root of unity, so $\mathrm{Hom}(G, \mathbb{C}^*)$ is also cyclic of order n. Any two cyclic groups of the same order are isomorphic, so G and $\mathrm{Hom}(G, \mathbb{C}^*)$ are isomorphic. □

Corollary 4.3. *We have $\#D(N, R) \mid \varphi(N)$, with equality if and only if the order of our choice of $\zeta \in R$ is a multiple of the exponent of the group $(\mathbb{Z}/N\mathbb{Z})^*$.*

Proof. This is because $\#(\mathbb{Z}/N\mathbb{Z})^* = \varphi(N)$. □

Fix a positive integer N. To find the set of "canonical" generators for the group $(Z/NZ)^*$, write $N = \prod_{i=0}^{n} p_i^{e_i}$ where $p_0 < p_1 < \cdots < p_n$ are the prime divisors of N. By Exercise 4.2, each factor $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ is a cyclic group $C_i = \langle g_i \rangle$, except if $p_0 = 2$ and $e_0 \geq 3$, in which case $(\mathbb{Z}/p_0^{e_0}\mathbb{Z})^*$ is a product of the cyclic subgroup $C_0 = \langle -1 \rangle$ of order 2 with the cyclic subgroup $C_1 = \langle 5 \rangle$. In all cases we have

$$(\mathbb{Z}/N\mathbb{Z})^* \cong \prod_{0 \leq i \leq n} C_i = \prod_{0 \leq i \leq n} \langle g_i \rangle.$$

For i such that $p_i > 2$, choose the generator g_i of C_i to be the element of $\{2, 3, \ldots, p_i^{e_i} - 1\}$ that is smallest and generates. Finally, use the Chinese Remainder Theorem (see [**Coh93**, §1.3.3]) to lift each g_i to an element in $(\mathbb{Z}/N\mathbb{Z})^*$, also denoted g_i, that is 1 modulo each $p_j^{e_j}$ for $j \neq i$.

Algorithm 4.4 (Minimal Generator for $(\mathbb{Z}/p^r\mathbb{Z})^*$)**.** *Given a prime power p^r with p odd, this algorithm computes the minimal generator of $(\mathbb{Z}/p^r\mathbb{Z})^*$.*

(1) [Factor Group Order] Factor $n = \phi(p^r) = p^{r-1} \cdot 2 \cdot ((p-1)/2)$ as a product $\prod p_i^{e_i}$ of primes. This is equivalent in difficulty to factoring $(p-1)/2$. (See, e.g., [**Coh93**, Ch.8, Ch. 10] for an excellent discussion of factorization algorithms, though of course much progress has been made since then.)
(2) [Initialize] Set $g = 2$.
(3) [Generator?] Using the binary powering algorithm (see [**Coh93**, §1.2]), compute $g^{n/p_i} \pmod{p^r}$, for each prime divisor p_i of n. If any of these powers are 1, then g is not a generator, so set $g = g+1$ and go to step (2). If no powers are 1, output g and terminate.

See Exercise 4.3 for a proof that this algorithm is correct.

Example 4.5. A minimal generator for $(\mathbb{Z}/49\mathbb{Z})^*$ is 3. We have $n = \varphi(49) = 42 = 2 \cdot 3 \cdot 7$ and

$$2^{n/2} \equiv 1, \qquad 2^{n/3} \equiv 18, \qquad 2^{n/7} \equiv 15 \pmod{49},$$

so 2 is not a generator for $(\mathbb{Z}/49\mathbb{Z})^*$. (We see this just from $2^{n/2} \equiv 1 \pmod{49}$.) However 3 is a generator since

$$3^{n/2} \equiv 48, \qquad 3^{n/3} \equiv 30, \qquad 3^{n/7} \equiv 43 \pmod{49}.$$

Example 4.6. In this example we compute minimal generators for $N = 25$, 100, and 200:

(1) The minimal generator for $(\mathbb{Z}/25\mathbb{Z})^*$ is 2.
(2) The minimal generators for $(\mathbb{Z}/100\mathbb{Z})^*$, lifted to numbers modulo 100, are $g_0 = 51$ and $g_1 = 77$. Notice that $g_0 \equiv -1 \pmod 4$ and $g_0 \equiv 1 \pmod{25}$ and that $g_1 \equiv 2 \pmod{25}$ is the minimal generator modulo 25.
(3) The minimal generators for $(\mathbb{Z}/200\mathbb{Z})^*$, lifted to numbers modulo 200, are $g_0 = 151$, $g_1 = 101$, and $g_2 = 177$. Note that $g_0 \equiv -1 \pmod 4$, that $g_1 \equiv 5 \pmod 8$ and $g_2 \equiv 2 \pmod{25}$.

In SAGE, the command `Integers(N)` creates $\mathbb{Z}/N\mathbb{Z}$.

```
sage: R = Integers(49)
sage: R
Ring of integers modulo 49
```

The unit_gens command computes the minimal generators for $(\mathbb{Z}/N\mathbb{Z})^*$, as defined above.

```
sage: R.unit_gens()
[3]
sage: Integers(25).unit_gens()
[2]
sage: Integers(100).unit_gens()
[51, 77]
sage: Integers(200).unit_gens()
[151, 101, 177]
sage: Integers(2005).unit_gens()
[402, 1206]
sage: Integers(200000000).unit_gens()
[174218751, 51562501, 187109377]
```

Fix an element ζ of finite multiplicative order in a ring R, and let $D(N,R)$ denote the group of Dirichlet characters of modulus N over R, with image in $\langle\zeta\rangle \cup \{0\}$. In most of this chapter, we specify an element $\varepsilon \in D(N,R)$ by giving the list

$$[\varepsilon(g_0), \varepsilon(g_1), \ldots, \varepsilon(g_n)] \tag{4.2.1}$$

of images of the generators of $(\mathbb{Z}/N\mathbb{Z})^*$. (Note that if N is even, the number of elements of the list (4.2.1) *does* depend on whether or not $8 \mid N$—there are two factors corresponding to 2 if $8 \mid N$, but only one if $8 \nmid N$.) This representation completely determines ε and is convenient for arithmetic operations. It is analogous to representing a linear transformation by a matrix.

Remark 4.7. In any actual implementation (e.g., the one in SAGE), it is better to represent the $\varepsilon(g_i)$ by recording an integer j such that $\varepsilon(g_i) = \zeta^j$, where $\zeta \in R$ is a fixed root of unity. Then (4.2.1) is internally represented as an element of $(\mathbb{Z}/m\mathbb{Z})^{n+1}$, where m is the multiplicative order of ζ. When the representation of (4.2.1) is needed for an algorithm, it can be quickly computed on the fly using a table of the powers of ζ. See Section 4.7 for further discussion about ways to represent characters.

Example 4.8. The group $D(5,\mathbb{C})$ has elements $\{[1],[i],[-1],[-i]\}$, so it is cyclic of order $\varphi(5) = 4$. In contrast, the group $D(5,\mathbb{Q})$ has only the two elements [1] and $[-1]$ and order 2. The command `DirichletGroup(N)` with no second argument creates the group of Dirichlet characters with values in the cyclotomic field $\mathbb{Q}(\zeta_n)$, where n is the exponent of the group $(\mathbb{Z}/N\mathbb{Z})^*$. Every element in $D(N, \mathbb{C})$ takes values in $\mathbb{Q}(\zeta_n)$, so $D(N, \mathbb{Q}(\zeta_n)) \approx D(N, \mathbb{C})$.

```
sage: list(DirichletGroup(5))
[[1], [zeta4], [-1], [-zeta4]]
sage: list(DirichletGroup(5, QQ))
[[1], [-1]]
```

4.3. Evaluation of Dirichlet Characters

This section is about how to compute $\varepsilon(n)$, where ε is a Dirichlet character and n is an integer. We begin with an example.

Example 4.9. If $N = 200$, then $g_0 = 151$, $g_1 = 101$ and $g_2 = 177$, as we saw in Example 4.6. The exponent of $(\mathbb{Z}/200\mathbb{Z})^*$ is 20, since that is the least common multiple of the exponents of $4 = \#(\mathbb{Z}/8\mathbb{Z})^*$ and $20 = \#(\mathbb{Z}/25\mathbb{Z})^*$. The orders of g_0, g_1, and g_2 are 2, 2, and 20. Let $\zeta = \zeta_{20}$ be a primitive 20th root of unity in $\mathbb{C}$. Then the following are generators for $D(200, \mathbb{C})$:

$$\varepsilon_0 = [-1, 1, 1], \qquad \varepsilon_1 = [1, -1, 1], \qquad \varepsilon_2 = [1, 1, \zeta],$$

and $\varepsilon = [1, -1, \zeta^5]$ is an example element of order 4. To evaluate $\varepsilon(3)$, we write 3 in terms of g_0, g_1, and g_2. First, reducing 3 modulo 8, we see that $3 \equiv g_0 \cdot g_1 \pmod{8}$. Next reducing 3 modulo 25 and trying powers of $g_2 = 2$, we find that $e \equiv g_2^7 \pmod{25}$. Thus

$$\begin{aligned}
\varepsilon(3) &= \varepsilon(g_0 \cdot g_1 \cdot g_2^7) \\
&= \varepsilon(g_0)\varepsilon(g_1)\varepsilon(g_2)^7 \\
&= 1 \cdot (-1) \cdot (\zeta^5)^7 \\
&= -\zeta^{35} = -\zeta^{15}.
\end{aligned}$$

We next illustrate the above computation of $\varepsilon(3)$ in SAGE. First we make the group $D(200, \mathbb{Q}(\zeta_8))$ and list its generators.

```
sage: G = DirichletGroup(200)
sage: G
Group of Dirichlet characters of modulus 200 over
Cyclotomic Field of order 20 and degree 8
sage: G.exponent()
20
sage: G.gens()
([-1, 1, 1], [1, -1, 1], [1, 1, zeta20])
```

We construct ε.

```
sage: K = G.base_ring()
sage: zeta = K.0
sage: eps = G([1,-1,zeta^5])
sage: eps
[1, -1, zeta20^5]
```

Finally, we evaluate ε at 3.

```
sage: eps(3)
zeta20^5
sage: -zeta^15
zeta20^5
```

Example 4.9 illustrates that if ε is represented using a list as described above, evaluation of ε is inefficient without extra information; it requires solving the discrete log problem in $(\mathbb{Z}/N\mathbb{Z})^*$.

Remark 4.10. For a general character ε, is calculation of ε at least as hard as finding discrete logarithms? Quadratic characters are easier—see Algorithm 4.23.

Algorithm 4.11 (Evaluate ε)**.** *Given a Dirichlet character ε of modulus N, represented by a list $[\varepsilon(g_0), \varepsilon(g_1), \ldots, \varepsilon(g_n)]$, and an integer a, this algorithm computes $\varepsilon(a)$.*

(1) [GCD] Compute $g = \gcd(a, N)$. If $g > 1$, output 0 and terminate.
(2) [Discrete Log] For each i, write $a \pmod{p_i^{e_i}}$ as a power m_i of g_i using some algorithm for solving the discrete log problem (see below). If $p_i = 2$, write $a \pmod{p_i^{e_i}}$ as $(-1)^{m_0} \cdot 5^{m_1}$. (This step is analogous to writing a vector in terms of a basis.)

(3) [Multiply] Output $\prod \varepsilon(g_i)^{m_i}$ as an element of R, and terminate. (This is analogous to multiplying a matrix times a vector.)

4.3.1. The Discrete Log Problem.

Exercise 4.4 gives an isomorphism of groups

$$(1 + p^{n-1}(\mathbb{Z}/p^n\mathbb{Z}), \times) \cong (\mathbb{Z}/p\mathbb{Z}, +),$$

so one sees by induction that step (2) is "about as difficult" as finding a discrete log in $(\mathbb{Z}/p\mathbb{Z})^*$. There is an algorithm called "baby-step giant-step", which solves the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$ in time $O(\sqrt{\ell})$, where ℓ is the largest prime factor of $p - 1 = \#(\mathbb{Z}/p\mathbb{Z})^*$ (note that the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$ reduces to a series of discrete log problems in each prime-order cyclic factor). This is unfortunately still exponential in the number of digits of ℓ; it also uses $O(\sqrt{\ell})$ memory. We now describe this algorithm without any specific optimizations.

Algorithm 4.12 (Baby-step Giant-step Discrete Log)**.** *Given a prime p, a generator g of $(\mathbb{Z}/p\mathbb{Z})^*$, and an element $a \in (\mathbb{Z}/p\mathbb{Z})^*$, this algorithm finds an n such that $g^n = a$. (Note that this algorithm works in any cyclic group, not just $(\mathbb{Z}/p\mathbb{Z})^*$.)*

(1) [Make Lists] Let $m = \lceil \sqrt{p} \rceil$ be the ceiling of $\sqrt{p}$, and construct two lists

$$1,\, g^m,\, \ldots,\, g^{(m-1)m} \qquad \text{(giant steps)}$$

and

$$a, ag,\, ag^2,\, \ldots,\, ag^{m-1} \qquad \text{(baby steps).}$$

(2) [Find Match] Sort the two lists and find a match $g^{im} = ag^j$. Then $a = g^{im-j}$.

Proof. We prove that there will always be a match. Since we know that $a = g^k$ for some k with $0 \leq k \leq p - 1$ and any such k can be written in the form $im - j$ for $0 \leq i, j \leq m - 1$, we will find such a match. □

Algorithm 4.12 uses nothing special about $(\mathbb{Z}/p\mathbb{Z})^*$, so it works in a generic group. It is a theorem that there is no faster algorithm to find discrete logs in a "generic group" (see [**Sho97, Nec94**]). There are much better subexponential algorithms for solving the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$, which use the special structure of this group. They use the number field sieve (see, e.g., [**Gor93**]), which is also the best-known algorithm for factoring integers. This class of algorithms has been very well studied by cryptographers; though sub-exponential, solving discrete log problems when p is large is still extremely difficult. For a more in-depth survey see [**Gor04**]. For computing Dirichlet characters in our context, p is not too large, so Algorithm 4.12 works well.

4.3.2. Enumeration of All Values. For many applications of Dirichlet characters to computing modular forms, N is fairly small, e.g., $N < 10^6$, and we evaluate ε on a *huge* number of random elements, inside inner loops of algorithms. Thus for such purposes it will often be better to make a table of all values of ε, so that evaluation of ε is extremely fast. The following algorithm computes a table of all values of ε, and it does not require computing any discrete logs since we are computing *all* values.

Algorithm 4.13 (Values of ε)**.** *Given a Dirichlet character ε represented by the list of values of ε on the minimal generators g_i of $(\mathbb{Z}/N\mathbb{Z})^*$, this algorithm creates a list of all the values of ε.*

(1) [Initialize] For each minimal generator g_i, set $a_i = 0$. Let $n = \prod g_i^{a_i}$, and set $z = 1$. Create a list v of N values, all initially set equal to 0. When this algorithm terminates, the list v will have the property that

$$v\ [x \pmod N] = \varepsilon(x).$$

Notice that we index v starting at 0.

(2) [Add Value to Table] Set $v[n] = z$.

(3) [Finished?] If each a_i is one less than the order of g_i, output v and terminate.

(4) [Increment] Set $a_0 = a_0 + 1$, $n = n \cdot g_0 \pmod N$, and $z = z \cdot \varepsilon(g_0)$. If $a_0 \geq \mathrm{ord}(g_0)$, set $a_0 \to 0$, and then set $a_1 = a_1 + 1$, $n = n \cdot g_1 \pmod N$, and $z = z \cdot \varepsilon(g_1)$. If $a_1 \geq \mathrm{ord}(g_1)$, do what you just did with a_0 but with all subscripts replaced by 1. Etc. (Imagine a car odometer.) Go to step (2).

4.4. Conductors of Dirichlet Characters

The following algorithm for computing the order of ε reduces the problem to computing the orders of powers of ζ in R.

Algorithm 4.14 (Order of Character)**.** *This algorithm computes the order of a Dirichlet character $\varepsilon \in D(N, R)$.*

(1) Compute the order r_i of each $\varepsilon(g_i)$, for each minimal generator g_i of $(\mathbb{Z}/N\mathbb{Z})^*$. The order of $\varepsilon(g_i)$ is a divisor of $n = \#(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ so we can compute its order by considering the divisors of n.

(2) Compute and output the least common multiple of the integers r_i.

Remark 4.15. Computing the order of $\varepsilon(g_i) \in R$ is potentially difficult. Simultaneously using a different representation of Dirichlet characters avoids having to compute the order of elements of R (see Section 4.7).

The next algorithm factors ε as a product of "local" characters, one for each prime divisor of N. It is useful for other algorithms, e.g., for explicit computations with trace formulas (see [**Hij74**]). This factorization is easy to compute because of how we represent ε.

Algorithm 4.16 (Factorization of Character)**.** *Given a Dirichlet character $\varepsilon \in D(N,R)$, with $N = \prod p_i^{e_i}$, this algorithm finds Dirichlet characters ε_i modulo $p_i^{e_i}$, such that for all $a \in (\mathbb{Z}/N\mathbb{Z})^*$, we have $\varepsilon(a) = \prod \varepsilon_i(a \pmod{p_i^{e_i}})$. If $2 \mid N$, the steps are as follows:*

(1) Let g_i be the minimal generators of $(\mathbb{Z}/N\mathbb{Z})^*$, so ε is given by a list
$$[\varepsilon(g_0), \ldots, \varepsilon(g_n)].$$
(2) For $i = 2, \ldots, n$, let ε_i be the element of $D(p_i^{e_i}, R)$ defined by the singleton list $[\varepsilon(g_i)]$.
(3) Let ε_1 be the element of $D(2^{e_1}, R)$ defined by the list $[\varepsilon(g_0), \varepsilon(g_1)]$ of length 2. Output the ε_i and terminate.

If $2 \nmid N$, then omit step (3), and include all i in step (2).

The factorization of Algorithm 4.16 is unique since each ε_i is determined by the image of the canonical map $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ in $(\mathbb{Z}/N\mathbb{Z})^*$, which sends $a \pmod{p_i^{e_i}}$ to the element of $(\mathbb{Z}/N\mathbb{Z})^*$ that is $a \pmod{p_i^{e_i}}$ and $1 \pmod{p_j^{e_j}}$ for $j \neq i$.

Example 4.17. If $\varepsilon = [1, -1, \zeta^5] \in D(200, \mathbb{C})$, then $\varepsilon_1 = [1, -1] \in D(8, \mathbb{C})$ and $\varepsilon_2 = [\zeta^5] \in D(25, \mathbb{C})$.

Definition 4.18 (Conductor)**.** The *conductor* of a Dirichlet character $\varepsilon \in D(N,R)$ is the smallest positive divisor $c \mid N$ such that there is a character $\varepsilon' \in D(c, R)$ for which $\varepsilon(a) = \varepsilon'(a)$ for all $a \in \mathbb{Z}$ with $(a, N) = 1$. A Dirichlet character is *primitive* if its modulus equals its conductor. The character ε' associated to ε with modulus equal to the conductor of ε is called the *primitive character associated to* ε.

We will be interested in conductors later, when computing new subspaces of spaces of modular forms with character. Also certain formulas for special values of L functions are only valid for primitive characters.

Algorithm 4.19 (Conductor)**.** *This algorithm computes the conductor of a Dirichlet character $\varepsilon \in D(N, R)$.*

(1) [Factor Character] Using Algorithm 4.16, find characters ε_i whose product is ε.
(2) [Compute Orders] Using Algorithm 4.14, compute the orders r_i of each ε_i.
(3) [Conductors of Factors] For each i, either set $c_i \to 1$ if ε_i is the trivial character (i.e., of order 1) or set $c_i = p_i^{\mathrm{ord}_{p_i}(r_i)+1}$, where $\mathrm{ord}_p(n)$ is the largest power of p that divides n.

(4) [Adjust at 2?] If $p_1 = 2$ and $\varepsilon_1(5) \neq 1$, set $c_1 = 2c_1$.
(5) [Finished] Output $c = \prod c_i$ and terminate.

Proof. Let ε_i be the local factors of ε, as in step (1). We first show that the product of the conductors f_i of the ε_i is the conductor f of ε. Since ε_i factors through $(\mathbb{Z}/f_i\mathbb{Z})^*$, the product ε of the ε_i factors through $(\mathbb{Z}/\prod f_i\mathbb{Z})^*$, so the conductor of ε divides $\prod f_i$. Conversely, if $\mathrm{ord}_{p_i}(f) < \mathrm{ord}_{p_i}(f_i)$ for some i, then we could factor ε as a product of local (prime power) characters differently, which contradicts that this factorization is unique.

It remains to prove that if ε is a nontrivial character of modulus p^n, where p is a prime, and if r is the order of ε, then the conductor of ε is $p^{\mathrm{ord}_p(r)+1}$, except possibly if $8 \mid p^n$. Since the order and conductor of ε and of the associated primitive character ε' are the same, we may assume ε is primitive, i.e., that p^n is the conductor of ε; note that $n > 0$, since ε is nontrivial.

First suppose p is odd. Then the abelian group $D(p^n, R)$ splits as a direct sum $D(p, R) \oplus D(p^n, R)'$, where $D(p^n, R)'$ is the p-power torsion subgroup of $D(p^n, R)$. Also ε has order $u \cdot p^m$, where u, which is coprime to p, is the order of the image of ε in $D(p, R)$ and p^m is the order of the image in $D(p^n, R)'$. If $m = 0$, then the order of ε is coprime to p, so ε is in $D(p, R)$, which means that $n = 1$, so $n = m + 1$, as required. If $m > 0$, then $\zeta \in R$ must have order divisible by p, so R has characteristic not equal to p. The conductor of ε does not change if we adjoin roots of unity to R, so in light of Lemma 4.2 we may assume that $D(N, R) \approx (\mathbb{Z}/N\mathbb{Z})^*$. It follows that for each $n' \leq n$, the p-power subgroup $D(p^{n'}, R)'$ of $D(p^{n'}, R)$ is the $p^{n'-1}$-torsion subgroup of $D(p^n, R)'$. Thus $m = n - 1$, since $D(p^n, R)'$ is by assumption the smallest such group that contains the projection of ε. This proves the formula of step (3). We leave the argument when $p = 2$ as an exercise (see Exercise 4.5). □

Example 4.20. If $\varepsilon = [1, -1, \zeta^5] \in D(200, \mathbb{C})$, then as in Example 4.17, ε is the product of $\varepsilon_1 = [1, -1]$ and $\varepsilon_2 = [\zeta^5]$. Because $\varepsilon_1(5) = -1$, the conductor of ε_1 is 8. The order of ε_2 is 4 (since ζ is a 20th root of unity), so the conductor of ε_2 is 5. Thus the conductor of ε is $40 = 8 \cdot 5$.

4.5. The Kronecker Symbol

In this section all characters have values in $\mathbb{C}$.

Frequently quadratic characters are described in terms of the Kronecker symbol $\left(\frac{a}{n}\right)$, which we define for any integer a and positive integer n as follows. First, if $n = p$ is an odd prime, then for any integer a,

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } \gcd(a,p) \neq 1, \\ 1 & \text{if } a \text{ is a square mod } p, \\ -1 & \text{if } a \text{ is not a square mod } p. \end{cases}$$

If $p = 2$, then

$$\left(\frac{a}{2}\right) = \begin{cases} 0 & \text{if } a \text{ is even}, \\ 1 & \text{if } a \equiv \pm 1 \pmod 8, \\ -1 & \text{if } a \equiv \pm 3 \pmod 8. \end{cases}$$

More generally, if $n = \prod p_i^{e_i}$ with the p_i prime, then

$$\left(\frac{a}{n}\right) = \prod \left(\frac{a}{p_i}\right)^{e_i}.$$

Remark 4.21. One can also extend $\left(\frac{a}{n}\right)$ to $n < 0$, but we will not need this. The extension is to set $\left(\frac{a}{-1}\right) = -1$ and $\left(\frac{a}{1}\right) = 1$, for $a \neq 0$, and to extend multiplicatively (in the denominator). Note that the map $\left(\frac{\bullet}{-1}\right)$ is not a Dirichlet character (see Exercise 4.1).

Let M be the product of the primes p such that $\mathrm{ord}_p(n)$ is odd. If M is odd, let $N = M$; otherwise, let $N = 8M$.

Lemma 4.22. *The function*

$$\varepsilon(a) = \begin{cases} \left(\frac{a}{n}\right) & \text{if } \gcd(a,N) = 1, \\ 0 & \text{otherwise} \end{cases}$$

is a Dirichlet character of modulus N. The function

$$\varepsilon(a) = \begin{cases} \left(\frac{-1}{a}\right) & \text{if } a \text{ is odd}, \\ 0 & \text{if } a \text{ is even} \end{cases}$$

is a Dirichlet character of modulus N.

Proof. When restricted to $(\mathbb{Z}/N\mathbb{Z})^*$, each map $\left(\frac{\bullet}{p}\right)$, for p prime, is a homomorphism, so ε a product of homomorphisms. The second statement follows from the definition and the fact that -1 is a square modulo an odd prime p if and only if $p \equiv 1 \pmod 4$. □

This section is about going between representing quadratic characters as row matrices and via Kronecker symbols. This is valuable because the algorithms in [**Coh93**, §1.1.4] for computing Kronecker symbols run in time quadratic in the number of digits of the input. They do not require computing discrete logarithms; instead, they use, e.g., that $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$, when p is an odd prime.

Algorithm 4.23 (Kronecker Symbol as Dirichlet Character)**.** *Given $n > 0$, this algorithm computes a representation of the Kronecker symbol $\left(\frac{\bullet}{n}\right)$ as a Dirichlet character.*

(1) [Modulus] Compute N as in Lemma 4.22.
(2) [Minimal Generators] Compute minimal generators g_i of $(\mathbb{Z}/N\mathbb{Z})^*$ using Algorithm 4.4.
(3) [Images] Compute $\left(\frac{g_i}{N}\right)$ for each g_i using one of the algorithms of [**Coh93**, §1.1.4].

Example 4.24. We compute the Dirichlet character associated to $\left(\frac{\bullet}{200}\right)$. Using SAGE, we compute the $\left(\frac{g_i}{200}\right)$, for $i = 0, 1, 2$, where the g_i are as in Example 4.9:

```
sage: kronecker(151,200)
1
sage: kronecker(101,200)
-1
sage: kronecker(177,200)
1
```

Thus the corresponding character is defined by $[1, -1, 1]$.

Example 4.25. We compute the character associated to $\left(\frac{\bullet}{420}\right)$. We have $420 = 4 \cdot 3 \cdot 5 \cdot 7$, and minimal generators are

$$g_0 = 211, \quad g_1 = 1, \quad g_2 = 281, \quad g_3 = 337, \quad g_4 = 241.$$

We have $g_0 \equiv -1 \pmod{4}$, $g_2 \equiv 2 \pmod{3}$, $g_3 \equiv 2 \pmod{5}$ and $g_4 \equiv 3 \pmod{7}$. We find $\left(\frac{g_0}{420}\right) = \left(\frac{g_1}{420}\right) = 1$ and $\left(\frac{g_2}{420}\right) = \left(\frac{g_3}{420}\right) = \left(\frac{g_4}{420}\right) = -1$. The corresponding character is $[1, 1, -1, -1, -1]$.

Using the following algorithm, we can go in the other direction, i.e., write any quadratic Dirichlet character as a Kronecker symbol.

Algorithm 4.26 (Dirichlet Character as Kronecker Symbol)**.** *Given ε of order 2 with modulus N, this algorithm writes ε as a Kronecker symbol.*

(1) [Conductor] Use Algorithm 4.19 to compute the conductor f of ε.
(2) [Odd] If f is odd, output $\left(\frac{\bullet}{f}\right)$.
(3) [Even] If $\varepsilon(-1) = 1$, output $\left(\frac{\bullet}{f}\right)$; if $\varepsilon(-1) = -1$, output $\left(\frac{\bullet}{f}\right) \cdot \left(\frac{-1}{\bullet}\right)$.

Proof. Since f is the conductor of a quadratic Dirichlet character, it is a square-free product g of odd primes times either 4 or 8, so the group $(\mathbb{Z}/f\mathbb{Z})^*$ does not inject into $(\mathbb{Z}/g\mathbb{Z})^*$ for any proper divisor g of f (see this by reducing to the prime power case). Since g is odd and square-free, the character $\left(\frac{\bullet}{g}\right)$ has conductor g. For each odd prime p, by step (3) of Algorithm 4.19 the factor at p of both ε and $\left(\frac{\bullet}{g}\right)$ is a quadratic character with modulus p. By Exercise 4.2 and Lemma 4.2 the group $D(p, \mathbb{C})$ is cyclic, so it has a unique element of order 2, so the factors of ε and $\left(\frac{\bullet}{g}\right)$ at p are equal.

The quadratic characters with conductor a power of 2 are $[-1]$, $[1, -1]$, and $[-1, -1]$. The character $[1, -1]$ is $\left(\frac{\bullet}{2}\right)$ and the character $[-1]$ is $\left(\frac{-1}{\bullet}\right)$. □

Example 4.27. Consider $\varepsilon = [-1, -1, -1, -1, -1]$ with modulus $840 = 8 \cdot 3 \cdot 5 \cdot 7$. It has conductor 840, and $\varepsilon(-1) = -1$, so for all a with $\gcd(a, 840) = 1$, we have $\varepsilon(a) = \left(\frac{a}{840}\right) \cdot \left(\frac{-1}{a}\right)$.

4.6. Restriction, Extension, and Galois Orbits

The following two algorithms restrict and extend characters to a compatible modulus. Using them, it is easy to define multiplication of two characters $\varepsilon \in D(N, R)$ and $\varepsilon' \in D(N', R')$, as long as R and R' are subrings of a common ring. To carry out the multiplication, extend both characters to a common base ring, and then extend them to characters modulo $\mathrm{lcm}(N, N')$ and multiply.

Algorithm 4.28 (Restriction of Character)**.** *Given a Dirichlet character $\varepsilon \in D(N, R)$ and a divisor N' of N that is a multiple of the conductor of ε, this algorithm finds a characters $\varepsilon' \in D(N', R)$, such that $\varepsilon'(a) = \varepsilon(a)$, for all $a \in \mathbb{Z}$ with $(a, N) = 1$.*

(1) [Conductor] Compute the conductor of ε using Algorithm 4.19, and verify that N' is divisible by the conductor and divides N.
(2) [Minimal Generators] Compute minimal generators g_i for $(\mathbb{Z}/N'\mathbb{Z})^*$.
(3) [Values of Restriction] For each i, compute $\varepsilon'(g_i)$ as follows. Find a multiple aN' of N' such that $(g_i + aN',\, N) = 1$; then $\varepsilon'(g_i) = \varepsilon(g_i + aN')$.
(4) [Output Character] Output the Dirichlet character of modulus N' defined by $[\varepsilon'(g_0), \ldots, \varepsilon'(g_n)]$.

Proof. The only part that is not clear is that in step (3) there is an a such that $(g_i + aN', N) = 1$. If we write $N = N_1 \cdot N_2$, with $(N_1, N_2) = 1$ and N_1 divisible by all primes that divide N', then $(g_i, N_1) = 1$ since $(g_i, N') = 1$. By the Chinese Remainder Theorem, there is an $x \in \mathbb{Z}$ such that $x \equiv g_i$

(mod N_1) and $x \equiv 1 \pmod{N_2}$. Then $x = g_i + bN_1 = g_i + (bN_1/N') \cdot N'$ and $(x, N) = 1$, which completes the proof. □

Algorithm 4.29 (Extension of Character)**.** *Given a Dirichlet character $\varepsilon \in D(N, R)$ and a multiple N' of N, this algorithm finds a character $\varepsilon' \in D(N', R)$, such that $\varepsilon'(a) = \varepsilon(a)$, for all $a \in \mathbb{Z}$ with $(a, N') = 1$.*

(1) [Minimal Generators] Compute minimal generators g_i for $(\mathbb{Z}/N'\mathbb{Z})^*$.
(2) [Evaluate] Compute $\varepsilon(g_i)$ for each i. Since $(g_i, N') = 1$, we also have $(g_i, N) = 1$.
(3) [Output Character] Output the character $[\varepsilon(g_0), \ldots, \varepsilon(g_n)]$.

Let F be the prime subfield of R, and assume that $R \subset \overline{F}$, where $\overline{F}$ is a separable closure of F. If $\sigma \in \mathrm{Gal}(\overline{F}/F)$ and $\varepsilon \in D(N, R)$, let $(\sigma\varepsilon)(n) = \sigma(\varepsilon(n))$; this defines an action of $\mathrm{Gal}(\overline{F}/F)$ on $D(N, R)$. Our next algorithm computes the orbits for the action of $\mathrm{Gal}(\overline{F}/F)$ on $D(N, R)$. This algorithm can provide huge savings for modular forms computations because the spaces $M_k(N, \varepsilon)$ and $M_k(N, \varepsilon')$ are canonically isomorphic if ε and ε' are conjugate.

Algorithm 4.30 (Galois Orbit)**.** *Given a Dirichlet character $\varepsilon \in D(N, R)$, this algorithm computes the orbit of ε under the action of $G = \mathrm{Gal}(\overline{F}/F)$, where F is the prime subfield of $\mathrm{Frac}(R)$, so $F = \mathbb{F}_p$ or $\mathbb{Q}$.*

(1) [Order of ζ] Let n be the order of the chosen root $\zeta \in R$.
(2) [Nontrivial Automorphisms] If $\mathrm{char}(R) = 0$, let

$$A = \{a : 2 \le a < n \text{ and } (a, n) = 1\}.$$

If $\mathrm{char}(R) = p > 0$, compute the multiplicative order r of $p \pmod{n}$, and let

$$A = \{p^m : 1 \le m < r\}.$$

(3) [Compute Orbit] Compute and output the *set* of unique elements ε^a for each $a \in A$ (there could be repeats, so we output unique elements only).

Proof. We prove that the nontrivial automorphisms of $\langle\zeta\rangle$ in characteristic p are as in step (2). It is well known that every automorphism in characteristic p on $\zeta \in \overline{\mathbb{F}}_p$ is of the form $x \mapsto x^{p^s}$, for some s. The images of ζ under such automorphisms are

$$\zeta, \zeta^p, \zeta^{p^2}, \ldots.$$

Suppose $r > 0$ is minimal such that $\zeta = \zeta^{p^r}$. Then the orbit of ζ is $\zeta, \ldots, \zeta^{p^{r-1}}$. Also $p^r \equiv 1 \pmod{n}$, where n is the multiplicative order of ζ, so r is the multiplicative order of p modulo n, which completes the proof. □

Example 4.31. The Galois orbits of characters in $D(20, \mathbb{C}^*)$ are as follows:

$$\begin{aligned} G_0 &= \{[1, 1, 1]\}, \\ G_1 &= \{[-1, 1, 1]\}, \\ G_2 &= \{[1, 1, \zeta_4],\ [1, 1, -\zeta_4]\} \\ G_3 &= \{[-1, 1, \zeta_4],\ [-1, 1, -\zeta_4]\} \\ G_4 &= \{[1, 1, -1]\}, \\ G_5 &= \{[-1, 1, -1]\}. \end{aligned}$$

The conductors of the characters in orbit G_0 are 1, in orbit G_1 they are 4, in orbit G_2 they are 5, in G_3 they are 20, in G_4 the conductor is 5, and in G_5 the conductor is 20. (You should verify this.)

SAGE computes Galois orbits as follows:

```
sage: G = DirichletGroup(20)
sage: G.galois_orbits()
[
[[1, 1]],
[[1, zeta4], [1, -zeta4]],
[[1, -1]],
[[-1, 1]],
[[-1, zeta4], [-1, -zeta4]],
[[-1, -1]]
]
```

4.7. Alternative Representations of Characters

Let N be a positive integer and R an integral domain, with fixed root of unity ζ of order n, and let $D(N, R) = D(N, R, \zeta)$. As in the rest of this chapter, write $N = \prod p_i^{e_i}$, and let $C_i = \langle g_i \rangle$ be the corresponding cyclic factors of $(\mathbb{Z}/N\mathbb{Z})^*$. In this section we discuss other ways to represent elements $\varepsilon \in D(N, R)$. Each representation has advantages and disadvantages, and no single representation is best. It is easy to convert between them, and some algorithms are much easier using one representation than when using another. In this section we present two other representations, each having advantages and disadvantages. There is no reason to restrict to only one representation; for example, SAGE internally uses both.

We could represent ε by giving a list $[b_0, \ldots, b_r]$, where each $b_i \in \mathbb{Z}/n\mathbb{Z}$ and $\varepsilon(g_i) = \zeta^{b_i}$. Then arithmetic in $D(N, R)$ is arithmetic in $(\mathbb{Z}/n\mathbb{Z})^{r+1}$, which is very efficient. A drawback to this approach (in practice) is that it is easy to accidentally consider sequences that do not actually correspond to elements of $D(N, R)$. Also the choice of ζ is less clear, which can cause confusion. Finally, the orders of the local factors is more opaque, e.g., compare $[-1, \zeta_{40}]$ with $[20, 1]$. Overall this representation is not too bad and is more like representing a linear transformation by a matrix. It has the *advantage* over the representation discussed earlier in this chapter that arithmetic in $D(N, R)$ is very efficient and does not require any operations in the ring R.

Another way to represent ε would be to give a list $[b_0, \ldots, b_r]$ of integers, but this time with $b_i \in \mathbb{Z}/\gcd(s_i, n)\mathbb{Z}$, where s_i is the order of g_i. Then

$$\varepsilon(g_i) = \zeta^{b_i \cdot n/(\gcd(s_i,n))},$$

which is already pretty complicated. With this representation we set up an identification

$$D(N, R) \cong \bigoplus_i \mathbb{Z}/\gcd(s_i, n)\mathbb{Z},$$

and arithmetic is efficient. This approach is seductive because every sequence of integers determines a character, and the sizes of the integers in the sequence nicely indicate the local orders of the character. However, giving analogues of many of the algorithms discussed in this chapter that operate on characters represented this way is tricky. For example, the representation depends very much on the order of ζ, so it is difficult to correctly compute natural maps $D(N, R) \to D(N, S)$, for $R \subset S$ rings.

4.8. Dirichlet Characters in SAGE

To create a Dirichlet character in SAGE, first create the group $D(N, R)$ of Dirichlet characters then construct elements of that group. First we make $D(11, \mathbb{Q})$:

```
sage: G = DirichletGroup(11, QQ); G
Group of Dirichlet characters of modulus 11 over
Rational Field
```

A Dirichlet character prints as a matrix that gives the values of the character on canonical generators of $(\mathbb{Z}/N\mathbb{Z})^*$ (as discussed below).

```
sage: list(G)
[[1], [-1]]
sage: eps = G.0      # 0th generator for Dirichlet group
sage: eps
[-1]
```

The character ε takes the value -1 on the unit generator.

```
sage: G.unit_gens()
[2]
sage: eps(2)
-1
sage: eps(3)
1
```

It is 0 on any integer not coprime to 11:

```
sage: [eps(11*n) for n in range(10)]
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

We can also create groups of Dirichlet characters taking values in other rings or fields. For example, we create the cyclotomic field $\mathbb{Q}(\zeta_4)$.

```
sage: R = CyclotomicField(4)
sage: CyclotomicField(4)
Cyclotomic Field of order 4 and degree 2
```

Then we define $G = D(15, \mathbb{Q}(\zeta_4))$.

```
sage: G = DirichletGroup(15, R)
sage: G
Group of Dirichlet characters of modulus 15 over
Cyclotomic Field of order 4 and degree 2
```

Next we list each of its elements.

```
sage: list(G)
[[1, 1], [-1, 1], [1, zeta4], [-1, zeta4], [1, -1],
[-1, -1], [1, -zeta4], [-1, -zeta4]]
```

Now we evaluate the second generator of G on various integers:

```
sage: e = G.1
sage: e(4)
-1
sage: e(-1)
-1
sage: e(5)
0
```

Finally we list all the values of e.

```
sage: [e(n) for n in range(15)]
[0, 1, zeta4, 0, -1, 0, 0, zeta4, -zeta4,
    0, 0, 1, 0, -zeta4, -1]
```

We can also compute with groups of Dirichlet characters with values in a finite field.

```
sage: G = DirichletGroup(15, GF(5)); G
Group of Dirichlet characters of modulus 15
      over Finite Field of size 5
```

We list all the elements of G, again represented by lists that give the images of each unit generator, as an element of $\mathbb{F}_5$.

```
sage: list(G)
  [[1, 1], [4, 1], [1, 2], [4, 2], [1, 4], [4, 4],
   [1, 3], [4, 3]]
```

We evaluate the second generator of G on several integers.

```
sage: e = G.1
sage: e(-1)
4
sage: e(2)
2
sage: e(5)
0
sage: print [e(n) for n in range(15)]
[0, 1, 2, 0, 4, 0, 0, 2, 3, 0, 0, 1, 0, 3, 4]
```

4.9. Exercises

4.1 Let $f : \mathbb{Z} \to \mathbb{C}$ be the map given by

$$f(a) = \begin{cases} 0 & \text{if } a = 0, \\ -1 & \text{if } a < 0, \\ 1 & \text{if } a > 0. \end{cases}$$

Prove that f is not a Dirichlet character of any modulus N.

4.2 This exercise is about the structure of the units of $\mathbb{Z}/p^n\mathbb{Z}$.

(a) If p is odd and n is a positive integer, prove that $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic.

(b) For $n \geq 3$, prove that $(\mathbb{Z}/2^n\mathbb{Z})^*$ is a direct sum of the cyclic subgroups $\langle -1 \rangle$ and $\langle 5 \rangle$, of orders 2 and 2^{n-2}, respectively.

4.3 Prove that Algorithm 4.4 works, i.e., that if $g \in (\mathbb{Z}/p^r\mathbb{Z})^*$ and $g^{n/p_i} \neq 1$ for all $p_i \mid n = \varphi(p^r)$, then g is a generator of $(\mathbb{Z}/p^r\mathbb{Z})^*$.

4.4 (a) Let p be an odd prime and $n \geq 2$ an integer, and prove that

$$\big((1 + p^{n-1}\mathbb{Z}/p^n\mathbb{Z}),\, \times\big) \cong (\mathbb{Z}/p\mathbb{Z},\, +).$$

(b) Use the first part to show that solving the discrete log problem in $(\mathbb{Z}/p^n\mathbb{Z})^*$ is "not much harder" than solving the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$.

4.5 Suppose ε is a nontrivial Dirichlet character of modulus 2^n of order r over the complex numbers $\mathbb{C}$. Prove that the conductor of ε is

$$c = \begin{cases} 2^{\mathrm{ord}_2(r)+1} & \text{if } \varepsilon(5) = 1, \\ 2^{\mathrm{ord}_2(r)+2} & \text{if } \varepsilon(5) \neq 1. \end{cases}$$

4.6 (a) Find an irreducible quadratic polynomial f over $\mathbb{F}_5$.

(b) Then $\mathbb{F}_{25} = \mathbb{F}_5[x]/(f)$. Find an element with multiplicative order 4 in $\mathbb{F}_{25}$.

(c) Make a list of all Dirichlet characters in $D(25, \mathbb{F}_{25}, \zeta)$.

(d) Divide these characters into orbits for the action of $\mathrm{Gal}(\overline{\mathbb{F}}_5/\mathbb{F}_5)$.

Chapter 5

Eisenstein Series and Bernoulli Numbers

We introduce generalized Bernoulli numbers attached to Dirichlet characters and give an algorithm to enumerate the Eisenstein series in $M_k(N, \varepsilon)$.

5.1. The Eisenstein Subspace

Let $M_k(\Gamma_1(N))$ be the space of modular forms of weight k for $\Gamma_1(N)$, and let $\mathbb{T}$ be the *Hecke algebra* acting on $M_k(\Gamma_1(N))$, which is the subring of $\mathrm{End}(M_k(\Gamma_1(N)))$ generated by all Hecke operators. Then there is a $\mathbb{T}$-module decomposition

$$M_k(\Gamma_1(N)) = E_k(\Gamma_1(N)) \oplus S_k(\Gamma_1(N)),$$

where $S_k(\Gamma_1(N))$ is the subspace of modular forms that vanish at all cusps and $E_k(\Gamma_1(N))$ is the *Eisenstein subspace*, which is uniquely determined by this decomposition. The above decomposition induces a decomposition of $M_k(\Gamma_0(N))$ and of $M_k(N, \varepsilon)$, for any Dirichlet character ε of modulus N.

5.2. Generalized Bernoulli Numbers

Suppose ε is a Dirichlet character of modulus N over $\mathbb{C}$. Leopoldt [**Leo58**] defined generalized Bernoulli numbers attached to ε.

Definition 5.1 (Generalized Bernoulli Number)**.** We define the *generalized Bernoulli numbers $B_{k,\varepsilon}$* attached to ε by the following identity of infinite series:

$$\sum_{a=1}^{N} \frac{\varepsilon(a) \cdot x \cdot e^{ax}}{e^{Nx} - 1} = \sum_{k=0}^{\infty} B_{k,\varepsilon} \cdot \frac{x^k}{k!}.$$

If ε is the trivial character of modulus 1 and B_k are as in Section 2.1, then $B_{k,\varepsilon} = B_k$, except when $k = 1$, in which case $B_{1,\varepsilon} = -B_1 = 1/2$ (see Exercise 5.2).

5.2.1. Algebraically Computing Generalized Bernoulli Numbers.

Let $\mathbb{Q}(\varepsilon)$ denote the field generated by the image of the character ε; thus $\mathbb{Q}(\varepsilon)$ is the cyclotomic extension $\mathbb{Q}(\zeta_n)$, where n is the order of ε.

Algorithm 5.2 (Generalized Bernoulli Numbers)**.** *Given an integer $k \geq 0$ and any Dirichlet character ε with modulus N, this algorithm computes the generalized Bernoulli numbers $B_{j,\varepsilon}$, for $j \leq k$.*

(1) Compute $g = x/(e^{Nx}-1) \in \mathbb{Q}[[x]]$ to precision $O(x^{k+1})$ by computing $e^{Nx} - 1 = \sum_{n\geq 1} N^n x^n/n!$ to precision $O(x^{k+2})$ and computing the inverse $1/(e^{Nx} - 1)$, then multiplying by x.

(2) For each $a = 1, \ldots, N$, compute $f_a = g \cdot e^{ax} \in \mathbb{Q}[[x]]$, to precision $O(x^{k+1})$. This requires computing $e^{ax} = \sum_{n\geq 0} a^n x^n/n!$ to precision $O(x^{k+1})$. (Omit computation of e^{Nx} if $N > 1$ since then $\varepsilon(N) = 0$.)

(3) Then for $j \leq k$, we have

$$B_{j,\varepsilon} = j! \cdot \sum_{a=1}^{N} \varepsilon(a) \cdot c_j(f_a),$$

where $c_j(f_a)$ is the coefficient of x^j in f_a.

Note that in steps (1) and (2) we compute the power series doing arithmetic only in $\mathbb{Q}[[x]]$, not in $\mathbb{Q}(\varepsilon)[[x]]$, which could be much less efficient if ε has large order. In step (1) if k is huge, we could compute the inverse $1/(e^{Nx} - 1)$ using asymptotically fast arithmetic and Newton iteration.

Example 5.3. The nontrivial character ε with modulus 4 has order 2 and takes values in $\mathbb{Q}$. The Bernoulli numbers $B_{k,\varepsilon}$ for k even are all 0 and for

k odd they are

$$\begin{aligned}
B_{1,\varepsilon} &= -1/2,\\
B_{3,\varepsilon} &= 3/2,\\
B_{5,\varepsilon} &= -25/2,\\
B_{7,\varepsilon} &= 427/2,\\
B_{9,\varepsilon} &= -12465/2,\\
B_{11,\varepsilon} &= 555731/2,\\
B_{13,\varepsilon} &= -35135945/2,\\
B_{15,\varepsilon} &= 2990414715/2,\\
B_{17,\varepsilon} &= -329655706465/2,\\
B_{19,\varepsilon} &= 45692713833379/2.
\end{aligned}$$

Example 5.4. The generalized Bernoulli numbers need not be in $\mathbb{Q}$. Suppose ε is the mod 5 character such that $\varepsilon(2) = i = \sqrt{-1}$. Then $B_{k,\varepsilon} = 0$ for k even and

$$\begin{aligned}
B_{1,\varepsilon} &= \frac{-i-3}{5},\\
B_{3,\varepsilon} &= \frac{6i+12}{5},\\
B_{5,\varepsilon} &= \frac{-86i-148}{5},\\
B_{7,\varepsilon} &= \frac{2366i+3892}{5},\\
B_{9,\varepsilon} &= \frac{-108846i-176868}{5},\\
B_{11,\varepsilon} &= \frac{7599526i+12309572}{5},\\
B_{13,\varepsilon} &= \frac{-751182406i-1215768788}{5},\\
B_{15,\varepsilon} &= \frac{99909993486i+161668772052}{5},\\
B_{17,\varepsilon} &= \frac{-17209733596766i-27846408467908}{5}.
\end{aligned}$$

Example 5.5. We use SAGE to compute some of the above generalized Bernoulli numbers. First we define the character and verify that $\varepsilon(2) = i$ (note that in SAGE `zeta4` is $\sqrt{-1}$).

```
sage: G = DirichletGroup(5)
sage: e = G.0
sage: e(2)
zeta4
```

We compute the Bernoulli number $B_{1,\varepsilon}$.

```
sage: e.bernoulli(1)
-1/5*zeta4 - 3/5
```

We compute $B_{9,\varepsilon}$.

```
sage: e.bernoulli(9)
-108846/5*zeta4 - 176868/5
```

Proposition 5.6. *If $\varepsilon(-1) \neq (-1)^k$ and $k \geq 2$, then $B_{k,\varepsilon} = 0$.*

Proof. See Exercise 5.3. □

5.2.2. Computing Generalized Bernoulli Numbers Analytically.

This section, which was written jointly with Kevin McGown, is about a way to compute generalized Bernoulli numbers, which is similar to the algorithm in Section 2.7.

Let χ be a primitive Dirichlet character modulo its conductor f. Note from the definition of Bernoulli numbers that if $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, then

$$\sigma(B_{n,\chi}) = B_{n,\sigma(\chi)}. \tag{5.2.1}$$

For any character χ, we define the Gauss sum $\tau(\chi)$ as

$$\tau(\chi) = \sum_{r=1}^{f-1} \chi(r)\,\zeta^r\,,$$

where $\zeta = \exp(2\pi i/f)$ is the principal fth root of unity. The Dirichlet L-function for χ for $\mathrm{Re}(s) > 1$ is

$$L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)\,n^{-s}\,.$$

In the right half plane $\{s \in \mathbb{C} \mid \mathrm{Re}(s) > 1\}$ this function is analytic, and because χ is multiplicative, we have the Euler product representation

$$L(s,\chi) = \prod_{p \text{ prime}} \left(1 - \chi(p)p^{-s}\right)^{-1}\,. \tag{5.2.2}$$

We note (but will not use) that through analytic continuation $L(s,\chi)$ can be extended to a meromorphic function on the entire complex plane.

If χ is a nonprincipal primitive Dirichlet character of conductor f such that $\chi(-1) = (-1)^n$, then (see, e.g., [**Wan82**])

$$L(n, \chi) = (-1)^{n-1} \frac{\tau(\chi)}{2} \left(\frac{2\pi i}{f}\right)^n \frac{B_{n,\overline{\chi}}}{n!}\,.$$

Solving for the Bernoulli number yields

$$B_{n,\chi} = (-1)^{n-1} \frac{2n!}{\tau(\overline{\chi})} \left(\frac{f}{2\pi i}\right)^n L(n, \overline{\chi})\,.$$

This allows us to give decimal approximations for $B_{n,\chi}$. It remains to compute $B_{n,\chi}$ exactly (i.e., as an algebraic integer). To simplify the above expression, we define

$$K_{n,\chi} = (-1)^{n-1}\, 2n! \left(\frac{f}{2i}\right)^n$$

and write

$$B_{n,\chi} = \frac{K_{n,\chi}}{\pi^n\, \tau(\overline{\chi})}\, L(n, \overline{\chi})\,. \tag{5.2.3}$$

Note that we can compute $K_{n,\chi}$ exactly in the field $\mathbb{Q}(i)$.

The following result identifies the denominator of $B_{n,\chi}$.

Theorem 5.7. *Let n and χ be as above, and define an integer d as follows:*

$$d = \begin{cases} 1 & \textit{if } f \textit{ is divisible by two distinct primes,} \\ 2 & \textit{if } f = 4, \\ 1 & \textit{if } f = 2^\mu,\ \mu > 2, \\ np & \textit{if } f = p,\ p > 2, \\ (1 - \chi(1+p)) & \textit{if } f = p^\mu,\ p > 2,\ \mu > 1. \end{cases}$$

Then $dn^{-1} B_{n,\chi}$ is integral.

Proof. See [**Car59a**] for the proof and [**Car59b**] for further details. □

To compute the algebraic integer $dn^{-1}B_{n,\chi}$, and we compute $L(n, \overline{\chi})$ to very high precision using the Euler product (5.2.2) and the formula (5.2.3). We carry out the same computation for each of the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ conjugates of χ, which by (5.2.1) yields the conjugates of $dn^{-1}B_{n,\chi}$. We can then write down the characteristic polynomial of $dn^{-1}B_{n,\chi}$ to very high precision and recognize the coefficients as rational integers. Finally, we determine which of the roots of the characteristic polynomial is $dn^{-1}B_{n,\chi}$ by approximating them all numerically to high precision and seeing which is closest to our numerical approximation to $dn^{-1}B_{n,\chi}$. The details are similar to what is explained in Section 2.7.

5.3. Explicit Basis for the Eisenstein Subspace

Suppose χ and ψ are primitive Dirichlet characters with conductors L and R, respectively. Let

$$E_{k,\chi,\psi}(q) = c_0 + \sum_{m\geq 1}\left(\sum_{n\mid m}\psi(n)\cdot\chi(m/n)\cdot n^{k-1}\right)q^m \in \mathbb{Q}(\chi,\psi)[[q]], \tag{5.3.1}$$

where

$$c_0 = \begin{cases} 0 & \text{if } L > 1,\\ -\dfrac{B_{k,\psi}}{2k} & \text{if } L = 1.\end{cases}$$

Note that when $\chi = \psi = 1$ and $k \geq 4$, then $E_{k,\chi,\psi} = E_k$, where E_k is from Chapter 1.

Miyake proves statements that imply the following in [**Miy89**, Ch. 7].

Theorem 5.8. *Suppose t is a positive integer and χ, ψ are as above and that k is a positive integer such that $\chi(-1)\psi(-1) = (-1)^k$. Except when $k = 2$ and $\chi = \psi = 1$, the power series $E_{k,\chi,\psi}(q^t)$ defines an element of $M_k(RLt, \chi\psi)$. If $\chi = \psi = 1$, $k = 2$, $t > 1$, and $E_2(q) = E_{k,\chi,\psi}(q)$, then $E_2(q) - tE_2(q^t)$ is a modular form in $M_2(\Gamma_0(t))$.*

Theorem 5.9. *The Eisenstein series in $M_k(N,\varepsilon)$ coming from Theorem 5.8 with $RLt \mid N$ and $\chi\psi = \varepsilon$ form a basis for the Eisenstein subspace $E_k(N,\varepsilon)$.*

Theorem 5.10. *The Eisenstein series $E_{k,\chi,\psi}(q) \in M_k(RL)$ defined above are eigenforms (i.e., eigenvectors for all Hecke operators T_n). Also $E_2(q) - tE_2(q^t)$, for $t > 1$, is an eigenform.*

Since $E_{k,\chi,\psi}(q)$ is normalized so the coefficient of q is 1, the eigenvalue of T_m is the coefficient

$$\sum_{n\mid m}\psi(n)\cdot\chi(m/n)\cdot n^{k-1}$$

of q^m (see Proposition 9.10). Also for $f = E_2(q) - tE_2(q^t)$ with $t > 1$ prime, the coefficient of q is 1, $T_m(f) = \sigma_1(m)\cdot f$ for $(m,t) = 1$, and $T_t(f) = ((t+1) - t)f = f$.

Algorithm 5.11 (Enumerating Eisenstein Series)**.** *Given a weight k and a Dirichlet character ε of modulus N, this algorithm computes a basis for the Eisenstein subspace $E_k(N,\varepsilon)$ of $M_k(N,\varepsilon)$ to precision $O(q^r)$.*

(1) [Weight 2 Trivial Character?] If $k = 2$ and $\varepsilon = 1$, output the Eisenstein series $E_2(q) - tE_2(q^t)$, for each divisor $t \mid N$ with $t \neq 1$, and then terminate.

(2) [Empty Space?] If $\varepsilon(-1) \neq (-1)^k$, output the empty list.
(3) [Compute Dirichlet Group] Let $G = D(N, \mathbb{Q}(\zeta_n))$ be the group of Dirichlet characters with values in $\mathbb{Q}(\zeta_n)$, where n is the exponent of $(\mathbb{Z}/N\mathbb{Z})^*$.
(4) [Compute Conductors] Compute the conductor of every element of G using Algorithm 4.19.
(5) [List Characters χ] Form a list V of all Dirichlet characters $\chi \in G$ such that $\mathrm{cond}(\chi) \cdot \mathrm{cond}(\chi/\varepsilon)$ divides N.
(6) [Compute Eisenstein Series] For each character χ in V, let $\psi = \chi/\varepsilon$ and compute $E_{k,\chi,\psi}(q^t) \pmod{q^r}$ for each divisor t of $N/(\mathrm{cond}(\chi) \cdot \mathrm{cond}(\psi))$. Here we compute $E_{k,\chi,\psi}(q^t) \pmod{q^r}$ using (5.3.1) and Algorithm 5.2.

Remark 5.12. Algorithm 5.11 is what is currently used in SAGE. It might be better to first reduce to the prime power case by writing all characters as a product of local characters and combine steps (4) and (5) into a single step that involves orders. However, this might make things more obscure.

Example 5.13. The following is a basis of Eisenstein series for $E_2(\Gamma_1(13))$.

$$\begin{aligned}
f_1 &= \frac{1}{2} + q + 3q^2 + 4q^3 + \cdots\,,\\
f_2 &= -\frac{7}{13}\zeta_{12}^2 - \frac{11}{13} + q + \left(2\zeta_{12}^2 + 1\right) q^2 + \left(-3\zeta_{12}^2 + 1\right) q^3 + \cdots\,,\\
f_3 &= q + \left(\zeta_{12}^2 + 2\right) q^2 + \left(-\zeta_{12}^2 + 3\right) q^3 + \cdots\,,\\
f_4 &= -\zeta_{12}^2 + q + \left(2\zeta_{12}^2 - 1\right) q^2 + \left(3\zeta_{12}^2 - 2\right) q^3 + \cdots\,,\\
f_5 &= q + \left(\zeta_{12}^2 + 1\right) q^2 + \left(\zeta_{12}^2 + 2\right) q^3 + \cdots\,,\\
f_6 &= -1 + q + -q^2 + 4q^3 + \cdots\,,\\
f_7 &= q + q^2 + 4q^3 + \cdots\,,\\
f_8 &= \zeta_{12}^2 - 1 + q + \left(-2\zeta_{12}^2 + 1\right) q^2 + \left(-3\zeta_{12}^2 + 1\right) q^3 + \cdots\,,\\
f_9 &= q + \left(-\zeta_{12}^2 + 2\right) q^2 + \left(-\zeta_{12}^2 + 3\right) q^3 + \cdots\,,\\
f_{10} &= \frac{7}{13}\zeta_{12}^2 - \frac{18}{13} + q + \left(-2\zeta_{12}^2 + 3\right) q^2 + \left(3\zeta_{12}^2 - 2\right) q^3 + \cdots\,,\\
f_{11} &= q + \left(-\zeta_{12}^2 + 3\right) q^2 + \left(\zeta_{12}^2 + 2\right) q^3 + \cdots\,.
\end{aligned}$$

We computed it as follows:

```
sage: E = EisensteinForms(Gamma1(13),2)
sage: E.eisenstein_series()
```

We can also compute the parameters χ, ψ, t that define each series:

```
sage: e = E.eisenstein_series()
sage: for e in E.eisenstein_series():
...       print e.parameters()
...
([1], [1], 13)
([1], [zeta6], 1)
([zeta6], [1], 1)
([1], [zeta6 - 1], 1)
([zeta6 - 1], [1], 1)
([1], [-1], 1)
([-1], [1], 1)
([1], [-zeta6], 1)
([-zeta6], [1], 1)
([1], [-zeta6 + 1], 1)
([-zeta6 + 1], [1], 1)
```

5.4. Exercises

5.1 Suppose A and B are diagonalizable linear transformations of a finite-dimensional vector space V over an algebraically closed field K and that $AB = BA$. Prove there is a basis for V so that the matrices of A and B with respect to that basis are both simultaneously diagonal.

5.2 If ε is the trivial character of modulus 1 and B_k are as in Section 2.1, then $B_{k,\varepsilon} = B_k$, except when $k = 1$, in which case $B_{1,\varepsilon} = -B_1 = 1/2$.

5.3 Prove that for $k \geq 2$ if $\varepsilon(-1) \neq (-1)^k$, then $B_{k,\varepsilon} = 0$.

5.4 Show that the dimension of the Eisenstein subspace $E_3(\Gamma_1(13))$ is 12 by finding a basis of series $E_{k,\chi,\psi}$. You do not have to write down the q-expansions of the series, but you do have to figure out which χ, ψ to use.

Chapter 6

Dimension Formulas

When computing with spaces of modular forms, it is helpful to have easy-to-compute formulas for dimensions of these spaces. Such formulas provide a check on the output of the algorithms from Chapter 8 that compute explicit bases for spaces of modular forms. We can also use dimension formulas to improve the efficiency of some of the algorithms in Chapter 8, since we can use them to determine the ranks of certain matrices without having to explicitly compute those matrices. Dimension formulas can also be used in generating bases of q-expansions; if we know the dimension of $M_k(N, \varepsilon)$ and if we have a process for computing q-expansions of elements of $M_k(N, \varepsilon)$, e.g., multiplying together q-expansions of certain forms of smaller weight, then we can tell when we are done generating $M_k(N, \varepsilon)$.

This chapter contains formulas for dimensions of spaces of modular forms, along with some remarks about how to evaluate these formulas. In some cases we give dimension formulas for spaces that we will define in later chapters. We also give many examples, some of which were computed using the modular symbols algorithms from Chapter 8.

Many of the dimension formulas and algorithms we give below grew out of Shimura's book [**Shi94**] and a program that Bruce Kaskel wrote (around 1996) in PARI, which Kevin Buzzard extended. That program codified dimension formulas that Buzzard and Kaskel found or extracted from the literature (mainly [**Shi94**, §2.6]). The algorithms for dimensions of spaces with nontrivial character are from [**CO77**], with some refinements suggested by Kevin Buzzard.

For the rest of this chapter, N denotes a positive integer and $k \geq 2$ is an integer. We will give *no simple formulas* for dimensions of spaces of weight 1 modular forms; in fact, it might not be possible to give such formulas since the methods used to derive the formulas below do not apply in the case $k = 1$. If $k = 0$, the only modular forms are the constants, and for $k < 0$ the dimension of $M_k(N, \varepsilon)$ is 0.

For a nonzero integer N and a prime p, let $v_p(N)$ be the largest integer e such that $p^e \mid N$. In the formulas in this chapter, p always denotes a prime number. Let $M_k(N, \varepsilon)$ be the space of modular forms of level N weight k and character ε, and let $S_k(N, \varepsilon)$ and $E_k(N, \varepsilon)$ be the cuspidal and Eisenstein subspaces, respectively.

The dimension formulas below for $S_k(\Gamma_0(N))$, $S_k(\Gamma_1(N))$, $E_k(\Gamma_0(N))$ and $E_k(\Gamma_1(N))$ can be found in [**DS05**, Ch. 3], [**Shi94**, §2.6][1] and [**Miy89**, §2.5]. They are derived using the Riemann-Roch Theorem applied to the covering $X_0(N) \to X_0(1)$ or $X_1(N) \to X_1(1)$ and appropriately chosen divisors. It would be natural to give a sample argument along these lines at this point, but we will not since it easy to find such arguments in other books and survey papers (see, e.g., [**DI95**]). So you will not learn much about how to derive dimension formulas from this chapter. What you will learn is precisely what the dimension formulas are, which is something that is often hard to extract from obscure references.

In addition to reading this chapter, the reader may wish to consult [**Mar05**] for proofs of similar dimension formulas, asymptotic results, and a nonrecursive formula for dimensions of certain new subspaces.

6.1. Modular Forms for $\Gamma_0(N)$

For any prime p and any positive integer N, let $v_p(N)$ be the power of p that divides N. Also, let

$$\mu_0(N) = \prod_{p \mid N} \left(p^{v_p(N)} + p^{v_p(N)-1}\right),$$

$$\mu_{0,2}(N) = \begin{cases} 0 & \text{if } 4 \mid N, \\ \prod_{p \mid N} \left(1 + \left(\frac{-4}{p}\right)\right) & \text{otherwise,} \end{cases}$$

$$\mu_{0,3}(N) = \begin{cases} 0 & \text{if } 2 \mid N \text{ or } 9 \mid N, \\ \prod_{p \mid N} \left(1 + \left(\frac{-3}{p}\right)\right) & \text{otherwise,} \end{cases}$$

$$c_0(N) = \sum_{d \mid N} \varphi(\gcd(d, N/d)),$$

$$g_0(N) = 1 + \frac{\mu_0(N)}{12} - \frac{\mu_{0,2}(N)}{4} - \frac{\mu_{0,3}(N)}{3} - \frac{c_0(N)}{2}.$$

[1] The formulas in [**Shi94**, §2.6] contain some minor mistakes.

Note that $\mu_0(N)$ is the index of $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ (see Exercise 6.1).

Proposition 6.1. *We have* $\dim S_2(\Gamma_0(N)) = g_0(N)$, *and for* $k \geq 4$ *even,*

$$\dim S_k(\Gamma_0(N)) = (k-1)\cdot(g_0(N)-1) \;+\; \left(\frac{k}{2}-1\right)\cdot c_0(N)$$
$$+\,\mu_{0,2}(N)\cdot\left\lfloor\frac{k}{4}\right\rfloor \;+\; \mu_{0,3}(N)\cdot\left\lfloor\frac{k}{3}\right\rfloor.$$

The dimension of the Eisenstein subspace is

$$\dim E_k(\Gamma_0(N)) = \begin{cases} c_0(N) & \textit{if } k \neq 2,\\ c_0(N)-1 & \textit{if } k = 2.\end{cases}$$

The following is a table of $\dim S_k(\Gamma_0(N))$ for some values of N and k:

N	$S_2(\Gamma_0(N))$	$S_4(\Gamma_0(N))$	$S_6(\Gamma_0(N))$	$S_{24}(\Gamma_0(N))$
1	0	0	0	2
10	0	3	5	33
11	1	2	4	22
100	7	36	66	336
389	32	97	161	747
1000	131	430	730	3430
2007	221	806	1346	6206
100000	14801	44800	74800	344800

Example 6.2. Use the commands `dimension_cusp_forms`, `dimension_eis`, and `dimension_modular_forms` to compute the dimensions of the three spaces $S_k(\Gamma_0(N))$, $E_k(\Gamma_0(N))$ and $M_k(\Gamma_0(N))$, respectively. For example,

```
sage: dimension_cusp_forms(Gamma0(2007),2)
221
sage: dimension_eis(Gamma0(2007),2)
7
sage: dimension_modular_forms(Gamma0(2007),2)
228
```

Remark 6.3. Csirik, Wetherell, and Zieve prove in [**CWZ01**] that a random positive integer has probability 0 of being a value of

$$g_0(N) = \dim S_2(\Gamma_0(N)),$$

and they give bounds on the size of the set of values of $g_0(N)$ below some given x. For example, they show that $150, 180, 210, 286, 304, 312, \ldots$ are the first few integers that are not of the form $g_0(N)$ for any N. See Figure 6.1.1 for a plot of the very erratic function $g_0(N)$. In contrast, the function $k \mapsto \dim S_{2k}(\Gamma_0(12))$ is very well behaved (see Figure 6.1.2).

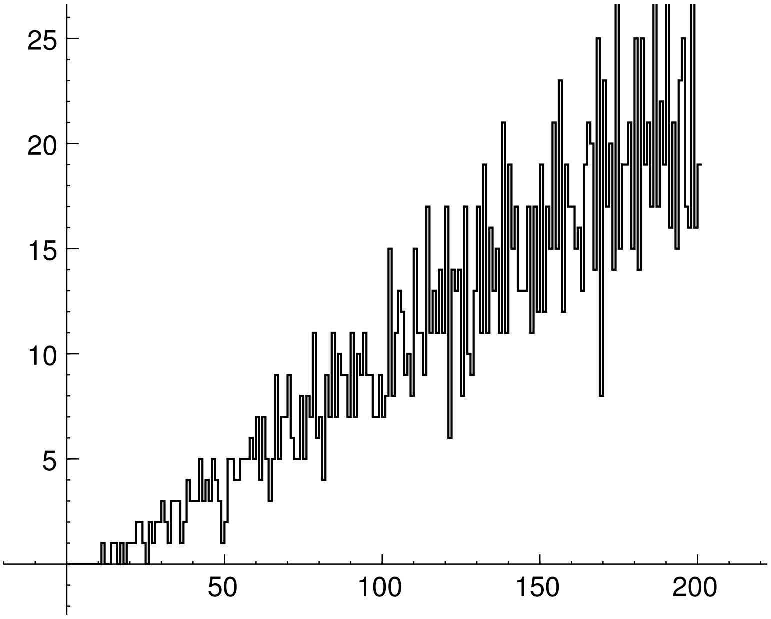


Figure 6.1.1. Dimension of $S_2(\Gamma_0(N))$ as a function of N.

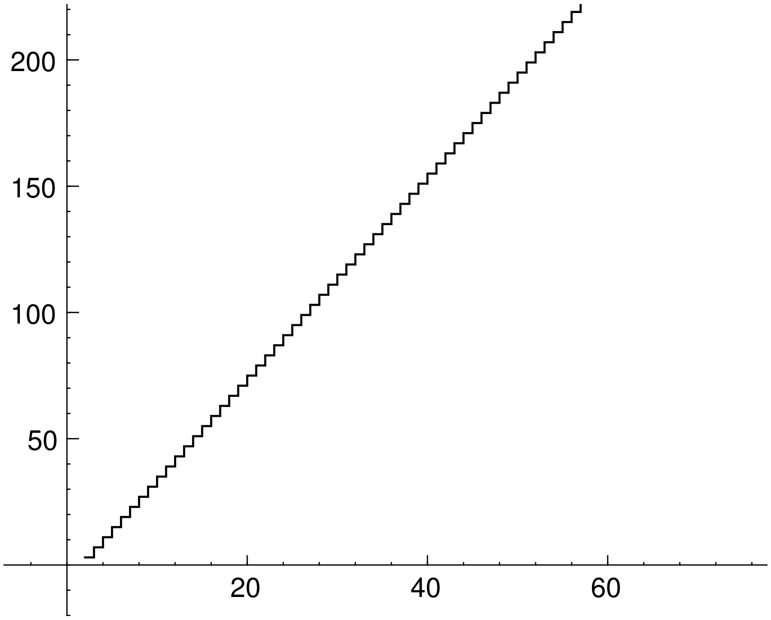


Figure 6.1.2. Dimension of $S_{2k}(\Gamma_0(12))$ as a function of k.

6.1.1. New and Old Subspaces. In this section we assume the reader is either familiar with newforms or has read Section 9.2.

For any integer R, let

$$\overline{\mu}(R) = \begin{cases} 0 & \text{if } p^3 \mid R \text{ for some } p, \\ \prod_{p\|R} -2 & \text{otherwise,} \end{cases}$$

where the product is over primes that exactly divide R. Note that $\overline{\mu}$ is *not* the Moebius function, but it has a similar flavor.

Proposition 6.4. *The dimension of the new subspace is*

$$\dim S_k(\Gamma_0(N))_{\text{new}} = \sum_{M|N} \overline{\mu}(N/M) \cdot \dim S_k(\Gamma_0(M)),$$

where the sum is over the positive divisors M of N. As a consequence of Theorem 9.4, we also have

$$\dim S_k(\Gamma_0(N)) = \sum_{M|N} \sigma_0(N/M) \dim S_k(\Gamma_0(M))_{\text{new}},$$

where $\sigma_0(N/M)$ is the number of divisors of N/M.

Example 6.5. We compute the dimension of the new subspace of $S_k(\Gamma_0(N))$ using the SAGE command `dimension_new_cusp_forms` as follows:

```
sage: dimension_new_cusp_forms(Gamma0(11),12)
8
sage: dimension_cusp_forms(Gamma0(11),12)
10
sage: dimension_new_cusp_forms(Gamma0(2007),12)
1017
sage: dimension_cusp_forms(Gamma0(2007),12)
2460
```

6.2. Modular Forms for $\Gamma_1(N)$

This section follows Section 6.1 closely, but with suitable modifications with $\Gamma_0(N)$ replaced by $\Gamma_1(N)$.

Define functions of a positive integer N by the following formulas:

$$\mu_1(N) = \begin{cases} \mu_0(N) & \text{if } N = 1, 2, \\ \dfrac{\phi(N) \cdot \mu_0(N)}{2} & \text{otherwise,} \end{cases}$$

$$\mu_{1,2}(N) = \begin{cases} 0 & \text{if } N \geq 4, \\ \mu_{0,2}(N) & \text{otherwise,} \end{cases}$$

$$\mu_{1,3}(N) = \begin{cases} 0 & \text{if } N \geq 4, \\ \mu_{0,3}(N) & \text{otherwise,} \end{cases}$$

$$c_1(N) = \begin{cases} c_0(N) & \text{if } N = 1, 2, \\ 3 & \text{if } N = 4, \\ \displaystyle\sum_{d \mid N} \frac{\phi(d)\phi(N/d)}{2} & \text{otherwise,} \end{cases}$$

$$g_1(N) = 1 + \frac{\mu_1(N)}{12} - \frac{\mu_{1,2}(N)}{4} - \frac{\mu_{1,3}(N)}{3} - \frac{c_1(N)}{2}.$$

Note that $g_1(N)$ is the genus of the modular curve $X_1(N)$ (associated to $\Gamma_1(N)$) and $c_1(N)$ is the number of cusps of $X_1(N)$.

Proposition 6.6. *We have* $\dim S_2(\Gamma_1(N)) = g_1(N)$. *If* $N \leq 2$, *then* $\Gamma_0(N) = \Gamma_1(N)$ *so*

$$\dim S_k(\Gamma_1(N)) = \dim S_k(\Gamma_0(N)),$$

where $\dim S_k(\Gamma_0(N))$ *is given by the formula of Proposition 6.1. If* $k \geq 3$, *let*

$$a(N,k) = (k-1)(g_1(N) - 1) + \left(\frac{k}{2} - 1\right) \cdot c_1(N).$$

Then for $N \geq 3$,

$$\dim S_k(\Gamma_1(N)) = \begin{cases} a + 1/2 & \textit{if } N = 4 \textit{ and } 2 \nmid k, \\ a + \lfloor k/3 \rfloor & \textit{if } N = 3, \\ a & \textit{otherwise.} \end{cases}$$

The dimension of the Eisenstein subspace is as follows:

$$\dim E_k(\Gamma_1(N)) = \begin{cases} c_1(N) & \textit{if } k \neq 2, \\ c_1(N) - 1 & \textit{if } k = 2. \end{cases}$$

The dimension of the new subspace of $M_k(\Gamma_1(N))$ *is*

$$\dim S_k(\Gamma_1(N))_{\text{new}} = \sum_{M \mid N} \overline{\mu}(N/M) \cdot \dim S_k(\Gamma_1(M)),$$

where $\overline{\mu}$ *is as in the statement of Proposition 6.4.*

Remark 6.7. Since $M_k = S_k \oplus E_k$, the formulas above for $\dim S_k$ and $\dim E_k$ also yield a formula for the dimension of M_k.

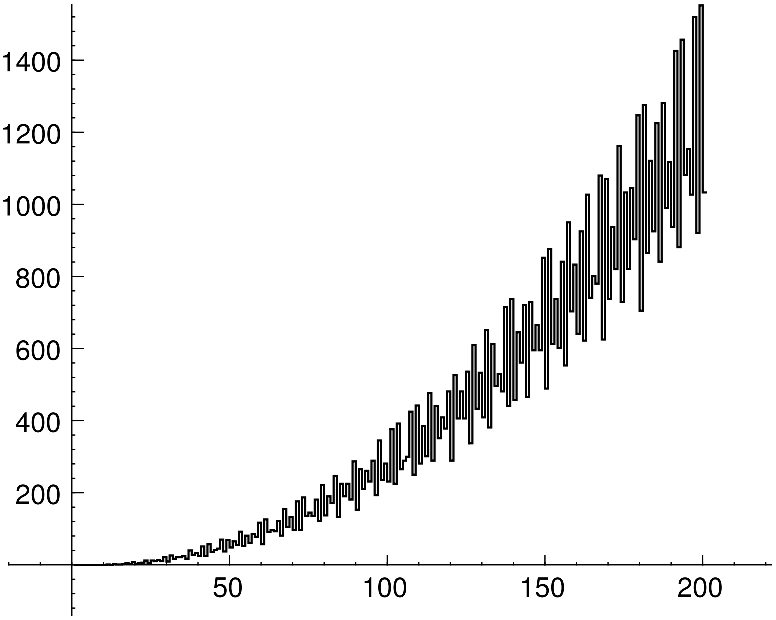


Figure 6.2.1. Dimension of $S_2(\Gamma_1(N))$ as a function of N.

The following table contains the dimension of $S_k(\Gamma_1(N))$ for some sample values of N and k:

N	$S_2(\Gamma_1(N))$	$S_3(\Gamma_1(N))$	$S_4(\Gamma_1(N))$	$S_{24}(\Gamma_1(N))$
1	0	0	0	2
10	0	2	5	65
11	1	5	10	110
100	231	530	830	6830
389	6112	12416	18721	144821
1000	28921	58920	88920	688920
2007	147409	296592	445776	3429456
100000	299792001	599792000	899792000	6899792000

Example 6.8. We compute dimensions of spaces of modular forms for $\Gamma_1(N)$:

```
sage: dimension_cusp_forms(Gamma1(2007),2)
147409
sage: dimension_eis(Gamma1(2007),2)
3551
sage: dimension_modular_forms(Gamma1(2007),2)
150960
```

6.3. Modular Forms with Character

Fix a Dirichlet character ε of modulus N, and let c be the conductor of ε (we do *not* assume that ε is primitive). Assume that $\varepsilon \neq 1$, since otherwise $M_k(N,\varepsilon) = M_k(\Gamma_0(N))$ and the formulas of Section 6.1 apply. Also, assume that $\varepsilon(-1) = (-1)^k$, since otherwise $\dim M_k(\Gamma_0(N)) = 0$. In this section we discuss formulas for computing each of $M_k(N,\varepsilon)$, $S_k(N,\varepsilon)$ and $E_k(N,\varepsilon)$.

In [**CO77**], Cohen and Oesterlé assert (without *published* proof; see Remark 6.11 below) that for any $k \in \mathbb{Z}$ and N, ε as above,

$$\begin{aligned}\dim S_k(N,\varepsilon) &- \dim M_{2-k}(N,\varepsilon)\\ &= \frac{k-1}{12}\cdot \mu_0(N) \;-\; \frac{1}{2}\cdot \prod_{p\mid N}\lambda(p,N,v_p(c))\\ &\quad + \gamma_4(k)\cdot \sum_{x\in A_4(N)}\varepsilon(x) \;+\; \gamma_3(k)\cdot\sum_{x\in A_3(N)}\varepsilon(x)\end{aligned}$$

where $\mu_0(N)$ is as in Section 6.1, $A_4(N) = \{x \in \mathbb{Z}/N\mathbb{Z} : x^2+1 = 0\}$ and $A_3(N) = \{x \in \mathbb{Z}/N\mathbb{Z} : x^2+x+1 = 0\}$, and γ_3, γ_4 are

$$\gamma_4(k) = \begin{cases} -1/4 & \text{if } k \equiv 2 \pmod 4,\\ 1/4 & \text{if } k \equiv 0 \pmod 4,\\ 0 & \text{if } k \text{ is odd.}\end{cases}$$

$$\gamma_3(k) = \begin{cases} -1/3 & \text{if } k \equiv 2 \pmod 3,\\ 1/3 & \text{if } k \equiv 0 \pmod 3,\\ 0 & \text{if } k \equiv 1 \pmod 3.\end{cases}$$

It remains to define λ. Fix a prime divisor $p \mid N$ and let $r = v_p(N)$. Then

$$\lambda(p,N,v_p(c)) = \begin{cases} p^{\frac{r}{2}} + p^{\frac{r}{2}-1} & \text{if } 2\cdot v_p(c) \le r \text{ and } 2 \mid r,\\ 2\cdot p^{\frac{r-1}{2}} & \text{if } 2\cdot v_p(c) \le r \text{ and } 2 \nmid r,\\ 2\cdot p^{r-v_p(c)} & \text{if } 2\cdot v_p(c) > r.\end{cases}$$

This flexible formula can be used to compute the dimension of $M_k(N,\varepsilon)$, $S_k(N,\varepsilon)$, and $E_k(N,\varepsilon)$ for any N, ε, $k \neq 1$, by using that

$$\begin{aligned}
\dim S_k(N,\varepsilon) &= 0 \qquad \text{if } k \leq 0,\\
\dim M_k(N,\varepsilon) &= 0 \qquad \text{if } k < 0,\\
\dim M_0(N,\varepsilon) &= 1 \qquad \text{if } k = 0.
\end{aligned}$$

One thing that is not straightforward when implementing an algorithm to compute the above dimension formulas is how to efficiently compute the sets $A_4(N)$ and $A_6(N)$. Kevin Buzzard suggested the following two algorithms. Note that if k is odd, then $\gamma_4(k) = 0$, so the sum over $A_4(N)$ is only needed when k is even.

Algorithm 6.9 (Sum over $A_4(N)$)**.** *Given a positive integer N and an even Dirichlet character ε of modulus N, this algorithm computes $\sum_{x\in A_4(N)} \varepsilon(x)$.*

```
(1) [Factor $N$] Compute the prime factorization $p_1^{e_1} \cdots p_n^{e_n}$ of $N$.
(2) [Initialize] Set $t = 1$ and $i = 0$.
(3) [Loop Over Prime Divisors] Set $i = i + 1$. If $i > n$, return $t$.
    Otherwise set $p = p_i$ and $e = e_i$.
    (a) If $p \equiv 3 \pmod 4$, return 0.
    (b) If $p = 2$ and $e > 1$, return 0.
    (c) If $p = 2$ and $e = 1$, go to step (3).
    (d) Compute a generator $a \in (\mathbb{Z}/p\mathbb{Z})^*$ using Algorithm 4.4.
    (e) Compute $\omega = a^{(p-1)/4}$.
    (f) Use the Chinese Remainder Theorem to find $x \in \mathbb{Z}/N\mathbb{Z}$ such
        that $x \equiv a \pmod p$ and $x \equiv 1 \pmod{N/p^e}$.
    (g) Set $x = x^{p^{r-1}}$.
    (h) Set $s = \varepsilon(x)$.
    (i) If $s = 1$, set $t = 2t$ and go to step (3).
    (j) If $s = -1$, set $t = -2t$ and go to step (3).
```

Proof. Note that $\varepsilon(-x) = \varepsilon(x)$, since ε is even. By the Chinese Remainder Theorem, the set $A_4(N)$ is empty if and only if there is no square root of -1 modulo some prime power divisor of p. If $A_4(N)$ is empty, the algorithm correctly detects this fact in steps (3a)–(3b). Thus assume $A_4(N)$ is nonempty. For each prime power $p_i^{e_i}$ that exactly divides N, let $x_i \in Z/N\mathbb{Z}$ be such that $x_i^2 = -1$ and $x_i \equiv 1 \pmod{p_j^{e_j}}$ for $i \neq j$. This is the value of x computed in steps (3d)–(3g) (as one sees using elementary number theory).

The next key observation is that

$$\prod_i (\varepsilon(x_i) + \varepsilon(-x_i)) = \sum_{x\in A_4(N)} \varepsilon(x), \tag{6.3.1}$$

since by the Chinese Remainder Theorem the elements of $A_4(N)$ are in bijection with the choices for a square root of -1 modulo each prime power divisors of N. The observation (6.3.1) is a huge gain from an efficiency point of view—if N had r prime factors, then $A_4(N)$ would have size 2^r, which could be prohibitive, where the product involves only r factors. To finish the proof, just note that steps (3h)–(3j) compute the local factors $\varepsilon(x_i) + \varepsilon(-x_i) = 2\varepsilon(x_i)$, where again we use that ε is even. Note that a solution of $x^2+1 \equiv 0 \pmod{p}$ lifts uniquely to a solution mod p^n for any n, because the kernel of the natural homomorphism $(\mathbb{Z}/p^n\mathbb{Z})^* \to (\mathbb{Z}/p\mathbb{Z})^*$ is a group of p-power order. □

The algorithm for computing the sum over $A_3(N)$ is similar.

For $k \geq 2$, to compute $\dim S_k(N,\varepsilon)$, use the formula directly and the fact that $\dim M_{2-k}(N,\varepsilon) = 0$, unless $\varepsilon = 1$ and $k = 2$. To compute $\dim M_k(N,\varepsilon)$ for $k \geq 2$, use the fact that the big formula at the beginning of this section is valid for any integer k to replace k by $2-k$ and that $\dim S_k(N,\varepsilon) = 0$ for $k \leq 0$ to rewrite the formula as

$$\begin{aligned}\dim M_k(N,\varepsilon) &= -(\dim S_{2-k}(N,\varepsilon) - \dim M_k(N,\varepsilon))\\ &= -\Big(\frac{1-k}{12}\cdot\mu_0(N) \;-\; \frac{1}{2}\cdot\prod_{p\mid N}\lambda(p,N,v_p(c))\\ &\qquad + \gamma_4(2-k)\cdot\sum_{x\in A_4(N)}\varepsilon(x) \;+\; \gamma_3(2-k)\cdot\sum_{x\in A_3(N)}\varepsilon(x)\Big).\end{aligned}$$

Note also that for $k = 0$, $\dim E_k(N,\varepsilon) = 1$ if and only if ε is trivial and it equals 0 otherwise. We then also obtain

$$\dim E_k(N,\varepsilon) = \dim M_k(N,\varepsilon) - \dim S_k(N,\varepsilon).$$

We can also compute $\dim E_k(N,\varepsilon)$ when $k = 1$ directly, since

$$\dim S_{2-1}(N,\varepsilon) = \dim S_1(N,\varepsilon).$$

The following table contains the dimension of $S_k(N,\varepsilon)$ for some sample values of N and k. In each case, ε is the product of characters ε_p of maximal order corresponding to the prime power factors of N (i.e., the product of the generators of the group $D(N,\mathbb{C}^*)$ of Dirichlet characters of modulus N).

N	$\dim S_2(N,\varepsilon)$	$\dim S_3(N,\varepsilon)$	$\dim S_4(N,\varepsilon)$	$\dim S_{24}(N,\varepsilon)$
1	0	0	0	2
10	0	1	0	0
11	0	1	0	0
100	13	0	43	343
389	0	64	0	0
1000	148	0	448	3448
2007	222	0	670	5150

Example 6.10. We compute the last line of the above table. First we create the character ε.

```
sage: G = DirichletGroup(2007)
sage: e = prod(G.gens(), G(1))
```

Next we compute the dimension of the four spaces.

```
sage: dimension_cusp_forms(e,2)
222
sage: dimension_cusp_forms(e,3)
0
sage: dimension_cusp_forms(e,4)
670
sage: dimension_cusp_forms(e,24)
5150
```

We can also compute dimensions of the corresponding spaces of Eisenstein series.

```
sage: dimension_eis(e,2)
4
sage: dimension_eis(e,3)
0
sage: dimension_eis(e,4)
4
sage: dimension_eis(e,24)
4
```

Remark 6.11. Cohen and Oesterlé also give dimension formulas for spaces of half-integral weight modular forms, which we do not give in this chapter. Note that [**CO77**] does not contain any *proofs* that their claimed formulas are correct, but instead they say only that "Les formules qui les donnent sont connues de beaucoup de gens et il existe plusieurs méthodes permettant de les obtenir (théorème de Riemann-Roch, application des formules de trace données par Shimura)."[2] Fortunately, in [**Que06**], Jordi Quer derives the (integral weight) formulas of [**CO77**] along with formulas for dimensions of spaces $S_k(G)$ and $M_k(G)$ for more general congruence subgroups.

Let f be the conductor of a Dirichlet character ε of modulus N. Then the dimension of the new subspace of $M_k(N,\varepsilon)$ is

$$\dim S_k(N,\varepsilon)_{\text{new}} = \sum_{M \text{ such that } f|M|N} \overline{\mu}(N/M) \cdot \dim S_k(M,\varepsilon'),$$

where $\overline{\mu}$ is as in the statement of Proposition 6.4, and ε' is the restriction of ε mod M.

Example 6.12. We compute the dimension of $S_2(2007,\varepsilon)_{\text{new}}$ for ε a quadratic character of modulus 2007.

```
sage: G = DirichletGroup(2007, QQ)
sage: e = prod(G.gens(), G(1))
sage: dimension_new_cusp_forms(e,2)
76
```

6.4. Exercises

6.1 Let μ_0 and μ_1 be as in this chapter.

(a) Prove that $\mu_0(N) = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)]$.

(b) Prove that for $N \geq 3$, $\mu_1(N) = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_1(N)]/2$, so $\mu_1(N)$ is the index of $\Gamma_1(N) \cdot \{\pm 1\}$ in $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$.

6.2 Use Proposition 6.4 to find a formula for $\dim S_k(\mathrm{SL}_2(\mathbb{Z}))$. Verify that this formula is the same as the one in Corollary 2.16.

6.3 Suppose either that $N = 1$ or that N is prime and $k = 2$. Prove that $M_k(\Gamma_0(N))_{\text{new}} = M_k(\Gamma_0(N))$.

6.4 Fill in the details of the proof of Algorithm 6.9.

6.5 Implement a computer program to compute $\dim S_k(\Gamma_0(N))$ as a function of k and N.

[2] The formulas that we give here are well known and there exist many methods to prove them, e.g., the Riemann-Roch theorem and applications of the trace formula of Shimura.

Chapter 7

Linear Algebra

This chapter is about several algorithms for matrix algebra over the rational numbers and cyclotomic fields. Algorithms for linear algebra over exact fields are necessary in order to implement the modular symbols algorithms that we will describe in Chapter 7. This chapter partly overlaps with [**Coh93**, Sections 2.1–2.4].

Note: We view all matrices as defining linear transformations by acting on row vectors from the right.

7.1. Echelon Forms of Matrices

Definition 7.1 (Reduced Row Echelon Form)**.** A matrix is in (reduced row) *echelon form* if each row in the matrix has more zeros at the beginning than the row above it, the first nonzero entry of every row is 1, and the first nonzero entry of any row is the only nonzero entry in its column.

Given a matrix A, there is another matrix B such that B is obtained from A by left multiplication by an invertible matrix and B is in reduced row echelon form. This matrix B is called the echelon form of A. It is unique.

A *pivot column* of A is a column of A such that the reduced row echelon form of A contains a leading 1.

Example 7.2. The following matrix is not in reduced row echelon form:

$$\begin{pmatrix} 14 & 2 & 7 & 228 & -224 \\ 0 & 0 & 3 & 78 & -70 \\ 0 & 0 & 0 & -405 & 381 \end{pmatrix}.$$

The reduced row echelon form of the above matrix is

$$\begin{pmatrix} 1 & \frac{1}{7} & 0 & 0 & -\frac{1174}{945} \\ 0 & 0 & 1 & 0 & \frac{152}{135} \\ 0 & 0 & 0 & 1 & -\frac{127}{135} \end{pmatrix}.$$

Notice that the entries of the reduced row echelon form can be rationals with large denominators even though the entries of the original matrix A are integers. Another example is the simple looking matrix

$$\begin{pmatrix} -9 & 6 & 7 & 3 & 1 & 0 & 0 & 0 \\ -10 & 3 & 8 & 2 & 0 & 1 & 0 & 0 \\ 3 & -6 & 2 & 8 & 0 & 0 & 1 & 0 \\ -8 & -6 & -8 & 6 & 0 & 0 & 0 & 1 \end{pmatrix}$$

whose echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{42}{1025} & -\frac{92}{1025} & \frac{1}{25} & -\frac{9}{205} \\ 0 & 1 & 0 & 0 & \frac{716}{3075} & -\frac{641}{3075} & -\frac{2}{75} & -\frac{7}{615} \\ 0 & 0 & 1 & 0 & -\frac{83}{1025} & \frac{133}{1025} & \frac{1}{25} & -\frac{23}{410} \\ 0 & 0 & 0 & 1 & \frac{184}{1025} & -\frac{159}{1025} & \frac{2}{25} & \frac{9}{410} \end{pmatrix}.$$

A basic fact is that two matrices A and B have the same reduced row echelon form if and only if there is an invertible matrix E such that $EA = B$. Also, many standard operations in linear algebra, e.g., computation of the kernel of a linear map, intersection of subspaces, membership checking, etc., can be encoded as a question about computing the echelon form of a matrix.

The following standard algorithm computes the echelon form of a matrix.

Algorithm 7.3 (Gauss Elimination)**.** *Given an $m \times n$ matrix A over a field, the algorithm outputs the reduced row echelon form of A. Write $a_{i,j}$ for the i, j entry of A, where $0 \leq i \leq m-1$ and $0 \leq j \leq n-1$.*

(1) [Initialize] Set $k = 0$.

(2) [Clear Each Column] For each column $c = 0, 1, \ldots, n-1$, clear the cth column as follows:

   (a) [First Nonzero] Find the smallest r such that $a_{r,c} \neq 0$, or if there is no such r, go to the next column.

   (b) [Rescale] Replace row r of A by $\frac{1}{a_{r,c}}$ times row r.

   (c) [Swap] Swap row r with row k.

   (d) [Clear] For each $i = 0, \ldots, m-1$ with $i \neq k$, if $a_{i,c} \neq 0$, add $-a_{i,c}$ times row k of A to row i to clear the leading entry of the ith row.

   (e) [Increment] Set $k = k + 1$.

This algorithm takes $O(mn^2)$ arithmetic operations in the base field, where A is an $m \times n$ matrix. If the base field is $\mathbb{Q}$, the entries can become huge and arithmetic operations are then very expensive. See Section 7.3 for ways to mitigate this problem.

To conclude this section, we mention how to convert a few standard problems into questions about reduced row echelon forms of matrices. Note that one can also phrase some of these answers in terms of the echelon form, which might be easier to compute, or an LUP decomposition (lower triangular times upper triangular times permutation matrix), which the numerical analysts use.

(1) **Kernel of A:** We explain how to compute the kernel of A acting on column vectors from the right (first transpose to obtain the kernel of A acting on row vectors). Since passing to the reduced row echelon form of A is the same as multiplying on the left by an invertible matrix, the kernel of the reduce row echelon form E of A is the same as the kernel of A. There is a basis vector of $\ker(E)$ that corresponds to each nonpivot column of E. That vector has a 1 at the nonpivot column, 0's at all other nonpivot columns, and for each pivot column, the negative of the entry of A at the nonpivot column in the row with that pivot element.

(2) **Intersection of Subspaces:** Suppose W_1 and W_2 are subspace of a finite-dimensional vector space V. Let A_1 and A_2 be matrices whose columns form a basis for W_1 and W_2, respectively. Let $A = [A_1|A_2]$ be the augmented matrix formed from A_1 and A_2. Let K be the kernel of the linear transformation defined by A. Then K is isomorphic to the desired intersection. To write down the intersection explicitly, suppose that $\dim(W_1) \leq \dim(W_2)$ and do the following: For each b in a basis for K, write down the linear combination of a basis for W_1 obtained by taking the first $\dim(W_1)$ entries of the vector b. The fact that b is in $\mathrm{Ker}(A)$ implies that the vector we just wrote down is also in W_2. This is because a linear relation

$$\sum a_i w_{1,i} + \sum b_j w_{2,j} = 0,$$

i.e., an element of that kernel, is the same as

$$\sum a_i w_{1,i} = \sum -b_j w_{2,j}.$$

For more details, see [**Coh93**, Alg. 2.3.9].

7.2. Rational Reconstruction

Rational reconstruction is a process that allows one to sometimes lift an integer modulo m uniquely to a bounded rational number.

Algorithm 7.4 (Rational Reconstruction)**.** *Given an integer $a \geq 0$ and an integer $m > 1$, this algorithm computes the numerator n and denominator d of the unique rational number n/d, if it exists, with*

$$|n|, d \leq \sqrt{\frac{m}{2}} \qquad \textit{and} \qquad n \equiv ad \pmod{m}, \tag{7.2.1}$$

or it reports that there is no such number.

(1) [Reduce mod m] Replace a with the least integer between 0 and $m-1$ that is congruent to a modulo m.

(2) [Trivial Cases] If $a = 0$ or $a = 1$, return a.

(3) [Initialize] Let $b = \sqrt{m/2}$, $u = m$, $v = a$, and set $U = (1, 0, u)$ and $V = (0, 1, v)$. Use the notation U_i and V_i to refer to the ith entries of U, V, for $i = 0, 1, 2$.

(4) [Iterate] Do the following as long as $|V_2| > b$: Set $q = \lfloor U_2/V_2 \rfloor$, set $T = U - qV$, and set $U = V$ and $V = T$.

(5) [Numerator and Denominator] Set $d = |V_1|$ and $n = V_2$.

(6) [Good?] If $d \leq b$ and $\gcd(n, d) = 1$, return n/d; otherwise report that there is no rational number as in (7.2.1).

Algorithm 7.4 for rational reconstruction is described (with proof) in [**Knu**, pgs. 656–657] as the solution to Exercise 51 on page 379 in that book. See, in particular, the paragraph right in the middle of page 657, which describes the algorithm. Knuth attributes this rational reconstruction algorithm to Wang, Kornerup, and Gregory from around 1983.

We now give an indication of why Algorithm 7.4 computes the rational reconstruction of $a \pmod{m}$, leaving the precise details and uniqueness to [**Knu**, pgs. 656–657]. At each step in Algorithm 7.4, the 3-tuple $V = (v_0, v_1, v_2)$ satisfies

$$m \cdot v_0 + a \cdot v_1 = v_2, \tag{7.2.2}$$

and similarly for U. When computing the usual extended gcd, at the end $v_2 = \gcd(a, m)$ and v_0, v_1 give a representation of the v_2 as a $\mathbb{Z}$-linear combination of m and a. In Algorithm 7.4, we are instead interested in finding a rational number n/d such that $n \equiv a \cdot d \pmod{m}$. If we set $n = v_2$ and $d = v_1$ in (7.2.2) and rearrange, we obtain

$$n = a \cdot d + m \cdot v_0.$$

Thus at *every* step of the algorithm we find a rational number n/d such that $n \equiv ad \pmod{m}$. The problem at intermediate steps is that, e.g., v_0 could be 0, or n or d could be too large.

Example 7.5. We compute an example using SAGE.

```
sage: p = 389
sage: k = GF(p)
sage: a = k(7/13); a
210
sage: a.rational_reconstruction()
7/13
```

7.3. Echelon Forms over $\mathbb{Q}$

A difficulty with computation of the echelon form of a matrix over the rational numbers is that arithmetic with large rational numbers is time-consuming; each addition potentially requires a gcd and numerous additions and multiplications of integers. Moreover, the entries of A during intermediate steps of Algorithm 7.3 can be huge even though the entries of A and the answer are small. For example, suppose A is an invertible square matrix. Then the echelon form of A is the identity matrix, but during intermediate steps the numbers involved could be quite large. One technique for mitigating this is to compute the echelon form using a multimodular method.

If A is a matrix with rational entries, let $H(A)$ be the *height* of A, which is the maximum of the absolute values of the numerators and denominators of all entries of A. If x, y are rational numbers and p is a prime, we write $x \equiv y \pmod{p}$ to mean that the denominators of x and y are not divisible by p but the numerator of the rational number $x - y$ (in reduced form) is divisible by p. For example, if $x = 5/7$ and $y = 2/11$, then $x - y = 41/77$, so $x \equiv y \pmod{41}$.

Algorithm 7.6 (Multimodular Echelon Form)**.** *Given an $m \times n$ matrix A with entries in $\mathbb{Q}$, this algorithm computes the reduced row echelon form of A.*

(1) Rescale the input matrix A to have integer entries. This does not change the echelon form and makes reduction modulo many primes easier. We may thus assume A has integer entries.
(2) Let c be a guess for the height of the echelon form.
(3) List successive primes $p_1, p_2, \ldots$ such that the product of the p_i is greater than $n \cdot c \cdot H(A) + 1$, where n is the number of columns of A.
(4) Compute the echelon forms B_i of the reduction $A \pmod{p_i}$ using, e.g., Algorithm 7.3 or any other echelon algorithm.
(5) Discard any B_i whose pivot column list is not maximal among pivot lists of all B_j found so far. (The pivot list associated to B_i is the ordered list of integers k such that the kth column of B_j is a pivot

column. We mean maximal with respect to the following ordering on integer sequences: shorter integer sequences are smaller, and if two sequences have the same length, then order in reverse lexicographic order. Thus $[1, 2]$ is smaller than $[1, 2, 3]$, and $[1, 2, 7]$ is smaller than $[1, 2, 5]$. Think of maximal as "optimal", i.e., best possible pivot columns.)

(6) Use the Chinese Remainder Theorem to find a matrix B with integer entries such that $B \equiv B_i \pmod{p_i}$ for all p_i.

(7) Use Algorithm 7.4 to try to find a matrix C whose coefficients are rational numbers n/r such that $|n|, r \leq \sqrt{M/2}$, where $M = \prod p_i$, and $C \equiv B_i \pmod{p_i}$ for each prime p. If rational reconstruction fails, compute a few more echelon forms mod the next few primes (using the above steps) and attempt rational reconstruction again. Let E be the matrix over $\mathbb{Q}$ so obtained. (A trick here is to keep track of denominators found so far to avoid doing very many rational reconstructions.)

(8) Compute the denominator d of E, i.e., the smallest positive integer such that dE has integer entries. If

$$H(dE) \cdot H(A) \cdot n < \prod p_i, \tag{7.3.1}$$

then E is the reduced row echelon form of A. If not, repeat the above steps with a few more primes.

Proof. We prove that if (7.3.1) is satisfied, then the matrix E computed by the algorithm really is the reduced row echelon form R of A. First note that E is in reduced row echelon form since the set of pivot columns of all matrices B_i used to construct E are the same, so the pivot columns of E are the same as those of any B_i and all other entries in the B_i pivot columns are 0, so the other entries of E in the pivot columns are also 0.

Recall from the end of Section 7.1 that a matrix whose columns are a basis for the kernel of A can be obtained from the reduced row echelon form R. Let K be the matrix whose columns are the vectors in the kernel algorithm applied to E, so $EK = 0$. Since the reduced row echelon form is obtained by left multiplying by an invertible matrix, for each i, there is an invertible matrix V_i mod p_i such that $A \equiv V_i B_i \pmod{p_i}$ so

$$A \cdot dK \equiv V_i B_i \cdot dK \equiv V_i \cdot dE \cdot K \equiv 0 \pmod{p_i}.$$

Since dK and A are integer matrices, the Chinese remainder theorem implies that

$$A \cdot dK \equiv 0 \left(\bmod \prod p_i\right).$$

The integer entries a of $A \cdot dK$ all satisfy $|a| \leq H(A) \cdot H(dK) \cdot n$, where n is the number of columns of A. Since $H(K) \leq H(E)$, the bound (7.3.1) implies that $A \cdot dK = 0$. Thus $AK = 0$, so $\mathrm{Ker}(E) \subset \mathrm{Ker}(A)$. On the other hand, the rank of E equals the rank of each B_i (since the pivot columns are the same), so

$$\mathrm{rank}(E) = \mathrm{rank}(B_i) = \mathrm{rank}(A \pmod{p_i}) \leq \mathrm{rank}(A).$$

Thus $\dim(\mathrm{Ker}(A)) \leq \dim(\mathrm{Ker}(E))$, and combining this with the bound obtained above, we see that $\mathrm{Ker}(E) = \mathrm{Ker}(A)$. This implies that E is the reduced row echelon form of A, since two matrices have the same kernel if and only if they have the same reduced row echelon form (the echelon form is an invariant of the row space, and the kernel is the orthogonal complement of the row space).

The reason for step (5) is that the matrices B_i need *not* be the reduction of R modulo p_i, and indeed this reduction might not even be defined, e.g., if p_i divides the denominator of some element of R, then this reduction makes no sense. For example, set $p = p_i$ and suppose $A = \left(\begin{smallmatrix} p & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Then $R = \begin{pmatrix} 1 & 1/p \\ 0 & 0 \end{pmatrix}$, which has no reduction modulo p; also, the reduction of A modulo B_i is $B_i = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \pmod{p}$, which is already in reduced row echelon form. However if we were to combine B_i with the echelon form of A modulo another prime, the result could never be lifted using rational reconstruction. Thus the reason we exclude all B_i with nonmaximal pivot column sequence is so that a rational reconstruction will exist. There are only finitely many primes that divide denominators of entries of R, so eventually all B_i will have maximal pivot column sequences, i.e., they are the reduction of the true reduced row echelon form R, so the algorithm terminates. □

Remark 7.7. Algorithm 7.6, with *sparse* matrices seems to work very well in practice. A simple but helpful modification to Algorithm 7.3 in the sparse case is to clear each column using a row with a minimal number of nonzero entries, so as to reduce the amount of "fill in" (denseness) of the matrix. There are much more sophisticated methods along these lines called "intelligent Gauss elimination". (Cryptographers are interested in linear algebra mod p with huge sparse matrices, since they come up in attacks on the discrete log problem and integer factorization.)

One can adapt Algorithm 7.6 to computation of echelon forms of matrices A over cyclotomic fields $\mathbb{Q}(\zeta_n)$. Assume A has denominator 1. Let p be a prime that splits completely in $\mathbb{Q}(\zeta_n)$. Compute the homomorphisms $f_i : \mathbb{Z}_p[\zeta_n] \to \mathbb{F}_p$ by finding the elements of order n in $\mathbb{F}_p^*$. Then compute the mod p matrix $f_i(A)$ for each i, and find its reduced row echelon form. Taken together, the maps f_i together induce an isomorphism $\Psi : \mathbb{F}_p[X]/\Phi_n(X) \cong \mathbb{F}_p^d$, where $\Phi_n(X)$ is the nth cyclotomic polynomial and d is its degree. It is easy to compute $\Psi(f(x))$ by evaluating $f(x)$ at each element of order n in $\mathbb{F}_p$. To compute Ψ^{-1}, simply use linear algebra over $\mathbb{F}_p$

to invert a matrix that represents Ψ. Use Ψ^{-1} to compute the reduced row echelon form of $A \pmod{p}$, where (p) is the nonprime ideal in $\mathbb{Z}[\zeta_n]$ generated by p. Do this for several primes p, and use rational reconstruction on each coefficient of each power of ζ_n, to recover the echelon form of A.

7.4. Echelon Forms via Matrix Multiplication

In this section we explain how to compute echelon forms using matrix multiplication. This is valuable because there are asymptotically fast, i.e., better than $O(n^3)$ field operations, algorithms for matrix multiplication, and implementations of linear algebra libraries often include highly optimized matrix multiplication algorithms. We only sketch the basic ideas behind these asymptotically fast algorithms (following [**Ste**]), since more detail would take us too far from modular forms.

The naive algorithm for multiplying two $m \times m$ matrices requires $O(m^3)$ arithmetic operations in the base ring. In [**Str69**], Strassen described a clever algorithm that computes the product of two $m \times m$ matrices in $O(m^{\log_2(7)}) = O(m^{2.807\cdots})$ arithmetic operations in the base ring. Because of numerical stability issues, Strassen's algorithm is rarely used in numerical analysis. But for matrix arithmetic over exact base rings (e.g., the rational numbers, finite fields, etc.) it is of extreme importance.

In [**Str69**], Strassen also sketched a new algorithm for computing the inverse of a square matrix using matrix multiplication. Using this algorithm, the number of operations to invert an $m \times m$ matrix is (roughly) the same as the number needed to multiply two $m \times m$ matrices. Suppose the input matrix is $2^n \times 2^n$ and we write it in block form as $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ where A, B, C, D are all $2^{n-1} \times 2^{n-1}$ matrices. Assume that any intermediate matrices below that we invert are invertible. Consider the augmented matrix

$$\left(\begin{array}{cc|cc} A & B & I & 0 \\ C & D & 0 & I \end{array}\right).$$

Multiply the top row by A^{-1} to obtain

$$\left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ C & D & 0 & I \end{array}\right),$$

and write $E = A^{-1}B$. Subtract C times the first row from the second row to get

$$\left(\begin{array}{cc|cc} I & E & A^{-1} & 0 \\ 0 & D - CE & -CA^{-1} & I \end{array}\right).$$

Set $F = D - CE$ and multiply the bottom row by F^{-1} on the left to obtain

$$\left(\begin{array}{cc|cc} I & E & A^{-1} & 0 \\ 0 & I & -F^{-1}CA^{-1} & F^{-1} \end{array}\right).$$

Set $G = -F^{-1}CA^{-1}$, and subtract E times the second from the first row to arrive at

$$\left(\begin{array}{cc|cc} I & 0 & A^{-1} - EG & -EF^{-1} \\ 0 & I & G & F^{-1} \end{array}\right).$$

The idea listed above can, with significant work, be extended to a general algorithm (as is done in [**Ste06**]).

Next we very briefly sketch how to compute echelon forms of matrices using matrix multiplication and inversion. Its complexity is comparable to the complexity of matrix multiplication.

As motivation, recall the standard algorithm from undergraduate linear algebra for inverting an invertible square matrix A: form the augmented matrix $[A|I]$, and then compute the echelon form of this matrix, which is $[I|A^{-1}]$. If T is the transformation matrix to echelon form, then $T[A|I] = [I|T]$, so $T = A^{-1}$. In particular, we could find the echelon form of $[A|I]$ by multiplying on the left by A^{-1}. Likewise, for any matrix B with the same number of rows as A, we could find the echelon form of $[A|B]$ by multiplying on the left by A^{-1}. Next we extend this idea to give an algorithm to compute echelon forms using only matrix multiplication (and echelon form modulo one prime).

Algorithm 7.8 (Asymptotically Fast Echelon Form)**.** *Given a matrix A over the rational numbers (or a number field), this algorithm computes the echelon form of A.*

(1) [Find Pivots] Choose a random prime p (coprime to the denominator of any entry of A) and compute the echelon form of $A \pmod p$, e.g., using Algorithm 7.3. Let $c_0, \ldots, c_{n-1}$ be the pivot columns of $A \pmod p$. When computing the echelon form, save the positions $r_0, \ldots, r_{n-1}$ of the rows used to clear each column.

(2) [Extract Submatrix] Extract the $n \times n$ submatrix B of A whose entries are A_{r_i,c_j} for $0 \leq i, j \leq n-1$.

(3) [Compute Inverse] Compute the inverse B^{-1} of B. Note that B must be invertible since its reduction modulo p is invertible.

(4) [Multiply] Let C be the matrix whose rows are the rows $r_0, \ldots, r_{n-1}$ of A. Compute $E = B^{-1}C$. If E is not in echelon form, go to step (1).

(5) [Done?] Write down a matrix D whose columns are a basis for $\ker(E)$ as explained on page 105. Let F be the matrix whose rows are the rows of A other than rows $r_0, \ldots, r_{n-1}$. Compute the product FD. If $FD = 0$, output E, which is the echelon form of A. If $FD \neq 0$, go to step (1) and run the whole algorithm again.

Proof. We prove both that the algorithm terminates and that when it terminates, the matrix E is the echelon form of A.

First we prove that the algorithm terminates. Let E be the echelon form of A. By Exercise 7.3, for all but finitely many primes p (i.e., any prime where $A \pmod{p}$ has the same rank as A) the echelon form of $A \pmod{p}$ equals $E \pmod{p}$. For any such prime p the pivot columns of $E \pmod{p}$ are the pivot columns of E, so the algorithm will terminate for that choice of p.

We next prove that when the algorithm terminates, E is the echelon form of A. By assumption, E is in echelon form and is obtained by multiplying C on the left by an invertible matrix, so E must be *the* echelon form of C. The rows of C are a subset of those of A, so the rows of E are a subset of the rows of the echelon form of A. Thus $\ker(A) \subset \ker(E)$. To show that E equals the echelon form of A, we just need to verify that $\ker(E) \subset \ker(A)$, i.e., that $AD = 0$, where D is as in step (5). Since E is the echelon form of C, we know that $CD = 0$. By step (5) we also know that $FD = 0$. Thus $AD = 0$, since the rows of A are the union of the rows of F and C.

□

Example 7.9. Let A be the 4×8 matrix

$$A = \begin{pmatrix} -9 & 6 & 7 & 3 & 1 & 0 & 0 & 0 \\ -10 & 3 & 8 & 2 & 0 & 1 & 0 & 0 \\ 3 & -6 & 2 & 8 & 0 & 0 & 1 & 0 \\ -8 & -6 & -8 & 6 & 0 & 0 & 0 & 1 \end{pmatrix}$$

from Example 7.2.

```
sage: M = MatrixSpace(QQ,4,8)
sage: A = M([[-9,6,7,3,1,0,0,0],[-10,3,8,2,0,1,0,0],
              [3,-6,2,8,0,0,1,0],[-8,-6,-8,6,0,0,0,1]])
```

First choose the "random" prime $p = 41$, which does not divide any of the entries of A, and compute the echelon form of the reduction of A modulo 41.

```
sage: A41 = MatrixSpace(GF(41),4,8)(A)
sage: E41 = A41.echelon_form()
```

The echelon form of $A \pmod{41}$ is

$$\begin{pmatrix} 1 & 0 & 0 & 2 & 0 & 20 & 33 & 18 \\ 0 & 1 & 0 & 40 & 0 & 30 & 7 & 1 \\ 0 & 0 & 1 & 39 & 0 & 19 & 13 & 17 \\ 0 & 0 & 0 & 0 & 1 & 31 & 0 & 37 \end{pmatrix}.$$

Thus we take $c_0 = 0$, $c_1 = 1$, $c_2 = 2$, and $c_3 = 4$. Also $r_i = i$ for $i = 0, 1, 2, 3$. Next extract the submatrix B.

```
sage: B = A.matrix_from_columns([0,1,2,4])
```

The submatrix B is

$$B = \begin{pmatrix} -9 & 6 & 7 & 1 \\ -10 & 3 & 8 & 0 \\ 3 & -6 & 2 & 0 \\ -8 & -6 & -8 & 0 \end{pmatrix}.$$

The inverse of B is

$$B^{-1} = \begin{pmatrix} 0 & -\frac{5}{92} & \frac{1}{46} & -\frac{9}{184} \\ 0 & -\frac{1}{138} & -\frac{3}{23} & -\frac{11}{276} \\ 0 & \frac{11}{184} & \frac{7}{92} & -\frac{17}{368} \\ 1 & -\frac{159}{184} & \frac{41}{92} & \frac{45}{368} \end{pmatrix}.$$

Multiplying by A yields

$$E = B^{-1}A = \begin{pmatrix} 1 & 0 & 0 & -\frac{21}{92} & 0 & -\frac{5}{92} & \frac{1}{46} & -\frac{9}{184} \\ 0 & 1 & 0 & -\frac{179}{138} & 0 & -\frac{1}{138} & -\frac{3}{23} & -\frac{11}{276} \\ 0 & 0 & 1 & \frac{83}{184} & 0 & \frac{11}{184} & \frac{7}{92} & -\frac{17}{368} \\ 0 & 0 & 0 & \frac{1025}{184} & 1 & -\frac{159}{184} & \frac{41}{92} & \frac{45}{368} \end{pmatrix}.$$

```
sage: E = B^(-1)*A
```

This is *not* the echelon form of A. Indeed, it is not even in echelon form, since the last row is not normalized so the leftmost nonzero entry is 1. We thus choose another random prime, say $p = 43$. The echelon form mod 43 has columns $0, 1, 2, 3$ as pivot columns. We thus extract the matrix

$$B = \begin{pmatrix} -9 & 6 & 7 & 3 \\ -10 & 3 & 8 & 2 \\ 3 & -6 & 2 & 8 \\ -8 & -6 & -8 & 6 \end{pmatrix}.$$

```
sage: B = A.matrix_from_columns([0,1,2,3])
```

This matrix has inverse

$$B^{-1} = \begin{pmatrix} \frac{42}{1025} & -\frac{92}{1025} & \frac{1}{25} & -\frac{9}{205} \\ \frac{716}{3075} & -\frac{641}{3075} & -\frac{2}{75} & -\frac{7}{615} \\ -\frac{83}{1025} & \frac{133}{1025} & \frac{1}{25} & -\frac{23}{410} \\ \frac{184}{1025} & -\frac{159}{1025} & \frac{2}{25} & \frac{9}{410} \end{pmatrix}.$$

Finally, the echelon form of A is $E = B^{-1}A$. No further checking is needed since the product so obtained is in echelon form, and the matrix F of the last step of the algorithm has 0 rows.

Remark 7.10. Above we have given only the barest sketch of asymptotically fast "block" algorithms for linear algebra. For optimized algorithms that work in the general case, please see the source code of [**Ste06**].

7.5. Decomposing Spaces under the Action of Matrix

Efficiently solving the following problem is a crucial step in computing a basis of eigenforms for a space of modular forms (see Sections 3.7 and 9.3.2).

Problem 7.11. Suppose T is an $n \times n$ matrix with entries in a field K (typically a number field or finite field) and that the minimal polynomial of T is square-free and has degree n. View T as acting on $V = K^n$. Find a simple module decomposition $W_0 \oplus \cdots \oplus W_m$ of V as a direct sum of simple $K[T]$-modules. Equivalently, find an invertible matrix A such that $A^{-1}TA$ is a block direct sum of matrices $T_0, \ldots, T_m$ such that the minimal polynomial of each T_i is irreducible.

Remark 7.12. A generalization of Problem 7.11 to arbitrary matrices with entries in $\mathbb{Q}$ is finding the *rational Jordan form* (or rational canonical form, or Frobenius form) of T. This is like the usual Jordan form, but the resulting matrix is over $\mathbb{Q}$ and the summands of the matrix corresponding to eigenvalues are replaced by matrices whose minimal polynomials are the minimal polynomials (over $\mathbb{Q}$) of the eigenvalues. The rational Jordan form was extensively studied by Giesbrecht in his Ph.D. thesis and many successive papers, where he analyzes the complexity of his algorithms and observes that the limiting step is factoring polynomials over K. The reason is that given a polynomial $f \in K[x]$, one can easily write down a matrix T such that one can read off the factorization of f from the rational Jordan form of T (see also [**Ste97**]).

7.5.1. Characteristic Polynomials.

The computation of characteristic polynomials of matrices is crucial to modular forms computations. There are many approaches to this problems: compute $\det(xI - A)$ symbolically (bad), compute the traces of the powers of A (bad), or compute the Hessenberg form modulo many primes and use CRT (bad; see for [**Coh93**, §2.2.4] the definition of Hessenberg form and the algorithm). A more sophisticated method is to compute the rational canonical form of A using Giesbrecht's algorithm[1] (see [**GS02**]), which involves computing Krylov subspaces (see Remark 7.13 below), and building up the whole space on which A acts. This

[1] Allan Steel also invented a similar algorithm.

latter method is a generalization of Wiedemann's algorithm for computing minimal polynomials (see Section 7.5.3), but with more structure to handle the case when the characteristic polynomial is not equal to the minimal polynomial.

7.5.2. Polynomial Factorization. Factorization of polynomials in $\mathbb{Q}[X]$ (or over number fields) is an important step in computing an explicit basis of Hecke eigenforms for spaces of modular forms. The best algorithm is the van Hoeij method [**BHKS06**], which uses the LLL lattice basis reduction algorithm [**LLL82**] in a novel way to solve the optimization problems that come up in trying to lift factorizations mod p to $\mathbb{Z}$. It has been generalized by Belebas, van Hoeij, Klüners, and Steel to number fields.

7.5.3. Wiedemann's Minimal Polynomial Algorithm. In this section we describe an algorithm due to Wiedemann for computing the minimal polynomial of an $n \times n$ matrix A over a field.

Choose a random vector v and compute the iterates

$$v_0 = v, \quad v_1 = A(v), \quad v_2 = A^2(v), \quad \ldots, \quad v_{2n-1} = A^{2n-1}(v). \tag{7.5.1}$$

If $f = x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0$ is the minimal polynomial of A, then

$$A^m + c_{m-1}A^{m-1} + \cdots + c_0 I_n = 0,$$

where I_n is the $n \times n$ identity matrix. For any $k \geq 0$, by multiplying both sides on the right by the vector $A^k v$, we see that

$$A^{m+k}v + c_{m-1}A^{m-1+k}v + \cdots + c_0 A^k v = 0;$$

hence

$$v_{m+k} + c_{m-1}v_{m-1+k} + \cdots + c_0 v_k = 0, \qquad \text{all } k \geq 0.$$

Wiedemann's idea is to observe that any single component of the vectors $v_0, \ldots, v_{2n-1}$ satisfies the linear recurrence with coefficients $1, c_{m-1}, \ldots, c_0$. The Berlekamp-Massey algorithm (see Algorithm 7.14 below) was introduced in the 1960s in the context of coding theory to find the minimal polynomial of a linear recurrence sequence $\{a_r\}$. The minimal polynomial of this linear recurrence is by definition the unique monic polynomial g, such that if $\{a_r\}$ satisfies a linear recurrence $a_{j+k} + b_{j-1}a_{j-1+k} + \cdots + b_0 a_k = 0$ (for all $k \geq 0$), then g divides the polynomial $x^j + \sum_{i=0}^{j-1} b_i x^i$. If we apply Berlekamp-Massey to the top coordinates of the v_i, we obtain a polynomial g_0, which divides f. We then apply it to the second to the top coordinates and find a polynomial g_1 that divides f, etc. Taking the least common multiple of the first few g_i, we find a divisor of the minimal polynomial of f. One can show that with "high probability" one quickly finds f, instead of just a proper divisor of f.

Remark 7.13. In the literature, techniques that involve iterating a vector as in (7.5.1) are often called *Krylov methods*. The subspace generated by the iterates of a vector under a matrix is called a *Krylov subspace*.

Algorithm 7.14 (Berlekamp-Massey)**.** *Suppose $a_0, \ldots, a_{2n-1}$ are the first $2n$ terms of a sequence that satisfies a linear recurrence of degree at most n. This algorithm computes the minimal polynomial f of the sequence.*

```
(1) Let $R_0 = x^{2n}$, $R_1 = \sum_{i=0}^{2n-1} a_i x^i$, $V_0 = 0$, $V_1 = 1$.
(2) While $\deg(R_1) \geq n$, do the following:
    (a) Compute $Q$ and $R$ such that $R_0 = QR_1 + R$.
    (b) Let $(V_0, V_1, R_0, R_1) = (V_1, V_0 - QV_1, R_1, R)$.
(3) Let $d = \max(\deg(V_1), 1 + \deg(R_1))$ and set $P = x^d V_1(1/x)$.
(4) Let $c$ be the leading coefficient of $P$ and output $f = P/c$.
```

The above description of Berlekamp-Massey is taken from [**ADT04**], which contains some additional ideas for improvements.

Now suppose T is an $n \times n$ matrix as in Problem 7.11. We find the minimal polynomial of T by computing the minimal polynomial of $T \pmod p$ using Wiedemann's algorithm, for many primes p, and using the Chinese Remainder Theorem. (One has to bound the number of primes that must be considered; see, e.g., [**Coh93**].)

One can also compute the characteristic polynomial of T directly from the Hessenberg form of T, which can be computed in $O(n^4)$ field operations, as described in [**Coh93**]. This is simple but slow. Also, the T we consider will often be sparse, and Wiedemann is particularly good when T is sparse.

Example 7.15. We compute the minimal polynomial of

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ -1 & 1/2 & -1 \end{pmatrix}$$

using Wiedemann's algorithm. Let $v = (1,0,0)^t$. Then

$$v = (1,0,0)^t, \quad Av = (3,0,-1)^t, \quad A^2v = (9,-2,-2)^t,$$
$$A^3v = (27,-4,-8)^t, \quad A^4v = (81,-16,-21)^t, \quad A^5v = (243,-42,-68)^t.$$

The linear recurrence sequence coming from the first entries is

$$1, 3, 9, 27, 81, 243.$$

This sequence satisfies the linear recurrence

$$a_{k+1} - 3a_k = 0, \qquad \text{all } k > 0,$$

so its minimal polynomial is $x - 3$. This implies that $x - 3$ divides the minimal polynomial of the matrix A. Next we use the sequence of second coordinates of the iterates of v, which is

$$0, 0, -2, -4, -16, -42.$$

The recurrence that this sequence satisfies is slightly less obvious, so we apply the Berlekamp-Massey algorithm to find it, with $n = 3$.

(1) We have $R_0 = x^6$, $R_1 = -42x^5 - 16x^4 - 4x^3 - 2x^2$, $V_0 = 0, V_1 = 1$.

(2) (a) Dividing R_0 by R_1, we find

$$R_0 = R_1\left(-\frac{1}{42}x + \frac{4}{441}\right) + \left(\frac{22}{441}x^4 - \frac{5}{441}x^3 + \frac{8}{441}x^2\right).$$

(b) The new V_0, V_1, R_0, R_1 are

$$\begin{aligned} V_0 &= 1, \\ V_1 &= \frac{1}{42}x - \frac{4}{441}, \\ R_0 &= -42x^5 - 16x^4 - 4x^3 - 2x^2, \\ R_1 &= \frac{22}{441}x^4 - \frac{5}{441}x^3 + \frac{8}{441}x^2. \end{aligned}$$

Since $\deg(R_1) \geq n = 3$, we do the above three steps again.

(3) We repeat the above three steps.

(a) Dividing R_0 by R_1, we find

$$R_0 = R_1\left(-\frac{9261}{11}x - \frac{123921}{242}\right) + \left(\frac{1323}{242}x^3 + \frac{882}{121}x^2\right).$$

(b) The new V_0, V_1, R_0, R_1 are:

$$\begin{aligned} V_0 &= \frac{1}{42}x - \frac{4}{441}, \\ V_1 &= \frac{441}{22}x^2 + \frac{2205}{484}x + \frac{441}{121}, \\ R_0 &= \frac{22}{441}x^4 - \frac{5}{441}x^3 + \frac{8}{441}x^2, \\ R_1 &= \frac{1323}{242}x^3 + \frac{882}{121}x^2. \end{aligned}$$

(4) We have to repeat the steps yet again:

$$\begin{aligned} V_0 &= \frac{441}{22}x^2 + \frac{2205}{484}x + \frac{441}{121}, \\ V_1 &= -\frac{242}{1323}x^3 + \frac{968}{3969}x^2 + \frac{484}{3969}x - \frac{242}{3969}, \\ R_0 &= \frac{1323}{242}x^3 + \frac{882}{121}x^2, \\ R_1 &= \frac{484}{3969}x^2. \end{aligned}$$

(5) We have $d = 3$, so $P = -\frac{242}{3969}x^3 + \frac{484}{3969}x^2 + \frac{968}{3969}x - \frac{242}{1323}$.

(6) Multiply through by $-3969/242$ and output

$$x^3 - 2x^2 - 4x + 3 = (x-3)(x^2 + x - 1).$$

The minimal polynomial of T_2 is $(x-3)(x^2+x-1)$, since the minimal polynomial has degree at most 3 and is divisible by $(x-3)(x^2+x-1)$.

7.5.4. p-adic Nullspace. We will use the following algorithm of Dixon [**Dix82**] to compute p-adic approximations to solutions of linear equations over $\mathbb{Q}$. Rational reconstruction modulo p^n then allows us to recover the corresponding solutions over $\mathbb{Q}$.

Algorithm 7.16 (p-adic Nullspace)**.** *Given a matrix A with integer entries and nonzero kernel, this algorithm computes a nonzero element of* $\ker(A)$ *using successive p-adic approximations.*

(1) [Prime] Choose a random prime p.

(2) [Echelon] Compute the echelon form of A modulo p.

(3) [Done?] If A has full rank modulo p, it has full rank, so we terminate the algorithm.

(4) [Setup] Let $b_0 = 0$.

(5) [Iterate] For each $m = 0, 1, 2, \ldots, k$, use the echelon form of A modulo p to find a vector y_m with integer entries such that $Ay_m \equiv b_m \pmod{p}$, and then set

$$b_{m+1} = \frac{b_m - Ay_m}{p}.$$

(If $m = 0$, choose $y_m \neq 0$.)

(6) [p-adic Solution] Let $x = y_0 + y_1 p + y_2 p^2 + y_3 p^3 + \cdots + y_k p^k$.

(7) [Lift] Use rational reconstruction (Algorithm 7.4) to find a vector z with rational entries such that $z \equiv x \pmod{p^{k+1}}$, if such a vector exists. If the vector does not exist, increase k or use a different p. Otherwise, output z.

Proof. We verify the case $k = 2$ only. We have $Ay_0 = 0 \pmod{p}$ and $Ay_1 = -\frac{Ay_0}{p} \pmod{p}$. Thus

$$Ay_0 + pAy_1 \equiv Ay_0 + (-Ay_0) \pmod{p^2}.$$

□

7.5.5. Decomposition Using Kernels. We now know enough to give an algorithm to solve Problem 7.11.

Algorithm 7.17 (Decomposition Using Kernels)**.** *Given an $n \times n$ matrix T over a field K as in Problem 7.11, this algorithm computes the decomposition of V as a direct sum of simple $K[T]$ modules.*

```
(1) [Minimal Polynomial] Compute the minimal polynomial $f$ of $T$,
    e.g., using the multimodular Wiedemann algorithm.
(2) [Factorization] Factor $f$ using the algorithm in Section 7.5.2.
(3) [Compute Kernels] For each irreducible factor $g_i$ of $f$, compute the
    following.
    (a) Compute the matrix $A_i = g_i(T)$.
    (b) Compute $W_i = \ker(A_i)$, e.g., using Algorithm 7.16.
(4) [Output Answer] Then $V = \bigoplus W_i$.
```

Remark 7.18. As mentioned in Remark 7.12, if one can compute such decompositions $V = \bigoplus W_i$, then one can easily factor polynomials f; hence the difficulty of polynomial factorization is a lower bound on the complexity of writing V as a direct sum of simples.

7.6. Exercises

7.1 Given a subspace W of k^n, where k is a field and $n \geq 0$ is an integer, give an algorithm to find a matrix A such that $W = \mathrm{Ker}(A)$.

7.2 If $\mathrm{rref}(A)$ denotes the row reduced echelon form of A and p is a prime not dividing any denominator of any entry of A or of $\mathrm{rref}(A)$, is $\mathrm{rref}(A \pmod{p}) = \mathrm{rref}(A) \pmod{p}$?

7.3 Let A be a matrix with entries in $\mathbb{Q}$. Prove that for all but finitely many primes p we have $\mathrm{rref}(A \pmod{p}) = \mathrm{rref}(A) \pmod{p}$.

7.4 Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

(a) Compute the echelon form of A using each of Algorithm 7.3, Algorithm 7.6, and Algorithm 7.8.

(b) Compute the kernel of A.

(c) Find the characteristic polynomial of A using the algorithm of Section 7.5.3.

7.5 The notion of echelon form extends to matrices whose entries come from certain rings other than fields, e.g., Euclidean domains. In the case of matrices over $\mathbb{Z}$ we define a matrix to be in echelon form (or *Hermite normal form*) if it satisfies

- $a_{ij} = 0$, for $i > j$,
- $a_{ii} \geq 0$,
- $a_{ij} < a_{ii}$ for all $j < i$ (unless $a_{ii} = 0$, in which case all $a_{ij} = 0$).

There are algorithms for computing with finitely generated modules over $\mathbb{Z}$ that are analogous to the ones in this chapter for vector spaces, which depend on computation of Hermite forms.

(a) Show that the Hermite form of $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ is $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{pmatrix}$.

(b) Describe an algorithm for transforming an $n \times n$ matrix A with integer entries into Hermite form using row operations and the Euclidean algorithm.

Chapter 8

General Modular Symbols

In this chapter we explain how to generalize the notion of modular symbols given in Chapter 3 to higher weight and more general level. We define Hecke operators on them and their relation to modular forms via the integration pairing. We omit many difficult proofs that modular symbols have certain properties and instead focus on how to compute with modular symbols. For more details see the references given in this section (especially [**Mer94**]) and [**Wie05**].

Modular symbols are a formalism that make it elementary to compute with homology or cohomology related to certain Kuga-Sato varieties (these are $\mathcal{E} \times_X \cdots \times_X \mathcal{E}$, where X is a modular curve and $\mathcal{E}$ is the universal elliptic curve over it). It is not necessary to know anything about these Kuga-Sato varieties in order to compute with modular symbols.

This chapter is about spaces of modular symbols and how to compute with them. It is by far the most important chapter in this book. The algorithms that build on the theory in this chapter are central to all the computations we will do later in the book.

This chapter closely follows Loïc Merel's paper [**Mer94**]. First we define modular symbols of weight $k \geq 2$. Then we define the corresponding Manin symbols and state a theorem of Merel-Shokurov, which gives all relations between Manin symbols. (The proof of the Merel-Shokurov theorem is beyond the scope of this book but is presented nicely in [**Wie05**].) Next we describe how the Hecke operators act on both modular and Manin symbols and how to compute trace and inclusion maps between spaces of modular symbols of different levels.

Not only are modular symbols useful for computation, but they have been used to prove theoretical results about modular forms. For example, certain technical calculations with modular symbols are used in Loïc Merel's proof of the uniform boundedness conjecture for torsion points on elliptic curves over number fields (modular symbols are used to understand linear independence of Hecke operators). Another example is [**Gri05**], which distills hypotheses about Kato's Euler system in K_2 of modular curves to a simple formula involving modular symbols (when the hypotheses are satisfied, one obtains a lower bound on the Shafarevich-Tate group of an elliptic curve).

8.1. Modular Symbols

We recall from Chapter 3 the free abelian group $\mathbb{M}_2$ of modular symbols. We view these as elements of the relative homology of the extended upper half plane $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q})$ relative to the cusps. The group $\mathbb{M}_2$ is the free abelian group on symbols $\{\alpha, \beta\}$ with

$$\alpha, \beta \in \mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$$

modulo the relations

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0,$$

for all $\alpha, \beta, \gamma \in \mathbb{P}^1(\mathbb{Q})$, and all torsion. More precisely,

$$\mathbb{M}_2 = (F/R)/(F/R)_{\mathrm{tor}},$$

where F is the free abelian group on all pairs (α, β) and R is the subgroup generated by all elements of the form $(\alpha, \beta) + (\beta, \gamma) + (\gamma, \alpha)$. Note that $\mathbb{M}_2$ is a huge free abelian group of countable rank.

For any integer $n \geq 0$, let $\mathbb{Z}[X, Y]_n$ be the abelian group of homogeneous polynomials of degree n in two variables X, Y.

Remark 8.1. Note that $\mathbb{Z}[X, Y]_n$ is isomorphic to $\mathrm{Sym}^n(\mathbb{Z} \times \mathbb{Z})$ as a group, but certain natural actions are different. In [**Mer94**], Merel uses the notation $\mathbb{Z}_n[X, Y]$ for what we denote by $\mathbb{Z}[X, Y]_n$.

Now fix an integer $k \geq 2$. Set

$$\mathbb{M}_k = \mathbb{Z}[X, Y]_{k-2} \otimes_{\mathbb{Z}} \mathbb{M}_2,$$

which is a torsion-free abelian group whose elements are sums of expressions of the form $X^i Y^{k-2-i} \otimes \{\alpha, \beta\}$. For example,

$$X^3 \otimes \{0, 1/2\} - 17XY^2 \otimes \{\infty, 1/7\} \in \mathbb{M}_5.$$

Fix a finite index subgroup G of $\mathrm{SL}_2(\mathbb{Z})$. Define a *left action of* G on $\mathbb{Z}[X,Y]_{k-2}$ as follows. If $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ and $P(X,Y) \in \mathbb{Z}[X,Y]_{k-2}$, let

$$(gP)(X,Y) = P(dX - bY, -cX + aY).$$

Note that if we think of $z = (X,Y)$ as a column vector, then

$$(gP)(z) = P(g^{-1}z),$$

since $g^{-1} = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$. The reason for the inverse is so that this is a left action instead of a right action, e.g., if $g, h \in G$, then

$$((gh)P)(z) = P((gh)^{-1}z) = P(h^{-1}g^{-1}z) = (hP)(g^{-1}z) = (g(hP))(z).$$

Recall that we let G act on the left on $\mathbb{M}_2$ by

$$g\{\alpha,\beta\} = \{g(\alpha), g(\beta)\},$$

where G acts via linear fractional transformations, so if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$g(\alpha) = \frac{a\alpha + b}{c\alpha + d}.$$

For example, useful special cases to remember are that if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$g(0) = \frac{b}{d} \qquad \text{and} \qquad g(\infty) = \frac{a}{c}.$$

(Here we view ∞ as $1/0$ in order to describe the action.)

We now combine these two actions to obtain a left action of G on $\mathbb{M}_k$, which is given by

$$g(P \otimes \{\alpha,\beta\}) = (gP) \otimes \{g(\alpha), g(\beta)\}.$$

For example,

$$\begin{aligned}\begin{pmatrix} 1 & 2 \\ -2 & -3 \end{pmatrix} (X^3 \otimes \{0, 1/2\}) &= (-3X - 2Y)^3 \otimes \left\{-\frac{2}{3}, -\frac{5}{8}\right\} \\ &= (-27X^3 - 54X^2Y - 36XY^2 - 8Y^3) \otimes \left\{-\frac{2}{3}, -\frac{5}{8}\right\}.\end{aligned}$$

We will often write $P(X,Y)\{\alpha,\beta\}$ for $P(X,Y) \otimes \{\alpha,\beta\}$.

Definition 8.2 (Modular Symbols)**.** Let $k \geq 2$ be an integer and let G be a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$. The space $\mathbb{M}_k(G)$ of *weight* k *modular symbols for* G is the quotient of $\mathbb{M}_k$ by all relations $gx - x$ for $x \in \mathbb{M}_k$, $g \in G$, and by any torsion.

Note that $\mathbb{M}_k$ is a torsion-free abelian group, and it is a nontrivial fact that $\mathbb{M}_k$ has finite rank. We denote modular symbols for G in exactly the same way we denote elements of $\mathbb{M}_k$; the group G will be clear from context.

The space of *modular symbols over a ring R* is

$$\mathbb{M}_k(G; R) = \mathbb{M}_k(G) \otimes_{\mathbb{Z}} R.$$

8.2. Manin Symbols

Let G be a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and $k \geq 2$ an integer. Just as in Chapter 3 it is possible to compute $\mathbb{M}_k(G)$ using a computer, despite that, as defined above, $\mathbb{M}_k(G)$ is the quotient of one infinitely generated abelian group by another one. This section is about Manin symbols, which are a distinguished subset of $\mathbb{M}_k(G)$ that lead to a finite presentation for $\mathbb{M}_k(G)$. Formulas written in terms of Manin symbols are frequently much easier to compute using a computer than formulas in terms of modular symbols.

Suppose $P \in \mathbb{Z}[X, Y]_{k-2}$ and $g \in \mathrm{SL}_2(\mathbb{Z})$. Then the *Manin symbol* associated to this pair of elements is

$$[P, g] = g(P\{0, \infty\}) \in \mathbb{M}_k(G).$$

Notice that if $Gg = Gh$, then $[P, g] = [P, h]$, since the symbol $g(P\{0, \infty\})$ is invariant by the action of G on the left (by definition, since it is a modular symbol for G). Thus for a right coset Gg it makes sense to write $[P, Gg]$ for the symbol $[P, h]$ for any $h \in Gg$. Since G has finite index in $\mathrm{SL}_2(\mathbb{Z})$, the abelian group generated by Manin symbols is of finite rank, generated by

$$\left\{[X^{k-2-i}Y^i,\, Gg_j] \,:\, i = 0, \ldots, k-2 \quad \text{and} \quad j = 0, \ldots, r\right\},$$

where $g_0, \ldots, g_r$ run through representatives for the right cosets $G\backslash \mathrm{SL}_2(\mathbb{Z})$.

We next show that every modular symbol can be written as a $\mathbb{Z}$-linear combination of Manin symbols, so they generate $\mathbb{M}_k(G)$.

Proposition 8.3. *The Manin symbols generate* $\mathbb{M}_k(G)$.

Proof. The proof if very similar to that of Proposition 3.11 except we introduce an extra twist to deal with the polynomial part. Suppose that we are given a modular symbol $P\{\alpha, \beta\}$ and wish to represent it as a sum of Manin symbols. Because

$$P\{a/b, c/d\} = P\{a/b, 0\} + P\{0, c/d\},$$

it suffices to write $P\{0, a/b\}$ in terms of Manin symbols. Let

$$0 = \frac{p_{-2}}{q_{-2}} = \frac{0}{1},\ \frac{p_{-1}}{q_{-1}} = \frac{1}{0},\ \frac{p_0}{1} = \frac{p_0}{q_0},\ \frac{p_1}{q_1},\ \frac{p_2}{q_2},\ \ldots,\ \frac{p_r}{q_r} = \frac{a}{b}$$

denote the continued fraction convergents of the rational number a/b. Then

$$p_j q_{j-1} - p_{j-1} q_j = (-1)^{j-1} \qquad \text{for } -1 \leq j \leq r.$$

If we let $g_j = \begin{pmatrix} (-1)^{j-1}p_j & p_{j-1} \\ (-1)^{j-1}q_j & q_{j-1} \end{pmatrix}$, then $g_j \in \mathrm{SL}_2(\mathbb{Z})$ and

$$\begin{aligned} P\{0, a/b\} &= P \sum_{j=-1}^{r} \left\{ \frac{p_{j-1}}{q_{j-1}}, \frac{p_j}{q_j} \right\} \\ &= \sum_{j=-1}^{r} g_j((g_j^{-1}P)\{0, \infty\}) \\ &= \sum_{j=-1}^{r} [g_j^{-1}P,\ g_j]. \end{aligned}$$

Since $g_j \in \mathrm{SL}_2(\mathbb{Z})$ and P has integer coefficients, the polynomial $g_j^{-1}P$ also has integer coefficients, so we introduce no denominators. □

Now that we know the Manin symbols generate $\mathbb{M}_k(G)$, we next consider the relations between Manin symbols. Fortunately, the answer is fairly simple (though the proof is not). Let

$$\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \tau = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \qquad J = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Define a *right action of* $\mathrm{SL}_2(\mathbb{Z})$ on Manin symbols as follows. If $h \in \mathrm{SL}_2(\mathbb{Z})$, let

$$[P, g]h = [h^{-1}P, gh].$$

This is a right action because both $P \mapsto h^{-1}P$ and $g \mapsto gh$ are right actions.

Theorem 8.4. *If x is a Manin symbol, then*

$$x + x\sigma = 0, \tag{8.2.1}$$

$$x + x\tau + x\tau^2 = 0, \tag{8.2.2}$$

$$x - xJ = 0. \tag{8.2.3}$$

Moreover, these are all the relations between Manin symbols, in the sense that the space $\mathbb{M}_k(G)$ of modular symbols is isomorphic to the quotient of the free abelian group on the finitely many symbols $[X^iY^{k-2-i}, Gg]$ (for $i = 0, \ldots, k-2$ and $Gg \in G\backslash \mathrm{SL}_2(\mathbb{Z})$) by the above relations and any torsion.

Proof. First we prove that the Manin symbols satisfy the above relations. We follow Merel's proof (see [**Mer94**, §1.2]). Note that

$$\sigma(0) = \sigma^2(\infty) = \infty \qquad \text{and} \qquad \tau(1) = \tau^2(0) = \infty.$$

Writing $x = [P, g]$, we have

$$\begin{aligned}
[P, g] + [P, g]\sigma &= [P, g] + [\sigma^{-1}P, g\sigma] \\
&= g(P\{0, \infty\}) + g\sigma(\sigma^{-1}P\{0, \infty\}) \\
&= (gP)\{g(0), g(\infty)\} + (g\sigma)(\sigma^{-1}P)\{g\sigma(0), g\sigma(\infty)\} \\
&= (gP)\{g(0), g(\infty)\} + (gP)\{g(\infty), g(0)\} \\
&= (gP)\{g(0), g(\infty)\} + \{g(\infty), g(0)\}) \\
&= 0.
\end{aligned}$$

Also,

$$\begin{aligned}
&[P, g] + [P, g]\tau + [P, g]\tau^2 = [P, g] + [\tau^{-1}P, g\tau] + [\tau^{-2}P, g\tau^2] \\
&\quad = g(P\{0, \infty\}) + g\tau(\tau^{-1}P\{0, \infty\}) + g\tau^2(\tau^{-2}P\{0, \infty\}) \\
&\quad = (gP)\{g(0), g(\infty)\} + (gP)\{g\tau(0), g\tau(\infty)\} + (gP)\{g\tau^2(0), \tau^2(\infty)\} \\
&\quad = (gP)\{g(0), g(\infty)\} + (gP)\{g(1), g(0)\}) + (gP)\{g(\infty), g(1)\} \\
&\quad = (gP)\big(\{g(0), g(\infty)\} + \{g(\infty), g(1)\} + \{g(1), g(0)\}\big) \\
&\quad = 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
[P, g] + [P, g]J &= g(P\{0, \infty\}) - gJ(J^{-1}P\{gJ(0), gJ(\infty)\} \\
&= (gP)\{g(0), g(\infty)\} - (gP)\{g(0), g(\infty)\} \\
&= 0,
\end{aligned}$$

where we use that J acts trivially via linear fractional transformations. This proves that the listed relations are all satisfied.

That the listed relations are all relations is more difficult to prove. One approach is to show (as in [**Mer94**, §1.3]) that the quotient of Manin symbols by the above relations and torsion is isomorphic to a space of Shokurov symbols, which is in turn isomorphic to $\mathbb{M}_k(G)$. A much better approach is to apply some results from group cohomology, as in [**Wie05**]. □

If G is a finite index subgroup and we have an algorithm to enumerate the right cosets $G\backslash \mathrm{SL}_2(\mathbb{Z})$ and to decide which coset an arbitrary element of $\mathrm{SL}_2(\mathbb{Z})$ belongs to, then Theorem 8.4 and the algorithms of Chapter 7 yield an algorithm to compute $\mathbb{M}_k(G; \mathbb{Q})$. Note that if $J \in G$, then the relation $x - xJ = 0$ is automatic.

Remark 8.5. The matrices σ and τ *do not commute*, so in computing $\mathbb{M}_k(G; \mathbb{Q})$, one cannot first quotient out by the two-term σ relations, then quotient out only the remaining free generators by the τ relations, and get the right answer in general.

8.2.1. Coset Representatives and Manin Symbols.

The following is analogous to Proposition 3.10:

Proposition 8.6. *The right cosets $\Gamma_1(N)\backslash \mathrm{SL}_2(\mathbf{Z})$ are in bijection with pairs (c, d) where $c, d \in \mathbf{Z}/N\mathbf{Z}$ and $\gcd(c, d, N) = 1$. The coset containing a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ corresponds to (c, d).*

Proof. This proof is copied from [**Cre92**, pg. 203], except in that paper Cremona works with the analogue of $\Gamma_1(N)$ in $\mathrm{PSL}_2(\mathbf{Z})$, so his result is slightly different. Suppose $\gamma_i = \left(\begin{smallmatrix} a_i & b_i \\ c_i & d_i \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$, for $i = 1, 2$. We have

$$\gamma_1\gamma_2^{-1} = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}\begin{pmatrix} d_2 & -b_2 \\ -c_2 & a_2 \end{pmatrix} = \begin{pmatrix} a_1d_2 - b_1c_2 & * \\ c_1d_2 - d_1c_2 & a_2d_1 - b_2c_1 \end{pmatrix},$$

which is in $\Gamma_1(N)$ if and only if

$$c_1d_2 - d_1c_2 \equiv 0 \pmod{N} \tag{8.2.4}$$

and

$$a_2d_1 - b_2c_1 \equiv a_1d_2 - b_1c_2 \equiv 1 \pmod{N}. \tag{8.2.5}$$

Since the γ_i have determinant 1, if $(c_1, d_1) = (c_2, d_2) \pmod{N}$, then the congruences (8.2.4)–(8.2.5) hold. Conversely, if (8.2.4)–(8.2.5) hold, then

$$\begin{aligned} c_2 &\equiv a_2d_1c_2 - b_2c_1c_2 \\ &\equiv a_2d_2c_1 - b_2c_2c_1 \quad \text{since } d_1c_2 \equiv d_2c_1 \pmod{N} \\ &\equiv c_1 \quad \text{since } a_2d_2 - b_2c_2 = 1, \end{aligned}$$

and likewise

$$d_2 \equiv a_2d_1d_2 - b_2c_1d_2 \equiv a_2d_1d_2 - b_2d_1c_2 \equiv d_1 \pmod{N}.$$

□

Thus we may view weight k Manin symbols for $\Gamma_1(N)$ as triples of integers (i, c, d), where $0 \leq i \leq k-2$ and $c, d \in \mathbf{Z}/N\mathbf{Z}$ with $\gcd(c, d, N) = 1$. Here (i, c, d) corresponds to the Manin symbol $[X^iY^{k-2-i}, \left(\begin{smallmatrix} a & b \\ c' & d' \end{smallmatrix}\right)]$, where c' and d' are congruent to $c, d \pmod{N}$. The relations of Theorem 8.4 become

$$(i,\, c, d) + (-1)^i(k-2-i,\, d, -c) = 0,$$

$$\begin{aligned}(i,\, c, d) \quad + \quad &(-1)^{k-2}\sum_{j=0}^{k-2-i}(-1)^j\binom{k-2-i}{j}(j,\, d, -c-d) \\ + \quad &(-1)^{k-2-i}\sum_{j=0}^{i}(-1)^j\binom{i}{j}(k-2-i+j,\, -c-d, c) = 0,\end{aligned}$$

$$(i,\, c, d) - (-1)^{k-2}(i,\, -c, -d) = 0.$$

Recall that Proposition 3.10 gives a similar description for $\Gamma_0(N)$.

8.2.2. Modular Symbols with Character. Suppose $G = \Gamma_1(N)$. Define an action of *diamond-bracket operators* $\langle d \rangle$o, with $\gcd(d, N) = 1$ on Manin symbols as follows:

$$\langle d \rangle([P,(u,v)]) = [P,(du,dv)] .$$

Let

$$\varepsilon : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{Q}(\zeta)^*$$

be a Dirichlet character, where ζ is an nth root of unity and n is the order of ε. Let $\mathbb{M}_k(N,\varepsilon)$ be the quotient of $\mathbb{M}_k(\Gamma_1(N);\mathbb{Z}[\zeta])$ by the relations (given in terms of Manin symbols)

$$\langle d \rangle x - \varepsilon(d)x = 0,$$

for all $x \in \mathbb{M}_k(\Gamma_1(N);\mathbb{Z}[\zeta])$, $d \in (\mathbb{Z}/N\mathbb{Z})^*$, and by any $\mathbb{Z}$-torsion. Thus $\mathbb{M}_k(N,\varepsilon)$ is a $\mathbb{Z}[\varepsilon]$-module that has no torsion when viewed as a $\mathbb{Z}$-module.

8.3. Hecke Operators

Suppose Γ is a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ of level N that contains $\Gamma_1(N)$. Just as for modular forms, there is a commutative *Hecke algebra* $\mathbb{T} = \mathbb{Z}[T_1, T_2, \ldots]$, which is the subring of $\mathrm{End}(\mathbb{M}_k(\Gamma))$ generated by all Hecke operators. Let

$$R_p = \left\{ \begin{pmatrix} 1 & r \\ 0 & p \end{pmatrix} : r = 0, 1, \ldots, p-1 \right\} \cup \left\{ \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

where we omit $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$ if $p \mid N$. Then the *Hecke operator* T_p on $\mathbb{M}_k(\Gamma)$ is given by

$$T_p(x) = \sum_{g \in R_p} gx.$$

Notice when $p \nmid N$ that T_p is defined by summing over $p+1$ matrices that correspond to the $p+1$ subgroups of $\mathbb{Z} \times \mathbb{Z}$ of index p. This is exactly how we defined T_p on modular forms in Definition 2.26.

8.3.1. General Definition of Hecke Operators. Let Γ be a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and suppose

$$\Delta \subset \mathrm{GL}_2(\mathbb{Q})$$

is a set such that $\Gamma\Delta = \Delta\Gamma = \Delta$ and $\Gamma\backslash\Delta$ is finite. For example, $\Delta = \Gamma$ satisfies this condition. Also, if $\Gamma = \Gamma_1(N)$, then for any positive integer n, the set

$$\Delta_n = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}) \,:\, ad - bc = n, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \pmod{N} \right\}$$

also satisfies this condition, as we will now prove.

Lemma 8.7. *We have*

$$\Gamma_1(N) \cdot \Delta_n = \Delta_n \cdot \Gamma_1(N) = \Delta_n$$

and

$$\Delta_n = \bigcup_{a,b} \Gamma_1(N) \cdot \sigma_a \begin{pmatrix} a & b \\ 0 & n/a \end{pmatrix},$$

where $\sigma_a \equiv \left(\begin{smallmatrix} 1/a & 0 \\ 0 & a \end{smallmatrix}\right) \pmod{N}$*, the union is disjoint and* $1 \leq a \leq n$ *with* $a \mid n$*,* $\gcd(a, N) = 1$*, and* $0 \leq b < n/a$*. In particular, the set of cosets* $\Gamma_1(N)\backslash\Delta_n$ *is finite.*

Proof. (This is Lemma 1 of [**Mer94**, §2.3].) If $\gamma \in \Gamma_1(N)$ and $\delta \in \Delta_n$, then

$$\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \cdot \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \pmod{N}.$$

Thus $\Gamma_1(N)\Delta_n \subset \Delta_n$, and since $\Gamma_1(N)$ is a group, $\Gamma_1(N)\Delta_n = \Delta_n$; likewise $\Delta_n\Gamma_1(N) = \Delta_n$.

For the coset decomposition, we first prove the statement for $N = 1$, i.e., for $\Gamma_1(N) = \mathrm{SL}_2(\mathbb{Z})$. If A is an arbitrary element of $\mathrm{Mat}_2(\mathbb{Z})$ with determinant n, then using row operators on the left with determinant 1, i.e., left multiplication by elements of $\mathrm{SL}_2(\mathbb{Z})$, we can transform A into the form $\left(\begin{smallmatrix} a & b \\ 0 & n/a \end{smallmatrix}\right)$, with $1 \leq a \leq n$ and $0 \leq b < n$. (Just imagine applying the Euclidean algorithm to the two entries in the first column of A. Then a is the gcd of the two entries in the first column, and the lower left entry is 0. Next subtract n/a from b until $0 \leq b < n/a$.)

Next suppose N is arbitrary. Let $g_1, \ldots, g_r$ be such that

$$g_1\Gamma_1(N) \cup \cdots \cup g_r\Gamma_1(N) = \mathrm{SL}_2(\mathbb{Z})$$

is a disjoint union. If $A \in \Delta_n$ is arbitrary, then as we showed above, there is some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, so that $\gamma \cdot A = \left(\begin{smallmatrix} a & b \\ 0 & n/a \end{smallmatrix}\right)$, with $1 \leq a \leq n$ and $0 \leq b < n/a$, and $a \mid n$. Write $\gamma = g_i \cdot \alpha$, with $\alpha \in \Gamma_1(N)$. Then

$$\alpha \cdot A = g_i^{-1} \cdot \begin{pmatrix} a & b \\ 0 & n/a \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \pmod{N}.$$

It follows that

$$g_i^{-1} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \cdot \begin{pmatrix} a & b \\ 0 & n/a \end{pmatrix}^{-1} \equiv \begin{pmatrix} 1/a & * \\ 0 & a \end{pmatrix} \pmod{N}.$$

Since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$ and $\gcd(a, N) = 1$, there is $\gamma' \in \Gamma_1(N)$ such that

$$\gamma' g_i^{-1} \equiv \begin{pmatrix} 1/a & 0 \\ 0 & a \end{pmatrix} \pmod{N}.$$

We may then choose $\sigma_a = \gamma' g_i^{-1}$. Thus every $A \in \Delta_n$ is of the form $\gamma\sigma_a \left(\begin{smallmatrix} a & b \\ 0 & n/a \end{smallmatrix}\right)$, with $\gamma \in \Gamma_1(N)$ and a, b suitably bounded. This proves the second claim. $\square$

Let any element $\delta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$ act on the left on modular symbols $\mathbb{M}_k \otimes \mathbb{Q}$ by

$$\delta(P\{\alpha, \beta\}) = P(dX - bY, -cX + aY)\{\delta(\alpha), \delta(\beta)\}.$$

(Until now we had only defined an action of $\mathrm{SL}_2(\mathbb{Z})$ on modular symbols.) For $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$, let

$$\tilde{g} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \det(g) \cdot g^{-1}. \tag{8.3.1}$$

Note that $\tilde{\tilde{g}} = g$. Also, $\delta P(X, Y) = (P \circ \tilde{g})(X, Y)$, where we set

$$\tilde{g}(X, Y) = (dX - bY, -cX + aY).$$

Suppose Γ and Δ are as above. Fix a finite set R of representatives for $\Gamma\backslash\Delta$. Let

$$T_\Delta : \mathbb{M}_k(\Gamma) \to \mathbb{M}_k(\Gamma)$$

be the linear map

$$T_\Delta(x) = \sum_{\delta \in R} \delta x.$$

This map is well defined because if $\gamma \in \Gamma$ and $x \in \mathbb{M}_k(\Gamma)$, then

$$\sum_{\delta \in R} \delta\gamma x = \sum_{\text{certain } \delta'} \gamma\delta' x = \sum_{\text{certain } \delta'} \delta' x = \sum_{\delta \in R} \delta x,$$

where we have used that $\Delta\Gamma = \Gamma\Delta$, and Γ acts trivially on $\mathbb{M}_k(\Gamma)$.

Let $\Gamma = \Gamma_1(N)$ and $\Delta = \Delta_n$. Then the nth Hecke operator T_n is T_{Δ_n}, and by Lemma 8.7,

$$T_n(x) = \sum_{a,b} \sigma_a \begin{pmatrix} a & b \\ 0 & n/a \end{pmatrix} \cdot x,$$

where a, b are as in Lemma 8.7.

Given this definition, we can compute the Hecke operators on $M_k(\Gamma_1(N))$ as follows. Write x as a modular symbol $P\{\alpha, \beta\}$, compute $T_n(x)$ as a modular symbol, and then convert to Manin symbols using continued fractions expansions. This is extremely inefficient; fortunately Loïc Merel (following [**Maz73**]) found a much better way, which we now describe (see [**Mer94**]).

8.3.2. Hecke Operators on Manin Symbols. If S is a subset of $\mathrm{GL}_2(\mathbb{Q})$, let

$$\tilde{S} = \{\tilde{g} : g \in S\},$$

where $\tilde{g}$ is as in (8.3.1). Also, for any ring R and any subset $S \subset \mathrm{Mat}_2(\mathbb{Z})$, let $R[S]$ denote the free R-module with basis the elements of S, so the elements of $R[S]$ are the finite R-linear combinations of the elements of S.

One of the main theorems of [**Mer94**] is that for any Γ, Δ satisfying the condition at the beginning of Section 8.3.1, if we can find $\sum u_M M \in \mathbb{C}[\mathrm{Mat}_2(\mathbb{Z})]$ and a map

$$\phi : \tilde{\Delta}\,\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z})$$

that satisfies certain conditions, then for any Manin symbol $[P, g] \in \mathbb{M}_k(\Gamma)$, we have

$$T_\Delta([P, g]) = \sum_{gM \in \tilde{\Delta}\,\mathrm{SL}_2(\mathbb{Z}) \text{ with } M \in \mathrm{SL}_2(\mathbb{Z})} u_M[\tilde{M} \cdot P,\ \phi(gM)].$$

The paper [**Mer94**] contains many examples of ϕ and $\sum u_M M \in \mathbb{C}[\mathrm{Mat}_2(\mathbb{Z})]$ that satisfy all of the conditions.

When $\Gamma = \Gamma_1(N)$, the complicated list of conditions becomes simpler. Let $\mathrm{Mat}_2(\mathbb{Z})_n$ be the set of 2×2 matrices with determinant n. An element

$$h = \sum u_M[M] \in \mathbb{C}[\mathrm{Mat}_2(\mathbb{Z})_n]$$

satisfies condition C_n if for every $K \in \mathrm{Mat}_2(\mathbb{Z})_n/\,\mathrm{SL}_2(\mathbb{Z})$, we have that

$$\sum_{M \in K} u_M([M\infty] - [M0]) = [\infty] - [0] \in \mathbb{C}[P^1(\mathbb{Q})]. \tag{8.3.2}$$

If h satisfies condition C_n, then for any Manin symbol $[P, g] \in M_k(\Gamma_1(N))$, Merel proves that

$$T_n([P, (u, v)]) = \sum_M u_M[P(aX + bY,\, cX + dY), (u, v)M]. \tag{8.3.3}$$

Here $(u, v) \in (\mathbb{Z}/N\mathbb{Z})^2$ corresponds via Proposition 8.6 to a coset of $\Gamma_1(N)$ in $\mathrm{SL}_2(\mathbb{Z})$, and if $(u', v') = (u, v)M \in (\mathbb{Z}/N\mathbb{Z})^2$ and $\gcd(u', v', N) \neq 1$, then we omit the corresponding summand.

For example, we will now check directly that the element

$$h_2 = \left[\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right] + \left[\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right] + \left[\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}\right] + \left[\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}\right]$$

satisfies condition C_2. We have, as in the proof of Lemma 8.7 (but using elementary column operations), that

$$\begin{aligned}\mathrm{Mat}_2(\mathbb{Z})_2/\operatorname{SL}_2(\mathbb{Z}) &= \left\{\begin{pmatrix} a & 0 \\ b & 2/a\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}) : a = 1, 2 \text{ and } 0 \le b < 2/a\right\} \\ &= \left\{\begin{pmatrix} 1 & 0 \\ 0 & 2\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}),\ \begin{pmatrix} 1 & 0 \\ 1 & 2\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}),\ \begin{pmatrix} 2 & 0 \\ 0 & 1\end{pmatrix}\operatorname{SL}_2(\mathbb{Z})\right\}.\end{aligned}$$

To verify condition C_2, we consider in turn each of the three elements of $\mathrm{Mat}_2(\mathbb{Z})_2/\operatorname{SL}_2(\mathbb{Z})$ and check that (8.3.2) holds. We have that

$$\begin{pmatrix} 1 & 0 \\ 0 & 2\end{pmatrix} \in \begin{pmatrix} 1 & 0 \\ 0 & 2\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}),$$

$$\begin{pmatrix} 2 & 1 \\ 0 & 1\end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 2\end{pmatrix} \in \begin{pmatrix} 1 & 0 \\ 1 & 2\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}),$$

and

$$\begin{pmatrix} 2 & 0 \\ 0 & 1\end{pmatrix} \in \begin{pmatrix} 2 & 0 \\ 0 & 1\end{pmatrix}\operatorname{SL}_2(\mathbb{Z}).$$

Thus if $K = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2\end{smallmatrix}\right)\operatorname{SL}_2(\mathbb{Z})$, the left sum of (8.3.2) is $[\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2\end{smallmatrix}\right)(\infty)] - [\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2\end{smallmatrix}\right)(0)] = [\infty] - [0]$, as required. If $K = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 2\end{smallmatrix}\right)\operatorname{SL}_2(\mathbb{Z})$, then the left side of (8.3.2) is

$$\begin{aligned}&[\left(\begin{smallmatrix} 2 & 1 \\ 0 & 1\end{smallmatrix}\right)(\infty)] - [\left(\begin{smallmatrix} 2 & 1 \\ 0 & 1\end{smallmatrix}\right)(0)] + [\left(\begin{smallmatrix} 1 & 0 \\ 1 & 2\end{smallmatrix}\right)(\infty)] - [\left(\begin{smallmatrix} 1 & 0 \\ 1 & 2\end{smallmatrix}\right)(0)] \\ &\qquad = [\infty] - [1] + [1] - [0] = [\infty] - [0].\end{aligned}$$

Finally, for $K = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1\end{smallmatrix}\right)\operatorname{SL}_2(\mathbb{Z})$ we also have $[\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1\end{smallmatrix}\right)(\infty)] - [\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1\end{smallmatrix}\right)(0)] = [\infty] - [0]$, as required. Thus by (8.3.3) we can compute T_2 on *any* Manin symbol, by summing over the action of the four matrices $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1\end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2\end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 1 \\ 0 & 1\end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 1 & 2\end{smallmatrix}\right)$.

Proposition 8.8 (Merel)**.** *The element*

$$\sum_{\substack{a>b\geq 0\\ d>c\geq 0\\ ad-bc=n}} \left[\begin{pmatrix} a & b \\ c & d\end{pmatrix}\right] \in \mathbb{Z}[\mathrm{Mat}_2(\mathbb{Z})_n]$$

satisfies condition C_n.

Merel's two-page proof of Proposition 8.8 is fairly elementary.

Remark 8.9. In [**Cre97a**, §2.4], Cremona discusses the work of Merel and Mazur on Heilbronn matrices in the special cases $\Gamma = \Gamma_0(N)$ and weight 2. He gives a simple proof that the action of T_p on Manin symbols can be computed by summing the action of some set R_p of matrices of determinant p. He then describes the set R_p and gives an efficient continued fractions algorithm for computing it (but he does not prove that his R_p satisfy Merel's hypothesis).

8.3.3. Remarks on Complexity. Merel gives another family $\mathcal{S}_n$ of matrices that satisfy condition C_n, and he proves that as $n \to \infty$,

$$\#\mathcal{S}_n \sim \frac{12\log(2)}{\pi^2} \cdot \sigma_1(n)\log(n),$$

where $\sigma_1(n)$ is the sum of the divisors of n. Thus for a fixed space $\mathbb{M}_k(\Gamma)$ of modular symbols, one can compute T_n using $O(\sigma_1(n)\log(n))$ arithmetic operations. Note that we have fixed $\mathbb{M}_k(\Gamma)$, so we ignore the linear algebra involved in computation of a presentation; also, adding elements takes a bounded number of field operations when the space is fixed. Thus, using Manin symbols the complexity of computing T_p, for p prime, is $O((p+1)\log(p))$ field operations, which is *exponential* in the number of digits of p.

8.3.4. Basmaji's Trick. There is a trick of Basmaji (see [**Bas96**]) for computing a matrix of T_n on $\mathbb{M}_k(\Gamma)$, when n is very large, and it is more efficient than one might naively expect. Basmaji's trick does not improve the big-oh complexity for a fixed space, but it does improve the complexity by a constant factor of the dimension of $\mathbb{M}_k(\Gamma;\mathbb{Q})$. Suppose we are interested in computing the matrix for T_n for some massive integer n and that $\mathbb{M}_k(\Gamma;\mathbb{Q})$ has large dimension. The trick is as follows. Choose a list

$$x_1 = [P_1, g_1], \ldots, x_r = [P_r, g_r] \in V = \mathbb{M}_k(\Gamma;\mathbb{Q})$$

of Manin symbols such that the map $\Psi : \mathbb{T} \to V^r$ given by

$$t \mapsto (tx_1, \ldots, tx_r)$$

is injective. In practice, it is often possible to do this with r "very small". Also, we emphasize that V^r is a $\mathbb{Q}$-vector space of dimension $r \cdot \dim(V)$.

Next find Hecke operators T_i, with i small, whose images form a basis for the image of Ψ. Now with the above data precomputed, which only required working with Hecke operators T_i for small i, we are ready to compute T_n with n huge. Compute $y_i = T_n(x_i)$, for each $i = 1, \ldots, r$, which we can compute using Heilbronn matrices since each $x_i = [P_i, g_i]$ is a Manin symbol. We thus obtain $\Psi(T_n) \in V^r$. Since we have precomputed Hecke operators T_j such that $\Psi(T_j)$ generate V^r, we can find a_j such that $\sum a_j \Psi(T_j) = \Psi(T_n)$. Then since Ψ is injective, we have $T_n = \sum a_j T_j$, which gives the full matrix of T_n on $M_k(\Gamma;\mathbb{Q})$.

8.4. Cuspidal Modular Symbols

Let $\mathbb{B}$ be the free abelian group on symbols $\{\alpha\}$, for $\alpha \in \mathbb{P}^1(\mathbb{Q})$, and set

$$\mathbb{B}_k = \mathbb{Z}[X,Y]_{k-2} \otimes \mathbb{B}.$$

Define a *left action of* $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{B}_k$ by

$$g(P\{\alpha\}) = (gP)\{g(\alpha)\},$$

for $g \in \mathrm{SL}_2(\mathbb{Z})$. For any finite index subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$, let $\mathbb{B}_k(\Gamma)$ be the quotient of $\mathbb{B}_k$ by the relations $x - gx$ for all $g \in \Gamma$ and by any torsion. Thus $\mathbb{B}_k(\Gamma)$ is a torsion-free abelian group.

The *boundary map* is the map

$$b : \mathbb{M}_k(\Gamma) \to \mathbb{B}_k(\Gamma)$$

given by extending the map

$$b(P\{\alpha, \beta\}) = P\{\beta\} - P\{\alpha\}$$

linearly. The space $\mathbb{S}_k(\Gamma)$ of *cuspidal modular symbols* is the kernel

$$\mathbb{S}_k(\Gamma) = \ker(\mathbb{M}_k(\Gamma) \to \mathbb{B}_k(\Gamma)),$$

so we have an exact sequence

$$0 \to \mathbb{S}_k(\Gamma) \to \mathbb{M}_k(\Gamma) \to \mathbb{B}_k(\Gamma).$$

One can prove that when $k > 2$, this sequence is exact on the right.

Next we give a presentation of $\mathbb{B}_k(\Gamma)$ in terms of “boundary Manin symbols”.

8.4.1. Boundary Manin Symbols.

We give an explicit description of the boundary map in terms of Manin symbols for $\Gamma = \Gamma_1(N)$, then describe an efficient way to compute the boundary map.

Let $\mathcal{R}$ be the equivalence relation on $\Gamma\backslash\mathbb{Q}^2$ given by

$$[\Gamma \left(\begin{smallmatrix} \lambda u \\ \lambda v \end{smallmatrix}\right)] \sim \mathrm{sign}(\lambda)^k [\Gamma \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)],$$

for any $\lambda \in \mathbb{Q}^*$. Denote by $B_k(\Gamma)$ the finite-dimensional $\mathbb{Q}$-vector space with basis the equivalence classes $(\Gamma\backslash\mathbb{Q}^2)/\mathcal{R}$. The following two propositions are proved in [**Mer94**].

Proposition 8.10. *The map*

$$\mu : \mathbb{B}_k(\Gamma) \to B_k(\Gamma), \qquad P\left\{\frac{u}{v}\right\} \mapsto P(u,v)\left[\Gamma\begin{pmatrix} u \\ v \end{pmatrix}\right]$$

is well defined and injective. Here u and v are assumed coprime.

Thus the kernel of $\delta : \mathbb{S}_k(\Gamma) \to \mathbb{B}_k(\Gamma)$ is the same as the kernel of $\mu \circ \delta$.

Proposition 8.11. *Let $P \in V_{k-2}$ and $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. We have*

$$\mu \circ \delta([P, (c,d)]) = P(1,0)[\Gamma \left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right)] - P(0,1)[\Gamma \left(\begin{smallmatrix} b \\ d \end{smallmatrix}\right)].$$

We next describe how to explicitly compute $\mu \circ \delta : \mathbb{M}_k(N, \varepsilon) \to B_k(N, \varepsilon)$ by generalizing the algorithm of [**Cre97a**, §2.2]. To compute the image of $[P, (c,d)]$, with $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we must compute the class of $[\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right)]$ and of $[\left(\begin{smallmatrix} b \\ d \end{smallmatrix}\right)]$. Instead of finding a canonical form for cusps, we use a quick test for equivalence modulo scalars. In the following algorithm, by the ith symbol we mean the ith basis vector for a basis of $B_k(N, \varepsilon)$. This basis is constructed as the algorithm is called successively. We first give the algorithm, and then prove the facts used by the algorithm in testing equivalence.

Algorithm 8.12 (Cusp Representation)**.** *Given a boundary Manin symbol $[(\begin{smallmatrix} u \\ v \end{smallmatrix})]$, this algorithm outputs an integer i and a scalar α such that $[(\begin{smallmatrix} u \\ v \end{smallmatrix})]$ is equivalent to α times the ith symbol found so far. (We call this algorithm repeatedly and maintain a running list of cusps seen so far.)*

(1) Use Proposition 3.21 to check whether or not $[(\begin{smallmatrix} u \\ v \end{smallmatrix})]$ is equivalent, modulo scalars, to any cusp found. If so, return the representative, its index, and the scalar. If not, record $(\begin{smallmatrix} u \\ v \end{smallmatrix})$ in the representative list.

(2) Using Proposition 8.16, check whether or not $[(\begin{smallmatrix} u \\ v \end{smallmatrix})]$ is forced to equal 0 by the relations. If it does not equal 0, return its position in the list and the scalar 1. If it equals 0, return the scalar 0 and the position 1; keep $(\begin{smallmatrix} u \\ v \end{smallmatrix})$ in the list, and record that it is equivalent to 0.

The case considered in Cremona's book [**Cre97a**] only involve the trivial character, so no cusp classes are forced to vanish. Cremona gives the following two criteria for equivalence.

Proposition 8.13 (Cremona)**.** *Consider $(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix})$, $i = 1, 2$, with u_i, v_i integers such that $\gcd(u_i, v_i) = 1$ for each i.*

(1) *There exists $g \in \Gamma_0(N)$ such that $g\,(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}) = (\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix})$ if and only if*

$$s_1 v_2 \equiv s_2 v_1 \pmod{\gcd(v_1 v_2, N)}, \text{ where } s_j \text{ satisfies } u_j s_j \equiv 1 \pmod{v_j}.$$

(2) *There exists $g \in \Gamma_1(N)$ such that $g\,(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}) = (\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix})$ if and only if*

$$v_2 \equiv v_1 \pmod{N} \text{ and } u_2 \equiv u_1 \pmod{\gcd(v_1, N)}.$$

Proof. The first statement is [**Cre97a**, Prop. 2.2.3], and the second is [**Cre92**, Lem. 3.2]. □

Algorithm 8.14 (Explicit Cusp Equivalence)**.** *Suppose $(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix})$ and $(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix})$ are equivalent modulo $\Gamma_0(N)$. This algorithm computes a matrix $g \in \Gamma_0(N)$ such that $g\,(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix}) = (\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix})$.*

(1) Let s_1, s_2, r_1, r_2 be solutions to $s_1 u_1 - r_1 v_1 = 1$ and $s_2 u_2 - r_2 v_2 = 1$.
(2) Find integers x_0 and y_0 such that $x_0 v_1 v_2 + y_0 N = 1$.
(3) Let $x = -x_0(s_1 v_2 - s_2 v_1)/(v_1 v_2, N)$ and $s_1' = s_1 + x v_1$.
(4) Output $g = (\begin{smallmatrix} u_2 & r_2 \\ v_2 & s_2 \end{smallmatrix}) \cdot \left(\begin{smallmatrix} u_1 & r_1 \\ v_1 & s_1' \end{smallmatrix}\right)^{-1}$, which sends $(\begin{smallmatrix} u_1 \\ v_1 \end{smallmatrix})$ to $(\begin{smallmatrix} u_2 \\ v_2 \end{smallmatrix})$.

Proof. See the proof of [**Cre97a**, Prop. 8.13]. □

The ε relations can make the situation more complicated, since it is possible that $\varepsilon(\alpha) \neq \varepsilon(\beta)$ but

$$\varepsilon(\alpha)\left[\begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\gamma_\alpha \begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\gamma_\beta \begin{pmatrix} u \\ v \end{pmatrix}\right] = \varepsilon(\beta)\left[\begin{pmatrix} u \\ v \end{pmatrix}\right].$$

One way out of this difficulty is to construct the cusp classes for $\Gamma_1(N)$, and then quotient out by the additional ε relations using Gaussian elimination. This is far too inefficient to be useful in practice because the number of $\Gamma_1(N)$ cusp classes can be unreasonably large. Instead, we give a quick test to determine whether or not a cusp vanishes modulo the ε-relations.

Lemma 8.15. *Suppose α and α' are integers such that $\gcd(\alpha, \alpha', N) = 1$. Then there exist integers β and β', congruent to α and α' modulo N, respectively, such that $\gcd(\beta, \beta') = 1$.*

Proof. By Exercise 8.2 the map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective. By the Euclidean algorithm, there exist integers x, y and z such that $x\alpha + y\alpha' + zN = 1$. Consider the matrix $\left(\begin{smallmatrix} y & -x \\ \alpha & \alpha' \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ and take β, β' to be the bottom row of a lift of this matrix to $\mathrm{SL}_2(\mathbb{Z})$. □

Proposition 8.16. *Let N be a positive integer and ε a Dirichlet character of modulus N. Suppose $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ is a cusp with u and v coprime. Then $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ vanishes modulo the relations*

$$[\gamma\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)] = \varepsilon(\gamma)\,[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)], \qquad \textit{all } \gamma \in \Gamma_0(N),$$

if and only if there exists $\alpha \in (\mathbb{Z}/N\mathbb{Z})^*$, with $\varepsilon(\alpha) \neq 1$, such that

$$\begin{aligned} v &\equiv \alpha v \pmod{N}, \\ u &\equiv \alpha u \pmod{\gcd(v, N)}. \end{aligned}$$

Proof. First suppose such an α exists. By Lemma 8.15 there exists $\beta, \beta' \in \mathbb{Z}$ lifting α, α^{-1} such that $\gcd(\beta, \beta') = 1$. The cusp $\left(\begin{smallmatrix} \beta u \\ \beta' v \end{smallmatrix}\right)$ has coprime coordinates so, by Proposition 8.13 and our congruence conditions on α, the cusps $\left(\begin{smallmatrix} \beta u \\ \beta' v \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ are equivalent by an element of $\Gamma_1(N)$. This implies that $\left[\left(\begin{smallmatrix} \beta u \\ \beta' v \end{smallmatrix}\right)\right] = [\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)]$. Since $\left[\left(\begin{smallmatrix} \beta u \\ \beta' v \end{smallmatrix}\right)\right] = \varepsilon(\alpha)\,[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)]$ and $\varepsilon(\alpha) \neq 1$, we have $[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)] = 0$.

Conversely, suppose $[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)] = 0$. Because all relations are two-term relations and the $\Gamma_1(N)$-relations identify $\Gamma_1(N)$-orbits, there must exists α and β with

$$\left[\gamma_\alpha \begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\gamma_\beta \begin{pmatrix} u \\ v \end{pmatrix}\right] \qquad \text{and } \varepsilon(\alpha) \neq \varepsilon(\beta).$$

Indeed, if this did not occur, then we could mod out by the ε relations by writing each $[\gamma_\alpha \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)]$ in terms of $[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)]$, and there would be no further relations left to kill $[\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)]$. Next observe that

$$\begin{aligned}\left[\gamma_{\beta^{-1}\alpha}\begin{pmatrix} u \\ v \end{pmatrix}\right] &= \left[\gamma_{\beta^{-1}}\gamma_\alpha\begin{pmatrix} u \\ v \end{pmatrix}\right] \\ &= \varepsilon(\beta^{-1})\left[\gamma_\alpha\begin{pmatrix} u \\ v \end{pmatrix}\right] = \varepsilon(\beta^{-1})\left[\gamma_\beta\begin{pmatrix} u \\ v \end{pmatrix}\right] = \left[\begin{pmatrix} u \\ v \end{pmatrix}\right].\end{aligned}$$

Applying Proposition 8.13 and noting that $\varepsilon(\beta^{-1}\alpha) \neq 1$ shows that $\beta^{-1}\alpha$ satisfies the properties of the "α" in the statement of the proposition. □

We enumerate the possible α appearing in Proposition 8.16 as follows. Let $g = (v, N)$ and list the $\alpha = v \cdot \frac{N}{g} \cdot a + 1$, for $a = 0, \ldots, g-1$, such that $\varepsilon(\alpha) \neq 0$.

8.5. Pairing Modular Symbols and Modular Forms

In this section we define a pairing between modular symbols and modular forms that the Hecke operators respect. We also define an involution on modular symbols and study its relationship with the pairing. This pairing is crucial in much that follows, because it gives rise to period maps from modular symbols to certain complex vector spaces.

Fix an integer weight $k \geq 2$ and a finite index subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$. Let $M_k(\Gamma)$ denote the space of holomorphic modular forms of weight k for Γ, and let $S_k(\Gamma)$ denote its cuspidal subspace. Following [**Mer94**, §1.5], let

$$\overline{S}_k(\Gamma) = \{\overline{f} : f \in S_k(\Gamma)\} \tag{8.5.1}$$

denote the space of *antiholomorphic* cusp forms. Here $\overline{f}$ is the function on $\mathfrak{h}^*$ given by $\overline{f}(z) = \overline{f(z)}$.

Define a pairing

$$(S_k(\Gamma) \oplus \overline{S}_k(\Gamma)) \times \mathbb{M}_k(\Gamma) \to \mathbb{C} \tag{8.5.2}$$

by letting

$$\langle (f_1, f_2), P\{\alpha, \beta\}\rangle = \int_\alpha^\beta f_1(z)P(z,1)dz + \int_\alpha^\beta f_2(z)P(\overline{z},1)d\overline{z}$$

and extending linearly. Here the integral is a complex path integral along a simple path from α to β in $\mathfrak{h}$ (so, e.g., write $z(t) = x(t) + iy(t)$, where $(x(t), y(t))$ traces out the path and consider two real integrals).

Proposition 8.17. *The integration pairing is well defined, i.e., if we replace $P\{\alpha, \beta\}$ by an equivalent modular symbol (equivalent modulo the left action of Γ), then the integral is the same.*

Proof. This follows from the change of variables formulas for integration and the fact that $f_1 \in S_k(\Gamma)$ and $f_2 \in \overline{S}_k(\Gamma)$. For example, if $k = 2$, $g \in \Gamma$ and $f \in S_k(\Gamma)$, then

$$\begin{aligned}\langle f, g\{\alpha, \beta\}\rangle &= \langle f, \{g(\alpha), g(\beta)\}\rangle \\ &= \int_{g(\alpha)}^{g(\beta)} f(z)dz \\ &= \int_{\alpha}^{\beta} f(g(z))dg(z) \\ &= \int_{\alpha}^{\beta} f(z)dz = \langle f, \{\alpha, \beta\}\rangle,\end{aligned}$$

where $f(g(z))dg(z) = f(z)dz$ because f is a weight 2 modular form. For the case of arbitrary weight $k > 2$, see Exercise 8.4. □

The integration pairing is very relevant to the study of special values of L-functions. The L-function of a cusp form $f = \sum a_n q^n \in S_k(\Gamma_1(N))$ is

$$L(f, s) = (2\pi)^s \Gamma(s)^{-1} \int_0^\infty f(it)t^s \frac{dt}{t} \tag{8.5.3}$$

$$= \sum_{n=1}^{\infty} \frac{a_n}{n^s} \qquad \text{for } \mathrm{Re}(s) \gg 0. \tag{8.5.4}$$

The equality of the integral and the Dirichlet series follows by switching the order of summation and integration, which is justified using standard estimates on $|a_n|$ (see, e.g., [**Kna92**, Section VIII.5]).

For each integer j with $1 \leq j \leq k-1$, we have, setting $s = j$ and making the change of variables $t \mapsto -it$ in (8.5.3), that

$$L(f, j) = \frac{(-2\pi i)^j}{(j-1)!} \cdot \left\langle f,\ X^{j-1}Y^{k-2-(j-1)}\{0, \infty\}\right\rangle. \tag{8.5.5}$$

The integers j as above are called *critical integers*. When f is an eigenform, they have deep conjectural significance (see [**BK90, Sch90**]). One can approximate $L(f, j)$ to any desired precision by computing the above pairing explicitly using the method described in Chapter 10. Alternatively, [**Dok04**] contains methods for computing special values of very general L-functions, which can be used for approximating $L(f, s)$ for arbitrary s, in addition to just the critical integers $1, \ldots, k-1$.

Theorem 8.18 (Shokoruv)**.** *The pairing*

$$\langle \cdot\,, \cdot\rangle : (S_k(\Gamma) \oplus \overline{S}_k(\Gamma)) \times \mathbb{S}_k(\Gamma, \mathbb{C}) \to \mathbb{C}$$

is a nondegenerate pairing of complex vector spaces.

Proof. This is [**Sho80b**, Thm. 0.2] and [**Mer94**, §1.5]. □

Corollary 8.19. *We have*

$$\dim_{\mathbb{C}} \mathbb{S}_k(\Gamma, \mathbb{C}) = 2 \dim_{\mathbb{C}} S_2(\Gamma).$$

The pairing is also compatible with Hecke operators. Before proving this, we define an *action of Hecke operators* on $M_k(\Gamma_1(N))$ and on $\overline{S}_k(\Gamma_1(N))$. The definition is similar to the one we gave in Section 2.4 for modular forms of level 1. For a positive integer n, let R_n be a set of coset representatives for $\Gamma_1(N)\backslash\Delta_n$ from Lemma 8.7. For any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$ and $f \in M_k(\Gamma_1(N))$ set

$$f^{[\gamma]_k} = \det(\gamma)^{k-1}(cz+d)^{-k} f(\gamma(z)).$$

Also, for $f \in \overline{S}_k(\Gamma_1(N))$, set

$$f^{[\gamma]'_k} = \det(\gamma)^{k-1}(c\overline{z}+d)^{-k} f(\gamma(z)).$$

Then for $f \in M_k(\Gamma_1(N))$,

$$T_n(f) = \sum_{\gamma \in R_n} f^{[\gamma]_k}$$

and for $f \in \overline{S}_k(\Gamma_1(N))$,

$$T_n(f) = \sum_{\gamma \in R_n} f^{[\gamma]'_k}.$$

This agrees with the definition from Section 2.4 when $N = 1$.

Remark 8.20. If Γ is an arbitrary finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$, then we can define operators T_Δ on $M_k(\Gamma)$ for any Δ with $\Delta\Gamma = \Gamma\Delta = \Delta$ and $\Gamma\backslash\Delta$ finite. For concreteness we do not do the general case here or in the theorem below, but the proof is exactly the same (see [**Mer94**, §1.5]).

Finally we prove the promised Hecke compatibility of the pairing. This proof should convince you that the definition of modular symbols is sensible, in that they are natural objects to integrate against modular forms.

Theorem 8.21. *If*

$$f = (f_1, f_2) \in S_k(\Gamma_1(N)) \oplus \overline{S}_k(\Gamma_1(N))$$

and $x \in \mathbb{M}_k(\Gamma_1(N))$*, then for any* n*,*

$$\langle T_n(f), x\rangle = \langle f, T_n(x)\rangle.$$

Proof. We follow [**Mer94**, §2.1] (but with more details) and will only prove the theorem when $f = f_1 \in S_k(\Gamma_1(N))$, the proof in the general case being the same.

Let $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$, $P \in \mathbb{Z}[X,Y]_{k-2}$, and for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})$, set

$$j(g, z) = cz + d.$$

Let n be any positive integer, and let R_n be a set of coset representatives for $\Gamma_1(N)\backslash\Delta_n$ from Lemma 8.7.

We have

$$\begin{aligned}\langle T_n(f), P\{\alpha,\beta\}\rangle &= \int_\alpha^\beta T_n(f)P(z,1)dz\\ &= \sum_{\delta\in R}\int_\alpha^\beta \det(\delta)^{k-1} f(\delta(z))j(\delta,z)^{-k}P(z,1)dz.\end{aligned}$$

Now for each summand corresponding to the $\delta \in R$, make the change of variables $u = \delta z$. Thus we make $\#R$ change of variables. Also, we will use the notation

$$\tilde{g} = \begin{pmatrix} d & -b\\ -c & a\end{pmatrix} = \det(g)\cdot g^{-1}$$

for $g \in \mathrm{GL}_2(\mathbb{Q})$. We have

$$\langle T_n(f), P\{\alpha,\beta\}\rangle = \\ \sum_{\delta\in R}\int_{\delta(\alpha)}^{\delta(\beta)} \det(\delta)^{k-1} f(u)j(\delta,\delta^{-1}(u))^{-k}P(\delta^{-1}(u),1)d(\delta^{-1}(u)).$$

We have $\delta^{-1}(u) = \tilde{\delta}(u)$, since a linear fractional transformation is unchanged by a nonzero rescaling of a matrix that induces it. Thus by the quotient rule, using that $\tilde{\delta}$ has determinant $\det(\delta)$, we see that

$$d(\delta^{-1}(u)) = d(\tilde{\delta}(u)) = \frac{\det(\delta)du}{j(\tilde{\delta},u)^2}.$$

We next show that

$$j(\delta,\delta^{-1}(u))^{-k}P(\delta^{-1}(u),1) = j(\tilde{\delta},u)^k\det(\delta)^{-k}P(\tilde{\delta}(u),1). \tag{8.5.6}$$

From the definitions, and again using that $\delta^{-1}(u) = \tilde{\delta}(u)$, we see that

$$j(\delta,\delta^{-1}(u)) = \frac{\det(\delta)}{j(\tilde{\delta},u)},$$

which proves that (8.5.6) holds. Thus

$$\langle T_n(f), P\{\alpha,\beta\}\rangle = \\ \sum_{\delta\in R}\int_{\delta(\alpha)}^{\delta(\beta)} \det(\delta)^{k-1} f(u)j(\tilde{\delta},u)^k\det(\delta)^{-k}P(\tilde{\delta}(u),1)\frac{\det(\delta)du}{j(\tilde{\delta},u)^2}.$$

Next we use that

$$(\delta P)(u,1) = j(\tilde{\delta},u)^{k-2}P(\tilde{\delta}(u),1).$$

To see this, note that $P(X,Y) = P(X/Y,1)\cdot Y^{k-2}$. Using this we see that

$$\begin{aligned}(\delta P)(X,Y) &= (P\circ\tilde{\delta})(X,Y)\\ &= P\left(\tilde{\delta}\left(\frac{X}{Y}\right),1\right)\left(-c\cdot\frac{X}{Y}+a\right)^{k-2}\cdot Y^{k-2}.\end{aligned}$$

Now substituting $(u,1)$ for $(X,1)$, we see that

$$(\delta P)(u,1) = P(\tilde{\delta}(u),1)\cdot(-cu+a)^{k-2},$$

as required. Thus finally

$$\begin{aligned}\langle T_n(f), P\{\alpha,\beta\}\rangle &= \sum_{\delta\in R}\int_{\delta(\alpha)}^{\delta(\beta)} f(u)j(\tilde{\delta},u)^{k-2}P(\tilde{\delta}(u),1)du\\ &= \sum_{\delta\in R}\int_{\delta(\alpha)}^{\delta(\beta)} f(u)\cdot((\delta P)(u,1))du\\ &= \langle f,\, T_n(P\{\alpha,\beta\})\rangle.\end{aligned}$$

□

Suppose that Γ is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ such that if $\eta = \left(\begin{smallmatrix}-1&0\\0&1\end{smallmatrix}\right)$, then

$$\eta\Gamma\eta = \Gamma.$$

For example, $\Gamma = \Gamma_1(N)$ satisfies this condition. There is an involution ι^* on $\mathbb{M}_k(\Gamma)$ given by

$$\iota^*(P(X,Y)\{\alpha,\beta\}) = -P(X,-Y)\{-\alpha,-\beta\}, \tag{8.5.7}$$

which we call the *star involution*. On Manin symbols, ι^* is

$$\iota^*[P,(u,v)] = -[P(-X,Y),(-u,v)]. \tag{8.5.8}$$

Let $\mathbb{S}_k(\Gamma)^+$ be the $+1$ eigenspace for ι^* on $\mathbb{S}_k(\Gamma)$, and let $\mathbb{S}_k(\Gamma)^-$ be the -1 eigenspace. There is also a map ι on modular forms, which is adjoint to ι^*.

Remark 8.22. Notice the minus sign in front of $-P(X,-Y)\{-\alpha,-\beta\}$ in (8.5.7). This sign is missing in [**Cre97a**], which is a potential source of confusion (this is not a mistake, but a different choice of convention).

We now state the final result about the pairing, which explains how modular symbols and modular forms are related.

Theorem 8.23. *The integration pairing $\langle\cdot\,,\,\cdot\rangle$ induces nondegenerate Hecke compatible bilinear pairings*

$$\mathbb{S}_k(\Gamma)^+\times S_k(\Gamma)\to\mathbb{C}\qquad\textit{and}\qquad\mathbb{S}_k(\Gamma)^-\times\overline{S}_k(\Gamma)\to\mathbb{C}.$$

Remark 8.24. We make some remarks about computing the boundary map of Section 8.4.1 when working in the ±1 quotient. Let s be a sign, either $+1$ or -1. To compute $\mathbb{S}_k(N,\varepsilon)_s$, it is necessary to replace $B_k(N,\varepsilon)$ by its quotient modulo the additional relations $[\left(\begin{smallmatrix}-u\\v\end{smallmatrix}\right)] = s\,[\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)]$ for all cusps $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$. Algorithm 8.12 can be modified to deal with this situation as follows. Given a cusp $x = \left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$, proceed as in Algorithm 8.12 and check if either $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}-u\\v\end{smallmatrix}\right)$ is equivalent (modulo scalars) to any cusp seen so far. If not, use the following trick to determine whether the ε and s-relations kill the class of $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$: use the unmodified Algorithm 8.12 to compute the scalars α_1, α_2 and indices i_1, i_2 associated to $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}-u\\v\end{smallmatrix}\right)$, respectively. The s-relation kills the class of $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right)$ if and only if $i_1 = i_2$ but $\alpha_1 \neq s\alpha_2$.

8.6. Degeneracy Maps

In this section, we describe natural maps between spaces of modular symbols with character of different levels. We consider spaces with character, since they are so important in applications.

Fix a positive integer N and a Dirichlet character $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$. Let M be a positive divisor of N that is divisible by the conductor of ε, in the sense that ε factors through $(\mathbb{Z}/M\mathbb{Z})^*$ via the natural map $(\mathbb{Z}/N\mathbb{Z})^* \to (\mathbb{Z}/M\mathbb{Z})^*$ composed with some uniquely defined character $\varepsilon' : (\mathbb{Z}/M\mathbb{Z})^* \to \mathbb{C}^*$. For any positive divisor t of N/M, let $T = \left(\begin{smallmatrix}1&0\\0&t\end{smallmatrix}\right)$ and fix a choice $D_t = \{T\gamma_i : i = 1, \ldots, n\}$ of coset representatives for $\Gamma_0(N)\backslash T\Gamma_0(M)$.

Remark 8.25. Note that [**Mer94**, §2.6] contains a typo: The quotient "$\Gamma_1(N)\backslash\Gamma_1(M)T$" should be replaced by "$\Gamma_1(N)\backslash T\Gamma_1(M)$".

Proposition 8.26. *For each divisor t of N/M there are well-defined linear maps*

$$\alpha_t : \mathbb{M}_k(N,\varepsilon) \to \mathbb{M}_k(M,\varepsilon'), \qquad \alpha_t(x) = (tT^{-1})x = \begin{pmatrix}t&0\\0&1\end{pmatrix}x,$$

$$\beta_t : \mathbb{M}_k(M,\varepsilon') \to \mathbb{M}_k(N,\varepsilon), \qquad \beta_t(x) = \sum_{T\gamma_i\in D_t} \varepsilon'(\gamma_i)^{-1}T\gamma_i x.$$

Furthermore, $\alpha_t \circ \beta_t$ is multiplication by $t^{k-2}\cdot[\Gamma_0(M) : \Gamma_0(N)]$.

Proof. To show that α_t is well defined, we must show that for each $x \in \mathbb{M}_k(N,\varepsilon)$ and $\gamma = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$\alpha_t(\gamma x - \varepsilon(\gamma)x) = 0 \in \mathbb{M}_k(M,\varepsilon').$$

We have

$$\alpha_t(\gamma x) = \begin{pmatrix}t&0\\0&1\end{pmatrix}\gamma x = \begin{pmatrix}a&tb\\c/t&d\end{pmatrix}\begin{pmatrix}t&0\\0&1\end{pmatrix}x = \varepsilon'(a)\begin{pmatrix}t&0\\0&1\end{pmatrix}x,$$

so

$$\alpha_t(\gamma x - \varepsilon(\gamma)x) = \varepsilon'(a)\alpha_t(x) - \varepsilon(\gamma)\alpha_t(x) = 0.$$

We next verify that β_t is well defined. Suppose that $x \in \mathbb{M}_k(M, \varepsilon')$ and $\gamma \in \Gamma_0(M)$; then $\varepsilon'(\gamma)^{-1}\gamma x = x$, so

$$\begin{aligned}\beta_t(x) &= \sum_{T\gamma_i \in D_t} \varepsilon'(\gamma_i)^{-1} T\gamma_i \varepsilon'(\gamma)^{-1}\gamma x \\ &= \sum_{T\gamma_i\gamma \in D_t} \varepsilon'(\gamma_i\gamma)^{-1} T\gamma_i\gamma x.\end{aligned}$$

This computation shows that the definition of β_t does not depend on the choice D_t of coset representatives. To finish the proof that β_t is well defined, we must show that, for $\gamma \in \Gamma_0(M)$, we have $\beta_t(\gamma x) = \varepsilon'(\gamma)\beta_t(x)$ so that β_t respects the relations that define $\mathbb{M}_k(M, \varepsilon)$. Using that β_t does not depend on the choice of coset representative, we find that for $\gamma \in \Gamma_0(M)$,

$$\begin{aligned}\beta_t(\gamma x) &= \sum_{T\gamma_i \in D_t} \varepsilon'(\gamma_i)^{-1} T\gamma_i\gamma x \\ &= \sum_{T\gamma_i\gamma^{-1} \in D_t} \varepsilon'(\gamma_i\gamma^{-1})^{-1} T\gamma_i\gamma^{-1}\gamma x \\ &= \varepsilon'(\gamma)\beta_t(x).\end{aligned}$$

To compute $\alpha_t \circ \beta_t$, we use that $\#D_t = [\Gamma_0(N) : \Gamma_0(M)]$:

$$\begin{aligned}\alpha_t(\beta_t(x)) &= \alpha_t\left(\sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1} T\gamma_i x\right) \\ &= \sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1}(tT^{-1})T\gamma_i x \\ &= t^{k-2}\sum_{T\gamma_i} \varepsilon'(\gamma_i)^{-1}\gamma_i x \\ &= t^{k-2}\sum_{T\gamma_i} x \\ &= t^{k-2} \cdot [\Gamma_0(N) : \Gamma_0(M)] \cdot x.\end{aligned}$$

The scalar factor of t^{k-2} appears instead of t, because t is acting on x as an element of $\mathrm{GL}_2(\mathbb{Q})$ and *not* as an an element of $\mathbb{Q}$. □

Definition 8.27 (New and Old Modular Symbols). The space $\mathbb{M}_k(N, \varepsilon)_{\mathrm{new}}$ of *new modular symbols* is the intersection of the kernels of the α_t as t runs through all positive divisors of N/M and M runs through positive divisors of M strictly less than N and divisible by the conductor of ε. The subspace

$\mathbb{M}_k(N,\varepsilon)_{\text{old}}$ of *old modular symbols* is the subspace generated by the images of the β_t where t runs through all positive divisors of N/M and M runs through positive divisors of M strictly less than N and divisible by the conductor of ε. The new and old subspaces of cuspidal modular symbols are the intersections of the above spaces with $\mathbb{S}_k(N,\varepsilon)$.

Example 8.28. The new and old subspaces need not be disjoint, as the following example illustrates! (This contradicts [**Mer94**, pg. 80].) Consider, for example, the case $N = 6$, $k = 2$, and trivial character. The spaces $\mathbb{M}_2(\Gamma_0(2))$ and $\mathbb{M}_2(\Gamma_0(3))$ are each of dimension 1, and each is generated by the modular symbol $\{\infty, 0\}$. The space $\mathbb{M}_2(\Gamma_0(6))$ is of dimension 3 and is generated by the three modular symbols $\{\infty, 0\}$, $\{-1/4, 0\}$, and $\{-1/2, -1/3\}$. The space generated by the two images of $\mathbb{M}_2(\Gamma_0(2))$ under the two degeneracy maps has dimension 2, and likewise for $\mathbb{M}_2(\Gamma_0(3))$. Together these images generate $\mathbb{M}_2(\Gamma_0(6))$, so $\mathbb{M}_2(\Gamma_0(6))$ is equal to its old subspace. However, the new subspace is nontrivial because the two degeneracy maps $\mathbb{M}_2(\Gamma_0(6)) \to \mathbb{M}_2(\Gamma_0(2))$ are equal, as are the two degeneracy maps

$$\mathbb{M}_2(\Gamma_0(6)) \to \mathbb{M}_2(\Gamma_0(3)).$$

In particular, the intersection of the kernels of the degeneracy maps has dimension at least 1 (in fact, it equals 1). We verify some of the above claims using SAGE.

```
sage: M = ModularSymbols(Gamma0(6)); M
Modular Symbols space of dimension 3 for Gamma_0(6)
of weight 2 with sign 0 over Rational Field
sage: M.new_subspace()
Modular Symbols subspace of dimension 1 of Modular
Symbols space of dimension 3 for Gamma_0(6) of weight
2 with sign 0 over Rational Field
sage: M.old_subspace()
Modular Symbols subspace of dimension 3 of Modular
Symbols space of dimension 3 for Gamma_0(6) of weight
2 with sign 0 over Rational Field
```

8.7. Explicitly Computing $\mathbb{M}_k(\Gamma_0(N))$

In this section we explicitly compute $\mathbb{M}_k(\Gamma_0(N))$ for various k and N. We represent Manin symbols for $\Gamma_0(N)$ as triples of integers (i, u, v), where $(u, v) \in \mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, and (i, u, v) corresponds to $[X^iY^{k-2-i}, (u, v)]$ in the usual notation. Also, recall from Proposition 3.10 that (u, v) corresponds to the right coset of $\Gamma_0(N)$ that contains a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $(u, v) \equiv (c, d)$ as elements of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, i.e., up to rescaling by an element of $(\mathbb{Z}/N\mathbb{Z})^*$.

8.7.1. Computing $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. In this section we give an algorithm to compute a canonical representative for each element of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. This algorithm is extremely important because modular symbols implementations use it a huge number of times. A more naive approach would be to store all pairs $(u, v) \in (\mathbb{Z}/N\mathbb{Z})^2$ and a fixed reduced representative, but this wastes a huge amount of memory. For example, if $N = 10^5$, we would store an array of

$$2 \cdot 10^5 \cdot 10^5 = 20 \text{ billion integers.}$$

Another approach to enumerating $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ is described at the end of [**Cre97a**, §2.2]. It uses the fact that is easy to test whether two pairs $(u_0, v_0), (u_1, v_1)$ define the same element of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$; they do if and only if we have equality of cross terms $u_0v_1 = v_0u_1 \pmod N$ (see [**Cre97a**, Prop. 2.2.1]). So we consider the 0-based list of elements

$$(1, 0), (1, 1), \ldots, (1, N - 1), (0, 1) \tag{8.7.1}$$

concatenated with the list of nonequivalent elements (d, a) for $d \mid N$ and $a = 1, \ldots, N - 1$, checking each time we add a new element to our list (of (d, a)) whether we have already seen it.

Given a random pair (u, v), the problem is then to find the index of the element of our list of the equivalent representative in $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. We use the following algorithm, which finds a canonical representative for each element of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. Given an arbitrary (u, v), we first find the canonical equivalent elements (u', v'). If $u' = 1$, then the index is v'. If $u' \neq 1$, we find the corresponding element in an explicit sorted list, e.g., using binary search.

In the following algorithm, $a \pmod N$ denotes the residue of a modulo N that satisfies $0 \leq a < N$. Note that we *never* create and store the list (8.7.1) itself in memory.

Algorithm 8.29 (Reduction in $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ to Canonical Form)**.** *Given u and v and a positive integer N, this algorithm outputs a pair u_0, v_0 such that $(u, v) \equiv (u_0, v_0)$ as elements of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ and $s \in \mathbb{Z}$ such that $(u, v) = (su_0, sv_0) \pmod{\mathbb{Z}/n\mathbb{Z}}$. Moreover, the element (u_0, v_0) does not depend on the class of (u, v), i.e., for any s with $\gcd(N, s) = 1$ the input (su, sv) also outputs (u_0, v_0). If (u, v) is not in $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, this algorithm outputs $(0, 0), 0$.*

(1) [Reduce] Reduce both u and v modulo N.

(2) [Easy $(0, 1)$ Case] If $u = 0$, check that $\gcd(v, N) = 1$. If so, return $s = 1$ and $(0, 1)$; otherwise return 0.

(3) [GCD] Compute $g = \gcd(u, N)$ and $s, t \in \mathbb{Z}$ such that $g = su + tN$.

(4) [Not in P^1?] We have $\gcd(u, v, N) = \gcd(g, v)$, so if $\gcd(g, v) > 1$, then $(u, v) \notin \mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, and we return 0.

(5) [Pseudo-Inverse] Now $g = su+tN$, so we may think of s as "pseudo-inverse" of $u \pmod N$, in the sense that su is as close as possible to being 1 modulo N. Note that since $g \mid u$, changing s modulo N/g does not change $su \pmod N$. We can adjust s modulo N/g so it is coprime to N (by adding multiples of N/g to s). (This is because $1 = su/g + tN/g$, so s is a unit mod N/g, and the map $(\mathbb{Z}/N\mathbb{Z})^* \to (\mathbb{Z}/(N/g)\mathbb{Z})^*$ is surjective, e.g., as we saw in the proof of Algorithm 4.28.)

(6) [Multiply by s] Multiply (u, v) by s, and replace (u, v) by the equivalent element (g, sv) of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$.

(7) [Normalize] Compute the unique pair (g, v') equivalent to (g, v) that minimizes v, as follows:

   (a) [Easy Case] If $g = 1$, this pair is $(1, v)$.

   (b) [Enumerate and Find Best] Otherwise, note that if $1 \neq t \in (\mathbb{Z}/N\mathbb{Z})^*$ and $tg \equiv g \pmod N$, then $(t-1)g \equiv 0 \pmod N$, so $t - 1 = kN/g$ for some k with $1 \leq k \leq g-1$. Then for $t = 1+kN/g$ coprime to N, we have $(gt, vt) = (g, v+kvN/g)$. So we compute all pairs $(g, v+kvN/g)$ and pick out the one that minimizes the least nonnegative residue of vt modulo N.

   (c) [Invert s and Output] The s that we computed in the above steps multiplies the input (u, v) to give the output (u_0, v_0). Thus we invert it, since the scalar we output multiplies (u_0, v_0) to give (u, v).

Remark 8.30. In the above algorithm, there are many gcd's with N so one should create a table of the gcd's of $0, 1, 2, \ldots, N-1$ with N.

Remark 8.31. Another approach is to instead use that

$$\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) \cong \prod_{p \mid N} \mathbb{P}^1(\mathbb{Z}/p^{\nu_p}\mathbb{Z}),$$

where $\nu_p = \mathrm{ord}_p(N)$, and that it is relatively easy to enumerate the elements of $\mathbb{P}^1(\mathbb{Z}/p^n\mathbb{Z})$ for a prime power p^n.

Algorithm 8.32 (List $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$)**.** *Given an integer $N > 1$, this algorithm makes a sorted list of the distinct representatives (c, d) of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ with $c \neq 0, 1$, as output by Algorithm 8.29.*

(1) For each $c = 1, \ldots, N-1$ with $g = \gcd(c, N) > 1$ do the following:

   (a) Use Algorithm 8.29 to compute the canonical representative (u', v') equivalent to $(c, 1)$, and include it in the list.

   (b) If $g = c$, for each $d = 2, \dots, N-1$ with $\gcd(d, N) > 1$ and $\gcd(c, d) = 1$, append the normalized representative of (c, d) to the list.

(2) Sort the list.

(3) Pass through the sorted list and delete any duplicates.

8.8. Explicit Examples

Explicit detailed examples are crucial when implementing modular symbols algorithms from scratch. This section contains a number of such examples.

8.8.1. Examples of Computation of $\mathbb{M}_k(\Gamma_0(N))$. In this section, we compute $\mathbb{M}_k(\Gamma_0(N))$ explicitly in a few cases.

Example 8.33. We compute $V = \mathbb{M}_4(\Gamma_0(1))$. Because $S_k(\Gamma_0(1)) = 0$ and $M_k(\Gamma_0(1)) = \mathbb{C}E_4$, we expect V to have dimension 1 and for each integer n the Hecke operator T_n to have eigenvalue the sum $\sigma_3(n)$ of the cubes of positive divisors of n.

The Manin symbols are

$$x_0 = (0,0,0), \quad x_1 = (1,0,0), \quad x_2 = (2,0,0).$$

The relation matrix is

$$\left(\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 0 & 0 \\ \hline 2 & -2 & 2 \\ 1 & -1 & 1 \\ 2 & -2 & 2 \end{array}\right),$$

where the first two rows correspond to S-relations and the second three to T-relations. Note that we do not include all S-relations, since it is obvious that some are redundant, e.g., $x + xS = 0$ and $(xS) + (xS)S = xS + x = 0$ are the same since S has order 2.

The echelon form of the relation matrix is

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

where we have deleted the zero rows from the bottom. Thus we may replace the above complicated list of relations with the following simpler list of relations:

$$\begin{aligned} x_0 + x_2 &= 0, \\ x_1 &= 0 \end{aligned}$$

from which we immediately read off that the second generator x_1 is 0 and $x_0 = -x_2$. Thus $\mathbb{M}_4(\Gamma_0(1))$ has dimension 1, with basis the equivalence class of x_2 (or of x_0).

Next we compute the Hecke operator T_2 on $\mathbb{M}_4(\Gamma_0(1))$. The Heilbronn matrices of determinant 2 from Proposition 8.8 are

$$\begin{aligned}
h_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},\\
h_1 &= \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix},\\
h_2 &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},\\
h_3 &= \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

To compute T_2, we apply each of these matrices to x_0, then reduce modulo the relations. We have

$$\begin{aligned}
x_2 \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} &= [X^2,(0,0)] \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} x_2,\\
x_2 \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} &= [X^2,(0,0)] = x_2,\\
x_2 \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} &= [(2X)^2,(0,0)] = 4x_2,\\
x_2 \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} &= [(2X+1)^2,(0,0)] = x_0 + 4x_1 + 4x_2 \sim 3x_2.
\end{aligned}$$

Summing we see that $T_2(x_2) \sim 9x_2$ in $\mathbb{M}_4(\Gamma_0(1))$. Notice that

$$9 = 1^3 + 2^3 = \sigma_3(2).$$

The Heilbronn matrices of determinant 3 from Proposition 8.8 are

$$\begin{aligned}
h_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, & h_1 &= \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix},\\
h_2 &= \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}, & h_3 &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},\\
h_4 &= \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, & h_5 &= \begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix},\\
h_6 &= \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

We have

$$\begin{aligned}
x_2 \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} &= [X^2, (0,0)] \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = x_2, \\
x_2 \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix} &= [X^2, (0,0)] = x_2, \\
x_2 \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} &= [X^2, (0,0)] = x_2, \\
x_2 \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} &= [(2X+1)^2, (0,0)] = x_0 + 4x_1 + 4x_2 \sim 3x_2, \\
x_2 \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} &= [(3X)^2, (0,0)] = 9x_2, \\
x_2 \begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix} &= [(3X+1)^2, (0,0)] = x_0 + 6x_1 + 9x_2 \sim 8x_2, \\
x_2 \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} &= [(3X+2)^2, (0,0)] = 4x_0 + 12x_1 + 9x_2 \sim 5x_2.
\end{aligned}$$

Summing we see that

$$T_3(x_2) \sim x_2 + x_2 + x_2 + 3x_2 + 9x_2 + 8x_2 + 5x_2 = 28x_2.$$

Notice that

$$28 = 1^3 + 3^3 = \sigma_3(3).$$

Example 8.34. Next we compute $\mathbb{M}_2(\Gamma_0(11))$ explicitly. The Manin symbol generators are

$$x_0 = (0,1),\ x_1 = (1,0),\ x_2 = (1,1),\ x_3 = (1,2),\ x_4 = (1,3),\ x_5 = (1,4),$$

$$x_6 = (1,5),\ x_7 = (1,6),\ x_8 = (1,7),\ x_9 = (1,8),\ x_{10} = (1,9),\ x_{11} = (1,10).$$

The relation matrix is as follows, where the S-relations are above the line and the T-relations are below it:

$$\left(\begin{array}{cccccccccccc}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
\hline
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0
\end{array}\right).$$

In weight 2, two out of three T-relations are redundant, so we do not include them. The reduced row echelon form of the relation matrix is

$$\left(\begin{array}{cccccccccccc}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

From the echelon form we see that every symbol is equivalent to a combination of $x_1 = (1,0)$, $x_9 = (1,8)$, and $x_{10} = (1,9)$. (Notice that columns 1, 9, and 10 are the pivot columns, where we index columns starting at 0.)

To compute T_2, we apply each of the Heilbronn matrices of determinant 2 from Proposition 8.8 to x_1, then to x_9, and finally to x_{10}. The matrices are as in Example 8.33 above. We have

$$T_2(x_1) = 3(1,0) + (1,6) \sim 3x_1 - x_{10}.$$

Applying T_2 to $x_9 = (1,8)$, we get

$$T_2(x_9) = (1,3) + (1,4) + (1,5) + (1,10) \sim -2x_9.$$

Applying T_2 to $x_{10} = (1,9)$, we get

$$T_2(x_{10}) = (1,4) + (1,5) + (1,7) + (1,10) \sim -x_1 - 2x_{10}.$$

Thus the matrix of T_2 with respect to this basis is

$$T_2 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ -1 & 0 & -2 \end{pmatrix},$$

where we write the matrix as an operator on the left on vectors written in terms of x_1, x_9, and x_{10}. The matrix T_2 has characteristic polynomial

$$(x-3)(x+2)^2.$$

The $(x-3)$ factor corresponds to the weight 2 Eisenstein series, and the $x+2$ factor corresponds to the elliptic curve $E = X_0(11)$, which has

$$a_2 = -2 = 2 + 1 - \#E(\mathbb{F}_2).$$

Example 8.35. In this example, we compute $\mathbb{M}_6(\Gamma_0(3))$, which illustrates both weight greater than 2 and level greater than 1. We have the following generating Manin symbols:

$$\begin{aligned} x_0 &= [XY^4,(0,1)], & x_1 &= [XY^4,(1,0)],\\ x_2 &= [XY^4,(1,1)], & x_3 &= [XY^4,(1,2)],\\ x_4 &= [XY^3,(0,1)], & x_5 &= [XY^3,(1,0)],\\ x_6 &= [XY^3,(1,1)], & x_7 &= [XY^3,(1,2)],\\ x_8 &= [X^2Y^2,(0,1)], & x_9 &= [X^2Y^2,(1,0)],\\ x_{10} &= [X^2Y^2,(1,1)], & x_{11} &= [X^2Y^2,(1,2)],\\ x_{12} &= [X^3Y,(0,1)], & x_{13} &= [X^3Y,(1,0)],\\ x_{14} &= [X^3Y,(1,1)], & x_{15} &= [X^3Y,(1,2)],\\ x_{16} &= [X^4Y,(0,1)], & x_{17} &= [X^4Y,(1,0)],\\ x_{18} &= [X^4Y,(1,1)], & x_{19} &= [X^4Y,(1,2)]. \end{aligned}$$

The relation matrix is already very large for $\mathbb{M}_6(\Gamma_0(3))$. It is as follows, where the S-relations are before the line and the T-relations after it:

$$\left(\begin{array}{cccccccccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\hline
1 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 1 & 0 & 1\\
1 & 1 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 1\\
0 & 0 & 2 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 2 & 0\\
0 & 1 & 0 & 1 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 1 & 1 & 0 & 0\\
0 & 0 & 0 & 1 & 1 & 0 & 0 & -3 & 0 & 0 & 0 & 3 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 3 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 0 & 0 & -3 & 0 & 1 & 0 & 3 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -2 & 1 & 1 & 0 & 1 & 0 & -2 & 0 & 0 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 0 & 0 & -2 & 0 & 0 & 1 & 1 & 0 & 1 & -2 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 3 & 0 & 0 & 1 & -3 & 0 & 0 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 3 & 0 & 1 & 0 & -3 & 0 & 0 & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 3 & 0 & 0 & 0 & -3 & 0 & 0 & 1 & 1 & 0 & 0 & 0\\
0 & 1 & 0 & 1 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 1 & 1 & 0 & 0\\
1 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 1 & 0 & 1\\
0 & 0 & 2 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 2 & 0\\
1 & 1 & 0 & 0 & -4 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}\right).$$

The reduced row echelon form of the relations matrix, with zero rows removed is

$$\left(\begin{array}{cccccccccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/16 & -3/16 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/16 & 1/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -5/16 & -3/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/2 & 3/16 & 5/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/6 & 1/12 & 1/12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & -1/12 & -1/12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & -1/4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1/4 & 1/4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1/16 & 1/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 3/16 & -3/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1/2 & 3/16 & 5/16 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1/2 & -5/16 & -3/16
\end{array}\right).$$

Since these relations are equivalent to the original relations, we see how $x_0, \ldots, x_{15}$ can be expressed in terms of x_{16}, x_{17}, x_{18}, and x_{19}. Thus $\mathbb{M}_6(\Gamma_0(3))$ has dimension 4. For example,

$$x_{15} \sim \frac{1}{2}x_{17} - \frac{5}{16}x_{18} - \frac{3}{16}x_{19}.$$

Notice that the number of relations is already quite large. It is perhaps surprising how complicated the presentation is already for $\mathbb{M}_6(\Gamma_0(3))$. Because there are denominators in the relations, the above calculation is only a computation of $\mathbb{M}_6(\Gamma_0(3);\mathbb{Q})$. Computing $\mathbb{M}_6(\Gamma_0(3);\mathbb{Z})$ involves finding a $\mathbb{Z}$-basis for the kernel of the relation matrix (see Exercise 7.5).

As before, we find that with respect to the basis x_{16}, x_{17}, x_{18}, and x_{19}

$$T_2 = \begin{pmatrix} 33 & 0 & 0 & 0 \\ 3 & 6 & 12 & 12 \\ -3/2 & 27/2 & 15/2 & 27/2 \\ -3/2 & 27/2 & 27/2 & 15/2 \end{pmatrix}.$$

Notice that there are denominators in the matrix for T_2 with respect to this basis. It is clear from the definition of T_2 acting on Manin symbols that T_2 preserves the $\mathbb{Z}$-module $\mathbb{M}_6(\Gamma_0(3))$, so there is some basis for $\mathbb{M}_6(\Gamma_0(3))$ such that T_2 is given by an integer matrix. Thus the characteristic polynomial f_2 of T_2 will have integer coefficients; indeed,

$$f_2 = (x-33)^2 \cdot (x+6)^2.$$

Note the factor $(x-33)^2$, which comes from the two images of the Eisenstein series E_4 of level 1. The factor $x+6$ comes from the cusp form

$$g = q - 6q^2 + \cdots \in S_6(\Gamma_0(3)).$$

By computing more Hecke operators T_n, we can find more coefficients of g. For example, the charpoly of T_3 is $(x-1)(x-243)(x-9)^2$, and the matrix of T_5 is

$$T_5 = \begin{pmatrix} 3126 & 0 & 0 & 0 \\ 240 & 966 & 960 & 960 \\ -120 & 1080 & 1086 & 1080 \\ -120 & 1080 & 1080 & 1086 \end{pmatrix},$$

which has characteristic polynomial

$$f_5 = (x - 3126)^2(x - 6)^2.$$

The matrix of T_7 is

$$T_7 = \begin{pmatrix} 16808 & 0 & 0 & 0 \\ 1296 & 5144 & 5184 & 5184 \\ -648 & 5832 & 5792 & 5832 \\ -648 & 5832 & 5832 & 5792 \end{pmatrix},$$

with characteristic polynomial

$$f_7 = (x - 16808)^2(x + 40)^2.$$

One can put this information together to deduce that

$$g = q - 6q^2 + 9q^3 + 4q^4 + 6q^5 - 54q^6 - 40q^7 + \cdots .$$

Example 8.36. Consider $\mathbb{M}_2(\Gamma_0(43))$, which has dimension 7. With respect to the symbols

$$x_1 = (1, 0), \quad x_{32} = (1, 31), \quad x_{33} = (1, 32),$$
$$x_{39} = (1, 38), \quad x_{40} = (1, 39), \quad x_{41} = (1, 40), \quad x_{42} = (1, 41),$$

the matrix of T_2 is

$$T_2 = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -2 & -1 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 2 & 1 \\ 0 & 0 & -1 & -1 & -1 & -2 & 0 \\ -1 & 0 & 0 & 1 & 1 & 1 & -1 \end{pmatrix},$$

which has characteristic polynomial

$$(x - 3)(x + 2)^2(x^2 - 2)^2.$$

There is one Eisenstein series and there are three cusp forms with $a_2 = -2$ and $a_2 = \pm\sqrt{2}$.

Example 8.37. To compute $\mathbb{M}_2(\Gamma_0(2004); \mathbb{Q})$, we first make a list of the

$$4032 = (2^2 + 2) \cdot (3 + 1) \cdot (167 + 1)$$

elements $(a, b) \in \mathbb{P}^1(\mathbb{Z}/2004\mathbb{Z})$ using Algorithm 8.29. The list looks like this:

$$(0, 1), (1, 0), (1, 1), (1, 2), \ldots, (668, 1), (668, 3), (668, 5), (1002, 1).$$

For each of the symbols x_i, we consider the S-relations and T-relations. Ignoring the redundant relations, we find 2016 S-relations and 1344 T-relations. It is simple to quotient out by the S-relations, e.g., by identifying x_i with $-x_iS = -x_j$ for some j (or setting $x_i = 0$ if $x_iS = x_i$). Once we have taken the quotient by the S-relations, we take the *image* of all 1344 of the T-relations modulo the S-relations and quotient out by those relations. Because S and T do not commutate, we cannot only quotient out by T-relations $x_i + x_iT + x_iT^2 = 0$ where the x_i are the basis after quotienting out by the S-relations. The relation matrix has rank 3359, so $\mathbb{M}_2(\Gamma_0(2004);\mathbb{Q})$ has dimension 673.

If we instead compute the quotient $\mathbb{M}_2(\Gamma_0(2004);\mathbb{Q})^+$ of $\mathbb{M}_2(\Gamma_0(2004);\mathbb{Q})$ by the subspace of elements $x - \eta^*(x)$, we include relations $x_i + x_iI = 0$, where $I = \left(\begin{smallmatrix}-1 & 0\\ 0 & 1\end{smallmatrix}\right)$. There are now 2016 S-relations, 2024 I-relations, and 1344 T-relations. Again, it is relatively easy to quotient out by the S-relations by identifying x_i and $-x_iS$. We then take the image of all 2024 I-relations modulo the S-relations, and again directly quotient out by the I-relations by identifying $[x_i]$ with $-[x_iI] = -[x_j]$ for some j, where by $[x_i]$ we mean the class of x_i modulo the S-relations. Finally, we quotient out by the 1344 T-relations, which involves sparse Gauss elimination on a matrix with 1344 rows and at most three nonzero entries per row. The dimension of $\mathbb{M}_2(\Gamma_0(2004);\mathbb{Q})^+$ is 331.

8.9. Refined Algorithm for the Presentation

Algorithm 8.38 (Modular Symbols Presentation)**.** *This is an algorithm to compute $\mathbb{M}_k(\Gamma_0(N);\mathbb{Q})$ or $\mathbb{M}_k(\Gamma_0(N);\mathbb{Q})^\pm$, which only requires doing generic sparse linear algebra to deal with the three term T-relations.*

(1) Let $x_0, \ldots, x_n$ by a list of all Manin symbols.

(2) Quotient out the two-term S-relations and if the ± quotient is desired, by the two-term η-relations. (Note that this is more subtle than just "identifying symbols in pairs", since complicated relations can cause generators to surprisingly equal 0.) Let $[x_i]$ denote the class of x_i after this quotienting process.

(3) Create a sparse matrix A with m columns, whose rows encode the relations

$$[x_i] + [x_iT] + [x_iT^2] = 0.$$

For example, there are about $n/3$ such rows when $k = 2$. The number of nonzero entries per row is at most $3(k-1)$. Note that we must include rows for *all* i, since even if $[x_i] = [x_j]$, it need not be the case that $[x_iT] = [x_jT]$, since the matrices S and T do not commute. However, we have an a priori formula for the dimension of the quotient by all these relations, so we could omit many relations and just check that there are enough at the end—if there are not, we add in more.

(4) Compute the reduced row echelon form of A using Algorithm 7.6. For $k = 2$, this is the echelon form of a matrix with size about $n/3 \times n/4$.

(5) Use what we have done above to read off a sparse matrix R that expresses each of the n Manin symbols in terms of a basis of Manin symbols, modulo the relations.

8.10. Applications

8.10.1. Later in This Book. We sketch some of the ways in which we will apply the modular symbols algorithms of this chapter later in this book.

Cuspidal modular symbols are in Hecke-equivariant duality with cuspidal modular forms, and as such we can compute modular forms by computing systems of eigenvalues for the Hecke operators acting on modular symbols. By the Atkin-Lehner-Li theory of newforms (see, e.g., Theorem 9.4), we can construct $S_k(N, \varepsilon)$ for any N, any ε, and $k \geq 2$ using this method. See Chapter 1 for more details.

Once we can compute spaces of modular symbols, we move to computing the corresponding modular forms. We define inclusion and trace maps from modular symbols of one level N to modular symbols of level a multiple or divisor of N. Using these, we compute the quotient V of the new subspace of cuspidal modular symbols on which a "star involution" acts as $+1$. The Hecke operators act by diagonalizable commuting matrices on this space, and computing the systems of Hecke eigenvalues is equivalent to computing newforms $\sum a_n q^n$. In this way, we obtain a list of *all* newforms (normalized eigenforms) in $S_k(N, \varepsilon)$ for any N, ε, and $k \geq 2$.

In Chapter 10, we compute with the period mapping from modular symbols to $\mathbb{C}$ attached to a newform $f \in S_k(N, \varepsilon)$. When $k = 2, \varepsilon = 1$ and f has rational Fourier coefficients, this gives a method to compute the period lattice associated to a modular elliptic curve attached to a newform (see Section 10.7). In general, computation of this map is important when finding equations for modular $\mathbb{Q}$-curves, CM curves, and curves with a given modular Jacobian. It is also important for computing special values of the L-function $L(f, s)$ at integer points in the critical strip.

8.10.2. Discussion of the Literature and Research. Modular symbols were introduced by Birch [**Bir71**] for computations in support of the Birch and Swinnerton-Dyer conjecture. Manin [**Man72**] used modular symbols to prove rationality results about special values of L-functions.

Merel's paper [**Mer94**] builds on work of Shokurov (mainly [**Sho80a**]), which develops a higher-weight generalization of Manin's work partly to understand rationality properties of special values of L-functions. Cremona's book [**Cre97a**] discusses how to compute the space of weight 2 modular symbols for $\Gamma_0(N)$, in connection with the problem of enumerating all elliptic curves of given conductor, and his article [**Cre92**] discusses the $\Gamma_1(N)$ case and computation of modular symbols with character.

There have been several Ph.D. theses about modular symbols. Basmaji's thesis [**Bas96**] contains tricks to efficiently compute Hecke operators T_p, with p very large (see Section 8.3.4), and also discusses how to compute spaces of half integral weight modular forms building on what one can get from modular symbols of integral weight. The author's Ph.D. thesis [**Ste00**] discusses higher-weight modular symbols and applies modular symbols to study Shafarevich-Tate groups (see also [**Aga00**]). Martin's thesis [**Mar01**] is about an attempt to study an analogue of analytic modular symbols for weight 1. Gabor Wiese's thesis [**Wie05**] uses modular symbols methods to study weight 1 modular forms modulo p. Lemelin's thesis [**Lem01**] discusses modular symbols for quadratic imaginary fields in the context of p-adic analogues of the Birch and Swinnerton-Dyer conjecture. See also the survey paper [**FM99**], which discusses computation with weight 2 modular symbols in the context of modular abelian varieties.

The appendix of this book is about analogues of modular symbols for groups besides finite index subgroups of $\mathrm{SL}_2(\mathbb{Z})$, e.g., for subgroup of higher rank groups such as $\mathrm{SL}_3(\mathbb{Z})$. There has also been work on computing Hilbert modular forms, e.g., by Lassina Dembelé [**Dem05**] Hilbert modular forms are functions on a product of copies of $\mathfrak{h}$, and $\mathrm{SL}_2(\mathbb{Z})$ is replaced by a group of matrices with entries in a totally real field.

Glenn Stevens, Robert Pollack and Henri Darmon (see [**DP04**]) have worked for many years to develop an analogue of modular symbols in a rigid analytic context, which is helpful for questions about computing with overconvergent p-adic modular forms or proving results about p-adic L-functions.

Finally we mention that Barry Mazur and some other authors use the term "modular symbol" in a different way than we do. They use the term in a way that is dual to our usage; for example, they attach a "modular symbol" to a modular form or elliptic curve. See [**MTT86**] for an extensive discussion of modular symbols from this point of view, where they are used to construct p-adic L-functions.

8.11. Exercises

8.1 Suppose M is an integer multiple of N. Prove that the natural map $(\mathbb{Z}/M\mathbb{Z})^* \to (\mathbb{Z}/N\mathbb{Z})^*$ is surjective.

8.2 Prove that $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective (see Lemma 8.15).

8.3 Compute $\mathbb{M}_3(\Gamma_1(3))$. List each Manin symbol the relations they satisfy, compute the quotient, etc. Find the matrix of T_2. (Check: The dimension of $\mathbb{M}_3(\Gamma_1(3))$ is 2, and the characteristic polynomial of T_2 is $(x-3)(x+3)$.)

8.4 Finish the proof of Proposition 8.17.

8.5 (a) Show that if $\eta = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, then $\eta\Gamma\eta = \Gamma$ for $\Gamma = \Gamma_0(N)$ and $\Gamma = \Gamma_1(N)$.

(b) (*) Give an example of a finite index subgroup Γ such that $\eta\Gamma\eta \neq \Gamma$.

Chapter 9

Computing with Newforms

In this chapter we pull together results and algorithms from Chapter 3, 4, 7, and 8 and explain how to use linear algebra techniques to compute cusp forms and eigenforms using modular symbols.

We first discuss in Section 9.1 how to decompose $M_k(\Gamma_1(N))$ as a direct sum of subspaces corresponding to Dirichlet characters. Next in Section 9.2 we state the main theorems of Atkin-Lehner-Li theory, which decomposes $S_k(\Gamma_1(N))$ into subspaces on which the Hecke operators act diagonalizably with "multiplicity one". In Section 9.3 we describe two algorithms for computing modular forms. One algorithm finds a basis of q-expansions, and the other computes eigenvalues of newforms.

9.1. Dirichlet Character Decomposition

The group $(\mathbb{Z}/N\mathbb{Z})^*$ acts on $M_k(\Gamma_1(N))$ through *diamond-bracket operators* $\langle d\rangle$, as follows. For $d \in (\mathbb{Z}/N\mathbb{Z})^*$, define

$$f|\langle d\rangle = f^{\left[\left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right)\right]_k},$$

where $\left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ is congruent to $\left(\begin{smallmatrix} d^{-1} & 0 \\ 0 & d \end{smallmatrix}\right) \pmod{N}$. Note that the map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective (see Exercise 8.2), so the matrix $\left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right)$ exists. To prove that $\langle d\rangle$ preserves $M_k(\Gamma_1(N))$, we prove the more general fact that $\Gamma_1(N)$ is a normal subgroup of

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\}.$$

This will imply that $\langle d\rangle$ preserves $M_k(\Gamma_1(N))$ since $\left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right) \in \Gamma_0(N)$.

Lemma 9.1. *The group $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$, and the quotient $\Gamma_0(N)/\Gamma_1(N)$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^*$.*

Proof. See Exercise 9.1. □

Alternatively, one can prove that $\langle d\rangle$ preserves $M_k(\Gamma_1(N))$ by showing that $\langle d\rangle \in \mathbb{T}$ and noting that $M_k(\Gamma_1(N))$ is preserved by $\mathbb{T}$ (see Remark 9.11).

The *diamond-bracket action* is the action of $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^*$ on $M_k(\Gamma_1(N))$. Since $M_k(\Gamma_1(N))$ is a finite-dimensional vector space over $\mathbb{C}$, the $\langle d\rangle$ action breaks $M_k(\Gamma_1(N))$ up as a direct sum of factors corresponding to the Dirichlet characters $D(N,\mathbb{C})$ of modulus N.

Proposition 9.2. *We have*

$$M_k(\Gamma_1(N)) = \bigoplus_{\varepsilon \in D(N,\mathbb{C})} M_k(N,\varepsilon),$$

where

$$M_k(N,\varepsilon) = \left\{ f \in \mathbb{M}_k(\Gamma_1(N)) : f|\langle d\rangle = \varepsilon(d) f, \text{ all } d \in (\mathbb{Z}/N\mathbb{Z})^* \right\}.$$

Proof. The linear transformations $\langle d\rangle$, for the $d \in (\mathbb{Z}/N\mathbb{Z})^*$, all commute, since $\langle d\rangle$ acts through the abelian group $\Gamma_0(N)/\Gamma_1(N)$. Also, if e is the exponent of $(\mathbb{Z}/N\mathbb{Z})^*$, then $\langle d\rangle^e = \langle d^e\rangle = \langle 1\rangle = 1$, so the matrix of $\langle d\rangle$ is diagonalizable. It is a standard fact from linear algebra that any commuting family of diagonalizable linear transformations is simultaneously diagonalizable (see Exercise 5.1), so there is a basis $f_1, \ldots, f_n$ for $M_k(\Gamma_1(N))$ such that all $\langle d\rangle$ act by diagonal matrices. The system of eigenvalues of the action of $(\mathbb{Z}/N\mathbb{Z})^*$ on a fixed f_i defines a Dirichlet character, i.e., each f_i has the property that $f_i|\langle d\rangle = \varepsilon_i(d)$, for all $d \in (\mathbb{Z}/N\mathbb{Z})^*$ and some Dirichlet character ε_i. The f_i for a given ε then span $M_k(N,\varepsilon)$, and taken together the $M_k(N,\varepsilon)$ must span $M_k(\Gamma_1(N))$. □

Definition 9.3 (Character of Modular Form)**.** If $f \in M_k(N,\varepsilon)$, we say that ε is the *character of the modular form f*.

The spaces $M_k(N,\varepsilon)$ are a direct sum of subspaces $S_k(N,\varepsilon)$ and $E_k(N,\varepsilon)$, where $S_k(N,\varepsilon)$ is the subspace of cusp forms, i.e., forms that vanish at *all* cusps (elements of $\mathbb{Q} \cup \{\infty\}$), and $E_k(N,\varepsilon)$ is the subspace of Eisenstein series, which is the unique subspace of $M_k(N,\varepsilon)$ that is invariant under all Hecke operators and is such that $M_k(N,\varepsilon) = S_k(N,\varepsilon) \oplus E_k(N,\varepsilon)$. The space $E_k(N,\varepsilon)$ can also be defined as the space spanned by all Eisenstein series of weight k and level N, as defined in Chapter 5. The space $E_k(N,\varepsilon)$ can be defined in a third way using the Petersson inner product (see [**Lan95**, §VII.5]).

The diamond-bracket operators preserve cusp forms, so the isomorphism of Proposition 9.2 restricts to an isomorphism of the corresponding cuspidal subspaces. We illustrate how to use SAGE to make a table of dimension of $M_k(\Gamma_1(N))$ and $M_k(N, \varepsilon)$ for $N = 13$.

```
sage: G = DirichletGroup(13)
sage: G
Group of Dirichlet characters of modulus 13 over
Cyclotomic Field of order 12 and degree 4
sage: dimension_modular_forms(Gamma1(13),2)
13
sage: [dimension_modular_forms(e,2) for e in G]
[1, 0, 3, 0, 2, 0, 2, 0, 2, 0, 3, 0]
```

Next we do the same for $N = 100$.

```
sage: G = DirichletGroup(100)
sage: G
Group of Dirichlet characters of modulus 100 over
Cyclotomic Field of order 20 and degree 8
sage: dimension_modular_forms(Gamma1(100),2)
370
sage: v = [dimension_modular_forms(e,2) for e in G]; v
[24, 0, 0, 17, 18, 0, 0, 17, 18, 0, 0, 21, 18, 0, 0, 17,
  18, 0, 0, 17, 24, 0, 0, 17, 18, 0, 0, 17, 18, 0, 0, 21,
  18, 0, 0, 17, 18, 0, 0, 17]
sage: sum(v)
370
```

9.2. Atkin-Lehner-Li Theory

In Section 8.6 we defined maps between modular symbols spaces of different level. There are similar maps between spaces of cusp forms. Suppose N and M are positive integers with $M \mid N$ and that t is a divisor of N/M. Let

$$\alpha_t : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)) \tag{9.2.1}$$

be the *degeneracy map*, which is given by $f(q) \mapsto f(q^t)$. There are also maps β_t in the other direction; see [**Lan95**, Ch. VIII].

The *old subspace* of $S_k(\Gamma_1(N))$, denoted $S_k(\Gamma_1(N))_{\mathrm{old}}$, is the sum of the images of all maps α_t with M a proper divisor of N and t any divisor of N/M (note that α_t depends on t, N, and M, so there is a slight abuse of notation). The *new subspace* of $S_k(\Gamma_1(N))$, which we denote by $S_k(\Gamma_1(N))_{\mathrm{new}}$, is the intersection of the kernel of all maps β_t with M a proper divisor of N. One can use the Petersson inner product to show that

$$S_k(\Gamma_1(N)) = S_k(\Gamma_1(N))_{\mathrm{new}} \oplus S_k(\Gamma_1(N))_{\mathrm{old}}.$$

Moreover, the new and old subspaces are preserved by all Hecke operators.

Let $\mathbb{T} = \mathbb{Z}[T_1, T_2, \ldots]$ be the commutative polynomial ring in infinitely many indeterminates T_n. This ring acts (via T_n acting as the nth Hecke operator) on $S_k(\Gamma_1(N))$ for every integer N. Let $\mathbb{T}^{(N)}$ be the subring of $\mathbb{T}$ generated by the T_n with $\gcd(n, N) = 1$.

Theorem 9.4 (Atkin, Lehner, Li)**.** *We have a decomposition*

$$S_k(\Gamma_1(N)) = \bigoplus_{M \mid N} \bigoplus_{d \mid N/M} \alpha_d(S_k(\Gamma_1(M))_{\mathrm{new}}). \tag{9.2.2}$$

Each space $S_k(\Gamma_1(M))_{\mathrm{new}}$ is a direct sum of distinct (nonisomorphic) simple $\mathbb{T}_{\mathbb{C}}^{(N)}$-modules.

Proof. The complete proof is in [**Li75**]. See also [**DS05**, Ch. 5] for a beautiful modern treatment of this and related results. □

The analogue of Theorem 9.4 with Γ_1 replaced by Γ_0 is also true (this is what was proved in [**AL70**]). The analogue for $S_k(N, \varepsilon)$ is also valid, as long as we omit the spaces $S_k(\Gamma_1(M), \varepsilon)$ for which $\mathrm{cond}(\varepsilon) \nmid M$.

Example 9.5. If N is prime and $k \leq 11$, then $S_k(\Gamma_1(N))_{\mathrm{new}} = S_k(\Gamma_1(N))$, since $S_k(\Gamma_1(1)) = 0$.

One can prove using the Petersson inner product that the operators T_n on $S_k(\Gamma_1(N))$, with $\gcd(n, N) = 1$, are diagonalizable. Another result of Atkin-Lehner-Li theory is that the ring of endomorphisms of $S_k(\Gamma_1(N))_{\mathrm{new}}$ generated by all Hecke operators equals the ring generated by the Hecke operators T_n with $(n, N) = 1$. This statement need not be true if we do not restrict to the new subspace, as the following example shows.

Example 9.6. We have

$$S_2(\Gamma_0(22)) = S_2(\Gamma_0(11)) \oplus \alpha_2(S_2(\Gamma_0(11))),$$

where each of the spaces $S_2(\Gamma_0(11))$ has dimension 1. Thus $S_2(\Gamma_0(22))_{\mathrm{new}} = 0$. The Hecke operator T_2 on $S_2(\Gamma_0(22))$ has characteristic polynomial $x^2 + 2x + 2$, which is irreducible. Since α_2 commutes with all Hecke operators T_n, with $\gcd(n, 2) = 1$, the subring $\mathbb{T}^{(22)}$ of the Hecke algebra generated by operators T_n with n odd is isomorphic to $\mathbb{Z}$ (the 2×2 scalar matrices). Thus on the full space $S_2(\Gamma_0(22))$, we do not have $\mathbb{T}^{(22)} = \mathbb{T}$. However, on the new subspace we do have this equality, since the new subspace has dimension 0.

Example 9.7. The space $S_2(\Gamma_0(45))$ has dimension 3 and basis

$$\begin{aligned} f_0 &= q - q^4 - q^{10} - 2q^{13} - q^{16} + 4q^{19} + \cdots , \\ f_1 &= q^2 - q^5 - 3q^8 + 4q^{11} - 2q^{17} + \cdots , \\ f_2 &= q^3 - q^6 - q^9 - q^{12} + q^{15} + q^{18} + \cdots . \end{aligned}$$

The new subspace $S_2(\Gamma_0(45))_{\text{new}}$ is spanned by the single cusp form

$$q + q^2 - q^4 - q^5 - 3q^8 - q^{10} + 4q^{11} - 2q^{13} + \cdots .$$

We have $S_2(\Gamma_0(45/5)) = 0$ and $S_2(\Gamma_0(15))$ has dimension 1 with basis

$$q - q^2 - q^3 - q^4 + q^5 + q^6 + 3q^8 + q^9 - q^{10} - 4q^{11} + q^{12} - 2q^{13} + \cdots .$$

We use SAGE to verify the above assertions.

```
sage: S = CuspForms(Gamma0(45), 2, prec=14); S
Cuspidal subspace of dimension 3 of Modular Forms space
of dimension 10 for Congruence Subgroup Gamma0(45) of
weight 2 over Rational Field
sage: S.basis()
[
q - q^4 - q^10 - 2*q^13 + O(q^14),
q^2 - q^5 - 3*q^8 + 4*q^11 + O(q^14),
q^3 - q^6 - q^9 - q^12 + O(q^14)
]
sage: S.new_subspace().basis()
(q - q^4 - q^10 - 2*q^13 + O(q^14),)
sage: CuspForms(Gamma0(9),2)
Cuspidal subspace of dimension 0 of Modular Forms space
of dimension 3 for Congruence Subgroup Gamma0(9) of
weight 2 over Rational Field
sage: CuspForms(Gamma0(15),2, prec=10).basis()
[
q - q^2 - q^3 - q^4 + q^5 + q^6 + 3*q^8 + q^9 + O(q^10)
]
```

Example 9.8. This example is similar to Example 9.6, except that there are newforms. We have

$$S_2(\Gamma_0(55)) = S_2(\Gamma_0(11)) \oplus \alpha_5(S_2(\Gamma_0(11))) \oplus S_2(\Gamma_0(55))_{\text{new}},$$

where $S_2(\Gamma_0(11))$ has dimension 1 and $S_2(\Gamma_0(55))_{\mathrm{new}}$ has dimension 3. The Hecke operator T_5 on $S_2(\Gamma_0(55))_{\mathrm{new}}$ acts via the matrix

$$\begin{pmatrix} -2 & 2 & -1 \\ -1 & 1 & -1 \\ 1 & -2 & 0 \end{pmatrix}$$

with respect to some basis. This matrix has eigenvalues 1 and -1. Atkin-Lehner theory asserts that T_5 must be a linear combination of T_n, with $\gcd(n, 55) = 1$. Upon computing the matrix for T_2, we find by simple linear algebra that $T_5 = 2T_2 - T_4$.

Definition 9.9 (Newform)**.** A *newform* is a $\mathbb{T}$-eigenform $f \in S_k(\Gamma_1(N))_{\mathrm{new}}$ that is normalized so that the coefficient of q is 1.

We now motivate this definition by explaining why any $\mathbb{T}$-eigenform can be normalized so that the coefficient of q is 1 and how such an eigenform has the property that its Fourier coefficients are exactly the Hecke eigenvalues.

Proposition 9.10. *If $f = \sum_{n=0}^{\infty} a_n q^n \in M_k(N, \varepsilon)$ is an eigenvector for all Hecke operators T_n normalized so that $a_1 = 1$, then $T_n(f) = a_n f$.*

Proof. If $\varepsilon = 1$, then $f \in M_k(\Gamma_0(N))$ and this is Lemma 3.22. However, we have not yet considered Hecke operators on q-expansions for more general spaces of modular forms.

The Hecke operators T_p, for p prime, act on $S_k(N, \varepsilon)$ by

$$T_p\left(\sum_{n=0}^{\infty} a_n q^n\right) = \sum_{n=0}^{\infty}\left(a_{np} q^n + \varepsilon(p) p^{k-1} a_n q^{np}\right),$$

and there is a similar formula for T_m with m composite. If $f = \sum_{n=0}^{\infty} a_n q^n$ is an eigenform for all T_p, with eigenvalues λ_p, then by the above formula

$$(9.2.3) \qquad \lambda_p f = \lambda_p a_1 q + \lambda_p a_2 q^2 + \cdots = T_p(f) = a_p q + \text{higher terms.}$$

Equating coefficients of q, we see that if $a_1 = 0$, then $a_p = 0$ for all p; hence $a_n = 0$ for all n, because of the multiplicativity of Fourier coefficients and the recurrence

$$a_{p^r} = a_{p^{r-1}} a_p - \varepsilon(p) p^{k-1} a_{p^{r-2}}.$$

This would mean that $f = 0$, a contradiction. Thus $a_1 \neq 0$, and it makes sense to normalize f so that $a_1 = 1$. With this normalization, (9.2.3) implies that $\lambda_p = a_p$, as desired. □

Remark 9.11. The Hecke algebra $\mathbb{T}_{\mathbb{Q}}$ on $M_k(\Gamma_1(N))$ contains the operators $\langle d\rangle$, since they satisfy the relation $T_{p^2} = T_p^2 - \langle p\rangle p^{k-1}$. Thus any $\mathbb{T}$-eigenform in $M_k(\Gamma_1(N))$ lies in a subspace $M_k(N, \varepsilon)$ for some Dirichlet character ε. Also, one can even prove that $\langle d\rangle \in \mathbb{Z}[\ldots, T_n, \ldots]$ (see Exercise 9.2).

9.3. Computing Cusp Forms

Let $\mathbb{S}_k(N,\varepsilon;\mathbb{C})$ be the space of cuspidal modular symbols as in Chapter 8. Let ι^* be the map of (8.5.8), and let $\mathbb{S}_k(N,\varepsilon;\mathbb{C})^+$ be the *plus one quotient* of cuspidal modular symbols, i.e., the quotient of $\mathbb{S}_k(N,\varepsilon;\mathbb{C})$ by the image of ι^*-1. It follows from Theorem 8.23 and compatibility of the degeneracy maps (for modular symbols they are defined in Section 8.6) that the $\mathbb{T}$-modules $S_k(N,\varepsilon)_{\text{new}}$ and $\mathbb{S}_k(N,\varepsilon;\mathbb{C})^+_{\text{new}}$ are dual as $\mathbb{T}$-modules. Thus finding the systems of $\mathbb{T}$-eigenvalues on cusp forms is the same as finding the systems of $\mathbb{T}$-eigenvalues on cuspidal modular symbols.

Our strategy to compute $S_k(N,\varepsilon)$ is to first compute spaces $S_k(N,\varepsilon)_{\text{new}}$ using the Atkin-Lehner-Li decomposition (9.2.2). To compute $S_k(N,\varepsilon)_{\text{new}}$ to a given precision, we compute the systems of eigenvalues of the Hecke operators T_p on the space $V=\mathbb{S}_k(N,\varepsilon;\mathbb{C})^+_{\text{new}}$, which we will define below. Using Proposition 9.10, we then recover a basis of q-expansions for newforms. Note that we only need to compute Hecke eigenvalues T_p, for p prime, not the T_n for n composite, since the a_n can be quickly recovered in terms of the a_p using multiplicativity and the recurrence.

For some problems, e.g., construction of models for modular curves, having a basis of q-expansions is enough. For many other problems, e.g., enumeration of modular abelian varieties, one is really interested in the newforms. We next discuss algorithms aimed at each of these problems.

9.3.1. A Basis of q-Expansions.

The following algorithm generalizes Algorithm 3.26. It computes $S_k(N,\varepsilon)$ without finding any eigenspaces.

Algorithm 9.12 (Merel's Algorithm for Computing a Basis)**.** *Given integers m, N and k and a Dirichlet character ε with modulus N, this algorithm computes a basis of q-expansions for $S_k(N,\varepsilon)$ to precision $O(q^{m+1})$.*

(1) [Compute Modular Symbols] Use Algorithm 8.38 to compute

$$V=\mathbb{S}_k(N,\varepsilon)^+\otimes\mathbb{Q}(\varepsilon),$$

viewed as a $K=\mathbb{Q}(\varepsilon)$ vector space, with an action of the T_n.

(2) [Basis for Linear Dual] Write down a basis for $V^*=\text{Hom}(V,\mathbb{Q}(\varepsilon))$. E.g., if we identify V with K^n viewed as column vectors, then V^* is the space of row vectors of length n, and the pairing is the row × column product.

(3) [Find Generator] Find $x\in V$ such that $\mathbb{T}x=V$ by choosing random x until we find one that generates. The set of x that fail to generate lie in a union of a finite number of proper subspaces.

(4) [Compute Basis] The set of power series

$$f_i = \sum_{n=1}^{m} \psi_i(T_n(x))q^n + O(q^{m+1})$$

forms a basis for $S_k(N, \varepsilon)$ to precision m.

In practice Algorithm 9.12 seems slower than the eigenspace algorithm that we will describe in the rest of this chapter. The theoretical complexity of Algorithm 9.12 *may* be better, because it is not necessary to factor any polynomials. Polynomial factorization is difficult from the worst-case complexity point of view, though it is usually fast in practice. The eigenvalue algorithm only requires computing a few images $T_p(x)$ for p prime and x a Manin symbol on which T_p can easily be computed. The Merel algorithm involves computing $T_n(x)$ for all n and for a fairly easy x, which is potentially more work.

Remark 9.13. By "easy x", I mean that computing $T_n(x)$ is easier on x than on a completely random element of $\mathbb{S}_k(N, \varepsilon)^+$, e.g., x could be a Manin symbol.

9.3.2. Newforms: Systems of Eigenvalues. In this section we describe an algorithm for computing the system of Hecke eigenvalues associated to a simple subspace of a space of modular symbols. This algorithm is better than doing linear algebra directly over the number field generated by the eigenvalues. It only involves linear algebra over the base field and also yields a compact representation for the answer, which is better than writing the eigenvalues in terms of a power basis for a number field. In order to use this algorithm, it is necessary to decompose the space of cuspidal modular symbols as a direct sum of simples, e.g., using Algorithm 7.17.

Fix N and a Dirichlet character ε of modulus N, and let

$$V = \mathbb{M}_k(N, \varepsilon)^+$$

be the $+1$ quotient of modular symbols (see equation (8.5.8)).

Algorithm 9.14 (System of Eigenvalues)**.** *Given a $\mathbb{T}$-simple subspace $W \subset V$ of modular symbols, this algorithm outputs maps ψ and e, where $\psi : \mathbb{T}_K \to W$ is a K-linear map and $e : W \cong L$ is an isomorphism of W with a number field L, such that $a_n = e(\psi(T_n))$ is the eigenvalue of the nth Hecke operator acting on a fixed $\mathbb{T}$-eigenvector in $W \otimes \overline{\mathbb{Q}}$. (Thus $f = \sum_{n=1}^{\infty} e(\psi(T_n))q^n$ is a newform.)*

(1) [Compute Projection] Let $\varphi : V \to W'$ be any surjective linear map such that $\ker(\varphi)$ equals the kernel of the $\mathbb{T}$-invariant projection onto W. For example, compute φ by finding a simple submodule

of $V^* = \mathrm{Hom}(V, K)$ that is isomorphic to W, e.g., by applying Algorithm 7.17 to V^* with T replaced by the transpose of T.

(2) [Choose v] Choose a nonzero element $v \in V$ such that $\pi(v) \neq 0$ and computation of $T_n(v)$ is "easy", e.g., choose v to be a Manin symbol.

(3) [Map from Hecke Ring] Let ψ be the map $\mathbb{T} \to W'$, given by $\psi(t) = \pi(tv)$. Note that computation of ψ is relatively easy, because v was chosen so that tv is relatively easy to compute. In particular, if $t = T_p$, we do not need to compute the full matrix of T_p on V; instead we just compute $T_p(v)$.

(4) [Find Generator] Find a random $T \in \mathbb{T}$ such that the iterates

$$\psi(T^0), \quad \psi(T), \quad \psi(T^2), \quad \ldots, \quad \psi(T^{d-1})$$

are a basis for W', where W has dimension d.

(5) [Characteristic Polynomial] Compute the characteristic polynomial f of $T|_W$, and let $L = K[x]/(f)$. Because of how we chose T in step (4), the minimal and characteristic polynomials of $T|_W$ are equal, and both are irreducible, so L is an extension of K of degree $d = \dim(W)$.

(6) [Field Structure] In this step we endow W' with a field structure. Let $e : W' \to L$ be the unique K-linear isomorphism such that

$$e(\psi(T^i)) \equiv x^i \pmod{f}$$

for $i = 0, 1, 2, ..., \deg(f) - 1$. The map e is uniquely determined since the $\psi(T^i)$ are a basis for W'. To compute e, we compute the change of basis matrix from the standard basis for W' to the basis $\{\psi(T^i)\}$. This change of basis matrix is the inverse of the matrix whose rows are the $\psi(T^i)$ for $i = 0, ..., \deg(f) - 1$.

(7) [Hecke Eigenvalues] Finally for each integer $n \geq 1$, we have

$$a_n = e(\psi(T_n)) = e(\pi(T_n(v))),$$

where a_n is the eigenvalue of T_n. Output the maps ψ and e and terminate.

One reason we separate ψ and e is that when $\dim(W)$ is large, the values $\psi(T_n)$ take less space to store and are easier to compute, whereas each one of the values $e(\psi(n))$ is huge.[1] The function e typically involves large numbers if $\dim(W)$ is large, since e is obtained from the iterates of a single vector. For many applications, e.g., databases, it is better to store a matrix that defines e and the images under ψ of many T_n.

[1] John Cremona initially suggested to me the idea of separating these two maps.

Example 9.15. The space $S_2(\Gamma_0(23))$ of cusp forms has dimension 2 and is spanned by two $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugate newforms, one of which is

$$f = q + aq^2 + (-2a-1)q^3 + (-a-1)q^4 + 2aq^5 + \cdots,$$

where $a = (-1+\sqrt{5})/2$. We will use Algorithm 9.14 to compute a few of these coefficients.

The space $\mathbb{M}_2(\Gamma_0(23))^+$ of modular symbols has dimension 3. It has the following basis of Manin symbols:

$$[(0,0)], \quad [(1,0)], \quad [(0,1)],$$

where we use square brackets to differentiate Manin symbols from vectors. The Hecke operator

$$T_2 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ -1 & 1/2 & -1 \end{pmatrix}$$

has characteristic polynomial $(x-3)(x^2+x-1)$. The kernel of $T_2 - 3$ corresponds to the span of the Eisenstein series of level 23 and weight 2, and the kernel V of $T_2^2 + T_2 - 1$ corresponds to $S_2(\Gamma_0(23))$. (We could also have computed V as the kernel of the boundary map $\mathbb{M}_2(\Gamma_0(23))^+ \to \mathbb{B}_2(\Gamma_0(23))^+$.) Each of the following steps corresponds to the step of Algorithm 9.14 with the same number.

(1) [Compute Projection] We compute projection onto V (this will suffice to give us a map ϕ as in the algorithm). The matrix whose first two columns are the echelon basis for V and whose last column is the echelon basis for the Eisenstein subspace is

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -2/11 \\ 0 & 1 & -3/11 \end{pmatrix}$$

and

$$B^{-1} = \begin{pmatrix} 2/11 & 1 & 0 \\ 3/11 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

so projection onto V is given by the first two rows:

$$\pi = \begin{pmatrix} 2/11 & 1 & 0 \\ 3/11 & 0 & 1 \end{pmatrix}.$$

(2) [Choose v] Let $v = (0,1,0)^t$. Notice that $\pi(v) = (1,0)^t \neq 0$, and $v = [(1,0)]$ is a sum of only one Manin symbol.

(3) [Map from Hecke Ring] This step is purely conceptual, since no actual work needs to be done. We illustrate it by computing $\psi(T_1)$ and $\psi(T_2)$. We have

$$\psi(T_1) = \pi(v) = (1,0)^t$$

and

$$\psi(T_2) = \pi(T_2(v)) = \pi((0,0,1/2)^t) = (0,1/2)^t.$$

(4) [Find Generator] We have

$$\psi(T_2^0) = \psi(T_1) = (1,0)^t,$$

which is clearly independent from $\psi(T_2) = (0,1/2)^t$. Thus we find that the image of the powers of $T = T_2$ generate V.

(5) [Characteristic Polynomial] The matrix of $T_2|_V$ is $\begin{pmatrix} 0 & 2 \\ 1/2 & -1 \end{pmatrix}$, which has characteristic polynomial $f = x^2 + x - 1$. Of course, we already knew this because we computed V as the kernel of $T_2^2 + T_2 - 1$.

(6) [Field Structure] We have

$$\psi(T_2^0) = \pi(v) = (1,0)^t \text{ and } \psi(T_2) = (0,1/2).$$

The matrix with rows the $\psi(T_2^i)$ is $\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$, which has inverse $e = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$. The matrix e defines an isomorphism between V and the field

$$L = \mathbb{Q}[x]/(f) = \mathbb{Q}((-1+\sqrt{5})/2).$$

I.e., $e((1,0)) = 1$ and $e((0,1)) = 2x$, where $x = (-1+\sqrt{5})/2$.

(7) [Hecke Eigenvalues] We have $a_n = e(\Psi(T_n))$. For example,

$$\begin{aligned}
a_1 &= e(\Psi(T_1)) = e((1,0)) = 1,\\
a_2 &= e(\Psi(T_2)) = e((0,1/2)) = x,\\
a_3 &= e(\Psi(T_3)) = e(\pi(T_3(v))) = e(\pi((0,-1,-1)^t))\\
&= e((-1,-1)^t) = -1-2x,\\
a_4 &= e(\Psi(T_4)) = e(\pi((0,-1,-1/2)^t)) = e((-1,-1/2)^t) = -1-x,\\
a_5 &= e(\Psi(T_5)) = e(\pi((0,0,1)^t)) = e((0,1)^t) = 2x,\\
a_{23} &= e(\Psi(T_{23})) = e(\pi((0,1,0)^t)) = e((1,0)^t) = 1,\\
a_{97} &= e(\Psi(T_{23})) = e(\pi((0,14,3)^t)) = e((14,3)^t) = 14+6x.
\end{aligned}$$

Example 9.16. It is easier to appreciate Algorithm 9.14 after seeing how big the coefficients of the power series expansion of a newform typically are, when the newform is defined over a large field. For example, there is a newform

$$f = \sum_{n=1}^{\infty} a_n q^n \in S_2(\Gamma_0(389))$$

such that if $\alpha = a_2$, then

$$\begin{aligned}1097385680{\cdot}a_3(f) = &-20146763x^{19} + 102331615x^{18} + 479539092x^{17}\\ &- 3014444212x^{16} - 3813583550x^{15} + 36114755350x^{14}\\ &+ 6349339639x^{13} - 227515736964x^{12} + 71555185319x^{11}\\ &+ 816654992625x^{10} - 446376673498x^{9} - 1698789732650x^{8}\\ &+ 1063778499268x^{7} + 1996558922610x^{6} - 1167579836501x^{5}\\ &- 1238356001958x^{4} + 523532113822x^{3} + 352838824320x^{2}\\ &- 58584308844x - 25674258672.\end{aligned}$$

In contrast, if we take $v = \{0, \infty\} = (0, 1) \in \mathbb{M}_2(\Gamma_0(389))^+$, then

$$T_3(v) = -4(1,0) + 2(1,291) - 2(1,294) - 2(1,310) + 2(1,313) + 2(1,383).$$

Storing $T_3(v), T_5(v), \ldots$ as vectors is more compact than storing $a_3(f)$, $a_5(f)$, $\ldots$ directly as polynomials in a_2!

9.4. Congruences between Newforms

This section is about congruences between modular forms. Understanding congruences is crucial for studying Serre's conjectures, Galois representations, and explicit construction of Hecke algebras. We assume more background in algebraic number theory here than elsewhere in this book.

9.4.1. Congruences between Modular Forms. Let Γ be an arbitrary congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and suppose $f \in M_k(\Gamma)$ is a modular form of integer weight k for Γ. Since $\left(\begin{smallmatrix}1 & N\\ 0 & 1\end{smallmatrix}\right) \in \Gamma$ for some integer N, the form f has a Fourier expansion in nonnegative powers of $q^{1/N}$. For a rational number n, let $a_n(f)$ be the coefficient of q^n in the Fourier expansion of f. Put

$$\mathrm{ord}_q(f) = \min\{n \in \mathbb{Q} : a_n \neq 0\},$$

where by convention we take $\min \emptyset = +\infty$, so $\mathrm{ord}_q(0) = +\infty$.

9.4.1.1. *The j-invariant.* Let

$$j = \frac{1}{q} + 744 + 196884q + \cdots$$

be the j-function, which is a weight 0 modular function that is holomorphic except for a simple pole at ∞ and has integer Fourier coefficients (see, e.g., [**Ser73**, Section VIII.3.3]).

Lemma 9.17. *Suppose g is a weight 0 level 1 modular function that is holomorphic except possibly with a pole of order n at ∞. Then g is a polynomial in j of degree at most n. Moreover, the coefficients of this polynomial lie in the ideal I generated by the coefficients $a_m(g)$ with $m \leq 0$.*

Proof. If $n = 0$, then $g \in M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$, so g is constant with constant term in I, so the statement is true. Next suppose $n > 0$ and the lemma has been proved for all functions with smaller order poles. Let $\alpha = a_n(g)$, and note that

$$\mathrm{ord}_q(g - \alpha j^n) = \mathrm{ord}_q\left(g - \alpha \cdot \left(\frac{1}{q} + 744 + 196884q + \cdots\right)^n\right) > -n.$$

Thus by induction $h = g - \alpha j^n$ is a polynomial in j of degree $< n$ with coefficients in the ideal generated by the coefficients $a_m(g)$ with $m < 0$. It follows that $g = \alpha \cdot j^n - h$ satisfies the conclusion of the lemma. □

9.4.1.2. *Sturm's Theorem.* If $\mathcal{O}$ is the ring of integers of a number field, $\mathfrak{m}$ is a maximal ideal of $\mathcal{O}$, and $f = \sum a_n q^n \in \mathcal{O}[[q^{1/N}]]$ for some integer N, let

$$\mathrm{ord}_{\mathfrak{m}}(f) = \mathrm{ord}_q(f \mod \mathfrak{m}) = \min\{n \in \mathbb{Q} : a_n \notin \mathfrak{m}\}.$$

Note that $\mathrm{ord}_{\mathfrak{m}}(fg) = \mathrm{ord}_{\mathfrak{m}}(f) + \mathrm{ord}_{\mathfrak{m}}(g)$. The following theorem was first proved in [**Stu87**].

Theorem 9.18 (Sturm)**.** *Let $\mathfrak{m}$ be a prime ideal in the ring of integers $\mathcal{O}$ of a number field K, and let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ of index m and level N. Suppose $f \in M_k(\Gamma, \mathcal{O})$ is a modular form and*

$$\mathrm{ord}_{\mathfrak{m}}(f) > \frac{km}{12}$$

or $f \in S_k(\Gamma, \mathcal{O})$ is a cusp form and

$$\mathrm{ord}_{\mathfrak{m}}(f) > \frac{km}{12} - \frac{m-1}{N}.$$

Then $f \equiv 0 \pmod{\mathfrak{m}}$.

Proof. Case 1: First we assume $\Gamma = \mathrm{SL}_2(\mathbb{Z})$.
Let

$$\Delta = q + 24q^2 + \cdots \in S_{12}(\mathrm{SL}_2(\mathbb{Z}), \mathbb{Z})$$

be the Δ function. Since $\mathrm{ord}_{\mathfrak{m}}(f) > k/12$, we have $\mathrm{ord}_{\mathfrak{m}}(f^{12}) > k$. We have

$$\mathrm{ord}_q(f^{12} \cdot \Delta^{-k}) = 12 \cdot \mathrm{ord}_q(f) - k \cdot \mathrm{ord}_q(\Delta) \geq -k, \tag{9.4.1}$$

since f is holomorphic at infinity and Δ has a zero of order 1. Also

$$\mathrm{ord}_{\mathfrak{m}}(f^{12} \cdot \Delta^{-k}) = \mathrm{ord}_{\mathfrak{m}}(f^{12}) - k \cdot \mathrm{ord}_{\mathfrak{m}}(\Delta) > k - k = 0. \tag{9.4.2}$$

Combining (9.4.1) and (9.4.2), we see that

$$f^{12}\cdot\Delta^{-k}=\sum_{n\geq -k} b_n q^n,$$

with $b_n\in\mathcal{O}$ and $b_n\in\mathfrak{m}$ if $n\leq 0$.

By Lemma 9.17,

$$f^{12}\cdot\Delta^{-k}\in\mathfrak{m}[j]$$

is a polynomial in j of degree at most k with coefficients in $\mathfrak{m}$. Thus

$$f^{12}\in\mathfrak{m}[j]\cdot\Delta^{k},$$

so since the coefficients of Δ are integers, every coefficient of f^{12} is in $\mathfrak{m}$. Thus $\mathrm{ord}_{\mathfrak{m}}(f^{12})=+\infty$, hence $\mathrm{ord}_{\mathfrak{m}}(f)=+\infty$, so $f=0$, as claimed.

Case 2: Γ Arbitrary

Let N be such that $\Gamma(N)\subset\Gamma$, so also $f\in M_k(\Gamma(N))$. If $g\in M_k(\Gamma(N))$ is arbitrary, then because $\Gamma(N)$ is a normal subgroup of $\mathrm{SL}_2(\mathbb{Z})$, we have that for any $\gamma\in\Gamma(N)$ and $\delta\in\mathrm{SL}_2(\mathbb{Z})$,

$$(g^{[\delta]_k})^{[\gamma]_k}=g^{[\delta\gamma]_k}=g^{[\gamma'\delta]_k}=(g^{[\gamma']_k})^{[\delta]_k}=g^{[\delta]_k},$$

where $\gamma'\in\mathrm{SL}_2(\mathbb{Z})$. Thus for any $\delta\in\mathrm{SL}_2(\mathbb{Z})$, we have that $g^{[\delta]_k}\in M_k(\Gamma(N))$, so $\mathrm{SL}_2(\mathbb{Z})$ acts on $M_k(\Gamma(N))$.

It is a standard (but nontrivial) fact about modular forms, which comes from the geometry of the modular curve $X(N)$ over $\mathbb{Q}(\zeta_N)$ and $\mathbb{Z}[\zeta_N]$, that $M_k(\Gamma(N))$ has a basis with Fourier expansions in $\mathbb{Z}[\zeta_N][[q^{1/N}]]$ and that the action of $\mathrm{SL}_2(\mathbb{Z})$ on $M_k(\Gamma(N))$ preserves

$$M_k(\Gamma(N),\mathbb{Q}(\zeta_N))=M_k(\Gamma(N))\cap(\mathbb{Q}(\zeta_N)[[q^{1/N}]])$$

and the cuspidal subspace $S_k(\Gamma(N),\mathbb{Q}(\zeta_N))$. In particular, for any $\gamma\in\mathrm{SL}_2(\mathbb{Z})$,

$$f^{[\gamma]_k}\in M_k(\Gamma(N),K(\zeta_N))$$

Moreover, the denominators of $f^{[\gamma]_k}$ are bounded, since f is an $\mathcal{O}[\zeta_N]$-linear combination of a basis for $M_k(\Gamma(N),\mathbb{Z}[\zeta_N])$, and the denominators of $f^{[\gamma]_k}$ divide the product of the denominators of the images of each of these basis vectors under $[\gamma]_k$.

Let $L=K(\zeta_N)$. Let $\mathfrak{M}$ be a prime of $\mathcal{O}_L$ that divides $\mathfrak{m}\mathcal{O}_L$. We will now show that for each $\gamma\in\mathrm{SL}_2(\mathbb{Z})$, the Chinese Remainder Theorem implies that there is an element $A_\gamma\in L^*$ such that

$$A_\gamma\cdot f^{[\gamma]_k}\in M_k(\Gamma(N),\mathcal{O}_L)\qquad\text{and}\qquad \mathrm{ord}_{\mathfrak{M}}(A_\gamma\cdot f^{[\gamma]_k})<\infty.\tag{9.4.3}$$

First find $A\in L^*$ such that $A\cdot f^{[\gamma]_k}$ has coefficients in $\mathcal{O}_L$. Choose $\alpha\in\mathfrak{M}$ with $\alpha\notin\mathfrak{M}^2$, and find a negative power α^t such that $\alpha^t\cdot A\cdot f^{[\gamma]_k}$ has $\mathfrak{M}$-integral coefficients and finite valuation. This is possible because we assumed that f is nonzero. Use the Chinese Remainder Theorem to find $\beta \in \mathcal{O}_L$ such that $\beta \equiv 1 \pmod{\mathfrak{M}}$ and $\beta \equiv 0 \pmod{\wp}$ for each prime $\wp \neq \mathfrak{M}$ that divides (α). Then for some s we have

$$\beta^s \cdot \alpha^t \cdot A \cdot f^{[\gamma]_k} = A_\gamma \cdot f^{[\gamma]_k} \in M_k(\Gamma(N), \mathcal{O}_L)$$

and $\mathrm{ord}_{\mathfrak{M}}(A_\gamma \cdot f^{[\gamma]_k}) < \infty$.

Write

$$\mathrm{SL}_2(\mathbb{Z}) = \bigcup_{i=1}^{m} \Gamma\gamma_i$$

with $\gamma_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and let

$$F = f \cdot \prod_{i=2}^{m} A_{\gamma_i} \cdot f^{[\gamma_i]_k}.$$

Then $F \in M_{km}(\mathrm{SL}_2(\mathbb{Z}))$ and since $\mathfrak{M} \cap \mathcal{O}_K = \mathfrak{m}$, we have $\mathrm{ord}_{\mathfrak{M}}(f) = \mathrm{ord}_{\mathfrak{m}}(f)$, so

$$\mathrm{ord}_{\mathfrak{M}}(F) \geq \mathrm{ord}_{\mathfrak{M}}(f) = \mathrm{ord}_{\mathfrak{m}}(f) > \frac{km}{12}.$$

Thus we can apply Case 1 to conclude that

$$\mathrm{ord}_{\mathfrak{M}}(F) = +\infty.$$

Thus

$$\infty = \mathrm{ord}_{\mathfrak{M}}(F) = \mathrm{ord}_{\mathfrak{m}}(f) + \sum_{i=2}^{m} \mathrm{ord}_{\mathfrak{M}}(A_{\gamma_i} f^{[\gamma]_k}), \tag{9.4.4}$$

so $\mathrm{ord}_{\mathfrak{m}}(f) = +\infty$, because of (9.4.3).

We next obtain a better bound when f is a cusp form. Since $[\gamma]_k$ preserves cusp forms, $\mathrm{ord}_{\mathfrak{M}}(A_{\gamma_i} f^{[\gamma]_k}) \geq \frac{1}{N}$ for each i. Thus

$$\mathrm{ord}_{\mathfrak{M}}(F) \geq \mathrm{ord}_{\mathfrak{M}}(f) + \frac{m-1}{N} = \mathrm{ord}_{\mathfrak{m}}(f) + \frac{m-1}{N} > \frac{km}{12},$$

since now we are merely assuming that

$$\mathrm{ord}_{\mathfrak{m}}(f) > \frac{km}{12} - \frac{m-1}{N}.$$

Thus we again apply Case 1 to conclude that $\mathrm{ord}_{\mathfrak{M}}(F) = +\infty$, and using (9.4.4), conclude that $\mathrm{ord}_{\mathfrak{m}}(f) = +\infty$. □

Corollary 9.19. *Let $\mathfrak{m}$ be a prime ideal in the ring of integers $\mathcal{O}$ of a number field. Suppose $f, g \in M_k(\Gamma, \mathcal{O})$ are modular forms and*

$$a_n(f) \equiv a_n(g) \pmod{\mathfrak{m}}$$

for all

$$n \leq \begin{cases} \dfrac{km}{12} - \dfrac{m-1}{N} & \text{if } f - g \in S_k(\Gamma, \mathcal{O}), \\ \dfrac{km}{12} & \text{otherwise,} \end{cases}$$

where $m = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$. *Then* $f \equiv g \pmod{\mathfrak{m}}$.

Buzzard proved the following corollary, which is extremely useful in practical computations. It asserts that the Sturm bound for modular forms with character is the same as the Sturm bound for $\Gamma_0(N)$.

Corollary 9.20 (Buzzard). *Let* $\mathfrak{m}$ *be a prime ideal in the ring of integers* $\mathcal{O}$ *of a number field. Suppose* $f, g \in M_k(N, \varepsilon, \mathcal{O})$ *are modular forms with Dirichlet character* $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$ *and assume that*

$$a_n(f) \equiv a_n(g) \pmod{\mathfrak{m}} \qquad \text{for all} \qquad n \leq \frac{km}{12},$$

where

$$m = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)] = \#\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) = N \cdot \prod_{p \mid N} \left(1 + \frac{1}{p}\right).$$

Then $f \equiv g \pmod{\mathfrak{m}}$.

Proof. Let $h = f - g$ and let $r = km/12$, so $\mathrm{ord}_{\mathfrak{m}}(h) > r$. Let s be the order of the Dirichlet character ε. Then $h^s \in M_{ks}(\Gamma_0(N))$ and

$$\mathrm{ord}_{\mathfrak{m}}(h^s) > sr = \frac{ksm}{12}.$$

By Theorem 9.18, we have $\mathrm{ord}_{\mathfrak{m}}(h^s) = \infty$, so $\mathrm{ord}_{\mathfrak{m}}(h) = \infty$. It follows that $f \equiv g \pmod{\mathfrak{m}}$. □

9.4.1.3. *Congruence for Newforms.* Sturm's paper [**Stu87**] also applies some results of Asai on q-expansions at various cusps to obtain a more refined result for newforms.

Theorem 9.21 (Sturm). *Let* N *be a positive integer that is square-free, and suppose* f *and* g *are two newforms in* $S_k(N, \varepsilon, \mathcal{O})$, *where* $\mathcal{O}$ *is the ring of integers of a number field, and suppose that* $\mathfrak{m}$ *is a maximal ideal of* $\mathcal{O}$. *Let* I *be an arbitrary subset of the prime divisors of* N. *If* $a_p(f) = a_p(g)$ *for all* $p \in I$ *and if*

$$a_p(f) \equiv a_p(g) \pmod{\mathfrak{m}}$$

for all primes

$$p \leq \frac{k \cdot [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)]}{12 \cdot 2^{\#I}},$$

then $f \equiv g \pmod{\mathfrak{m}}$.

The paper [**BS02**] contains a similar result about congruences between newforms, which does not require that the level be square-free. Recall from Definition 4.18 that the conductor of a Dirichlet character ε is the largest divisor c of N such that ε factors through $(\mathbb{Z}/c\mathbb{Z})^\times$.

Theorem 9.22. *Let $N > 4$ be any integer, and suppose f and g are two normalized eigenforms in $S_k(N, \varepsilon; \mathcal{O})$, where $\mathcal{O}$ is the ring of integers of a number field, and suppose that $\mathfrak{m}$ is a maximal ideal of $\mathcal{O}$. Let I be the set of prime divisors of N that do not divide $\frac{N}{\mathrm{cond}(\varepsilon)}$. If*

$$a_p(f) \equiv a_p(g) \pmod{\mathfrak{m}}$$

for all primes $p \in I$ and for all primes

$$p \leq \frac{k \cdot [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)]}{12 \cdot 2^{\#I}},$$

then $f \equiv g \pmod{\mathfrak{m}}$.

For the proof, see Lemma 1.4 and Corollary 1.7 in [**BS02**, §1.3].

9.4.2. Generating the Hecke Algebra.

The following theorem appeared in [**LS02**, Appendix], except that we give a better bound here. It is a nice application of the congruence result above, which makes possible explicit computations with Hecke rings $\mathbb{T}$.

Theorem 9.23. *Suppose Γ is a congruence subgroup that contains $\Gamma_1(N)$ and let*

$$r = \frac{km}{12} - \frac{m-1}{N}, \tag{9.4.5}$$

where $m = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$. Then the Hecke algebra

$$\mathbb{T} = \mathbb{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_k(\Gamma))$$

is generated as a $\mathbb{Z}$-module by the Hecke operators T_n for $n \leq r$.

Proof. For any ring R, let $S_k(N, R) = S_k(N; \mathbb{Z}) \otimes R$, where $S_k(N; \mathbb{Z}) \subset \mathbb{Z}[[q]]$ is the submodule of cusp forms with integer Fourier expansion at the cusp ∞, and let $\mathbb{T}_R = \mathbb{T} \otimes_{\mathbb{Z}} R$. For any ring R, there is a perfect pairing

$$S_k(N, R) \otimes_R \mathbb{T}_R \to R$$

given by $\langle f, T\rangle \mapsto a_1(T(f))$ (this is true for $R = \mathbb{Z}$, hence for any R).

Let M be the submodule of $\mathbb{T}$ generated by $T_1, T_2, \ldots, T_r$, where r is the largest integer $\leq \frac{kN}{12} \cdot [\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$. Consider the exact sequence of additive abelian groups

$$0 \to M \overset{i}{\to} \mathbb{T} \to \mathbb{T}/M \to 0.$$

Let p be a prime and use the fact that tensor product is right exact to obtain an exact sequence

$$M \otimes \mathbb{F}_p \xrightarrow{\bar{i}} \mathbb{T} \otimes \mathbb{F}_p \to (\mathbb{T}/M) \otimes \mathbb{F}_p \to 0.$$

Suppose that $f \in S_k(N, \mathbb{F}_p)$ pairs to 0 with each of $T_1, \ldots, T_r$. Then

$$a_m(f) = a_1(T_m f) = \langle f, T_m \rangle = 0$$

in $\mathbb{F}_p$ for each $m \leq r$. By Theorem 9.18, it follows that $f = 0$. Thus the pairing restricted to the image of $M \otimes \mathbb{F}_p$ in $\mathbb{T}_{\mathbb{F}_p}$ is nondegenerate, so because (9.4.5) is perfect, it follows that

$$\dim_{\mathbb{F}_p} \bar{i}(M \otimes \mathbb{F}_p) = \dim_{\mathbb{F}_p} S_k(N, \mathbb{F}_p).$$

Thus $(\mathbb{T}/M) \otimes \mathbb{F}_p = 0$. Repeating the argument for all primes p shows that $\mathbb{T}/M = 0$, as claimed. □

Remark 9.24. In general, the conclusion of Theorem 9.23 is not true if one considers only T_n where n runs over the primes less than the bound. Consider, for example, $S_2(11)$, where the bound is 1 and there are no primes ≤ 1. However, the Hecke algebra is generated as an algebra by operators T_p with $p \leq r$.

9.5. Exercises

9.1 Prove that the group $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and that the quotient $\Gamma_0(N)/\Gamma_1(N)$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^*$.

9.2 Prove that the operators $\langle d \rangle$ are elements of $\mathbb{Z}[\ldots, T_n, \ldots]$. [Hint: Use Dirichlet's theorem on primes in arithmetic progression.]

9.3 Find an example like Example 9.6 but in which the new subspace is nonzero. More precisely, find an integer N such that the Hecke ring on $S_2(\Gamma_0(N))$ is not equal to the ring generated by Hecke operators T_n with $\gcd(n, N) = 1$ and $S_2(\Gamma_0(N))_{\text{new}} \neq 0$.

9.4 (a) Following Example 9.15, compute a basis for $S_2(\Gamma_0(31))$.
(b) Use Algorithm 9.12 to compute a basis for $S_2(\Gamma_0(31))$.

Chapter 10

Computing Periods

This chapter is about computing period maps associated to newforms. We assume you have read Chapters 8 and 9 and that you are familiar with abelian varieties at the level of [**Ros86**].

In Section 10.1 we introduce the period map and give some examples of situations in which computing it is relevant. Section 10.2 is about how to use the period mapping to attach an abelian variety to any newform. In Section 10.3, we introduce extended modular symbols, which are the key computational tool for quickly computing periods of modular symbols. We turn to numerical computation of period integrals in Section 10.4, and in Section 10.5 we explain how to use Atkin-Lehner operators to speed convergence. In Section 10.6 we explain how to compute the full period map with a minimum amount of work.

Section 10.7 briefly sketches three approaches to computing all elliptic curves of a given conductor.

This chapter was inspired by [**Cre97a**], which contains similar algorithms in the special case of a newform $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ with $a_n \in \mathbb{Z}$.

See also [**Dok04**] for algorithmic methods to compute special values of very general L-functions, which can be used for approximating $L(f, s)$ for arbitrary s.

10.1. The Period Map

Let Γ be a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma_1(N)$ for some N, and suppose

$$f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma)$$

is a newform (see Definition 9.9). In this chapter we describe how to approximately compute the complex period mapping

$$\Phi_f : \mathbb{M}_k(\Gamma) \to \mathbb{C},$$

given by

$$\Phi_f(P\{\alpha,\beta\}) = \langle f, P\{\alpha,\beta\}\rangle = \int_\alpha^\beta f(z)P(z,1)dz,$$

as in Section 8.5. As an application, we can approximate the special values $L(f,j)$, for $j = 1, 2, \ldots, k-1$ using (8.5.5). We can also compute the period lattice attached to a modular abelian variety, which is an important step, e.g., in enumeration of $\mathbb{Q}$-curves (see, e.g., [**GLQ04**]) or computation of a curve whose Jacobian is a modular abelian variety A_f (see, e.g., [**Wan95**]).

10.2. Abelian Varieties Attached to Newforms

Fix a newform $f \in S_k(\Gamma)$, where $\Gamma_1(N) \subset \Gamma$ for some N. Let $f_1, \ldots, f_d$ be the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f, where $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts via its action on the Fourier coefficients, which are algebraic integers (since they are the eigenvalues of matrices with integer entries). Let

$$V_f = \mathbb{C}f_1 \oplus \cdots \oplus \mathbb{C}f_d \subset S_k(\Gamma) \tag{10.2.1}$$

be the subspace of cusp forms spanned by the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f. One can show using the results discussed in Section 9.2 that the above sum is direct, i.e., that V_f has dimension d.

The integration pairing induces a $\mathbb{T}$-equivariant homomorphism

$$\Phi_f : \mathbb{M}_k(\Gamma) \to V_f^* = \mathrm{Hom}_{\mathbb{C}}(V_f, \mathbb{C})$$

from modular symbols to the $\mathbb{C}$-linear dual V_f^* of V_f. Here $\mathbb{T}$ acts on V_f^* via $(\varphi t)(x) = \varphi(tx)$, and this homomorphism is $\mathbb{T}$-stable by Theorem 8.21. The *abelian variety attached to* f is the quotient

$$A_f(\mathbb{C}) = V_f^*/\Phi_f(\mathbb{S}_k(\Gamma;\mathbb{Z})).$$

Here $\mathbb{S}_k(\Gamma;\mathbb{Z}) = \mathbb{S}_k(\Gamma)$, and we include the $\mathbb{Z}$ in the notation to emphasize that these are integral modular symbols. See [**Shi59**] for a proof that $A_f(\mathbb{C})$ is an abelian variety (in particular, $\Phi_f(\mathbb{S}_k(\Gamma;\mathbb{Z}))$ is a lattice, and V_f^* is equipped with a nondegenerate Riemann form).

When $k = 2$, we can also construct A_f as a quotient of the modular Jacobian $\mathrm{Jac}(X_\Gamma)$, so A_f is an abelian variety canonically defined over $\mathbb{Q}$.

In general, we have an exact sequence

$$0 \to \mathrm{Ker}(\Phi_f) \to \mathbb{S}_k(\Gamma) \to V_f^* \to A_f(\mathbb{C}) \to 0.$$

Remark 10.1. When $k = 2$, the abelian variety A_f has a canonical structure of abelian variety over $\mathbb{Q}$. Moreover, there is a conjecture of Ribet and Serre in [**Rib92**] that describes the simple abelian varieties A over $\mathbb{Q}$ that should arise via this construction. In particular, the conjecture is that A is isogenous to some abelian variety A_f if and only if $\mathrm{End}(A/\mathbb{Q}) \otimes \mathbb{Q}$ is a number field of degree $\dim(A)$. The abelian varieties A_f have this property since $\mathbb{Q}(\ldots, a_n(f), \ldots)$ embeds in $\mathrm{End}(A/\mathbb{Q})\otimes\mathbb{Q}$ and the endomorphism ring over $\mathbb{Q}$ has degree at most $\dim(A)$ (see [**Rib92**] for details). Ribet proves that his conjecture is a consequence of Serre's conjecture [**Ser87**] on modularity of mod p odd irreducible Galois representations (see Section 1.5). Much of Serre's conjecture has been proved by Khare and Wintenberger (not published). In particular, it is a theorem that if A is a simple abelian variety over $\mathbb{Q}$ with $\mathrm{End}(A/\mathbb{Q}) \otimes \mathbb{Q}$ a number field of degree $\dim(A)$ and if A has good reduction at 2, then A is isogenous to some abelian variety A_f.

Remark 10.2. When $k > 2$, there is an object called a *Grothendieck motive* that is attached to f and has a canonical "structure over $\mathbb{Q}$". See [**Sch90**].

10.3. Extended Modular Symbols

In this section, we extend the notion of modular symbols to allows symbols of the form $P\{w, z\}$ where w and z are arbitrary elements of $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q})$.

Definition 10.3 (Extended Modular Symbols)**.** The abelian group $\overline{\mathbb{M}}_k$ of *extended modular symbols* of weight k is the $\mathbb{Z}$-span of symbols $P\{w, z\}$, with $P \in V_{k-2}$ a homogeneous polynomial of degree $k-2$ with integer coefficients, modulo the relations

$$P \cdot (\{w, y\} + \{y, z\} + \{z, w\}) = 0$$

and modulo any torsion.

Fix a finite index subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$. Just as for usual modular symbols, $\overline{\mathbb{M}}_k$ is equipped with an action of Γ, and we define the space of *extended modular symbols* of weight k for Γ to be the quotient

$$\overline{\mathbb{M}}_k(\Gamma) = (\overline{\mathbb{M}}_k / \langle \gamma x - x : \gamma \in \Gamma, x \in \overline{\mathbb{M}}_k \rangle)/\,\mathrm{tor}\,.$$

The quotient $\overline{\mathbb{M}}_k(\Gamma)$ is torsion-free and fixed by Γ.

The integration pairing extends naturally to a pairing

$$\big(S_k(\Gamma) \oplus \overline{S}_k(\Gamma)\big) \times \overline{\mathbb{M}}_k(\Gamma) \to \mathbb{C}, \tag{10.3.1}$$

where we recall from (8.5.1) that $\overline{S}_k(\Gamma)$ denotes the space of antiholomorphic cusp forms. Moreover, if

$$\iota : \mathbb{M}_k(\Gamma) \to \overline{\mathbb{M}}_k(\Gamma)$$

is the natural map, then ι respects (10.3.1) in the sense that for all $f \in S_k(\Gamma) \oplus \overline{S}_k(\Gamma)$ and $x \in \mathbb{M}_k(\Gamma)$, we have

$$\langle f, x\rangle = \langle f, \iota(x)\rangle.$$

As we will see soon, it is often useful to replace $x \in \mathbb{M}_k(\Gamma)$ first by $\iota(x)$ and then by an equivalent sum $\sum y_i$ of symbols $y_i \in \overline{\mathbb{M}}_k(N, \varepsilon)$ such that $\langle f, \sum y_i\rangle$ is easier to compute numerically than $\langle f, x\rangle$.

Let ε be a Dirichlet character of modulus N. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, let $\varepsilon(\gamma) = \varepsilon(d)$. Let $\overline{\mathbb{M}}_k(N, \varepsilon)$ be the quotient of $\overline{\mathbb{M}}_k(N, \mathbb{Z}[\varepsilon])$ by the relations $\gamma(x) - \varepsilon(\gamma)x$, for all $x \in \mathbb{M}_k(N, \mathbb{Z}[\varepsilon])$, $\gamma \in \Gamma_0(N)$, and modulo any torsion.

10.4. Approximating Period Integrals

In this section we assume Γ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma_1(N)$ for some N. Suppose $\alpha \in \mathfrak{h}$, so $\mathrm{Im}(\alpha) > 0$ and m is an integer such that $0 \le m \le k-2$, and consider the extended modular symbol $X^mY^{k-2-m}\{\alpha, \infty\}$. Let $\langle \cdot, \cdot \rangle$ denote the integration pairing from Section 8.5. Given an arbitrary cusp form $f = \sum_{n=1}^{\infty} a_n q^n \in S_k(\Gamma)$, we have

$$\Phi_f(X^mY^{k-2-m}\{\alpha,\infty\}) = \left\langle f,\, X^mY^{k-2-m}\{\alpha,\infty\}\right\rangle \tag{10.4.1}$$

$$= \int_{\alpha}^{\infty} f(z) z^m dz \tag{10.4.2}$$

$$= \sum_{n=1}^{\infty} a_n \int_{\alpha}^{\infty} e^{2\pi i n z} z^m dz. \tag{10.4.3}$$

The reversal of summation and integration is justified because the imaginary part of α is positive so that the sum converges absolutely. The following lemma is useful for computing the above infinite sum.

Lemma 10.4.

$$\int_{\alpha}^{\infty} e^{2\pi i n z} z^m dz \;=\; e^{2\pi i n \alpha} \sum_{s=0}^{m} \left(\frac{(-1)^s \alpha^{m-s}}{(2\pi i n)^{s+1}} \prod_{j=(m+1)-s}^{m} j \right). \tag{10.4.4}$$

Proof. See Exercise 10.1 □

In practice we will usually be interested in computing the period map Φ_f when $f \in S_k(\Gamma)$ is a newform. Since f is a newform, there is a Dirichlet character ε such that $f \in S_k(N, \varepsilon)$. The period map $\Phi_f : \mathbb{M}_k(\Gamma) \to \mathbb{C}$ then factors through the quotient $\mathbb{M}_k(N,\varepsilon)$, so it suffices to compute the period map on modular symbols in $\mathbb{M}_k(N,\varepsilon)$.

The following proposition is an analogue of [**Cre97a**, Prop. 2.1.1(5)].

Proposition 10.5. *For any $\gamma \in \Gamma_0(N)$, $P \in V_{k-2}$ and $\alpha \in \mathfrak{h}^*$, we have the following relation in $\overline{\mathbb{M}}_k(N,\varepsilon)$:*

$$\begin{aligned} P\{\infty,\gamma(\infty)\} &= P\{\alpha,\gamma(\alpha)\} + (P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\} && (10.4.5)\\ &= \varepsilon(\gamma)(\gamma^{-1}P)\{\alpha,\infty\} - P\{\gamma(\alpha),\infty\}. && (10.4.6)\end{aligned}$$

Proof. By definition, if $x \in \mathbb{M}_k(N,\varepsilon)$ is a modular symbol and $\gamma \in \Gamma_0(N)$, then $\gamma x = \varepsilon(\gamma)x$. Thus $\varepsilon(\gamma)\gamma^{-1}x = x$, so

$$\begin{aligned} P\{\infty,\gamma(\infty)\} &= P\{\infty,\alpha\} + P\{\alpha,\gamma(\alpha)\} + P\{\gamma(\alpha),\gamma(\infty)\}\\ &= P\{\infty,\alpha\} + P\{\alpha,\gamma(\alpha)\} + \varepsilon(\gamma)\gamma^{-1}(P\{\gamma(\alpha),\gamma(\infty)\})\\ &= P\{\infty,\alpha\} + P\{\alpha,\gamma(\alpha)\} + \varepsilon(\gamma)(\gamma^{-1}P)\{\alpha,\infty\}\\ &= P\{\alpha,\gamma(\alpha)\} + P\{\infty,\alpha\} - \varepsilon(\gamma)(\gamma^{-1}P)\{\infty,\alpha\}\\ &= P\{\alpha,\gamma(\alpha)\} + (P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\}.\end{aligned}$$

The second equality in the statement of the proposition now follows easily. □

In the case of weight 2 and trivial character, the "error term"

$$(P-\varepsilon(\gamma)\gamma^{-1}P)\{\infty,\alpha\} \tag{10.4.7}$$

vanishes since P is constant and $\varepsilon(\gamma)=1$. In general this term does not vanish. However, we can suitably modify the formulas found in [**Cre97a**, 2.10] and still obtain an algorithm for computing period integrals.

Algorithm 10.6 (Period Integrals)**.** *Given $\gamma \in \Gamma_0(N)$, $P \in V_{k-2}$ and $f \in S_k(N,\varepsilon)$ presented as a q-expansion to some precision, this algorithm outputs an approximation to the period integral $\langle f,\, P\{\infty,\gamma(\infty)\}\rangle$.*

(1) Write $\gamma = \left(\begin{smallmatrix} a & b \\ cN & d\end{smallmatrix}\right) \in \Gamma_0(N)$, with $a,b,c,d \in \mathbb{Z}$, and set $\alpha = \frac{-d+i}{cN}$ in Proposition 10.5.
(2) Replacing γ by $-\gamma$ if necessary, we find that the imaginary parts of α and $\gamma(\alpha) = \frac{a+i}{cN}$ are both equal to the positive number $\frac{1}{cN}$.
(3) Use (10.4.3) and Lemma 10.4 to compute the integrals that appear in Proposition 10.5.

It would be nice if the modular symbols of the form $P\{\infty,\gamma(\infty)\}$ for $P \in V_{k-2}$ and $\gamma \in \Gamma_0(N)$ were to generate a large subspace of $\mathbb{M}_k(N,\varepsilon)\otimes\mathbb{Q}$. When $k=2$ and $\varepsilon=1$, Manin proved in [**Man72**] that the map $\Gamma_0(N) \to H_1(X_0(N),\mathbb{Z})$ sending γ to $\{0,\gamma(0)\}$ is a surjective group homomorphism. When $k>2$, the author does not know a similar group-theoretic statement. However, we have the following theorem.

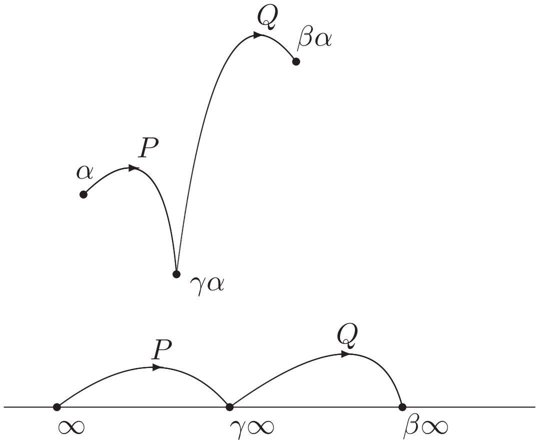


Figure 10.4.1. “Transporting” a transportable modular symbol.

Theorem 10.7. *Any element of* $\mathbb{S}_k(N,\varepsilon)$ *can be written in the form*

$$\sum_{i=1}^{n} P_i\{\infty,\gamma_i(\infty)\}$$

for some $P_i \in V_{k-2}$ *and* $\gamma_i \in \Gamma_0(N)$. *Moreover,* P_i *and* γ_i *can be chosen so that* $\sum P_i = \sum \varepsilon(\gamma_i)\gamma_i^{-1}(P_i)$, *so the error term* (10.4.7) *vanishes.*

The author and Helena Verrill prove this theorem in [**SV01**]. The condition that the error term vanishes means that one can replace ∞ by any α in the expression for the modular symbol and obtain an equivalent modular symbol. For this reason, we call such modular symbols *transportable*, as illustrated in Figure 10.4.1.

Note that in general not every element of the form $P\{\infty,\gamma(\infty)\}$ must lie in $\mathbb{S}_k(N,\varepsilon)$. However, if $\gamma P = P$, then $P\{\infty,\gamma(\infty)\}$ does lie in $\mathbb{S}_k(N,\varepsilon)$. It would be interesting to know under what circumstances $\mathbb{S}_k(N,\varepsilon)$ is generated by symbols of the form $P\{\infty,\gamma(\infty)\}$ with $\gamma P = P$. This sometimes fails for k odd; for example, when $k = 3$, the condition $\gamma P = P$ implies that $\gamma \in \Gamma_0(N)$ has an eigenvector with eigenvalue 1, and hence is of finite order. When k is even, the author can see no obstruction to generating $\mathbb{S}_k(N,\varepsilon)$ using such symbols.

10.5. Speeding Convergence Using Atkin-Lehner

Let $w_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) \in \mathrm{Mat}_2(\mathbb{Z})$. Consider the Atkin-Lehner involution W_N on $M_k(\Gamma_1(N))$, which is defined by

$$\begin{aligned} W_N(f) &= N^{(2-k)/2} \cdot f|_{[w_N]_k} \\ &= N^{(2-k)/2} \cdot f\left(-\frac{1}{Nz}\right) \cdot N^{k-1} \cdot (Nz)^{-k} \\ &= N^{-k/2} \cdot z^{-k} \cdot f\left(-\frac{1}{Nz}\right). \end{aligned}$$

Here we take the positive square root if k is odd. Then $W_N^2 = (-1)^k$ is an involution when k is even.

There is an operator on modular symbols, which we also denote W_N, which is given by

$$\begin{aligned} W_N(P\{\alpha,\beta\}) &= N^{(2-k)/2} \cdot w_N(P)\{w_N(\alpha), w_N(\beta)\} \\ &= N^{(2-k)/2} \cdot P(-Y, NX)\left\{-\frac{1}{\alpha N}, -\frac{1}{\beta N}\right\}, \end{aligned}$$

and one has that if $f \in S_k(\Gamma_1(N))$ and $x \in \mathbb{M}_k(\Gamma_1(N))$, then

$$\langle W_N(f), x\rangle = \langle f, W_N(x)\rangle.$$

If ε is a Dirichlet character of modulus N, then the operator W_N sends $S_k(N,\varepsilon)$ to $S_k(\Gamma_1(N),\overline{\varepsilon})$. Thus if $\varepsilon^2 = 1$, then W_N preserves $S_k(N,\varepsilon)$. In particular, W_N acts on $S_k(\Gamma_0(N))$.

The next proposition shows how to compute the pairing $\langle f, P\{\infty,\gamma(\infty)\}\rangle$ under certain restrictive assumptions. It generalizes a result of [**Cre97b**] to higher weight.

Proposition 10.8. *Let $f \in S_k(N,\varepsilon)$ be a cusp form which is an eigenform for the Atkin-Lehner operator W_N having eigenvalue $w \in \{\pm 1\}$ (thus $\varepsilon^2 = 1$ and k is even). Then for any $\gamma \in \Gamma_0(N)$ and any $P \in V_{k-2}$, with the property that $\gamma P = \varepsilon(\gamma) P$, we have the following formula, valid for any $\alpha \in \mathfrak{h}$:*

$$\begin{aligned} \langle f, P\{\infty,\gamma(\infty)\}\rangle = \Big\langle f, \quad & w\frac{P(Y,-NX)}{N^{k/2-1}}\{w_N(\alpha),\infty\} \\ &+ \left(P - w\frac{P(Y,-NX)}{N^{k/2-1}}\right)\left\{i/\sqrt{N},\infty\right\} - P\left\{\gamma(\alpha),\infty\right\}\Big\rangle. \end{aligned}$$

Here $w_N(\alpha) = -\dfrac{1}{N\alpha}$.

Proof. By Proposition 10.5 our condition on P implies that $P\{\infty,\gamma(\infty)\} = P\{\alpha,\gamma(\alpha)\}$. We describe the steps of the following computation below.

$$\begin{aligned}&\Big\langle f,\quad P\{\alpha,\gamma(\alpha)\}\Big\rangle\\&=\Big\langle f,\quad P\{\alpha,i/\sqrt{N}\}+P\{i/\sqrt{N},W(\alpha)\}+P\{W(\alpha),\gamma(\alpha)\}\Big\rangle\\&=\Big\langle f,\quad w\frac{W(P)}{N^{k/2-1}}\{W(\alpha),i/\sqrt{N}\}+P\{i/\sqrt{N},W(\alpha)\}+P\{W(\alpha),\gamma(\alpha)\}\Big\rangle.\end{aligned}$$

For the first equality, we break the path into three paths, and in the second, we apply the W-involution to the first term and use that the action of W is compatible with the pairing $\langle\, ,\,\rangle$ and that f is an eigenvector with eigenvalue w. In the following sequence of equalities we combine the first two terms and break up the third; then we replace $\{W(\alpha),i/\sqrt{N}\}$ by $\{W(\alpha),\infty\}+\{\infty,i/\sqrt{N}\}$ and regroup:

$$\begin{aligned}&w\frac{W(P)}{N^{k/2-1}}\{W(\alpha),i/\sqrt{N}\}+P\{i/\sqrt{N},W(\alpha)\}+P\{W(\alpha),\gamma(\alpha)\}\\&=\left(w\frac{W(P)}{N^{k/2-1}}-P\right)\{W(\alpha),i/\sqrt{N}\}+P\{W(\alpha),\infty\}-P\{\gamma(\alpha),\infty\}\\&=w\frac{W(P)}{N^{k/2-1}}\{W(\alpha),\infty\}+\left(P-w\frac{W(P)}{N^{k/2-1}}\right)\{i/\sqrt{N},\infty\}-P\{\gamma(\alpha),\infty\}.\end{aligned}$$

□

A good choice for α is $\alpha=\gamma^{-1}\left(\frac{b}{d}+\frac{i}{d\sqrt{N}}\right)$, so that $W(\alpha)=\frac{c}{d}+\frac{i}{d\sqrt{N}}$. This maximizes the minimum of the imaginary parts of α and $W(\alpha)$, which results in series that converge more quickly.

Let $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma_0(N)$. The polynomial

$$P(X,Y)=(cX^2+(d-a)XY-bY^2)^{\frac{k-2}{2}}$$

satisfies $\gamma(P)=P$. We obtained this formula by viewing V_{k-2} as the $(k-2)$th symmetric product of the 2-dimensional space on which $\Gamma_0(N)$ acts naturally. For example, observe that since $\det(\gamma)=1$, the symmetric product of two eigenvectors for γ is an eigenvector in V_2 having eigenvalue 1. For the same reason, if $\varepsilon(\gamma)\neq 1$, there need not be a polynomial $P(X,Y)$ such that $\gamma(P)=\varepsilon(\gamma)P$. One remedy is to choose another γ so that $\varepsilon(\gamma)=1$.

Since the imaginary parts of the terms $i/\sqrt{N}$, α and $W(\alpha)$ in the proposition are all relatively large, the sums appearing at the beginning of Section 10.4 converge quickly if d is small. It is *important* to choose γ in Proposition 10.8 with d small; otherwise the series will converge very slowly.

Remark 10.9. Is there a generalization of Proposition 10.8 without the restrictions that $\varepsilon^2=1$ and k is even?

10.5.1. Another Atkin-Lehner Trick. Suppose E is an elliptic curve and let $L(E,s)$ be the corresponding L-function. Let $\varepsilon \in \{\pm 1\}$ be the root number of E, i.e., the sign of the functional equation for $L(E,s)$, so $\Lambda(E,s) = \varepsilon\Lambda(E,2-s)$, where $\Lambda(E,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(E,s)$. Let $f = f_E$ be the modular form associated to E (which exists by [**Wil95, BCDT01**]). If $W_N(f) = wf$, then $\varepsilon = -w$ (see Exercise 10.2). We have

$$\begin{aligned} L(E,1) &= -2\pi \int_0^\infty f(z)\,dz \\ &= -2\pi i\,\langle f,\,\{0,\infty\}\rangle \\ &= -2\pi i\,\left\langle f, \{0, i/\sqrt{N}\} + \{i/\sqrt{N},\infty\}\right\rangle \\ &= -2\pi i\,\left\langle wf, \{w_N(0), w_N(i/\sqrt{N})\} + \{i/\sqrt{N},\infty\}\right\rangle \\ &= -2\pi i\,\left\langle wf, \{\infty, i/\sqrt{N}\} + \{i/\sqrt{N},\infty\}\right\rangle \\ &= -2\pi i\,(w-1)\left\langle f, \{\infty, i/\sqrt{N}\}\right\rangle. \end{aligned}$$

If $w = 1$, then $L(E,1) = 0$. If $w = -1$, then

$$L(E,1) = 4\pi i \left\langle f, \{\infty, i/\sqrt{N}\}\right\rangle = 2\sum_{n=1}^{\infty} \frac{a_n}{n} e^{-2\pi n/\sqrt{N}}. \tag{10.5.1}$$

For more about computing with L-functions of elliptic curves, including a trick for computing ε quickly without directly computing W_N, see [**Coh93**, §7.5] and [**Cre97a**, §2.11]. One can also find higher derivatives $L^{(r)}(E,1)$ by a formula similar to (10.5.1) (see [**Cre97a**, §2.13]). The methods in this chapter for obtaining rapidly converging series are not just of computational interest; see, e.g., [**Gre83**] for a nontrivial theoretical application to the Birch and Swinnerton-Dyer conjecture.

10.6. Computing the Period Mapping

Fix a newform $f = \sum a_n q^n \in S_k(\Gamma)$, where $\Gamma_1(N) \subset \Gamma$ for some N. Let V_f be as in (10.2.1).

Let $\Theta_f : M_k(\Gamma;\mathbb{Q}) \to V$ be *any* $\mathbb{Q}$-linear map with the same kernel as Φ_f; we call any such map a *rational period mapping* associated to f. Let Φ_f be the period mapping associated to the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f. We have a commutative diagram

$$\begin{tikzcd}
\mathbb{M}_k(\Gamma;\mathbb{Q}) \arrow[rr, "\Phi_f"] \arrow[dr, "\Theta_f"'] & & \mathrm{Hom}_{\mathbb{C}}(V_f,\mathbb{C}) \\
& V \arrow[ur, hook, "i_f"] &
\end{tikzcd}$$

Recall from Section 10.2 that the cokernel of Φ_f is the abelian variety $A_f(\mathbb{C})$.

The Hecke algebra $\mathbb{T}$ acts on the linear dual

$$\mathbb{M}_k(\Gamma;\mathbb{Q})^* = \mathrm{Hom}(\mathbb{M}_k(\Gamma),\mathbb{Q})$$

by $(t\varphi)(x) = \varphi(tx)$. Let $I = I_f \subset \mathbb{T}$ be the kernel of the ring homomorphism $\mathbb{T} \to \mathbb{Z}[a_2, a_3, \ldots]$ that sends T_n to a_n. Let

$$\mathbb{M}_k(\Gamma;\mathbb{Q})^*[I] = \{\varphi \in \mathbb{M}_k(\Gamma;\mathbb{Q})^* : t\varphi = 0 \text{ all } t \in I\}.$$

Since f is a newform, one can show that $\mathbb{M}_k(\Gamma;\mathbb{Q})^*[I]$ has dimension d. Let $\theta_1, \ldots, \theta_d$ be a basis for $\mathbb{M}_k(\Gamma;\mathbb{Q})^*[I]$, so

$$\mathrm{Ker}(\Phi_f) = \mathrm{Ker}(\theta_1) \oplus \cdots \oplus \mathrm{Ker}(\theta_d).$$

We can thus compute $\mathrm{Ker}(\Phi_f)$, hence a choice of Θ_f. To compute Φ_f, it remains to compute i_f.

Let $S_k(\Gamma;\mathbb{Q})$ denote the space of cusp forms with q-expansion in $\mathbb{Q}[[q]]$. By Exercise 10.3

$$S_k(\Gamma;\mathbb{Q})[I] = S_k(\Gamma)[I] \cap \mathbb{Q}[[q]]$$

is a $\mathbb{Q}$-vector space of dimension d. Let $g_1, \ldots, g_d$ be a basis for this $\mathbb{Q}$-vector space. We will compute Φ_f with respect to the basis of $\mathrm{Hom}_{\mathbb{Q}}(S_k(\Gamma;\mathbb{Q})[I];\mathbb{C})$ dual to this basis. Choose elements $x_1, \ldots, x_d \in \mathbb{M}_k(\Gamma)$ with the following properties:

(1) Using Proposition 10.5 or Proposition 10.8, it is possible to compute the period integrals $\langle g_i, x_j\rangle$, $i, j \in \{1, \ldots, d\}$, efficiently.

(2) The $2d$ elements $v + \eta(v)$ and $v - \eta(v)$ for $v = \Theta_f(x_1), \ldots, \Theta_f(x_d)$ span a space of dimension $2d$ (i.e., they span $\mathbb{M}_k(\Gamma)/\operatorname{Ker}(\Phi_f)$).

Given this data, we can compute

$$i_f(v + \eta(v)) = 2\mathrm{Re}(\langle g_1, x_i\rangle, \ldots, \langle g_d, x_i\rangle)$$

and

$$i_f(v - \eta(v)) = 2i\mathrm{Im}(\langle g_1, x_i\rangle, \ldots, \langle g_d, x_i\rangle).$$

We break the integrals into real and imaginary parts because this increases the precision of our answers. Since the vectors $v_n + \eta(v_n)$ and $v_n - \eta(v_n)$, $n = 1, \ldots, d$, span $\mathbb{M}_k(N, \varepsilon;\mathbb{Q})/\operatorname{Ker}(\Phi_f)$, we have computed i_f.

Remark 10.10. We want to find symbols x_i satisfying the conditions of Proposition 10.8. This is usually possible when d is very small, but in practice it is difficult when d is large.

Remark 10.11. The above strategy was motivated by [**Cre97a**, §2.10].

10.7. All Elliptic Curves of Given Conductor

Using modular symbols and the period map, we can compute all elliptic curves over $\mathbb{Q}$ of conductor N, up to isogeny. The algorithm in this section gives all *modular elliptic curves* (up to isogeny), i.e., elliptic curves attached to modular forms, of conductor N. Fortunately, it is now known by [**Wil95, BCDT01, TW95**] that every elliptic curve over $\mathbb{Q}$ is modular, so the procedure of this section gives all elliptic curves (up to isogeny) of given conductor. See [**Cre06**] for a nice historical discussion of this problem.

Algorithm 10.12 (Elliptic Curves of Conductor N)**.** *Given $N > 0$, this algorithm outputs equations for all elliptic curves of conductor N, up to isogeny.*

(1) [Modular Symbols] Compute $\mathbb{M}_2(\Gamma_0(N))$ using Section 8.7.

(2) [Find Rational Eigenspaces] Find the 2-dimensional eigenspaces V in $\mathbb{M}_2(\Gamma_0(N))_{\mathrm{new}}$ that correspond to elliptic curves. Do *not* use the algorithm for decomposition from Section 7.5, which is too complicated and gives more information than we need. Instead, for the first few primes $p \nmid N$, compute all eigenspaces $\mathrm{Ker}(T_p - a)$, where a runs through integers with $-2\sqrt{p} < a < 2\sqrt{p}$. Intersect these eigenspaces to find the eigenspaces that correspond to elliptic curves. To find just the new ones, either compute the degeneracy maps to lower level or find all the rational eigenspaces of all levels that strictly divide N and exclude them.

(3) [Find Newforms] Use Algorithm 9.14 to compute to some precision each newform $f = \sum_{n=1}^{\infty} a_n q^n \in \mathbb{Z}[[q]]$ associated to each eigenspace V found in step (2).

(4) [Find Each Curve] For each newform f found in step (3), do the following:

   (a) [Period Lattice] Compute the corresponding period lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ by computing the image of Φ_f, as described in Section 10.6.

   (b) [Compute τ] Let $\tau = \omega_1/\omega_2$. If $\mathrm{Im}(\tau) < 0$, swap ω_1 and ω_2, so $\mathrm{Im}(\tau) > 0$. By successively applying generators of $\mathrm{SL}_2(\mathbb{Z})$, we find an $\mathrm{SL}_2(\mathbb{Z})$ equivalent element τ' in $\mathcal{F}$, i.e., $|\mathrm{Re}(\tau')| \leq 1/2$ and $|\tau| \geq 1$.

(c) [c-invariants] Compute the invariants c_4 and c_6 of the lattice Λ using the following rapidly convergent series:

$$c_4 = \left(\frac{2\pi}{\omega_2}\right)^4 \cdot \left(1 + 240\sum_{n=1}^{\infty} \frac{n^3 q^n}{1-q^n}\right),$$
$$c_6 = \left(\frac{2\pi}{\omega_2}\right)^6 \cdot \left(1 - 504\sum_{n=1}^{\infty} \frac{n^5 q^n}{1-q^n}\right),$$

where $q = e^{2\pi i \tau'}$, where τ' is as in step (4b). A theorem of Edixhoven (that the Manin constant is an integer) implies that the invariants c_4 and c_6 of Λ are integers, so it is only necessary to compute Λ to large precision to completely determine them.

(d) [Elliptic Curve] An elliptic curve with invariants c_4 and c_6 is

$$E: \quad y^2 = x^3 - \frac{c_4}{48}x - \frac{c_6}{864}.$$

(e) [Prove Correctness] Using Tate's algorithm, find the conductor of E. If the conductor is not N, then recompute c_4 and c_6 using more terms of f and real numbers to larger precision, etc. If the conductor is N, compute the coefficients b_p of the modular form $g = g_E$ attached to the elliptic curve E, for $p \leq \#\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})/6$. Verify that $a_p = b_p$, where a_p are the coefficients of f. If this equality holds, then E must be isogenous to the elliptic curve attached to f, by the Sturm bound (Theorem 9.18) and Faltings's isogeny theorem. If the equality fails for some p, recompute c_4 and c_6 to larger precision.

There are numerous tricks to optimize the above algorithm. For example, often one can work separately with $\mathbb{M}_k(\Gamma_0(N))^+_{\text{new}}$ and $\mathbb{M}_k(\Gamma_0(N))^-_{\text{new}}$ and get enough information to find E, up to isogeny (see [**Cre97b**]).

Once we have one curve from each isogeny class of curves of conductor N, we find each curve in each isogeny class (which is another interesting problem discussed in [**Cre97a**]), hence all curves of conductor N. If $E/\mathbb{Q}$ is an elliptic curve, then any curve isogenous to E is isogenous via a chain of isogenies of prime degree. There is an *a priori* bound on the degrees of these isogenies due to Mazur. Also, there are various methods for finding all isogenies of a given degree with domain E. See [**Cre97a**, §3.8] for more details.

10.7.1. Finding Curves: S-Integral Points.

In this section we briefly survey an alternative approach to finding curves of a given conductor by finding integral points on other elliptic curves.

Cremona and others have developed a complementary approach to the problem of computing all elliptic curves of given conductor (see [**CL04**]).

Instead of computing all curves of given conductor, we instead consider the seemingly more difficult problem of finding all curves with good reduction outside a finite set S of primes. Since one can compute the conductor of a curve using Tate's algorithm [**Tat75, Cre97a**, §3.2], if we know all curves with good reduction outside S, we can find all curves of conductor N by letting S be the set of prime divisors of N.

There is a strategy for finding all curves with good reduction outside S. It is not an algorithm, in the sense that it is always guaranteed to terminate (the modular symbols method above *is* an algorithm), but in practice it often works. Also, this strategy makes sense over any number field, whereas the modular symbols method does not (there are generalizations of modular symbols to other number fields).

Fix a finite set S of primes of a number field K. It is a theorem of Shafarevich that there are only finitely many elliptic curves with good reduction outside S (see [**Sil92**, Section IX.6]). His proof uses that the group of S-units in K is finite and Siegel's theorem that there are only finitely many S-integral points on an elliptic curve. One can make all this explicit, and sometimes in practice one can compute all these S-integral points.

The problem of finding all elliptic curves with good reduction outside of S can be broken into several subproblems, the main ones being

(1) determine the following finite subgroup of $K^*/(K^*)^m$:

$$K(S,m) = \{x \in K^*/(K^*)^m \,:\, m \mid \mathrm{ord}_{\mathfrak{p}}(x) \text{ all } \mathfrak{p} \notin S\};$$

(2) find all S-integral points on certain elliptic curves $y^2 = x^3 + k$.

In [**CL04**], there is one example, where they find all curves of conductor $N = 2^8 \cdot 17^2 = 73984$ by finding all curves with good reduction outside $\{2, 17\}$. They finds 32 curves of conductor 73984 that divide into 16 isogeny classes. (Note that $\dim S_2(\Gamma_0(N)) = 9577$.)

10.7.2. Finding Curves: Enumeration.

One can also find curves by simply enumerating Weierstrass equations. For example, the paper [**SW02**] discusses a database that the author and Watkins created that contains hundreds of millions of elliptic curves. It was constructed by enumerating Weierstrass equations of a certain form. This database does not contain *every* curve of each conductor included in the database. It is, however, fairly complete in some cases. For example, using the Mestre method of graphs [**Mes86**], we verified in [**JBS03**] that the database contains all elliptic curve of prime conductor < 234446, which implies that the smallest conductor rank 4 curve is composite.

10.8. Exercises

10.1 Prove Lemma 10.4.

10.2 Suppose $f \in S_2(\Gamma_0(N))$ is a newform and that $W_N(f) = wf$. Let $\Lambda(E, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(E, s)$. Prove that

$$\Lambda(E, s) = -w\Lambda(E, 2 - s).$$

[Hint: Show that $\Lambda(f, s) = \int_{0,\infty} f(iy/\sqrt{N})y^{s-1}\,dy$. Then substitute $1/y$ for y.]

10.3 Let $f = \sum a_n q^n \in \mathbb{C}[[q]]$ be a power series whose coefficients a_n together generate a number field K of degree d over $\mathbb{Q}$. Let V_f be the complex vector space spanned by the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f.

(a) Give an example to show that V_f need not have dimension d.

(b) Suppose V_f has dimension d. Prove that $V_f \cap \mathbb{Q}[[q]]$ is a $\mathbb{Q}$-vector space of dimension d.

10.4 Find an elliptic curve of conductor 11 using Section 10.7.

Chapter 11

Solutions to Selected Exercises

11.1. Chapter 1

(1) Exercise 1.1. Suppose $\gamma = \left(\begin{smallmatrix} a & nb \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ is a matrix with positive determinant. Then γ is a linear fractional transformation that fixes the real line, so it must either fix or swap the upper and lower half planes. Now

$$\gamma(i) = \frac{ai+b}{ci+d} = \frac{ac+bd+(ad-bc)i}{d^2+c^2},$$

so since $\det\gamma = ad-bc > 0$, the imaginary part of $\gamma(i)$ is positive; hence γ sends the upper half plane to itself.

(2) Exercise 1.2. Avoiding poles, the quotient rule for differentiation goes through exactly as in the real case, so any rational function $f(z) = p(z)/q(z)$ $(p, q \in \mathbb{C}[z])$ is holomorphic on $\mathbb{C} - \{\alpha : q(\alpha) = 0\}$. By the fundamental theorem of algebra, this set of poles is finite, and hence it is discrete. Write $q(z) = a_n(z-\alpha_1)^{r_1} \cdots (z-\alpha_k)^{r_k}$ for each α_i and let $q_i(z) = q(z)/(z-\alpha_i)^{r_i}$ which is a polynomial nonzero at α_i. Thus for each i we have $(z-\alpha_i)^{r_i} f(z) = p(z)/q'(z)$ is holomorphic at α_i and hence $f(z)$ is meromorphic on $\mathbb{C}$.

(3) Exercise 1.3.

(a) The product fg of two meromorphic functions on the upper half plane is itself meromorphic. Also, for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$(fg)^{[\gamma]_{k+j}} = \frac{1}{(cz+d)^{k+j}}((fg)\circ\gamma)$$
$$= \frac{1}{(cz+d)^{k}}(f\circ\gamma)\frac{1}{(cz+d)^{j}}(g\circ\gamma) = fg,$$

so fg is weakly modular.

(b) If f is meromorphic on the upper half plane, then so is $1/f$. Now

$$\frac{1}{f} = \frac{1}{(cz+d)^{-k}f\circ\gamma} = (cz+d)^{k}((1/f)\circ\gamma) = \frac{1}{f}^{[\gamma]_{-k}},$$

so $1/f$ is a weakly modular form of weight $-k$.

(c) Let f and g be modular functions. Then, as above, fg is a weakly modular function. Let $\sum_{n=m}^{\infty} a_n q^n$ and $\sum_{n=m'}^{\infty} b_n q^n$ be their q-expansions around any $\alpha \in \mathbb{P}^1(\mathbb{Q})$; then their formal product is the q-expansion of fg. But the formal product of two Laurent series about the same point is itself a Laurent series with convergence in the intersection of the convergent domains of the original series, so fg has a meromorphic q-expansion at each $\alpha \in \mathbb{P}^1(\mathbb{Q})$ and hence at each cusp.

(d) We are in exactly the same case as in part (c), but because f and g are modular functions, $m, m' \geq 0$ and hence the function is holomorphic at each of its cusps.

(4) Exercise 1.4. Let f be a weakly modular function of odd weight k. Since $\gamma = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we have $f(z) = (-1)^{-k}f(\gamma(z)) = -f(z)$ so $f = 0$.

(5) Exercise 1.5. Because $\mathrm{SL}_2(\mathbb{Z}/1\mathbb{Z})$ is the trivial group, $\Gamma(1) = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/1\mathbb{Z}))$ must be all of $\mathrm{SL}_2(\mathbb{Z})$. As $\mathrm{SL}_2(\mathbb{Z}) = \Gamma(1) \subset \Gamma_1(1) \subset \Gamma_0(1) \subset \mathrm{SL}_2(\mathbb{Z})$, we must have $\Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$.

(6) Exercise 1.6.

(a) The group $\Gamma_1(N)$ is the inverse image of the subgroup of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ generated by $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and the inverse image of a group (under a group homomorphism) is a group.

(b) The group contains the kernel of the homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, and that kernel has finite index since the quotient is contained in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, which is finite.

(c) Same argument as previous part.

(d) The level is at most N since both groups contain $\Gamma(N)$. It can be no greater than N since $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$ is in both groups.

(7) Exercise 1.7. See [**DS05**, Lemma 1.2.2].

(8) Exercise 1.8. Let $\alpha = p/q \in \mathbb{Q}$, where p and q are relatively prime. By the Euclidean algorithm, we can find $x, y \in \mathbb{Z}$ such that $px + qy = 1$. Let $\gamma_\alpha = \left(\begin{smallmatrix} p & -y \\ q & x \end{smallmatrix}\right)$. Note that $\gamma_\alpha \in \mathrm{SL}_2(\mathbb{Z})$ and $\gamma_\alpha(\infty) = \alpha$. Also let γ_∞ be the identity map on $\mathbb{P}^1(\mathbb{Q})$. Now γ_β^{-1} sends β to ∞ so we have $\gamma_\alpha \circ \gamma_\beta^{-1}$ which sends α to β.

11.2. Chapter 2

(1) Exercise 2.1. We have

$$\zeta(26) = \frac{1315862 \cdot \pi^{26}}{11094481976030578125}.$$

Variation: Compute $\zeta(28)$.

(2) Exercise 2.2. Omitted.

(3) Exercise 2.3.

$$\begin{aligned} E_8 &= -\frac{B_8}{16} + q + \sum_{n=2}^{\infty} \sigma_7(n) q^n \\ &= \frac{1}{480} + q + 129q^2 + 2188q^3 + \cdots . \end{aligned}$$

Variation: Compute E_{10}.

(4) Exercise 2.4. Omitted.

(5) Exercise 2.5. We have $d = \dim S_{28} = 2$. A choice of a, b with $4a + 6b \leq 14$ and $4a + 6b \equiv 4 \pmod{12}$ is $a = 1, b = 0$. A basis for S_{28} is then

$$\begin{aligned} g_1 &= \Delta F_6^{2(2-1)+0} F_4 = q - 792q^2 - 324q^3 + 67590208q^4 + \cdots , \\ g_2 &= \Delta^2 F_6^{2(2-2)+0} F_4 = q^2 + 192q^3 - 8280q^4 + \cdots . \end{aligned}$$

The Victor Miller basis is then

$$\begin{aligned} f_1 &= g_1 + 729 g_2 = q + 151740q^3 + 61032448q^4 + \cdots , \\ f_2 &= g_2 = q^2 + 192q^3 - 8280q^4 + \cdots . \end{aligned}$$

Variation: Compute the Victor Miller basis for S_{30}.

(6) Exercise 2.6. From the previous exercise we have $f = \Delta^2 F_4$. Then

$$\begin{aligned} f = \Delta^2 F_4 &= \left(\frac{F_4^3 - F_6^2}{-1728}\right)^2 \cdot F_4 \\ &= \left(\frac{\left(-\frac{8}{B_4}E_4\right)^3 - \left(-\frac{12}{B_6}E_6\right)^2}{-1728}\right)^2 \cdot \left(-\frac{8}{B_4}E_4\right) \\ &= 5186160 E_4 E_6^4 - 564480000 E_4^4 E_6^2 + 15360000000 E_4^7. \end{aligned}$$

(7) Exercise 2.7. No, it is not always integral. For example, for $k = 12$, the coefficient of q is $-2 \cdot 12/B_{12} = 65520/691 \notin \mathbb{Z}$. Variation: Find, with proof, the set of all k such that the normalized series F_k *is* integral (use that B_k is eventually very large compared to $2k$).

(8) Exercise 2.8. We compute the Victor Miller basis to precision great enough to determine T_2. This means we need up to $O(q^5)$.

$$\begin{aligned} f_0 &= 1 + 2611200q^3 + 19524758400q^4 + \cdots, \\ f_1 &= q + 50220q^3 + 87866368q^4 + \cdots, \\ f_2 &= q^2 + 432q^3 + 39960q^4 + \cdots. \end{aligned}$$

Then the matrix of T_2 on this basis is

$$\begin{pmatrix} 2147483649 & 0 & 19524758400 \\ 0 & 0 & 2235350016 \\ 0 & 1 & 39960 \end{pmatrix}.$$

(The rows of this matrix are the linear combinations that give the images of the f_i under T_2.) This matrix has characteristic polynomial

$$(x - 2147483649) \cdot (x^2 - 39960x - 2235350016).$$

11.3. Chapter 3

(1) Exercise 3.1. Write $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, so $\lambda' = \frac{a\lambda+b}{c\lambda+d}$. Let f be the isomorphism $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ given by $f(z) = z/(c\lambda + d)$. We have

$$f\left(\frac{1}{N}\right) = \frac{1}{N(c\lambda+d)} = \frac{a}{N} - \frac{c}{N} \cdot \frac{a\lambda+b}{c\lambda+d} \cong \frac{a}{N} \pmod{\mathbb{Z} + \mathbb{Z}\lambda'},$$

where the second equality can be verified easily by expanding out each side, and for the congruence we use that $N \mid c$. Thus the subgroup of $\mathbb{C}/\Lambda$ generated by $\frac{1}{N}$ is taken isomorphically to the subgroup of $\mathbb{C}/\Lambda'$ generated by $\frac{1}{N}$.

(2) Exercise 3.2. For any integer r, we have $\left(\begin{smallmatrix} 1 & r \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$, so $\{0, \infty\} = \{r, \infty\}$. Thus

$$0 = \{0,\infty\} - \{0,\infty\} = \{n,\infty\} - \{m,\infty\} = \{n,\infty\} + \{\infty, m\} = \{n, m\}.$$

(3) Exercise 3.3.
   (a) $(0 : 1), (1 : 0), (1 : 1), \dots, (1, p-1)$.
   (b) $p + 1$.
   (c) See [**Cre97a**, Prop. 2.2.1].

(4) Exercise 3.4. We start with $b = 4$, $a = 7$. Then $4 \cdot 2 \equiv 1 \pmod 7$. Let $\delta_1 = \left(\begin{smallmatrix} 4 & 1 \\ 7 & 2 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Since $\delta_1 \in \Gamma_0(7)$, we use the right coset representative $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and see that

$$\{0, 4/7\} = \{0, 1/2\} + \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \{0, \infty\}.$$

Repeating the process, we have $\delta_2 = \left(\begin{smallmatrix} 1 & 1 \\ 2 & 0 \end{smallmatrix}\right)$, which is in the same coset at $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Thus

$$\{0, 1/2\} = \left(\begin{smallmatrix} 0 & 6 \\ 1 & 0 \end{smallmatrix}\right) \{0, \infty\} + \{0, 0\}.$$

Putting it together gives

$$\{0, 4/7\} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \{0, \infty\} + \left(\begin{smallmatrix} 0 & 6 \\ 1 & 0 \end{smallmatrix}\right) \{0, \infty\} = [(0, 1)] + [(1, 0)].$$

(5) Exercise 3.5.
   (a) Coset representatives for $\Gamma_0(3)$ in $\mathrm{SL}_2(\mathbb{Z})$ are

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

   which we refer to below as $[r_0], [r_1], [r_2]$, and $[r_3]$, respectively.
   (b) In terms of representatives we have

$$\begin{aligned} [r_0] + [r_3] &= 0, & [r_0] + [r_3] + [r_2] &= 0, \\ [r_1] + [r_2] &= 0, & [r_1] + [r_1] + [r_1] &= 0, \\ [r_2] + [r_1] &= 0, & [r_2] + [r_0] + [r_3] &= 0, \\ [r_3] + [r_0] &= 0, & [r_3] + [r_2] + [r_0] &= 0. \end{aligned}$$

   (c) By the first three relations we have $[r_2] = [r_1] = 0 = 0[r_0]$ and $[r_3] = -1[r_0]$.
   (d)

$$\begin{aligned} T_2([r_0]) &= [r_0] \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right) + [r_0] \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right) + [r_0] \left(\begin{smallmatrix} 2 & 1 \\ 0 & 1 \end{smallmatrix}\right) + [r_0] \left(\begin{smallmatrix} 1 & 0 \\ 1 & 2 \end{smallmatrix}\right) \\ &= [\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)] + [\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right)] + [\left(\begin{smallmatrix} 2 & 1 \\ 0 & 1 \end{smallmatrix}\right)] + [\left(\begin{smallmatrix} 1 & 0 \\ 1 & 2 \end{smallmatrix}\right)] \\ &= [r_0] + [r_0] + [r_0] + [r_2] \\ &= 3[r_0]. \end{aligned}$$

11.4. Chapter 4

(1) Exercise 4.1. Suppose f is a Dirichlet character with modulus N. Then $-1 = f(-1) = f(-1+N) = 1$, a contradiction.

(2) Exercise 4.2.

(a) Any finite subgroup of the multiplicative group of a field is cyclic (since the number of roots of a polynomial over a field is at most its degree), so $(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic. Let g be an integer that reduces to a generator of $(\mathbb{Z}/p\mathbb{Z})^*$. Let $x = 1 + p \in (\mathbb{Z}/p^n\mathbb{Z})^*$; by the binomial theorem

$$x^{p^{n-2}} = 1 + p^{n-2} \cdot p + \cdots \equiv 1 + p^{n-1} \not\equiv 0 \pmod{p^n},$$

so x has order p^{n-1}. Since p is odd, $\gcd(p^{n-1}, p-1) = 1$, so xg has order $p^{n-1} \cdot (p-1) = \varphi(p^n)$; hence $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic.

(b) By the binomial theorem $(1+2^2)^{2^{n-3}} \not\equiv 1 \pmod{2^n}$, so 5 has order 2^{n-2} in $(\mathbb{Z}/2^n\mathbb{Z})^*$, and clearly -1 has order 2. Since $5 \equiv 1 \pmod 4$, -1 is not a power of 5 in $(\mathbb{Z}/2^n\mathbb{Z})^*$. Thus the subgroups $\langle -1 \rangle$ and $\langle 5 \rangle$ have trivial intersection. The product of their orders is $2^{n-1} = \varphi(2^n) = \#(\mathbb{Z}/2^n\mathbb{Z})^*$, so the claim follows.

(3) Exercise 4.3. Write $n = \prod p_i^{e_i}$. The order of g divides n, so the condition implies that $p_i^{e_i}$ divides the order of g for each i. Thus the order of g is divisible by the least common multiple of the $p_i^{e_i}$, i.e., by n.

(4) Exercise 4.4.

(a) The bijection given by $1 + p^{n-1}a \pmod{p^n} \mapsto a \pmod p$ is a homomorphism since

$$(1 + p^{n-1}a)(1 + p^{n-1}b) \equiv 1 + p^{n-1}(a+b) \pmod{p^n}.$$

(b) We have an exact sequence

$$1 \to 1 + p\mathbb{Z}/p^n\mathbb{Z} \to (\mathbb{Z}/p^n\mathbb{Z})^* \to (\mathbb{Z}/p\mathbb{Z})^* \to 1,$$

so it suffices to solve the discrete log problem in the kernel and cokernel. We prove by induction on n that we can solve the discrete log problem in the kernel easily (compared to known methods for solving the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$). We have an exact sequence

$$1 \to 1 + p^{n-1}\mathbb{Z}/p^n\mathbb{Z} \to (\mathbb{Z}/p^n\mathbb{Z})^* \to (\mathbb{Z}/p^{n-1}\mathbb{Z})^* \to 1.$$

The first part of this problem shows that we can solve the discrete log problem in the kernel, and by induction we can solve it in the cokernel. This completes the proof.

(5) Exercise 4.5. If $\varepsilon(5) = 1$, then since ε is nontrivial, Exercise 4.2 implies that ε factors through $(\mathbb{Z}/4\mathbb{Z})^*$, hence has conductor $4 = 2^{1+1}$, as claimed. If $\varepsilon(5) \neq 1$, then again from Exercise 4.2 we see that if ε has order r, then ε factors through $(\mathbb{Z}/2^{r+2}\mathbb{Z})^*$ but nothing smaller.

(6) Exercise 4.6.
   (a) Take $f = x^2 + 2$.
   (b) The element 2 has order 4.
   (c) A minimal generator for $(\mathbb{Z}/25\mathbb{Z})^*$ is 2, and the characters are [1], [2], [3], [4].
   (d) Each of the four Galois orbits has size 1.

11.5. Chapter 5

(1) Exercise 5.1. The eigenspace E_λ of A with eigenvalue λ is preserved by B, since if $v \in E_\lambda$, then

$$ABv = BAv = B(\lambda v) = \lambda Bv.$$

Because B is diagonalizable, its minimal polynomial equals its characteristic polynomial; hence the same is true for the restriction of B to E_λ, i.e., the restriction of B is diagonalizable. Choose basis for all E_λ so that the restrictions of B to these eigenspaces is diagonal with respect to these bases. Then the concatenation of these bases is a basis that simultaneously diagonalizes A and B.

(2) Exercise 5.2. When ε is the trivial character, the $B_{k,\varepsilon}$ are defined by

$$\sum_{a=1}^{1} \frac{\varepsilon(a)xe^{ax}}{e^x - 1} = \frac{xe^x}{e^x - 1} = x + \frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_{k,\varepsilon}\frac{x^k}{k!}.$$

Thus $B_{1,\varepsilon} = 1 + B_1 = \frac{1}{2}$, and for $k > 1$, we have $B_{k,\varepsilon} = B_k$.

(3) Exercise 5.3. Omitted.

(4) Exercise 5.4. The Eisenstein series in our basis for $E_3(\Gamma_1(13))$ are of the form $E_{3,1,\varepsilon}$ or $E_{3,\varepsilon,1}$ with $\varepsilon(-1) = (-1)^3 = -1$. There are six characters ε with modulus 13 such that $\varepsilon(-1) = -1$, and we have the two series $E_{3,1,\varepsilon}$ and $E_{3,\varepsilon,1}$ associated to each of these. This gives a dimension of 12.

11.6. Chapter 6

(1) Exercise 6.1.

(a) By Proposition 3.10, we have $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)] = \#\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. By the Chinese Remainder Theorem,
$$\#\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) = \prod_{p|N} \#\mathbb{P}^1(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z}).$$
So we are reduced to computing $\#\mathbb{P}^1(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})$. We have $(a,b) \in (\mathbb{Z}/p^n\mathbb{Z})^2$ with $\gcd(a,b,p) = 1$ if and only if $(a,b) \notin (p\mathbb{Z}/p^n\mathbb{Z})^2$, so there are $p^{2n} - p^{2(n-1)}$ such pairs. The unit group $(\mathbb{Z}/p^n\mathbb{Z})^*$ has order $\varphi(p^n) = p^n - p^{n-1}$. It follows that
$$\#\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) = \frac{p^{2n} - p^{2(n-1)}}{p^n - p^{n-1}} = p^n + p^{n-1}.$$

(b) Omitted.

(2) Exercise 6.2. Omitted.

(3) Exercise 6.3. Omitted.

(4) Exercise 6.4. Omitted.

(5) Exercise 6.5. See the source code to SAGE.

11.7. Chapter 7

(1) Exercise 7.1. Take a basis of W and let G be the matrix whose rows are these basis elements. Let B be the row echelon form of G. After a permutation p of columns, we may write $B = p_i(I|C)$, where I is the identity matrix. The matrix $A = p^{-1}(-C^t|I)$, where I is a different sized identity matrix, has the property that $W = \mathrm{Ker}(A)$.

(2) Exercise 7.2. The answer is no. For example if $A = nI$ is n times the identity matrix and if $p \mid n$, then $\mathrm{rref}(A \pmod{p}) = 0$ but $\mathrm{rref}(A) \pmod{p} = I$.

(3) Exercise 7.3. Let $T = \prod E_i$ be an invertible matrix such that $TA = E$ is in (reduced) echelon form and the E_i are elementary matrices, i.e., the result of applying an elementary row operation to the identity matrix. If p is a prime that does not divide any of the nonzero numerators or denominators of the entries of A and any E_i, then $\mathrm{rref}(A \pmod{p}) = \mathrm{rref}(A) \pmod{p}$. This is because $E \pmod{p}$ is in echelon form and $A \pmod{p}$ can be transformed to $E \pmod{p}$ via a series of elementary row operations modulo p.

(4) Exercise 7.4.

(a) The echelon form (over $\mathbb{Q}$) is
$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

  - (b) The kernel is the 1-dimensional span of $(1, -2, 1)$.
  - (c) The characteristic polynomial is $x \cdot (x^2 - 15x - 18)$.

(5) Exercise 7.5.
  - (a) The answer is given in the problem.
  - (b) See [**Coh93**, §2.4].

11.8. Chapter 8

(1) Exercise 8.1. Using the Chinese Remainder Theorem we immediately reduce to proving the statement when both $M = p^r$ and $N = p^s$ are powers of a prime p. Then $[a] \in (\mathbb{Z}/p^s\mathbb{Z})^*$ is represented by an integer a with $\gcd(a, p) = 1$. That same integer a defines an element of $(\mathbb{Z}/p^r\mathbb{Z})^*$ that reduces modulo p^s to $[a]$.

(2) Exercise 8.2. See [**Shi94**, Lemma 1.38].

(3) Exercise 8.3. Coset representatives for $\Gamma_1(3)$ are in bijection with (c, d) where $c, d \in \mathbb{Z}/3\mathbb{Z}$ and $\gcd(c, d, N) = 1$, so the following are representatives:

$$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 0 \\ 1 & 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 \\ 2 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 & 0 \\ 2 & 2 \end{smallmatrix}\right),$$

which we call $r_1, \ldots, r_8$, respectively. Now our Manin symbols are of the form $[X, r_i]$ and $[Y, r_i]$ for $1 \leq i \leq 8$ modulo the relations

$$x + x\sigma = 0, \quad x + x\tau + x\tau^2 = 0, \quad \text{and } x - xJ = 0.$$

First, note that J acts trivially on Manin symbols of odd weight because it sends X to $-X$, Y to $-Y$ and r_i to $-r_i$, so

$$[z, g]J = [-z, -g] = [z, g].$$

Thus the last relation is trivially true.

Now $\sigma^{-1}X = -Y$ and $\sigma^{-1}Y = X$. Also $\tau^{-1}X = -Y, \tau^{-1}Y = X - Y, \tau^{-2}X = -X + Y$ and $\tau^{-2}Y = -X$.

The first relation on the first symbol says that

$$[X, r_1] = -[-Y, r_3] = [Y, r_3]$$

and the second relation tells us that

$$[X, r_1] + [-Y, r_5] + [-X + Y, r_6] = 0.$$

(4) Exercise 8.4. Let $f \in S_k(\Gamma)$ and $g \in \Gamma$. All that remains to be shown is that this pairing respects the relation $x = xg$ for all modular symbols x. By linearity it suffices to show the invariance

of $\langle f, X^{k-i-2}Y^i\{\alpha,\beta\}\rangle$. We have

$$\begin{aligned}
&\left\langle f, (X^{k-2-i}Y^i\{\alpha,\beta\})g^{-1}\right\rangle \\
&= \left\langle f, (aX+bY)^{k-i-2}(cX+dY)^i\{g^{-1}(\alpha), g^{-1}(\beta)\}\right\rangle \\
&= \int_{g^{-1}(\alpha)}^{g^{-1}(\beta)} f(z)(az+b)^{k-i-2}(cz+d)^i\,dz \\
&= \int_{g^{-1}(\alpha)}^{g^{-1}(\beta)} f(z)\frac{(az+b)^{k-i-2}}{(cz+d)^{k-i-2}}(cz+d)^{k-2}\,dz \\
&= \int_{g^{-1}(\alpha)}^{g^{-1}(\beta)} f(z)\,g(z)^{k-i-2}(cz+d)^{k-2}\,dz \\
&= \int_{\alpha}^{\beta} f(g^{-1}(z))\,g(g^{-1}(z))^{k-i-2}(cg^{-1}(z)+d)^{k-2}\,d(g^{-1}(z)) \\
&= \int_{\alpha}^{\beta} f(g^{-1}(z))\,z^{k-i-2}(cg^{-1}(z)+d)^{k-2}\,(cg^{-1}(z)+d)^2\,dz \\
&= \int_{\alpha}^{\beta} f(z)\,z^{k-i-2}\,dz \\
&= \left\langle f, X^{k-i-2}Y^i\{\alpha,\beta\}\right\rangle,
\end{aligned}$$

where the second to last simplification is due to invariance under $[g]_k$, i.e.,

$$f(g^{-1}(z)) = f^{[g]_k}(g^{-1}(z)) = (cg^{-1}(z)+d)^{-k}f(g(g^{-1}(z))).$$

(The proof for $f \in \overline{S}_k(\Gamma)$ works in exactly the same way.)

(5) Exercise 8.5.

(a) Let $\eta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. For any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have

$$\gamma\eta = \begin{pmatrix} -a & b \\ -c & d \end{pmatrix},\ \eta\gamma = \begin{pmatrix} -a & -b \\ c & d \end{pmatrix},\ \text{and } \eta\gamma\eta = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}.$$

First, if $\gamma \in SL_2(\mathbb{Z})$, then $\eta\gamma\eta \in GL_2(\mathbb{Z})$ and

$$\det(\eta\gamma\eta) = \det\eta\det\gamma\det\eta = (-1)(1)(-1) = 1$$

so $\eta\gamma\eta \in SL_2(\mathbb{Z})$. As $\eta^2 = 1$, conjugation by η is self-inverse, so it must be a bijection.

Now if $\gamma \in \Gamma_0(N)$, then $c \equiv 0 \pmod N$, so $-c \equiv 0 \pmod N$, and so $\eta\gamma\eta \in \Gamma_0(N)$. Thus $\eta\Gamma_0(N)\eta = \Gamma_0(N)$.

If $\gamma \in \Gamma_1(N)$, then $-c \equiv 0 \pmod N$ as before and also $a \equiv d \equiv 1 \pmod N$, so $\eta\gamma\eta \in \Gamma_1(N)$. Thus $\eta\Gamma_1(N)\eta = \Gamma_1(N)$.

(b) Omitted.

11.9. Chapter 9

(1) Exercise 9.1. Consider the surjective homomorphism

$$r : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

Notice that $\Gamma_1(N)$ is the exact inverse image of the subgroup H of matrices of the form $\left(\begin{smallmatrix}1 & * \\ 0 & 1\end{smallmatrix}\right)$ and $\Gamma_0(N)$ is the inverse image of the subgroup T of upper triangular matrices. It thus suffices to observe that H is normal in T, which is clear. Finally, the quotient T/H is isomorphic to the group of diagonal matrices in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})^*$, which is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^*$.

(2) Exercise 9.2. It is enough to show $\langle p\rangle \in \mathbb{Z}[\ldots, T_n, \ldots]$ for primes p, since each $\langle d\rangle$ can be written in terms of the $\langle p\rangle$. Since $p \nmid N$, we have that

$$T_{p^2} = T_p^2 - \langle p\rangle p^{k-1},$$

so

$$\langle p\rangle p^{k-1} = T_p^2 - T_{p^2}.$$

By Dirichlet's theorem on primes in arithmetic progression, there is a prime $q \neq p$ congruent to p mod N. Since p^{k-1} and q^{k-1} are relatively prime, there exist integers a and b such that $ap^{k-1} + bq^{k-1} = 1$. Then

$$\langle p\rangle = \langle p\rangle(ap^{k-1} + bq^{k-1}) = a({T_p}^2 - T_{p^2}) + b({T_q}^2 - T_{q^2}) \in \mathbb{Z}[\ldots, T_n, \ldots].$$

(3) Exercise 9.3. Take $N = 33$. The space $S_2(\Gamma_0(33))$ is a direct sum of the two old subspaces coming from $S_2(\Gamma_0(11))$ and the new subspace, which has dimension 1. If f is a basis for $S_2(\Gamma_0(11))$ and g is a basis for $S_2(\Gamma_0(33))_{\mathrm{new}}$, then $\alpha_1(f), \alpha_3(f), g$ is a basis for $S_2(\Gamma_0(33))$ on which all Hecke operators T_n, with $\gcd(n, 33) = 1$, have diagonal matrix. However, the operator T_3 on $S_2(\Gamma_0(33))$ does not act as a scalar on $\alpha_1(f)$, so it cannot be in the ring generated by all operators T_n with $\gcd(n, 33) = 1$.

(4) Exercise 9.4. Omitted.

11.10. Chapter 10

(1) Exercise 10.1. Hint: Use either repeated integration by parts or a change of variables that relates the integral to the Γ function.

(2) Exercise 10.2. See [**Cre97a**, §2.8].

(3) Exercise 10.3.

   (a) Let $f = \sqrt{-1}\sum q^n$. Then $d = 2$, but the nontrivial conjugate of f is $-f$, so V_f has dimension 1.

(b) Choose $\alpha \in K$ such that $K = \mathbb{Q}(\alpha)$. Write

$$f = \sum_{i=0}^{d-1} \alpha^i g_i \tag{11.10.1}$$

with $g_i \in \mathbb{Q}[[q]]$. Let W_g be the $\mathbb{Q}$-span of the g_i, and let $W_f = V_f \cap \mathbb{Q}[[q]]$. By considering the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ conjugates of (11.10.1), we see that the Galois conjugates of f are in the $\mathbb{C}$-span of the g_i, so

$$d = \dim_{\mathbb{C}} V_f \leq \dim_{\mathbb{Q}} W_g. \tag{11.10.2}$$

Likewise, taking the above modulo $O(q^n)$ for any n, we obtain a matrix equation

$$F = AG,$$

where the columns of F are the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates of f, the matrix A is the Vandermonde matrix corresponding to α (and its $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ conjugates), and G has columns g_i. Since A is a Vandermonde matrix, it is invertible, so $A^{-1}F = G$. Taking the limit as n goes to infinity, we see that each g_i is a linear combination of the f_i, hence an element of V_f. Thus $W_g \subset W_f$, so (11.10.2) implies that $\dim_{\mathbb{Q}} W_f \geq d$. But $W_f \otimes_{\mathbb{Q}} \mathbb{C} \subset V_f$ so finally

$$d \leq \dim_{\mathbb{Q}} W_f = \dim_{\mathbb{C}}(W_f \otimes_{\mathbb{Q}} \mathbb{C}) \leq \dim_{\mathbb{C}} V_f = d.$$

(4) Exercise 10.4. See the appendix to Chapter II in [**Cre97a**], where this example is worked out in complete detail.

Appendix A

Computing in Higher Rank

by Paul E. Gunnells

A.1. Introduction

This book has addressed the theoretical and practical problems of performing computations with modular forms. Modular forms are the simplest examples of the general theory of automorphic forms attached to a reductive algebraic group G with an arithmetic subgroup Γ; they are the case $G = \mathrm{SL}_2(\mathbb{R})$ with Γ a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. For such pairs (G, Γ) the Langlands philosophy asserts that there should be deep connections between automorphic forms and arithmetic, connections that are revealed through the action of the Hecke operators on spaces of automorphic forms. There have been many profound advances in recent years in our understanding of these phenomena, for example:

- the establishment of the modularity of elliptic curves defined over $\mathbb{Q}$ [**Wil95, TW95, Dia96, CDT99, BCDT01**],
- the proof by Harris–Taylor of the local Langlands correspondence [**HT01**], and
- Lafforgue's proof of the global Langlands correspondence for function fields [**Laf02**].

Nevertheless, we are still far from seeing that the links between automorphic forms and arithmetic hold in the broad scope in which they are generally believed. Hence one has the natural problem of studying spaces of automorphic forms computationally.

The goal of this appendix is to describe some computational techniques for automorphic forms. We focus on the case $G = \mathrm{SL}_n(\mathbb{R})$ and $\Gamma \subset \mathrm{SL}_n(\mathbb{Z})$, since the automorphic forms that arise are one natural generalization of modular forms, and since this is the setting for which we have the most tools available. In fact, we do not work directly with automorphic forms, but rather with the cohomology of the arithmetic group Γ with certain coefficient modules. This is the most natural generalization of the tools developed in previous chapters.

Here is a brief overview of the contents. Section A.2 gives background on automorphic forms and the cohomology of arithmetic groups and explains why the two are related. In Section A.3 we describe the basic topological tools used to compute the cohomology of Γ explicitly. Section A.4 defines the Hecke operators, describes the generalization of the modular symbols from Chapter 8 to higher rank, and explains how to compute the action of the Hecke operators on the top degree cohomology group. Section A.5 discusses computation of the Hecke action on cohomology groups below the top degree. Finally, Section A.6 briefly discusses some related material and presents some open problems.

A.1.1. The theory of automorphic forms is notorious for the difficulty of its prerequisites. Even if one is only interested in the cohomology of arithmetic groups—a small part of the full theory—one needs considerable background in algebraic groups, algebraic topology, and representation theory. This is somewhat reflected in our presentation, which falls far short of being self-contained. Indeed, a complete account would require a long book of its own. We have chosen to sketch the foundational material and to provide many pointers to the literature; good general references are [**BW00, Harb, LS90, Vog97**]. We hope that the energetic reader will follow the references and fill many gaps on his/her own.

The choice of topics presented here is heavily influenced (as usual) by the author's interests and expertise. There are many computational topics in the cohomology of arithmetic groups we have completely omitted, including the trace formula in its many incarnations [**GP05**], the explicit Jacquet–Langlands correspondence [**Dem04, SW05**], and moduli space techniques [**FvdG, vdG**]. We encourage the reader to investigate these extremely interesting and useful techniques.

A.1.2. Acknowledgements. I thank Avner Ash, John Cremona, Mark McConnell, and Dan Yasaki for helpful comments. I also thank the NSF for support.

A.2. Automorphic Forms and Arithmetic Groups

A.2.1. Let $\Gamma = \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$ be the usual Hecke congruence subgroup of matrices upper-triangular mod N. Let $Y_0(N)$ be the modular curve $\Gamma \backslash \mathfrak{h}$, and let $X_0(N)$ be its canonical compactification obtained by adjoining cusps. For any integer $k \geq 2$, let $S_k(N)$ be the space of weight k holomorphic cuspidal modular forms on Γ. According to Eichler–Shimura [**Shi94**, Chapter 8], we have the isomorphism

$$H^1(X_0(N); \mathbb{C}) \xrightarrow{\sim} S_2(N) \oplus \overline{S_2(N)}, \tag{A.2.1}$$

where the bar denotes complex conjugation and where the isomorphism is one of Hecke modules.

More generally, for any integer $n \geq 0$, let $M_n \subset \mathbb{C}[x, y]$ be the subspace of degree n homogeneous polynomials. The space M_n admits a representation of Γ by the "change of variables" map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot p(x, y) = p(dx - by, -cx + ay). \tag{A.2.2}$$

This induces a local system $\widetilde{M}_n$ on the curve $X_0(N)$.[1] Then the analogue of (A.2.1) for higher-weight modular forms is the isomorphism

$$H^1(X_0(N); \widetilde{M}_{k-2}) \xrightarrow{\sim} S_k(N) \oplus \overline{S_k(N)}. \tag{A.2.3}$$

Note that (A.2.3) reduces to (A.2.1) when $k = 2$.

Similar considerations apply if we work with the open curve $Y_0(N)$ instead, except that Eisenstein series also contribute to the cohomology. More precisely, let $E_k(N)$ be the space of weight k Eisenstein series on $\Gamma_0(N)$. Then (A.2.3) becomes

$$H^1(Y_0(N); \widetilde{M}_{k-2}) \xrightarrow{\sim} S_k(N) \oplus \overline{S_k(N)} \oplus E_k(N). \tag{A.2.4}$$

These isomorphisms lie at the heart of the modular symbols method.

A.2.2. The first step on the path to general automorphic forms is a reinterpretation of modular forms in terms of functions on $\mathrm{SL}_2(\mathbb{R})$. Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup. A weight k modular form on Γ is a holomorphic function $f \colon \mathfrak{h} \to \mathbb{C}$ satisfying the transformation property

$$f((az + b)/(cz + d)) = j(\gamma, z)^k f(z), \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \quad z \in \mathfrak{h}.$$

[1] The classic references for cohomology with local systems are [**Ste99a**, Section 31] and [**Eil47**, Ch. V]. A more recent exposition (in the language of Čech cohomology and locally constant sheaves) can be found in [**BT82**, II.13]. For an exposition tailored to our needs, see [**Harb**, Section 2.9].

Here $j(\gamma, z)$ is the *automorphy factor* $cz + d$. There are some additional conditions f must satisfy at the cusps of $\mathfrak{h}$, but these are not so important for our discussion.

The group $G = \mathrm{SL}_2(\mathbb{R})$ acts transitively on $\mathfrak{h}$, with the subgroup $K = \mathrm{SO}(2)$ fixing i. Thus $\mathfrak{h}$ can be written as the quotient G/K. From this, we see that f can be viewed as a function $G \to \mathbb{C}$ that is K-*invariant on the right* and that satisfies a certain symmetry condition with respect to the Γ-*action on the left*. Of course not every f with these properties is a modular form: some extra data is needed to take the role of holomorphicity and to handle the behavior at the cusps. Again, this can be ignored right now.

We can turn this interpretation around as follows. Suppose φ is a function $G \to \mathbb{C}$ that is Γ-*invariant on the left*, that is, $\varphi(\gamma g) = \varphi(g)$ for all $\gamma \in \Gamma$. Hence φ can be thought of as a function $\varphi\colon \Gamma\backslash G \to \mathbb{C}$. We further suppose that φ satisfies a certain symmetry condition with respect to the K-*action on the right.* In particular, any matrix $m \in K$ can be written

$$m = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in \mathbb{R}, \tag{A.2.5}$$

with θ uniquely determined modulo 2π. Let ζ_m be the complex number $e^{i\theta}$. Then the K-symmetry we require is

$$\varphi(gm) = \zeta_m^{-k}\varphi(z), \quad m \in K,$$

where k is some fixed nonnegative integer.

It turns out that such functions φ are very closely related to modular forms: *any* $f \in S_k(\Gamma)$ uniquely determines such a function $\varphi_f\colon \Gamma\backslash G \to \mathbb{C}$. The correspondence is very simple. Given a weight k modular form f, define

$$\varphi_f(g) := f(g \cdot i) j(g, i)^{-k}. \tag{A.2.6}$$

We claim φ_f is left Γ-invariant and satisfies the desired K-symmetry on the right. Indeed, since j satisfies the cocycle property

$$j(gh, z) = j(g, h \cdot z) j(h, z),$$

we have

$$\varphi_f(\gamma g) = f((\gamma g)\cdot i) j(\gamma g, i)^{-k} = j(\gamma, g\cdot i)^k f(g\cdot i) j(\gamma, g\cdot i)^{-k} j(g, i)^{-k} = \varphi_f(g).$$

Moreover, any $m \in K$ stabilizes i. Hence

$$\varphi_f(gm) = f((gm) \cdot i) j(gm, i)^{-k} = f(g \cdot i) j(m, i)^{-k} j(g, m \cdot i)^{-k}.$$

From (A.2.5) we have $j(m, i)^{-k} = (\cos\theta + i\sin\theta)^{-k} = \zeta_m^{-k}$, and thus $\varphi_f(gm) = \zeta_m^{-k}\varphi_f(g)$.

Hence in (A.2.6) the weight and the automorphy factor "untwist" the Γ-action to make φ_f left Γ-invariant. The upshot is that we can study modular forms by studying the spaces of functions that arise through the construction (A.2.6).

Of course, not every $\varphi\colon \Gamma\backslash G \to \mathbb{C}$ will arise as φ_f for some $f \in S_K(\Gamma)$: after all, f is holomorphic and satisfies rather stringent growth conditions. Pinning down all the requirements is somewhat technical and is (mostly) done in the sequel.

A.2.3. Before we define automorphic forms, we need to find the correct generalizations of our groups $\mathrm{SL}_2(\mathbb{R})$ and $\Gamma_0(N)$. The correct setup is rather technical, but this really reflects the power of the general theory, which handles so many different situations (e.g., Maass forms, Hilbert modular forms, Siegel modular forms, etc.).

Let G be a connected Lie group, and let $K \subset G$ be a maximal compact subgroup. We assume that G is the set of real points of a connected semisimple algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$. These conditions mean the following [**PR94**, §2.1.1]:

(1) The group $\mathbf{G}$ has the structure of an affine algebraic variety given by an ideal I in the ring $R = \mathbb{C}[x_{ij}, D^{-1}]$, where the variables $\{x_{ij} \mid 1 \leq i, j \leq n\}$ should be interpreted as the entries of an "indeterminate matrix," and D is the polynomial $\det(x_{ij})$. Both the group multiplication $\mathbf{G} \times \mathbf{G} \to \mathbf{G}$ and inversion $\mathbf{G} \to \mathbf{G}$ are required to be morphisms of algebraic varieties.

The ring R is the coordinate ring of the algebraic group GL_n. Hence this condition means that $\mathbf{G}$ can be essentially viewed as a subgroup of $\mathrm{GL}_n(\mathbb{C})$ defined by polynomial equations in the matrix entries of the latter.

(2) *Defined over* $\mathbb{Q}$ means that I is generated by polynomials with rational coefficients.

(3) *Connected* means that $\mathbf{G}$ is connected as an algebraic variety.

(4) *Set of real points* means that G is the set of real solutions to the equations determined by I. We write $G = \mathbf{G}(\mathbb{R})$.

(5) *Semisimple* means that the maximal connected solvable normal subgroup of $\mathbf{G}$ is trivial.

Example A.1. The most important example for our purposes is the *split form of* SL_n. For this choice we have

$$G = \mathrm{SL}_n(\mathbb{R}) \text{ and } K = \mathrm{SO}(n).$$

Example A.2. Let $F/\mathbb{Q}$ be a number field. Then there is a $\mathbb{Q}$-group $\mathbf{G}$ such that $\mathbf{G}(\mathbb{Q}) = \mathrm{SL}_n(F)$. The group $\mathbf{G}$ is constructed as $\mathbf{R}_{F/\mathbb{Q}}(\mathrm{SL}_n)$, where $\mathbf{R}_{F/\mathbb{Q}}$ denotes the *restriction of scalars* from F to $\mathbb{Q}$ [**PR94**, §2.1.2]. For example, if F is totally real, the group $\mathbf{R}_{F/\mathbb{Q}}(\mathrm{SL}_2)$ appears when one studies Hilbert modular forms.

Let (r, s) be the signature of the field F, so that $F \otimes \mathbb{R} \simeq \mathbb{R}^r \times \mathbb{C}^s$. Then $G = \mathrm{SL}_n(\mathbb{R})^r \times \mathrm{SL}_n(\mathbb{C})^s$ and $K = \mathrm{SO}(n)^r \times \mathrm{SU}(n)^s$.

Example A.3. Another important example is the *split symplectic group* Sp_{2n}. This is the group that arises when one studies Siegel modular forms. The group of real points $\mathrm{Sp}_{2n}(\mathbb{R})$ is the subgroup of $\mathrm{SL}_{2n}(\mathbb{R})$ preserving a fixed nondegenerate alternating bilinear form on $\mathbb{R}^{2n}$. We have $K = \mathrm{U}(n)$.

A.2.4. To generalize $\Gamma_0(N)$, we need the notion of an *arithmetic group*. This is a discrete subgroup Γ of the group of rational points $\mathbf{G}(\mathbb{Q})$ that is commensurable with the set of integral points $\mathbf{G}(\mathbb{Z})$. Here commensurable simply means that $\Gamma \cap \mathbf{G}(\mathbb{Z})$ is a finite index subgroup of both Γ and $\mathbf{G}(\mathbb{Z})$; in particular $\mathbf{G}(\mathbb{Z})$ itself is an arithmetic group.

Example A.4. For the split form of SL_n we have $\mathbf{G}(\mathbb{Z}) = \mathrm{SL}_n(\mathbb{Z}) \subset \mathbf{G}(\mathbb{Q}) = \mathrm{SL}_n(\mathbb{Q})$. A trivial way to obtain other arithmetic groups is by conjugation: if $g \in \mathrm{SL}_n(\mathbb{Q})$, then $g \cdot \mathrm{SL}_n(\mathbb{Z}) \cdot g^{-1}$ is also arithmetic.

A more interesting collection of examples is given by the congruence subgroups. The *principal congruence subgroup* $\Gamma(N)$ is the group of matrices congruent to the identity modulo N for some fixed integer $N \geq 1$. A *congruence subgroup* is a group containing $\Gamma(N)$ for some N.

In higher dimensions there are many candidates to generalize the Hecke subgroup $\Gamma_0(N)$. For example, one can take the subgroup of $\mathrm{SL}_n(\mathbb{Z})$ that is upper-triangular mod N. From a computational perspective, this choice is not so good since its index in $\mathrm{SL}_n(\mathbb{Z})$ is large. A better choice, and the one that usually appears in the literature, is to define $\Gamma_0(N)$ to be the subgroup of $\mathrm{SL}_n(\mathbb{Z})$ with bottom row congruent to $(0, \ldots, 0, *) \mod N$.

A.2.5. We are almost ready to define automorphic forms. Let $\mathfrak{g}$ be the Lie algebra of G, and let $U(\mathfrak{g})$ be its universal enveloping algebra over $\mathbb{C}$. Geometrically, $\mathfrak{g}$ is just the tangent space at the identity of the smooth manifold G. The algebra $U(\mathfrak{g})$ is a certain complex associative algebra canonically built from $\mathfrak{g}$. The usual definition would lead us a bit far afield, so we will settle for an equivalent characterization: $U(\mathfrak{g})$ can be realized as a certain subalgebra of the ring of differential operators on $C^\infty(G)$, the space of smooth functions on G.

In particular, G acts on $C^\infty(G)$ by *left translations*: given $g \in G$ and $f \in C^\infty(G)$, we define

$$L_g(f)(x) := f(g^{-1}x).$$

Then $U(\mathfrak{g})$ can be identified with the ring of all differential operators on $C^\infty(G)$ that are invariant under left translation. For our purposes the most important part of $U(\mathfrak{g})$ is its center $Z(\mathfrak{g})$. In terms of differential operators, $Z(\mathfrak{g})$ consists of those operators that are also invariant under *right translation*:

$$R_g(f)(x) := f(xg).$$

Definition A.5. An *automorphic form* on G with respect to Γ is a function $\varphi\colon G \to \mathbb{C}$ satisfying

(1) $\varphi(\gamma g) = \varphi(g)$ for all $\gamma \in \Gamma$,

(2) the right translates $\{\varphi(gk) \mid k \in K\}$ span a finite-dimensional space ξ of functions,

(3) there exists an ideal $J \subset Z(\mathfrak{g})$ of finite codimension such that J annihilates φ, and

(4) φ satisfies a certain growth condition that we do not wish to make precise. (In the literature, φ is said to be *slowly increasing*.)

For fixed ξ and J, we denote by $\mathscr{A}(\Gamma, \xi, J, K)$ the space of all functions satisfying the above four conditions. It is a basic theorem, due to Harish-Chandra [**HC68**], that $\mathscr{A}(\Gamma, \xi, J, K)$ is finite-dimensional.

Example A.6. We can identify the cuspidal modular forms $S_k(N)$ in the language of Definition A.5. Given a modular form f, let $\varphi_f \in C^\infty(\mathrm{SL}_2(\mathbb{R}))$ be the function from (A.2.6). Then the map $f \mapsto \varphi_f$ identifies $S_k(N)$ with the subspace $\mathscr{A}_k(N)$ of functions φ satisfying

(1) $\varphi(\gamma g) = \varphi(g)$ for all $\gamma \in \Gamma_0(N)$,

(2) $\varphi(gm) = \zeta_m^{-k}\varphi(g)$ for all $m \in \mathrm{SO}(2)$,

(3) $(\Delta + \lambda_k)\varphi = 0$, where $\Delta \in Z(\mathfrak{g})$ is the *Laplace–Beltrami–Casimir operator* and

$$\lambda_k = \frac{k}{2}\left(\frac{k}{2} - 1\right),$$

(4) φ is slowly increasing, and

(5) φ is *cuspidal*.

The first four conditions parallel Definition A.5. Item (1) is the Γ-invariance. Item (2) implies that the right translates of φ by SO(2) lie in a fixed finite-dimensional representation of SO(2). Item (3) is how holomorphicity appears, namely that φ is killed by a certain differential operator. Finally, item (4) is the usual growth condition.

The only condition missing from the general definition is (5), which is an extra constraint placed on φ to ensure that it comes from a cusp form. This condition can be expressed by the vanishing of certain integrals ("constant terms"); for details we refer to [**Bum97, Gel75**].

Example A.7. Another important example appears when we set $k = 0$ in (2) in Example A.6 and relax (3) by requiring only that $(\Delta - \lambda)\varphi = 0$ for *some* nonzero $\lambda \in \mathbb{R}$. Such automorphic forms cannot possibly arise from modular forms, since there are no nontrivial cusp forms of weight 0. However, there are plenty of solutions to these conditions: they correspond to *real-analytic* cuspidal modular forms of weight 0 and are known as *Maass forms*. Traditionally one writes $\lambda = (1 - s^2)/4$. The positivity of Δ implies that $s \in (-1, 1)$ or is purely imaginary.

Maass forms are highly elusive objects. Selberg proved that there are infinitely many linearly independent Maass forms of full level (i.e., on $\mathrm{SL}_2(\mathbb{Z})$), but to this date no explicit construction of a single one is known. (Selberg's argument is indirect and relies on the trace formula; for an exposition see [**Sar03**].) For higher levels some explicit examples can be constructed using theta series attached to indefinite quadratic forms [**Vig77**]. Numerically Maass forms have been well studied; see for example [**FL**].

In general the arithmetic nature of the eigenvalues λ that correspond to Maass forms is unknown, although a famous conjecture of Selberg states that for congruence subgroups they satisfy the inequality $\lambda \geq 1/4$ (in other words, only purely imaginary s appear above). The truth of this conjecture would have far-reaching consequences, from analytic number theory to graph theory [**Lub94**].

A.2.6. As Example A.6 indicates, there is a notion of *cuspidal automorphic form*. The exact definition is too technical to state here, but it involves an appropriate generalization of the notion of constant term familiar from modular forms.

There are also *Eisenstein series* [**Lan66, Art79**]. Again the complete definition is technical; we only mention that there are different types of Eisenstein series corresponding to certain subgroups of G. The Eisenstein series that are easiest to understand are those built from cusp forms on lower rank groups. Very explicit formulas for Eisenstein series on GL_3 can be seen in [**Bum84**]. For a down-to-earth exposition of some of the Eisenstein series on GL_n, we refer to [**Gol05**].

The decomposition of $M_k(\Gamma_0(N))$ into cusp forms and Eisenstein series also generalizes to a general group G, although the statement is much more complicated. The result is a theorem of Langlands [**Lan76**] known as the *spectral decomposition of* $L^2(\Gamma\backslash G)$. A thorough recent presentation of this can be found in [**MW94**].

A.2.7. Let $\mathscr{A} = \mathscr{A}(\Gamma, K)$ be the space of all automorphic forms, where ξ and J range over all possibilities. The space $\mathscr{A}$ is huge, and the arithmetic significance of much of it is unknown. This is already apparent for $G =$

$\mathrm{SL}_2(\mathbb{R})$. The automorphic forms directly connected with arithmetic are the holomorphic modular forms, not the Maass forms[2]. Thus the question arises: which automorphic forms in $\mathscr{A}$ are the most natural generalization of the modular forms?

One answer is provided by the isomorphisms (A.2.1), (A.2.3), (A.2.4). These show that modular forms appear naturally in the cohomology of modular curves. Hence a reasonable approach is to generalize the *left* of (A.2.1), (A.2.3), (A.2.4), and to study the resulting cohomology groups. This is the approach we will take. One drawback is that it is not obvious that our generalization has anything to do with automorphic forms, but we will see eventually that it certainly does. So we begin by looking for an appropriate generalization of the modular curve $Y_0(N)$.

Let G and K be as in Section A.2.3, and let X be the quotient G/K. This is a global Riemannian symmetric space [**Hel01**]. One can prove that X is contractible. Any arithmetic group $\Gamma \subset G$ acts on X properly discontinuously. In particular, if Γ is torsion-free, then the quotient $\Gamma\backslash X$ is a smooth manifold.

Unlike the modular curves, $\Gamma\backslash X$ will not have a complex structure in general[3]; nevertheless, $\Gamma\backslash X$ is a very nice space. In particular, if Γ is torsion-free, it is an *Eilenberg–Mac Lane* space for Γ, otherwise known as a $K(\Gamma, 1)$. This means that the only nontrivial homotopy group of $\Gamma\backslash X$ is its fundamental group, which is isomorphic to Γ, and that the universal cover of $\Gamma\backslash X$ is contractible. Hence $\Gamma\backslash X$ is in some sense a "topological incarnation"[4] of Γ.

This leads us to the notion of the *group cohomology* $H^*(\Gamma;\mathbb{C})$ of Γ with trivial complex coefficients. In the early days of algebraic topology, this was defined to be the complex cohomology of an Eilenberg–Mac Lane space for Γ [**Bro94**, Introduction, I.4]:

$$H^*(\Gamma;\mathbb{C}) = H^*(\Gamma\backslash X;\mathbb{C}). \tag{A.2.7}$$

Today there are purely algebraic approaches to $H^*(\Gamma;\mathbb{C})$ [**Bro94**, III.1], but for our purposes (A.2.7) is exactly what we need. In fact, the group cohomology $H^*(\Gamma;\mathbb{C})$ can be identified with the cohomology of the quotient $\Gamma\backslash X$ even if Γ has torsion, since we are working with complex coefficients. The cohomology groups $H^*(\Gamma;\mathbb{C})$, where Γ is an arithmetic group, are our proposed generalization for the weight 2 modular forms.

What about higher weights? For this we must replace the trivial coefficient module $\mathbb{C}$ with local systems, just as we did in (A.2.3). For our

[2] However, Maass forms play a very important *indirect* role in arithmetic.

[3] The symmetric spaces that have a complex structure are known as *bounded domains*, or *Hermitian symmetric spaces* [**Hel01**].

[4] This apt phrase is due to Vogan [**Vog97**].

purposes it is enough to let $\mathscr{M}$ be a rational finite-dimensional representation of G over the complex numbers. Any such $\mathscr{M}$ gives a representation of $\Gamma \subset G$ and thus induces a local system $\widetilde{\mathscr{M}}$ on $\Gamma\backslash X$. As before, the group cohomology $H^*(\Gamma;\mathscr{M})$ is the cohomology $H^*(\Gamma\backslash X;\widetilde{\mathscr{M}})$. In (A.2.3) we took $\mathscr{M} = M_n$, the nth symmetric power of the standard representation. For a general group G there are many kinds of representations to consider. In any case, we contend that the cohomology spaces

$$H^*(\Gamma;\mathscr{M}) = H^*(\Gamma\backslash X;\widetilde{\mathscr{M}})$$

are a good generalization of the spaces of modular forms.

A.2.8. It is certainly not obvious that the cohomology groups $H^*(\Gamma;\mathscr{M})$ have *anything* to do with automorphic forms, although the isomorphisms (A.2.1), (A.2.3), (A.2.4) look promising.

The connection is provided by a deep theorem of Franke [**Fra98**], which asserts that

(1) the cohomology groups $H^*(\Gamma;\mathscr{M})$ can be directly computed in terms of certain automorphic forms (the automorphic forms of "cohomological type," also known as those with "nonvanishing $(\mathfrak{g}, K)$ cohomology" [**VZ84**]); and

(2) there is a direct sum decomposition

$$H^*(\Gamma;\mathscr{M}) = H^*_{\mathrm{cusp}}(\Gamma;\mathscr{M}) \oplus \bigoplus_{\{P\}} H^*_{\{P\}}(\Gamma;\mathscr{M}), \tag{A.2.8}$$

where the sum is taken over the set of classes of *associate proper* $\mathbb{Q}$*-parabolic subgroups of* G.

The precise version of statement (1) is known in the literature as the *Borel conjecture*. Statement (2) parallels Langlands's spectral decomposition of $L^2(\Gamma\backslash G)$.

Example A.8. For $\Gamma = \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$, the decomposition (A.2.8) is exactly (A.2.4). The cusp forms $S_k(N) \oplus \overline{S_k(N)}$ correspond to the summand $H^1_{\mathrm{cusp}}(\Gamma;\mathscr{M})$. There is one class of proper $\mathbb{Q}$-parabolic subgroups in $\mathrm{SL}_2(\mathbb{R})$, represented by the Borel subgroup of upper-triangular matrices. Hence only one term appears in big direct sum on the right of (A.2.8), which is the Eisenstein term E_k.

The summand $H^*_{\mathrm{cusp}}(\Gamma;\mathscr{M})$ of (A.2.8) is called the *cuspidal cohomology*; this is the subspace of classes represented by cuspidal automorphic forms. The remaining summands constitute the *Eisenstein cohomology* of Γ [**Har91**]. In particular the summand indexed by $\{P\}$ is constructed using Eisenstein series attached to certain cuspidal automorphic forms on lower rank groups. Hence $H^*_{\mathrm{cusp}}(\Gamma;\mathscr{M})$ is in some sense the most important part of the cohomology: all the rest can be built systematically from cuspidal cohomology on lower rank groups[5]. This leads us to our basic computational problem:

Problem A.9. Develop tools to compute explicitly the cohomology spaces $H^*(\Gamma;\mathscr{M})$ and to identify the cuspidal subspace $H^*_{\mathrm{cusp}}(\Gamma;\mathscr{M})$.

A.3. Combinatorial Models for Group Cohomology

A.3.1. In this section, we restrict attention to $G = \mathrm{SL}_n(\mathbb{R})$ and Γ, a congruence subgroup of $\mathrm{SL}_n(\mathbb{Z})$. By the previous section, we can study the group cohomology $H^*(\Gamma;\mathscr{M})$ by studying the cohomology $H^*(\Gamma\backslash X;\widetilde{\mathscr{M}})$. The latter spaces can be studied using standard topological techniques, such as taking the cohomology of complexes associated to cellular decompositions of $\Gamma\backslash X$. For $\mathrm{SL}_n(\mathbb{R})$, one can construct such decompositions using a version of explicit reduction theory of real positive-definite quadratic forms due to Voronoĭ [**Vor08**]. The goal of this section is to explain how this is done. We also discuss how the cohomology can be explicitly studied for congruence subgroups of $\mathrm{SL}_3(\mathbb{Z})$.

A.3.2. Let V be the $\mathbb{R}$-vector space of all symmetric $n \times n$ matrices, and let $C \subset V$ be the subset of positive-definite matrices. The space C can be identified with the space of all real positive-definite quadratic forms in n variables: in coordinates, if $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ (column vector), then the matrix $A \in C$ induces the quadratic form

$$x \longmapsto x^t A x,$$

and it is well known that any positive-definite quadratic form arises in this way. The space C is a cone, in that it is preserved by homotheties: if $x \in C$, then $\lambda x \in C$ for all $\lambda \in \mathbb{R}_{>0}$. It is also convex: if $x_1, x_2 \in C$, then $tx_1 + (1-t)x_2 \in C$ for $t \in [0,1]$. Let D be the quotient of C by homotheties.

Example A.10. The case $n = 2$ is illustrative. We can take coordinates on $V \simeq \mathbb{R}^3$ by representing any matrix in V as

$$\begin{pmatrix} x & y \\ y & z \end{pmatrix}, \quad x, y, z \in \mathbb{R}.$$

The subset of singular matrices $Q = \{xz - y^2 = 0\}$ is a quadric cone in V dividing the complement $V \smallsetminus Q$ into three connected components. The component containing the identity matrix is the cone C of positive-definite matrices. The quotient D can be identified with an open 2-disk.

[5] This is a bit of an oversimplification, since it is a highly nontrivial problem to decide when cusp cohomology from lower rank groups appears in Γ. However, many results are known; as a selection we mention [**Har91, Har87, LS04**]

The group G acts on C on the left by

$$(g, c) \longmapsto gcg^t.$$

This action commutes with that of the homotheties and thus descends to a G-action on D. One can show that G acts transitively on D and that the stabilizer of the image of the identity matrix is $K = \mathrm{SO}(n)$. Hence we may identify D with our symmetric space $X = \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}(n)$. We will do this in the sequel, using the notation D when we want to emphasize the coordinates coming from the linear structure of $C \subset V$ and using the notation X for the quotient G/K.

We can make the identification $D \simeq X$ more explicit. If $g \in \mathrm{SL}_n(\mathbb{R})$, then the map

$$g \longmapsto gg^t \tag{A.3.1}$$

takes g to a symmetric positive-definite matrix. Any coset gK is taken to the same matrix since $KK^t = \mathrm{Id}$. Thus (A.3.1) identifies G/K with a subset C_1 of C, namely those positive-definite symmetric matrices with determinant 1. It is easy to see that C_1 maps diffeomorphically onto D.

The inverse map $C_1 \to G/K$ is more complicated. Given a determinant 1 positive-definite symmetric matrix A, one must find $g \in \mathrm{SL}_n(\mathbb{R})$ such that $gg^t = A$. Such a representation always exists, with g determined uniquely up to right multiplication by an element of K. In computational linear algebra, such a g can be constructed through *Cholesky decomposition* of A.

The group $\mathrm{SL}_n(\mathbb{Z})$ acts on C via the G-action and does so properly discontinuously. This is the "unimodular change of variables" action on quadratic forms [**Ser73**, V.1.1]. Under our identification of D with X, this is the usual action of $\mathrm{SL}_n(\mathbb{Z})$ by left translation from Section A.2.7.

A.3.3. Now consider the group cohomology $H^*(\Gamma; \mathscr{M}) = H^*(\Gamma\backslash X; \widetilde{\mathscr{M}})$. The identification $D \simeq X$ shows that the dimension of X is $n(n+1)/2 - 1$. Hence $H^i(\Gamma; \mathscr{M})$ vanishes if $i > n(n+1)/2 - 1$. Since $\dim X$ grows quadratically in n, there are many potentially interesting cohomology groups to study.

However, it turns out that there is some additional vanishing of the cohomology for deeper (topological) reasons. For $n = 2$, this is easy to see. The quotient $\Gamma\backslash\mathfrak{h}$ is homeomorphic to a topological surface with punctures, corresponding to the cusps of Γ. Any such surface S can be retracted onto a finite graph simply by "stretching" S along its punctures. Thus $H^2(\Gamma; \mathscr{M}) = 0$, even though $\dim \Gamma\backslash\mathfrak{h} = 2$.

For $\Gamma \subset \mathrm{SL}_n(\mathbb{Z})$, a theorem of Borel–Serre implies that $H^i(\Gamma; \mathscr{M})$ vanishes if $i > \dim X - n + 1 = n(n-1)/2$ [**BS73**, Theorem 11.4.4]. The number $\nu = n(n-1)/2$ is called the *virtual cohomological dimension* of Γ

and is denoted $\operatorname{vcd}\Gamma$. Thus we only need to consider cohomology in degrees $i \leq \nu$.

Moreover we know from Section A.2.8 that the most interesting part of the cohomology is the cuspidal cohomology. In what degrees can it live? For $n = 2$, there is only one interesting cohomology group $H^1(\Gamma; \mathscr{M})$, and it contains the cuspidal cohomology. For higher dimensions, the situation is quite different: for most i, the subspace $H^i_{\mathrm{cusp}}(\Gamma; \mathscr{M})$ vanishes! In fact in the late 1970's Borel, Wallach, and Zuckerman observed that the cuspidal cohomology can only live in the cohomological degrees lying in an interval around $(\dim X)/2$ of size linear in n. An explicit description of this interval is given in [**Sch86**, Proposition 3.5]; one can also look at Table A.3.1, from which the precise statement is easy to determine.

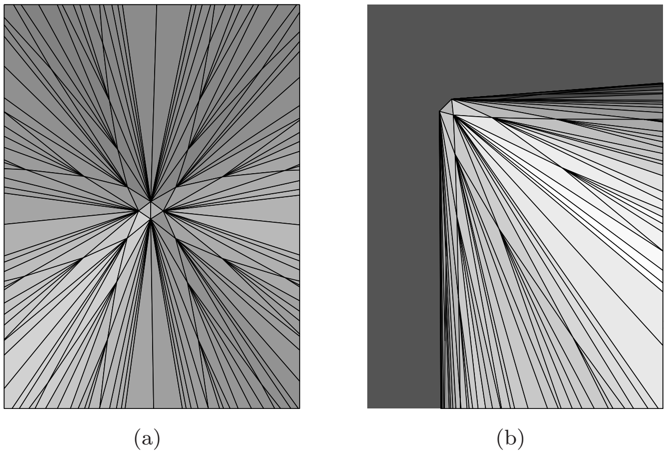
Another feature of Table A.3.1 deserves to be mentioned. There are exactly two values of n, namely $n = 2, 3$, such that virtual cohomological dimension equals the upper limit of the cuspidal range. This will have implications later, when we study the action of the Hecke operators on the cohomology.

n	2	3	4	5	6	7	8	9
$\dim X$	2	5	9	14	20	27	35	44
$\operatorname{vcd}\Gamma$	1	3	6	10	15	21	28	36
top degree of H^*_{cusp}	1	3	5	8	11	15	19	24
bottom degree of H^*_{cusp}	1	2	4	6	9	12	16	20

Table A.3.1. The virtual cohomological dimension and the cuspidal range for subgroups of $\mathrm{SL}_n(\mathbb{Z})$.

A.3.4. Recall that a point in $\mathbb{Z}^n$ is said to be *primitive* if the greatest common divisor of its coordinates is 1. In particular, a primitive point is nonzero. Let $\mathscr{P} \subset \mathbb{Z}^n$ be the set of primitive points. Any $v \in \mathscr{P}$, written as a column vector, determines a rank-1 symmetric matrix $q(v)$ in the closure $\bar{C}$ via $q(v) = vv^t$. The *Voronoĭ polyhedron* Π is defined to be the closed convex hull in $\bar{C}$ of the points $q(v)$, as v ranges over $\mathscr{P}$. Note that by construction, $\mathrm{SL}_n(\mathbb{Z})$ acts on Π, since $\mathrm{SL}_n(\mathbb{Z})$ preserves the set $\{q(v)\}$ and acts linearly on V.

Example A.11. Figure A.3.1 represents a crude attempt to show what Π looks like for $n = 2$. These images were constructed by computing a large subset of the points $q(v)$ and taking the convex hull (we took all points $v \in \mathscr{P}$ such that $\operatorname{Trace} q(v) < N$ for some large integer N). From a distance, the polyhedron Π looks almost indistinguishable from the cone C; this is somewhat conveyed by the right of Figure A.3.1. Unfortunately Π is not locally finite, so we really cannot produce an accurate picture. To get a more accurate image, the reader should imagine that each vertex meets infinitely many edges. On the other hand, Π is not hopelessly complex: each maximal face is a triangle, as the pictures suggest.



(a)

(b)

Figure A.3.1. The polyhedron Π for $\mathrm{SL}_2(\mathbb{Z})$. In (a) we see Π from the origin, in (b) from the side. The small triangle at the right center of (a) is the facet with vertices $\{q(e_1), q(e_2), q(e_1+e_2)\}$, where $\{e_1, e_2\}$ is the standard basis of $\mathbb{Z}^2$. In (b) the x-axis runs along the top from left to right, and the z-axis runs down the left side. The facet from (a) is the little triangle at the top left corner of (b).

A.3.5. The polyhedron Π is quite complicated: it has infinitely many faces and is not locally finite. However, one of Voronoĭ's great insights is that Π is actually not as complicated as it seems.

For any $A \in C$, let $\mu(A)$ be the minimum value attained by A on $\mathscr{P}$ and let $M(A) \subset \mathscr{P}$ be the set on which A attains $\mu(A)$. Note that $\mu(A) > 0$ and $M(A)$ is finite since A is positive-definite. Then A is called *perfect* if it is recoverable from the knowledge of the pair $(\mu(A), M(A))$. In other words, given $(\mu(A), M(A))$, we can write a system of linear equations

$$mZm^t = \mu(A), \quad m \in M(A), \tag{A.3.2}$$

where $Z = (z_{ij})$ is a symmetric matrix of variables. Then A is perfect if and only if A is the unique solution to the system (A.3.2).

Example A.12. The quadratic form $Q(x, y) = x^2 - xy + y^2$ is perfect. The smallest nontrivial value it attains on $\mathbb{Z}^2$ is $\mu(Q) = 1$, and it does so on the columns of

$$M(Q) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and their negatives. Letting $\alpha x^2+\beta xy+\gamma y^2$ be an undetermined quadratic form and applying the data $(\mu(Q), M(Q))$, we are led to the system of linear equations

$$\alpha = 1, \quad \gamma = 1, \quad \alpha+\beta+\gamma = 1.$$

From this we recover $Q(x,y)$.

Example A.13. The quadratic form $Q'(x,y) = x^2+y^2$ is not perfect. Again the smallest nontrivial value of Q' on $\mathbb{Z}^2$ is $m(Q') = 1$, attained on the columns of

$$M(Q') = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and their negatives. But every member of the one-parameter family of quadratic forms

$$x^2+\alpha xy+y^2, \quad \alpha \in (-1,1) \tag{A.3.3}$$

has the same set of minimal vectors, and so Q' cannot be recovered from the knowledge of $m(Q'), M(Q')$.

Example A.14. Example A.12 generalizes to all n. Define

$$A_n(x) := \sum_{i=1}^{n} x_i^2 - \sum_{1\le i<j\le n} x_i x_j. \tag{A.3.4}$$

Then A_n is perfect for all n. We have $\mu(A_n) = 1$, and $M(A_n)$ consists of all points of the form

$$\pm(e_i+e_{i+1}+\cdots+e_{i+k}), \quad 1\le i\le n, \quad i\le i+k\le n,$$

where $\{e_i\}$ is the standard basis of $\mathbb{Z}^n$. This quadratic form is closely related to the A_n root lattice [**FH91**], which explains its name. It is one of two infinite families of perfect forms studied by Voronoĭ (the other is related to the D_n root lattice).

We can now summarize Voronoĭ's main results:

(1) There are finitely many equivalence classes of perfect forms modulo the action of $\mathrm{SL}_n(\mathbb{Z})$. Voronoĭ even gave an explicit algorithm to determine all the perfect forms of a given dimension.

(2) The facets of Π, in other words the codimension 1 faces, are in bijection with the rank n perfect quadratic forms. Under this correspondence the minimal vectors $M(A)$ determine a facet F_A by taking the convex hull in $\bar{C}$ of the finite point set $\{q(m) \mid m \in M(A)\}$. Hence there are finitely many faces of Π modulo $\mathrm{SL}_n(\mathbb{Z})$ and thus finitely many modulo any finite index subgroup Γ.

(3) Let $\mathscr{V}$ be the set of cones over the faces of Π. Then $\mathscr{V}$ is a *fan*, which means (i) if $\sigma \in \mathscr{V}$, then any face of σ is also in $\mathscr{V}$; and (ii) if $\sigma, \sigma' \in \mathscr{V}$, then $\sigma \cap \sigma'$ is a common face of each[6]. The fan $\mathscr{V}$ provides a reduction theory for C in the following sense: any point $x \in C$ is contained in a unique $\sigma(x) \in \mathscr{V}$, and the set $\{\gamma \in \mathrm{SL}_n(\mathbb{Z}) \mid \gamma \cdot \sigma(x) = \sigma(x)\}$ is finite. Voronoĭ also gave an explicit algorithm to determine $\sigma(x)$ given x, the *Voronoĭ reduction algorithm.*

The number N_{perf} of equivalence classes of perfect forms modulo the action of $\mathrm{GL}_n(\mathbb{Z})$ grows rapidly with n (Table A.3.2); the complete classification is known only for $n \leq 8$. For a list of perfect forms up to $n = 7$, see [**CS88**]. For a recent comprehensive treatment of perfect forms, with many historical remarks, see [**Mar03**].

Dimension	N_{perf}	Authors
2	1	Voronoĭ [**Vor08**]
3	1	*ibid.*
4	2	*ibid.*
5	3	*ibid.*
6	7	Barnes [**Bar57**]
7	33	Jaquet-Chiffelle [**Jaq91, JC93**]
8	10916	Dutour–Schürmann–Vallentin [**DVS05**]

Table A.3.2. The number N_{perf} of equivalence classes of perfect forms.

A.3.6. Our goal now is to describe how the Voronoĭ fan $\mathscr{V}$ can be used to compute the cohomology $H^*(\Gamma; \mathscr{M})$. The idea is to use the cones in $\mathscr{V}$ to chop the quotient D into pieces.

For any $\sigma \in \mathscr{V}$, let σ° be the open cone obtained by taking the complement in σ of its proper faces. Then after taking the quotient by homotheties, the cones $\{\sigma^\circ \cap C \mid \sigma \in \mathscr{V}\}$ pass to locally closed subsets of D. Let $\mathscr{C}$ be the set of these images.

Any $c \in \mathscr{C}$ is a *topological cell*, i.e., it is homeomorphic to an open ball, since c is homeomorphic to a face of Π. Because $\mathscr{C}$ comes from the fan $\mathscr{V}$, the cells in $\mathscr{C}$ have good incidence properties: the closure in D of any $c \in \mathscr{C}$ can be written as a finite disjoint union of elements of $\mathscr{C}$. Moreover, $\mathscr{C}$ is locally finite: by taking quotients of all the σ° meeting C, we have eliminated the open cones lying in $\bar{C}$, and it is these cones that are responsible for the failure of local finiteness of $\mathscr{V}$. We summarize these properties by saying

[6] Strictly speaking, Voronoĭ actually showed that every codimension 1 cone is contained in two top-dimensional cones.

that $\mathscr{C}$ gives a *cellular decomposition* of D. Clearly $\mathrm{SL}_n(\mathbb{Z})$ acts on $\mathscr{C}$, since $\mathscr{C}$ is constructed using the fan $\mathscr{V}$. Thus we obtain a cellular decomposition[7] of $\Gamma\backslash D$ for any torsion-free Γ. We call $\mathscr{C}$ the *Voronoĭ decomposition* of D.

Some care must be taken in using these cells to perform topological computations. The problem is that even though the individual pieces are homeomorphic to balls and are glued together nicely, the boundaries of the closures of the pieces are not homeomorphic to spheres in general. (If they were, then the Voronoĭ decomposition would give rise to a *regular* cell complex [**CF67**], which can be used as a substitute for a simplicial or CW complex in homology computations.) Nevertheless, there is a way to remedy this.

Recall that a subspace A of a topological space B is a *strong deformation retract* if there is a continuous map $f\colon B \times [0,1] \to B$ such that $f(b,0) = b$, $f(b,1) \in A$, and $f(a,t) = a$ for all $a \in A$. For such pairs $A \subset B$ we have $H^*(A) = H^*(B)$. One can show that there is a strong deformation retraction from C to itself equivariant under the actions of both $\mathrm{SL}_n(\mathbb{Z})$ and the homotheties and that the image of the retraction modulo homotheties, denoted W, is naturally a locally finite regular cell complex of dimension ν. Moreover, the cells in W are in bijective, inclusion-reversing correspondence with the cells in $\mathscr{C}$. In particular, if a cell in $\mathscr{C}$ has *codimension* d, the corresponding cell in W has *dimension* d. Thus, for example, the vertices of W modulo $\mathrm{SL}_n(\mathbb{Z})$ are in bijection with the top-dimensional cells in $\mathscr{C}$, which are in bijection with equivalence classes of perfect forms.

In the literature W is called the *well-rounded retract*. The subspace $W \subset D \simeq X$ has a beautiful geometric interpretation. The quotient

$$\mathrm{SL}_n(\mathbb{Z})\backslash X = \mathrm{SL}_n(\mathbb{Z})\backslash \mathrm{SL}_n(\mathbb{R})/\,\mathrm{SO}(n)$$

can be interpreted as the moduli space of lattices in $\mathbb{R}^n$ modulo the equivalence relation of rotation and positive scaling (cf. [**AG00**]; for $n = 2$ one can also see [**Ser73**, VII, Proposition 3]). Then W corresponds to those lattices whose shortest nonzero vectors span $\mathbb{R}^n$. This is the origin of the name: the shortest vectors of such a lattice are "more round" than those of a generic lattice.

The space W was known classically for $n = 2$ and was constructed for $n \geq 3$ by Lannes and Soulé, although Soulé only published the case $n = 3$ [**Sou75**]. The construction for all n appears in work of Ash [**Ash80, Ash84**], who also generalized W to a much larger class of groups. Explicit computations of the cell structure of W have only been performed up to

[7] If Γ has torsion, then cells in $\mathscr{C}$ can have nontrivial stabilizers in Γ, and thus $\Gamma\backslash\mathscr{C}$ should be considered as an "orbifold" cellular decomposition.

$n = 6$ [**EVGS02**]. Certainly computing W explicitly for $n = 8$ seems very difficult, as Table A.3.2 indicates.

Example A.15. Figure A.3.2 illustrates $\mathscr{C}$ and W for $\mathrm{SL}_2(\mathbb{Z})$. As in Example A.11, the polyhedron Π is 3-dimensional, and so the Voronoĭ fan $\mathscr{V}$ has cones of dimensions $0, 1, 2, 3$. The 1-cones of $\mathscr{V}$, which correspond to the vertices of Π, pass to infinitely many points on the boundary $\partial\bar{D} = \bar{D} \smallsetminus D$. The 3-cones become triangles in $\bar{D}$ with vertices on $\partial\bar{D}$. In fact, the identifications $D \simeq \mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}(2) \simeq \mathfrak{h}$ realize D as the Klein model for the hyperbolic plane, in which geodesics are represented by Euclidean line segments. Hence, the images of the 1-cones of $\mathscr{V}$ are none other than the usual cusps of $\mathfrak{h}$, and the triangles are the $\mathrm{SL}_2(\mathbb{Z})$-translates of the ideal triangle with vertices $\{0, 1, \infty\}$. These triangles form a tessellation of $\mathfrak{h}$ sometimes known as the *Farey tessellation*. The edges of the Voronoĭ are the $\mathrm{SL}_2(\mathbb{Z})$-translates of the ideal geodesic between 0 and ∞. After adjoining cusps and passing to the quotient $X_0(N)$, these edges become the supports of the Manin symbols from Section 8.2 (cf. Figure 3.2.1). This example also shows how the Voronoĭ decomposition fails to be a regular cell complex: the boundaries of the closures of the triangles in D do not contain the vertices and thus are not homeomorphic to circles.

The virtual cohomological dimension of $\mathrm{SL}_2(\mathbb{Z})$ is 1. Hence the well-rounded retract W is a graph (Figures A.3.2 and A.3.3). Note that W is not a manifold. The vertices of W are in bijection with the Farey triangles—each vertex lies at the center of the corresponding triangle—and the edges are in bijection with the Manin symbols. Under the map $D \to \mathfrak{h}$, the graph W becomes the familiar "PSL_2-tree" embedded in $\mathfrak{h}$, with vertices at the order 3 elliptic points (Figure A.3.3).

A.3.7. We now discuss the example $\mathrm{SL}_3(\mathbb{Z})$ in some detail. This example gives a good feeling for how the general situation compares to the case $n = 2$.

We begin with the Voronoĭ fan $\mathscr{V}$. The cone C is 6-dimensional, and the quotient D is 5-dimensional. There is one equivalence class of perfect forms modulo the action of $\mathrm{SL}_3(\mathbb{Z})$, represented by the form (A.3.4). Hence there are 12 minimal vectors; six are the columns of the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \tag{A.3.5}$$

and the remaining six are the negatives of these. This implies that the cone σ corresponding to this form is 6-dimensional and simplicial. The latter implies that the faces of σ are the cones generated by $\{q(v) \mid v \in S\}$, where S ranges over all subsets of (A.3.5). To get the full structure of the fan, one

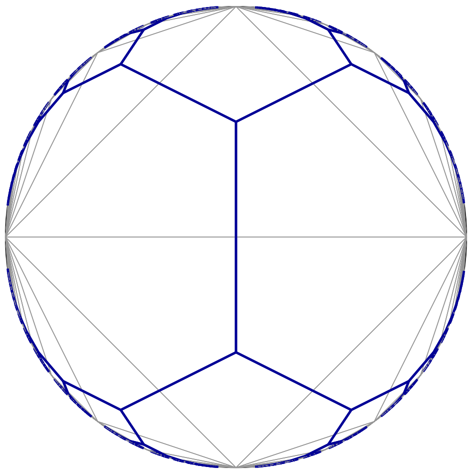


Figure A.3.2. The Voronoĭ decomposition and the retract in D.

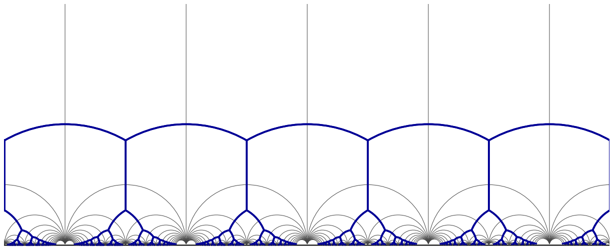


Figure A.3.3. The Voronoĭ decomposition and the retract in $\mathfrak{h}$.

must determine the $\mathrm{SL}_3(\mathbb{Z})$ orbits of faces, as well as which faces lie in the boundary $\partial\bar{C} = \bar{C} \smallsetminus C$. After some pleasant computation, one finds:

(1) There is one equivalence class modulo $\mathrm{SL}_3(\mathbb{Z})$ for each of the 6-, 5-, 2-, and 1-dimensional cones.

(2) There are two equivalence classes of the 4-dimensional cones, represented by the sets of minimal vectors

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

(3) There are two equivalence classes of the 3-dimensional cones, represented by the sets of minimal vectors

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The second type of 3-cone lies in $\partial\bar{C}$ and thus does not determine a cell in $\mathscr{C}$.

(4) The 2- and 1-dimensional cones lie entirely in $\partial\bar{C}$ and do not determine cells in $\mathscr{C}$.

After passing from C to D, the cones of dimension k determine cells of dimension $k-1$. Therefore, modulo the action of $\mathrm{SL}_3(\mathbb{Z})$ there are five types of cells in the Voronoĭ decomposition $\mathscr{C}$, with dimensions from 5 to 2. We denote these cell types by c_5, c_4, c_{3a}, c_{3b}, and c_2. Here c_{3a} corresponds to the first type of 4-cone in item (2) above, and c_{3b} to the second. For a beautiful way to index the cells of $\mathscr{C}$ using configurations in projective spaces, see [**McC91**].

The virtual cohomological dimension of $\mathrm{SL}_3(\mathbb{Z})$ is 3, which means that the retract W is a 3-dimensional cell complex. The closures of the top-dimensional cells in W, which are in bijection with the Voronoĭ cells of type c_2, are homeomorphic to solid cubes truncated along two pairs of opposite corners (Figure A.3.4). To compute this, one must see how many Voronoĭ cells of a given type contain a fixed cell of type c_2 (since the inclusions of cells in W are the *opposite* of those in $\mathscr{C}$).

A table of the incidence relations between the cells of $\mathscr{C}$ and W is given in Table A.3.3. To interpret the table, let $m = m(X, Y)$ be the integer in row X and column Y.

- If m is below the diagonal, then the boundary of a cell of type Y contains m cells of type X.
- If m is above the diagonal, then a cell of type Y appears in the boundary of m cells of type X.

For instance, the entry 16 in row c_5 and column c_2 means that a Voronoĭ cell of type c_2 meets the boundaries of 16 cells of type c_5. This is the same as the number of vertices in the Soulé cube (Figure A.3.4). Investigation of the table shows that the triangular (respectively, hexagonal) faces of the Soulé cube correspond to the Voronoĭ cells of type c_{3a} (resp., c_{3b}).

Figure A.3.5 shows a Schlegel diagram for the Soulé cube. One vertex is at infinity; this is indicated by the arrows on three of the edges. This Soulé cube is dual to the Voronoĭ cell C of type c_2 with minimal vectors given by the columns of the identity matrix. The labels on the 2-faces are additional minimal vectors that show which Voronoĭ cells contain C. For example, the central triangle labelled with $(1,1,1)^t$ is dual to the Voronoĭ cell of type c_{3a} with minimal vectors given by those of C together with $(1,1,1)^t$. Cells of type c_4 containing C in their closure correspond to the edges of the figure; the minimal vectors for a given edge are those of C together with the two vectors on the 2-faces containing the edge. Similarly, one can read off the minimal vectors of the top-dimensional Voronoĭ cells containing C, which correspond to the vertices of Figure A.3.5.

	c_5	c_4	c_{3a}	c_{3b}	c_2
c_5	•	2	3	6	16
c_4	6	•	3	6	24
c_{3a}	3	1	•	•	4
c_{3b}	12	4	•	•	6
c_2	12	8	4	3	•

Table A.3.3. Incidence relations in the Voronoĭ decomposition and the retract for $\mathrm{SL}_3(\mathbb{Z})$.

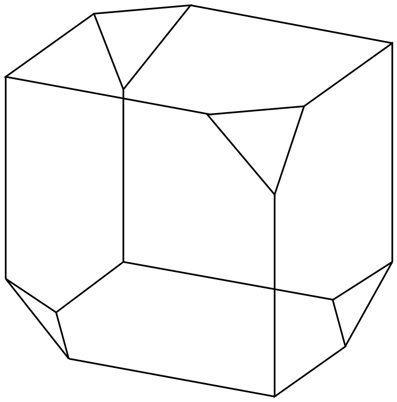


Figure A.3.4. The Soulé cube.

A.3.8. Now let p be a prime, and let $\Gamma = \Gamma_0(p) \subset \mathrm{SL}_3(\mathbb{Z})$ be the Hecke subgroup of matrices with bottom row congruent to $(0, 0, *) \mod p$ (Example A.4). The virtual cohomological dimension of Γ is 3, and the cusp cohomology with constant coefficients can appear in degrees 2 and 3. One can show that the cusp cohomology in degree 2 is dual to that in degree 3, so for computational purposes it suffices to focus on degree 3.

In terms of W, these will be cochains supported on the 3-cells. Unfortunately we cannot work directly with the quotient $\Gamma\backslash W$ since Γ has torsion: there will be cells taken to themselves by the Γ-action, and thus the cells of W need to be subdivided to induce the structure of a cell complex on $\Gamma\backslash W$. Thus when Γ has torsion, the "set of 3-cells modulo Γ" unfortunately makes no sense.

To circumvent this problem, one can mimic the idea of Manin symbols. The quotient $\Gamma\backslash \mathrm{SL}_3(\mathbb{Z})$ is in bijection with the finite projective plane $\mathbb{P}^2(\mathbb{F}_p)$,

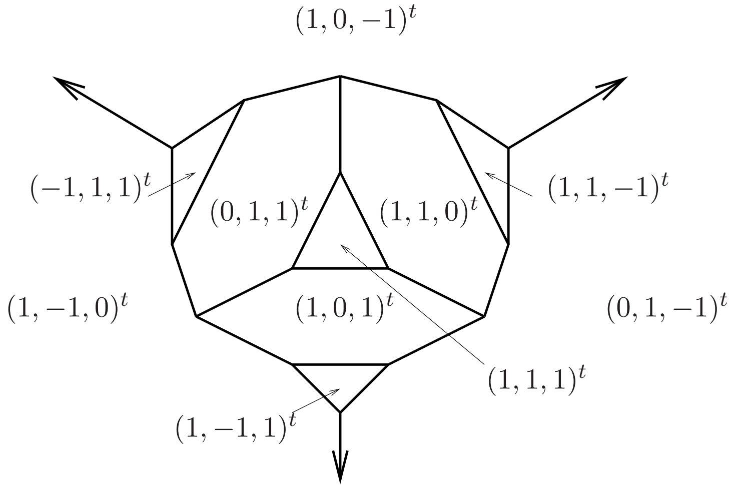


Figure A.3.5. A Schlegel diagram of a Soulé cube, showing the minimal vectors that correspond to the 2-faces.

where $\mathbb{F}_p$ is the field with p elements (cf. Proposition 3.10). The group $\mathrm{SL}_3(\mathbb{Z})$ acts transitively on the set of all 3-cells of W; if we fix one such cell w, its stabilizer $\mathrm{Stab}(w) = \{\gamma \in \mathrm{SL}_3(\mathbb{Z}) \mid \gamma w = w\}$ is a finite subgroup of $\mathrm{SL}_3(\mathbb{Z})$. Hence the set of 3-cells modulo Γ should be interpreted as the set of orbits in $\mathbb{P}^2(\mathbb{F}_p)$ of the finite group $\mathrm{Stab}(w)$. This suggests describing $H^3(\Gamma;\mathbb{C})$ in terms of the space $\mathscr{S}$ of complex-valued functions $f\colon \mathbb{P}^2(\mathbb{F}_p) \to \mathbb{C}$. To carry this out, there are two problems:

(1) How do we explicitly describe $H^3(\Gamma;\mathbb{C})$ in terms of $\mathscr{S}$?

(2) How can we isolate the cuspidal subspace $H^3_{\mathrm{cusp}}(\Gamma;\mathbb{C}) \subset H^3(\Gamma;\mathbb{C})$ in terms of our description?

Fully describing the solutions to these problems is rather complicated. We content ourselves with presenting the following theorem, which collects together several statements in [**AGG84**]. This result should be compared to Theorems 3.13 and 8.4.

Theorem A.16 (Theorem 3.19 and Summary 3.23 of [**AGG84**])**.** *We have*

$$\dim H^3(\Gamma_0(p);\mathbb{C}) = \dim H^3_{cusp}(\Gamma_0(p);\mathbb{C}) + 2S_p,$$

where S_p is the dimension of the space of weight 2 holomorphic cusp forms on $\Gamma_0(p) \subset \mathrm{SL}_2(\mathbb{Z})$. Moreover, the cuspidal cohomology $H^3_{cusp}(\Gamma_0(p);\mathbb{C})$ is isomorphic to the vector space of functions $f\colon \mathbb{P}^2(\mathbb{F}_p) \to \mathbb{C}$ satisfying

(1) $f(x,y,z) = f(z,x,y) = f(-x,y,z) = -f(y,x,z)$,

(2) $f(x,y,z) + f(-y,x-y,z) + f(y-x,-x,z) = 0$,

(3) $f(x,y,0) = 0$, *and*

(4) $\sum_{z=1}^{p-1} f(x,y,z) = 0$.

Unlike subgroups of $\mathrm{SL}_2(\mathbb{Z})$, cuspidal cohomology is apparently much rarer for $\Gamma_0(p) \subset \mathrm{SL}_3(\mathbb{Z})$. The computations of [**AGG84, vGvdKTV97**] show that the only prime levels $p \leq 337$ with nonvanishing cusp cohomology are 53, 61, 79, 89, and 223. In all these examples, the cuspidal subspace is 2-dimensional.

For more details of how to implement such computations, we refer to [**AGG84, vGvdKTV97**]. For further details about the additional complications arising for higher rank groups, in particular subgroups of $\mathrm{SL}_4(\mathbb{Z})$, see [**AGM02**, Section 3].

A.4. Hecke Operators and Modular Symbols

A.4.1. There is one ingredient missing so far in our discussion of the cohomology of arithmetic groups, namely the Hecke operators. These are an essential tool in the study of modular forms. Indeed, the forms with the most arithmetic significance are the Hecke eigenforms, and the connection with arithmetic is revealed by the Hecke eigenvalues.

In higher rank the situation is similar. There is an algebra of Hecke operators acting on the cohomology spaces $H^*(\Gamma; \mathscr{M})$. The eigenvalues of these operators are conjecturally related to certain representations of the Galois group. Just as in the case $G = \mathrm{SL}_2(\mathbb{R})$, we need tools to compute the Hecke action.

In this section we discuss this problem. We begin with a general description of the Hecke operators and how they act on cohomology. Then we focus on one particular cohomology group, namely the top degree $H^\nu(\Gamma; \mathbb{C})$, where $\nu = \mathrm{vcd}(\Gamma)$ and Γ has finite index in $\mathrm{SL}_n(\mathbb{Z})$. This is the setting that generalizes the modular symbols method from Chapter 8. We conclude by giving examples of Hecke eigenclasses in the cuspidal cohomology of $\Gamma_0(p) \subset \mathrm{SL}_3(\mathbb{Z})$.

A.4.2. Let $g \in \mathrm{SL}_n(\mathbb{Q})$. The group $\Gamma' = \Gamma \cap g^{-1}\Gamma g$ has finite index in both Γ and $g^{-1}\Gamma g$. The element g determines a diagram $C(g)$

$$\begin{tikzcd}
 & \Gamma'\backslash X \arrow[dl, "s"'] \arrow[dr, "t"] & \\
\Gamma\backslash X & & \Gamma\backslash X
\end{tikzcd}$$

called a *Hecke correspondence*. The map s is induced by the inclusion $\Gamma' \subset \Gamma$, while t is induced by the inclusion $\Gamma' \subset g^{-1}\Gamma g$ followed by the diffeomorphism $g^{-1}\Gamma g\backslash X \to \Gamma\backslash X$ given by left multiplication by g. Specifically,

$$s(\Gamma' x) = \Gamma x, \quad t(\Gamma' x) = \Gamma g x, \quad x \in X.$$

The maps s and t are finite-to-one, since the indices $[\Gamma' : \Gamma]$ and $[\Gamma' : g^{-1}\Gamma g]$ are finite. This implies that we obtain maps on cohomology

$$s^*\colon H^*(\Gamma\backslash X) \to H^*(\Gamma'\backslash X), \quad t_*\colon H^*(\Gamma'\backslash X) \to H^*(\Gamma\backslash X).$$

Here the map s^* is the usual induced map on cohomology, while the "wrong-way" map[8] t_* is given by summing a class over the finite fibers of t. These maps can be composed to give a map

$$T_g := t_* s^*\colon H^*(\Gamma\backslash X; \widetilde{\mathscr{M}}) \longrightarrow H^*(\Gamma\backslash X; \widetilde{\mathscr{M}}).$$

This is called the *Hecke operator* associated to g. There is an obvious notion of isomorphism of Hecke correspondences. One can show that up to isomorphism, the correspondence $C(g)$ and thus the Hecke operator T_g depend only on the double coset $\Gamma g \Gamma$. One can compose Hecke correspondences, and thus we obtain an algebra of operators acting on the cohomology, just as in the classical case.

Example A.17. Let $n = 2$, and let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. If we take $g = \mathrm{diag}(1, p)$, where p is a prime, then the action of T_g on $H^1(\Gamma; M_{k-2})$ is the same as the action of the classical Hecke operator T_p on the weight k holomorphic modular forms. If we take $\Gamma = \Gamma_0(N)$, we obtain an operator $T(p)$ for all p prime to N, and the algebra of Hecke operators coincides with the (semisimple) Hecke algebra generated by the T_p, $(p, N) = 1$. For $p|N$, one can also describe the U_p operators in this language.

Example A.18. Now let $n > 2$ and let $\Gamma = \mathrm{SL}_n(\mathbb{Z})$. The picture is very similar, except that now there are several Hecke operators attached to any prime p. In fact there are $n - 1$ operators $T(p, k)$, $k = 1, \ldots, n - 1$. The operator $T(p, k)$ is associated to the correspondence $C(g)$, where $g = \mathrm{diag}(1, \ldots, 1, p, \ldots, p)$ and where p occurs k times. If we consider the congruence subgroups $\Gamma_0(N)$, we have operators $T(p, k)$ for $(p, N) = 1$ and analogues of the U_p operators for $p|N$.

Just as in the classical case, any double coset $\Gamma g \Gamma$ can be written as a disjoint union of left cosets

$$\Gamma g \Gamma = \coprod_{h \in \Omega} \Gamma h$$

[8] Under the identification $H^*(\Gamma\backslash X; \widetilde{\mathscr{M}}) \simeq H^*(\Gamma; \mathscr{M})$, the map t_* becomes the transfer map in group cohomology [**Bro94**, III.9].

for a certain finite set of $n \times n$ integral matrices Ω. For the operator $T(p,k)$, the set Ω can be taken to be all upper-triangular matrices of the form [**Kri90**, Proposition 7.2]

$$\begin{pmatrix} p^{e_1} & & a_{ij} \\ & \ddots & \\ & & p^{e_n} \end{pmatrix},$$

where

- $e_i \in \{0,1\}$ and exactly k of the e_i are equal to 1 and
- $a_{ij} = 0$ unless $e_i = 0$ and $e_j = 1$, in which case a_{ij} satisfies $0 \le a_{ij} < p$.

Remark A.19. The number of coset representatives for the operator $T(p,k)$ is the same as the number of points in the finite Grassmannian $G(k,n)(\mathbb{F}_p)$. A similar phenomenon is true for the Hecke operators for any group G, although there are some subtleties [**Gro98**].

A.4.3. Recall that in Section A.3.6 we constructed the Voronoĭ decomposition $\mathscr{C}$ and the well-rounded retract W and that we can use them to compute the cohomology $H^*(\Gamma;\mathscr{M})$. Unfortunately, we cannot directly use them to compute the action of the Hecke operators on cohomology, since the Hecke operators do not act cellularly on $\mathscr{C}$ or W. The problem is that the Hecke image of a cell in $\mathscr{C}$ (or W) is usually not a union of cells in $\mathscr{C}$ (or W). This is already apparent for $n = 2$. The edges of $\mathscr{C}$ are the $\mathrm{SL}_2(\mathbb{Z})$-translates of the ideal geodesic τ from 0 to ∞ (Example A.15). Applying a Hecke operator takes such an edge to a union of ideal geodesics, each with vertices at a pair of cusps. In general such geodesics are not an $\mathrm{SL}_2(\mathbb{Z})$-translate of τ.

For $n = 2$, one solution is to work with all possible ideal geodesics with vertices at the cusps, in other words the space of modular symbols $\mathbb{M}_2$ from Section 3.2. Manin's trick (Proposition 3.11) shows how to write any modular symbol as a linear combination of unimodular symbols, by which we mean modular symbols supported on the edges of $\mathscr{C}$. These are the ideas we now generalize to all n.

Definition A.20. Let S_0 be the $\mathbb{Q}$-vector space spanned by the symbols $\mathbf{v} = [v_1,\dots,v_n]$, where $v_i \in \mathbb{Q}^n \smallsetminus \{0\}$, modulo the following relations:

(1) If τ is a permutation on n letters, then

$$[v_1,\dots,v_n] = \mathrm{sign}(\tau)[\tau(v_1),\dots,\tau(v_n)],$$

where $\mathrm{sign}(\tau)$ is the sign of τ.

(2) If $q \in \mathbb{Q}^\times$, then

$$[qv_1, v_2 \dots, v_n] = [v_1,\dots,v_n].$$

(3) If the points $v_1, \ldots, v_n$ are linearly dependent, then $\mathbf{v} = 0$.

Let $B \subset S_0$ be the subspace generated by linear combinations of the form

$$\sum_{i=0}^{n} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_n], \tag{A.4.1}$$

where $v_0, \ldots, v_n \in \mathbb{Q}^n \smallsetminus \{0\}$ and where $\hat{v}_i$ means to omit v_i.

We call S_0 the space of *modular symbols*. We caution the reader that there are some differences in what we call modular symbols and those found in Section 3.2 and Definition 8.2; we compare them in Section A.4.4. The group $\mathrm{SL}_n(\mathbb{Q})$ acts on S_0 by left multiplication: $g \cdot \mathbf{v} = [gv_1, \ldots, gv_n]$. This action preserves the subspace B and thus induces an action on the quotient $M = S_0/B$. For $\Gamma \subset \mathrm{SL}_n(\mathbb{Z})$ a finite index subgroup, let M_Γ be the space of Γ-coinvariants in M. In other words, M_Γ is the quotient of M by the subspace generated by $\{m - \gamma \cdot m \mid \gamma \in \Gamma\}$.

The relationship between modular symbols and the cohomology of Γ is given by the following theorem, first proved for SL_n by Ash and Rudolph [**AR79**] and by Ash for general G [**Ash86**]:

Theorem A.21 ([**Ash86, AR79**]). *Let $\Gamma \subset \mathrm{SL}_n(\mathbb{Z})$ be a finite index subgroup. There is an isomorphism*

$$M_\Gamma \xrightarrow{\sim} H^\nu(\Gamma; \mathbb{Q}), \tag{A.4.2}$$

where Γ acts trivially on $\mathbb{Q}$ and where $\nu = \mathrm{vcd}(\Gamma)$.

We remark that Theorem A.21 remains true if $\mathbb{Q}$ is replaced with nontrivial coefficients as in Section A.2.7. Moreover, if Γ is assumed to be torsion-free then we can replace $\mathbb{Q}$ with $\mathbb{Z}$.

The great virtue of M_Γ is that it admits an action of the Hecke operators. Given a Hecke operator T_g, write the double coset $\Gamma g \Gamma$ as a disjoint union of left cosets

$$\Gamma g \Gamma = \coprod_{h \in \Omega} \Gamma h \tag{A.4.3}$$

as in Example A.18. Any class in M_Γ can be lifted to a representative $\eta = \sum q(\mathbf{v})\mathbf{v} \in S_0$, where $q(\mathbf{v}) \in \mathbb{Q}$ and almost all $q(\mathbf{v})$ vanish. Then we define

$$T_g(\mathbf{v}) = \sum_{h \in \Omega} h \cdot \mathbf{v} \tag{A.4.4}$$

and extend to η by linearity. The right side of (A.4.4) depends on the choices of η and Ω, but after taking quotients and coinvariants, we obtain a well-defined action on cohomology via (A.4.2).

A.4.4. The space S_0 is closely related to the space $\mathbb{M}_2$ from Section 3.2 and Section 8.1. Indeed, $\mathbb{M}_2$ was defined to be the quotient $(F/R)/(F/R)_{\mathrm{tor}}$, where F is the free abelian group generated by ordered pairs

$$\{\alpha, \beta\}, \quad \alpha, \beta \in \mathbb{P}^1(\mathbb{Q}), \tag{A.4.5}$$

and R is the subgroup generated by elements of the form

$$\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\}, \quad \alpha, \beta, \gamma \in \mathbb{P}^1(\mathbb{Q}). \tag{A.4.6}$$

The only new feature in Definition A.20 is item (3). For $n = 2$ this corresponds to the condition $\{\alpha, \alpha\} = 0$, which follows from (A.4.6). We have

$$S_0/B \simeq \mathbb{M}_2 \otimes \mathbb{Q}.$$

Hence there are two differences between S_0 and $\mathbb{M}_2$: our notion of modular symbols uses rational coefficients instead of integral coefficients and is the space of symbols *before* dividing out by the subspace of relations B; we further caution the reader that this is somewhat at odds with the literature.

We also remark that the general arbitrary weight definition of modular symbols for a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ given in Section 8.1 also includes taking Γ-coinvariants, as well as extra data for a coefficient system. We have not included the latter data since our emphasis is trivial coefficients, although it would be easy to do so in the spirit of Section 8.1.

Elements of $\mathbb{M}_2$ also have a geometric interpretation: the symbol $\{\alpha, \beta\}$ corresponds to the ideal geodesic in $\mathfrak{h}$ with endpoints at the cusps α and β. We have a similar picture for the symbols $\mathbf{v} = [v_1, \ldots, v_n]$. We can assume that each v_i is primitive, which means that each v_i determines a vertex of the Voronoĭ polyhedron Π. The rational cone generated by these vertices determines a subset $\Delta(\mathbf{v}) \subset D$, where D is the linear model of the symmetric space $X = \mathrm{SL}_n(\mathbb{R})/\mathrm{SO}(n)$ from Section A.3.2. This subset $\Delta(\mathbf{v})$ is then an "ideal simplex" in X. There is also a connection between $\Delta(\mathbf{v})$ and torus orbits in X; we refer to [**Ash86**] for a related discussion.

A.4.5. Now we need a generalization of the Manin trick (Section 3.3.1). This is known in the literature as the *modular symbols algorithm.*

We can define a kind of norm function on S_0 as follows. Let $\mathbf{v} = [v_1, \ldots, v_n]$ be a modular symbol. For each v_i, choose $\lambda_i \in \mathbb{Q}^\times$ such that $\lambda_i v_i$ is primitive. Then we define

$$\|\mathbf{v}\| := |\det(\lambda_1 v_1, \ldots, \lambda_n v_n)| \in \mathbb{Z}.$$

Note that $\|\mathbf{v}\|$ is well defined, since the λ_i are unique up to sign, and permuting the v_i only changes the determinant by a sign. We extend $\|\ \|$ to all of S_0 by taking the maximum of $\|\ \|$ over the support of any $\eta \in S_0$: if

$\eta = \sum q(\mathbf{v})\mathbf{v}$, where $q(\mathbf{v}) \in \mathbb{Q}$ and almost all $q(\mathbf{v})$ vanish, then we put

$$\|\eta\| = \operatorname*{Max}_{q(\mathbf{v})\neq 0} \|\mathbf{v}\|.$$

We say a modular symbol η is *unimodular* if $\|\eta\| = 1$. It is clear that the images of the unimodular symbols generate a finite-dimensional subspace of M_Γ. The next theorem shows that this subspace is actually *all* of M_Γ.

Theorem A.22 ([**AR79, Bar94**]). *The space M_Γ is spanned by the images of the unimodular symbols. More precisely, given any symbol $\mathbf{v} \in S_0$ with $\|\mathbf{v}\| > 1$,*

(1) *in S_0/B we may write*

$$\mathbf{v} = \sum q(\mathbf{w})\mathbf{w}, \quad q(\mathbf{w}) \in \mathbb{Z}, \tag{A.4.7}$$

where if $q(\mathbf{w}) \neq 0$, then $\|\mathbf{w}\| = 1$, and

(2) *the number of terms on the right side of* (A.4.7) *is bounded by a polynomial in* $\log \|\mathbf{v}\|$ *that depends only on the dimension n.*

Proof. (Sketch) Given a modular symbol $\mathbf{v} = [v_1, \dots, v_n]$, we may assume that the points v_i are primitive. We will show that if $\|\mathbf{v}\| > 1$, we can find a point u such that when we apply the relation (A.4.1) using the points $u, v_1, \dots, v_n$, all terms other than $\mathbf{v}$ have norm less than $\|\mathbf{v}\|$. We call such a point a *reducing point* for $\mathbf{v}$.

Let $P \subset \mathbb{R}^n$ be the open parallelotope

$$P := \left\{ \sum \lambda_i v_i \;\middle|\; |\lambda_i| < \|\mathbf{v}\|^{-1/n} \right\}.$$

Then P is an n-dimensional centrally symmetric convex body with volume 2^n. By Minkowski's theorem from the geometry of numbers (cf. [**FT93**, IV.2.6]), $P \cap \mathbb{Z}^n$ contains a nonzero point u. Using (A.4.1), we find

$$\mathbf{v} = \sum_{i=1}^{n} (-1)^{i-1} \mathbf{v}_i(u), \tag{A.4.8}$$

where $\mathbf{v}_i(u)$ is the symbol

$$\mathbf{v}_i(u) = [v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_n].$$

Moreover, it is easy to see that the new symbols satisfy

$$0 \leq \|\mathbf{v}_i(u)\| < \|\mathbf{v}\|^{(n-1)/n}, \quad i = 1, \dots, n. \tag{A.4.9}$$

This completes the proof of the first statement.

To prove the second statement, we must estimate how many times relations of the form (A.4.8) need to be applied to obtain (A.4.7). A nonunimodular symbol produces at most n new modular symbols after (A.4.8) is performed; we potentially have to apply (A.4.8) again to each of the symbols that result, which in turn could produce as many as n new symbols for each. Hence we can visualize the process of constructing (A.4.7) as building a rooted tree, where the root is $\mathbf{v}$, the leaves are the symbols $\mathbf{w}$, and where each node has at most n children. It is not hard to see that the bound (A.4.9) implies that the depth of this tree (i.e., the longest length of a path from the root to a leaf) is $O(\log\log\|\mathbf{v}\|)$. From this the second statement follows easily. □

Statement (1) of Theorem A.22 is due to Ash and Rudolph [**AR79**]. Instead of P, they used the larger parallelotope P' defined by

$$P' := \Big\{ \sum \lambda_i v_i \;\Big|\; |\lambda_i| < 1 \Big\},$$

which has volume $2^n\|\mathbf{v}\|$. The observation that P' can be replaced by P and the proof of (2) are both due to Barvinok [**Bar94**].

A.4.6. The relationship between Theorem A.22 and Manin's trick should be clear. For $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$, the Manin symbols correspond exactly to the unimodular symbols mod Γ. So Theorem A.22 implies that every modular symbol (in the language of Section 8.1) is a linear combination of Manin symbols. This is exactly the conclusion of Proposition 8.3.

In higher rank the relationship between Manin symbols and unimodular symbols is more subtle. In fact there are two possible notions of "Manin symbol," which agree for $\mathrm{SL}_2(\mathbb{Z})$ but not in general. One possibility is the obvious one: a Manin symbol is a unimodular symbol.

The other possibility is to define a Manin symbol to be a modular symbol corresponding to a top-dimensional cell of the retract W. But for $n \geq 5$, such modular symbols need not be unimodular. In particular, for $n = 5$ there are two equivalence classes of top-dimensional cells. One class corresponds to the unimodular symbols, the other to a set of modular symbols of norm 2. However, Theorems A.21 and A.22 show that $H^\nu(\Gamma;\mathbb{Q})$ is spanned by unimodular symbols. Thus as far as this cohomology group is concerned, the second class of symbols is in some sense unnecessary.

A.4.7. We return to the setting of Section A.3.8 and give examples of Hecke eigenclasses in the cusp cohomology of $\Gamma = \Gamma_0(p) \subset \mathrm{SL}_3(\mathbb{Z})$. We closely follow [**AGG84, vGvdKTV97**]. Note that since the top of the cuspidal range for SL_3 is the same as the virtual cohomological dimension ν, we can use modular symbols to compute the Hecke action on cuspidal classes.

Given a prime l coprime to p, there are two Hecke operators of interest $T(l,1)$ and $T(l,2)$. We can compute the action of these operators on

$H^3_{\mathrm{cusp}}(\Gamma;\mathbb{C})$ as follows. Recall that $H^3_{\mathrm{cusp}}(\Gamma;\mathbb{C})$ can be identified with a certain space of functions $f\colon \mathbb{P}^2(\mathbb{F}_p)\to\mathbb{C}$ (Theorem A.16). Given $x\in\mathbb{P}^2(\mathbb{F}_p)$, let $Q_x\in\mathrm{SL}_3(\mathbb{Z})$ be a matrix such that $Q_x\mapsto x$ under the identification $\mathbb{P}^2(\mathbb{F}_p)\xrightarrow{\sim}\Gamma\backslash\mathrm{SL}_3(\mathbb{Z})$. Then Q_x determines a unimodular symbol $[Q_x]$ by taking the v_i to be the columns of Q_x. Given any Hecke operator T_g, we can find coset representatives h_i such that $\Gamma g\Gamma=\coprod\Gamma h_i$ (explicit representatives for $\Gamma=\Gamma_0(p)$ and $T_g=T(l,k)$ are given in [**AGG84, vGvdKTV97**]). The modular symbols $[h_iQ_x]$ are no longer unimodular in general, but we can apply Theorem A.22 to write

$$[h_iQ_x]=\sum_j[R_{ij}],\quad R_{ij}\in\mathrm{SL}_3(\mathbb{Z}).$$

Then for $f\colon\mathbb{P}^2(\mathbb{F}_p)\to\mathbb{C}$ as in Theorem A.16, we have

$$(T_gf)(x)=\sum_{i,j}f(\overline{R_{ij}}),$$

where $\overline{R_{ij}}$ is the class of R_{ij} in $\mathbb{P}^2(\mathbb{F}_p)$.

Now let $\xi\in H^3_{\mathrm{cusp}}(\Gamma;\mathbb{C})$ be a simultaneous eigenclass for all the Hecke operators $T(l,1)$, $T(l,2)$, as l ranges over all primes coprime with p. General considerations from the theory of automorphic forms imply that the eigenvalues $a(l,1)$, $a(l,2)$ are complex conjugates of one other. Hence it suffices to compute $a(l,1)$. We give two examples of cuspidal eigenclasses for two different prime levels.

Example A.23. Let $p=53$. Then $H^3_{\mathrm{cusp}}(\Gamma_0(53);\mathbb{C})$ is 2-dimensional. Let $\eta=(1+\sqrt{-11})/2$. One eigenclass is given by the data

l	2	3	5	7	11	13
$a(l,1)$	$-1-2\eta$	$-2+2\eta$	1	-3	1	$-2-12\eta$

and the other is obtained by complex conjugation.

Example A.24. Let $p=61$. Then $H^3_{\mathrm{cusp}}(\Gamma_0(61);\mathbb{C})$ is 2-dimensional. Let $\omega=(1+\sqrt{-3})/2$. One eigenclass is given by the data

l	2	3	5	7	11	13
$a(l,1)$	$1-2\omega$	$-5+4\omega$	$-2+4\omega$	-6ω	$-2+2\omega$	$-2-4\omega$

and the other is obtained by complex conjugation.

A.5. Other Cohomology Groups

A.5.1. In Section A.4 we saw how to compute the Hecke action on the top cohomology group $H^\nu(\Gamma;\mathbb{C})$. Unfortunately for $n\geq 4$, this cohomology group does not contain any cuspidal cohomology. The first case is $\Gamma\subset\mathrm{SL}_4(\mathbb{Z})$; we have $\mathrm{vcd}(\Gamma)=6$, and the cusp cohomology lives in degrees 4 and 5. One can show that the cusp cohomology in degree 4 is dual to that in degree 5, so for computational purposes it suffices to be able to compute the Hecke action on $H^5(\Gamma;\mathbb{C})$. But modular symbols do not help us here.

In this section we describe a technique to compute the Hecke action on $H^{\nu-1}(\Gamma;\mathbb{C})$, following [**Gun00a**]. The technique is an extension of the modular symbol algorithm to these cohomology groups. In principle the ideas in this section can be modified to compute the Hecke action on other cohomology groups $H^{\nu-k}(\Gamma;\mathbb{C})$, $k>1$, although this has not been investigated[9]. For $n=4$, we have applied the algorithm in joint work with Ash and McConnell to investigate computationally the cohomology $H^5(\Gamma;\mathbb{C})$, where $\Gamma_0(N)\subset \mathrm{SL}_4(\mathbb{Z})$ [**AGM02**].

A.5.2. To begin, we need an analogue of Theorem A.21 for lower degree cohomology groups. In other words, we need a generalization of the modular symbols for other cohomology groups. This is achieved by the *sharbly complex* S_*:

Definition A.25 ([**Ash94**])**.** Let $\{S_*,\partial\}$ be the chain complex given by the following data:

(1) For $k\geq 0$, S_k is the $\mathbb{Q}$-vector space generated by the symbols $\mathbf{u}=[v_1,\dots,v_{n+k}]$, where $v_i\in\mathbb{Q}^n\smallsetminus\{0\}$, modulo the relations:
   (a) If τ is a permutation on $(n+k)$ letters, then

$$[v_1,\dots,v_{n+k}]=\operatorname{sign}(\tau)[\tau(v_1),\dots,\tau(v_{n+k})],$$

   where $\operatorname{sign}(\tau)$ is the sign of τ.
   (b) If $q\in\mathbb{Q}^\times$, then

$$[qv_1,v_2\dots,v_{n+k}]=[v_1,\dots,v_{n+k}].$$

   (c) If the rank of the matrix $(v_1,\dots,v_{n+k})$ is less than n, then $\mathbf{u}=0$.

(2) For $k>0$, the boundary map $\partial\colon S_k\to S_{k-1}$ is

$$[v_1,\dots,v_{n+k}]\longmapsto\sum_{i=1}^{n+k}(-1)^i[v_1,\dots,\hat{v}_i,\dots,v_{n+k}].$$

We define ∂ to be identically zero on S_0.

The elements

$$\mathbf{u}=[v_1,\dots,v_{n+k}]$$

[9] The first interesting case is $n=5$, for which the cuspidal cohomology lives in $H^{\nu-2}$.

are called *k-sharblies*[10]. The 0-sharblies are exactly the modular symbols from Definition A.20, and the subspace $B \subset S_0$ is the image of the boundary map $\partial\colon S_1 \to S_0$.

There is an obvious left action of Γ on S_* commuting with ∂. For any $k \geq 0$, let $S_{k,\Gamma}$ be the space of Γ-coinvariants. Since the boundary map ∂ commutes with the Γ-action, we obtain a complex $(S_{*,\Gamma}, \partial_\Gamma)$. The following theorem shows that this complex computes the cohomology of Γ:

Theorem A.26 ([**Ash94**])**.** *There is a natural isomorphism*

$$H^{\nu-k}(\Gamma; \mathbb{C}) \xrightarrow{\sim} H_k(S_{*,\Gamma} \otimes \mathbb{C}).$$

A.5.3. We can extend our norm function $\|\ \|$ from modular symbols to all of S_k as follows. Let $\mathbf{u} = [v_1, \ldots, v_{n+k}]$ be a k-sharbly, and let $Z(\mathbf{u})$ be the set of all submodular symbols determined by $\mathbf{u}$. In other words, $Z(\mathbf{u})$ consists of the modular symbols of the form $[v_{i_1}, \ldots, v_{i_n}]$, where $\{i_1, \ldots, i_n\}$ ranges over all n-fold subsets of $\{1, \ldots, n+k\}$. Define $\|\mathbf{u}\|$ by

$$\|\mathbf{u}\| = \operatorname*{Max}_{\mathbf{v} \in Z(\mathbf{u})} \|\mathbf{v}\|.$$

Note that $\|\mathbf{u}\|$ is well defined modulo the relations in Definition A.25. As for modular symbols, we extend the norm to sharbly chains $\xi = \sum q(\mathbf{u})\mathbf{u}$ taking the maximum norm over the support. Formally, we let $\operatorname{supp}(\xi) = \{\mathbf{u} \mid q(\mathbf{u}) \neq 0\}$ and $Z(\xi) = \bigcup_{\mathbf{u} \in \operatorname{supp}(\xi)} Z(\mathbf{u})$, and then we define $\|\xi\|$ by

$$\|\xi\| = \operatorname*{Max}_{\mathbf{v} \in Z(\xi)} \|\mathbf{v}\|.$$

We say that ξ is *reduced* if $\|\xi\| = 1$. Hence ξ is reduced if and only if all its submodular symbols are unimodular or have determinant 0. Clearly there are only finitely many reduced k-sharblies modulo Γ for any k.

In general the cohomology groups $H^*(\Gamma; \mathbb{C})$ are *not* spanned by reduced sharblies. However, it is known (cf. [**McC91**]) that for $\Gamma \subset \mathrm{SL}_4(\mathbb{Z})$, the group $H^5(\Gamma; \mathbb{C})$ is spanned by reduced 1-sharbly cycles. The best one can say in general is that for each pair n, k, there is an integer $N = N(n,k)$ such that for $\Gamma \subset \mathrm{SL}_n(\mathbb{Z})$, $H^{\nu-k}(\Gamma; \mathbb{C})$ is spanned by k-sharblies of norm $\leq N$. This set of sharblies is also finite modulo Γ, although it is not known how large N must be for any given pair n, k.

A.5.4. Recall that the cells of the well-rounded retract W are indexed by sets of primitive vectors in $\mathbb{Z}^n$. Since any primitive vector determines a point in $\mathbb{Q}^n \smallsetminus \{0\}$ and since sets of such points index sharblies, it is clear that there is a close relationship between S_* and the chain complex associated to W,

[10] The terminology for S_* is due to Lee Rudolph, in honor of Lee and Szczarba. They introduced a very similar complex in [**LS76**] for $\mathrm{SL}_3(\mathbb{Z})$.

although of course S_* is much bigger. In any case, both complexes compute $H^*(\Gamma;\mathbb{C})$.

The main benefit of using the sharbly complex to compute cohomology is that it admits a Hecke action. Suppose $\xi = \sum q(\mathbf{u})\mathbf{u}$ is a sharbly cycle mod Γ, and consider a Hecke operator T_g. Then we have

$$T_g(\xi) = \sum_{h\in\Omega,\mathbf{u}} n(\mathbf{u})h\cdot\mathbf{u}, \tag{A.5.1}$$

where Ω is a set of coset representatives as in (A.4.3). Since $\Omega \not\subset \mathrm{SL}_n(\mathbb{Z})$ in general, the Hecke image of a reduced sharbly is not usually reduced.

A.5.5. We are now ready to describe our algorithm for the computation of the Hecke operators on $H^{\nu-1}(\Gamma;\mathbb{C})$. It suffices to describe an algorithm that takes as input a 1-sharbly cycle ξ and produces as output a cycle ξ' with

(a) the classes of ξ and ξ' in $H^{\nu-1}(\Gamma;\mathbb{C})$ the same, and

(b) $\|\xi'\| < \|\xi\|$ if $\|\xi\| > 1$.

Below, we will present an algorithm satisfying (a). In [**Gun00a**], we conjectured (and presented evidence) that the algorithm satisfies (b) for $n \le 4$. Further evidence is provided by the computations in [**AGM02**], which relied on the algorithm to compute the Hecke action on $H^5(\Gamma;\mathbb{C})$, where $\Gamma = \Gamma_0(N) \subset \mathrm{SL}_4(\mathbb{Z})$.

The idea behind the algorithm is simple: given a 1-sharbly cycle ξ that is not reduced, (i) simultaneously apply the modular symbol algorithm (Theorem A.22) to each of its submodular symbols, and then (ii) package the resulting data into a new 1-sharbly cycle. Our experience in presenting this algorithm is that most people find the geometry involved in (ii) daunting. Hence we will give details only for $n = 2$ and will provide a sketch for $n > 2$. Full details are contained in [**Gun00a**]. Note that $n = 2$ is topologically and arithmetically uninteresting, since we are computing the Hecke action on $H^0(\Gamma;\mathbb{C})$; nevertheless, the geometry faithfully represents the situation for all n.

A.5.6. Fix $n = 2$, let $\xi \in S_1$ be a 1-sharbly cycle mod Γ for some $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$, and suppose ξ is not reduced. Assume Γ is torsion-free to simplify the presentation.

Suppose first that all submodular symbols $\mathbf{v} \in Z(\xi)$ are nonunimodular. Select reducing points for each $\mathbf{v} \in Z(\xi)$ and make these choices Γ-equivariantly. This means the following. Suppose $\mathbf{u}, \mathbf{u}' \in \operatorname{supp}\xi$ and $\mathbf{v} \in \operatorname{supp}(\partial\mathbf{u})$ and $\mathbf{v}' \in \operatorname{supp}(\partial\mathbf{u}')$ are modular symbols such that $\mathbf{v} = \gamma\cdot\mathbf{v}'$ for some $\gamma \in \Gamma$. Then we select reducing points w for $\mathbf{v}$ and w' for $\mathbf{v}'$ such that $w = \gamma \cdot w'$. (Note that since Γ is torsion-free, no modular symbol can be identified to itself by an element of Γ; hence $\mathbf{v} \neq \mathbf{v}'$.) This is possible since if $\mathbf{v}$ is a modular symbol and w is a reducing point for $\mathbf{v}$, then $\gamma \cdot w$ is a reducing point for $\gamma \cdot \mathbf{v}$ for any $\gamma \in \Gamma$. Because there are only finitely many Γ-orbits in $Z(\xi)$, we can choose reducing points Γ-equivariantly by selecting them for some set of orbit representatives.

It is important to note that Γ-equivariance is the only global criterion we use when selecting reducing. In particular, there is a priori no relationship among the three reducing points chosen for any $\mathbf{u} \in \operatorname{supp} \xi$.

A.5.7. Now we want to use the reducing points and the 1-sharblies in ξ to build ξ'. Choose $\mathbf{u} = [v_1, v_2, v_3] \in \operatorname{supp} \xi$, and denote the reducing point for $[v_i, v_j]$ by w_k, where $\{i, j, k\} = \{1, 2, 3\}$. We use the v_i and the w_i to build a 2-sharbly chain $\eta(\mathbf{u})$ as follows.

Let P be an octahedron in $\mathbb{R}^3$. Label the vertices of P with the v_i and w_i such that the vertex labeled v_i is opposite the vertex labeled w_i (Figure A.5.1). Subdivide P into four tetrahedra by connecting two opposite vertices, say v_1 and w_1, with an edge (Figure A.5.2). For each tetrahedron T, take the labels of four vertices and arrange them into a quadruple. If we orient P, then we can use the induced orientation on T to order the four primitive points. In this way, each T determines a 2-sharbly, and $\eta(\mathbf{u})$ is defined to be the sum. For example, if we use the decomposition in Figure A.5.2, we have

(A.5.2)
$$\eta(\mathbf{u}) = [v_1, v_3, v_2, w_1] + [v_1, w_2, v_3, w_1] + [v_1, w_3, w_2, w_1] + [v_1, v_2, w_3, w_1].$$

Repeat this construction for all $\mathbf{u} \in \operatorname{supp} \xi$, and let $\eta = \sum q(\mathbf{u})\eta(\mathbf{u})$. Finally, let $\xi' = \xi + \partial\eta$.

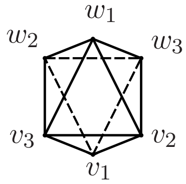


Figure A.5.1.

A.5.8. By construction, ξ' is a cycle mod Γ in the same class as ξ. We claim in addition that no submodular symbol from ξ appears in ξ'. To see this, consider $\partial\eta(\mathbf{u})$. From (A.5.2), we have

$$\begin{aligned}(\text{A.5.3}) \quad \partial\eta(\mathbf{u}) = &-[v_1, v_2, v_3] + [v_1, v_2, w_3] + [v_1, w_2, v_3] + [w_1, v_2, v_3] \\ &- [v_1, w_2, w_3] - [w_1, v_2, w_3] - [w_1, w_2, v_3] + [w_1, w_2, w_3].\end{aligned}$$

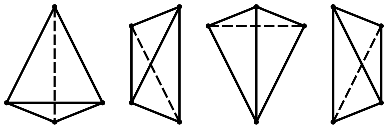


Figure A.5.2.

Note that this is the boundary in S_*, not in $S_{*,\Gamma}$. Furthermore, $\partial\eta(\mathbf{u})$ is independent of which pair of opposite vertices of P we connected to build $\eta(\mathbf{u})$.

From (A.5.3), we see that in $\xi + \partial\eta$ the 1-sharbly $-[v_1, v_2, v_3]$ is canceled by $\mathbf{u} \in \operatorname{supp}\xi$. We also claim that 1-sharblies in (A.5.3) of the form $[v_i, v_j, w_k]$ vanish in $\partial_\Gamma\eta$.

To see this, let $\mathbf{u}$,$\mathbf{u}' \in \operatorname{supp}\xi$, and suppose $\mathbf{v} = [v_1, v_2] \in \operatorname{supp}\partial\mathbf{u}$ equals $\gamma \cdot \mathbf{v}'$ for some $\mathbf{v}' = [v_1', v_2'] \in \operatorname{supp}\partial\mathbf{u}'$. Since the reducing points were chosen Γ-equivariantly, we have $w = \gamma \cdot w'$. This means that the 1-sharbly $[v_1, v_2, w] \in \partial\eta(\mathbf{u})$ will be canceled mod Γ by $[v_1', v_2', w'] \in \partial\eta(\mathbf{u}')$. Hence, in passing from ξ to ξ', the effect in $(S_*)_\Gamma$ is to replace $\mathbf{u}$ with *four* 1-sharblies in $\operatorname{supp}\xi'$:

$$[v_1, v_2, v_3] \longmapsto -[v_1, w_2, w_3] - [w_1, v_2, w_3] - [w_1, w_2, v_3] + [w_1, w_2, w_3]. \tag{A.5.4}$$

Note that in (A.5.4), there are no 1-sharblies of the form $[v_i, v_j, w_k]$.

Remark A.27. For implementation purposes, it is not necessary to explicitly construct η. Rather, one may work directly with (A.5.4).

A.5.9. Why do we expect ξ' to satisfy $\|\xi'\| < \|\xi\|$? First of all, in the right hand side of (A.5.4) there are no submodular symbols of the form $[v_i, v_j]$. In fact, any submodular symbol involving a point v_i also includes a reducing point for $[v_i, v_j]$.

On the other hand, consider the submodular symbols in (A.5.4) of the form $[w_i, w_j]$. Since there is no relationship among the w_i, one has no reason to believe that these modular symbols are closer to unimodularity than those in $\mathbf{u}$. Indeed, for certain choices of reducing points it can happen that $\|[w_i, w_j]\| \geq \|\mathbf{u}\|$.

The upshot is that some care must be taken in choosing reducing points. In [**Gun00a**, Conjectures 3.5 and 3.6] we describe two methods for finding reducing points for modular symbols, one using Voronoĭ reduction and one using LLL-reduction. Our experience is that if one selects reducing points using either of these conjectures, then $\|[w_i, w_j]\| < \|\mathbf{u}\|$ for each of the new modular symbols $[w_i, w_j]$. In fact, in practice these symbols are trivial or satisfy $\|[w_i, w_j]\| = 1$.

A.5.10. In the previous discussion we assumed that no submodular symbols of any $\mathbf{u} \in \operatorname{supp} \xi$ were unimodular. Now we say what to do if some are. There are three cases to consider.

First, all submodular symbols of $\mathbf{u}$ may be unimodular. In this case there are no reducing points, and (A.5.4) becomes

$$[v_1, v_2, v_3] \longmapsto [v_1, v_2, v_3]. \tag{A.5.5}$$

Second, one submodular symbol of $\mathbf{u}$ may be nonunimodular, say the symbol $[v_1, v_2]$. In this case, to build η, we use a tetrahedron P' and put $\eta(\mathbf{u}) = [v_1, v_2, v_3, w_3]$ (Figure A.5.3). Since $[v_1, v_2, w_3]$ vanishes in the boundary of η mod Γ, (A.5.4) becomes

$$[v_1, v_2, v_3] \mapsto -[v_1, v_3, w_3] + [v_2, v_3, w_3]. \tag{A.5.6}$$

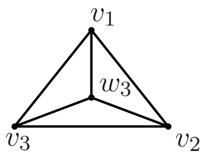


Figure A.5.3.

Finally, two submodular symbols of $\mathbf{u}$ may be nonunimodular, say $[v_1, v_2]$ and $[v_1, v_3]$. In this case we use the cone on a square P'' (Figure A.5.4). To construct $\eta(\mathbf{u})$, we must choose a decomposition of P'' into tetrahedra. Since P'' has a nonsimplicial face, this choice affects ξ' (in contrast to the previous cases). If we subdivide P'' by connecting the vertex labelled v_2 with the vertex labelled w_2, we obtain

$$[v_1, v_2, v_3] \longmapsto [v_2, w_2, w_3] + [v_2, v_3, w_2] + [v_1, v_3, w_2]. \tag{A.5.7}$$

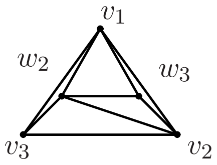


Figure A.5.4.

A.5.11. Now consider general n. The basic technique is the same, but the combinatorics become more complicated. Suppose $\mathbf{u} = [v_1, \ldots, v_{n+1}]$ satisfies $q(\mathbf{u}) \neq 0$ in a 1-sharbly cycle ξ, and for $i = 1, \ldots, n+1$ let $\mathbf{v}_i$ be the submodular symbol $[v_1, \ldots, \widehat{v_i}, \ldots, v_{n+1}]$. Assume that all $\mathbf{v}_i$ are nonunimodular, and for each i let w_i be a reducing point for $\mathbf{v}_i$.

For any subset $I \subset \{1, \dots, n+1\}$, let $\mathbf{u}_I$ be the 1-sharbly $[u_1, \dots, u_{n+1}]$, where $u_i = w_i$ if $i \in I$, and $u_i = v_i$ otherwise. The polytope P used to build $\eta(\mathbf{u})$ is the *cross polytope*, which is the higher-dimensional analogue of the octahedron [**Gun00a**, §4.4]. We suppress the details and give the final answer: (A.5.4) becomes

$$\mathbf{u} \longmapsto -\sum_I (-1)^{\#I} \mathbf{u}_I, \tag{A.5.8}$$

where the sum is taken over all subsets $I \subset \{1, \dots, n+1\}$ of cardinality *at least* 2.

More generally, if some $\mathbf{v}_i$ happen to be unimodular, then the polytope used to build η is an iterated cone on a lower-dimensional cross polytope. This is already visible for $n = 2$:

- The 2-dimensional cross polytope is a square, and the polytope P'' is a cone on a square.
- The 1-dimensional cross polytope is an interval, and the polytope P' is a double cone on an interval.

Altogether there are $n+1$ relations generalizing (A.5.5)–(A.5.7).

A.5.12. Now we describe how these computations are carried out in practice, focusing on $\Gamma = \Gamma_0(N) \subset \mathrm{SL}_4(\mathbb{Z})$ and $H^5(\Gamma; \mathbb{C})$. Besides discussing technical details, we also have to slightly modify some aspects of the construction in Section A.5.6, since Γ is not torsion-free.

Let W be the well-rounded retract. We can represent a cohomology class $\beta \in H^5(\Gamma; \mathbb{C})$ as $\beta = \sum q(\sigma)\sigma$, where σ denotes a codimension 1 cell in W. In this case there are three types of codimension 1 cells in W. Under the bijection $W \leftrightarrow \mathscr{C}$, these cells correspond to the Voronoĭ cells indexed by the columns of the matrices

(A.5.9)
$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Thus each σ in W modulo Γ corresponds to an $\mathrm{SL}_4(\mathbb{Z})$-translate of one of the matrices in (A.5.9). These translates determine basis 1-sharblies $\mathbf{u}$ (by taking the points u_i to be the columns), and hence we can represent β by a 1-sharbly chain $\xi = \sum q(\mathbf{u})\mathbf{u} \in S_1$ that is a cycle in the complex of coinvariants $(S_{*,\Gamma}, \partial_\Gamma)$.

To make later computations more efficient, we precompute more data attached to ξ. Given a 1-sharbly $\mathbf{u} = [u_1, \dots, u_{n+1}]$, a *lift* $M(\mathbf{u})$ of $\mathbf{u}$ is defined to be an integral matrix with primitive columns M_i such that $\mathbf{u} =$

$[M_1, \dots, M_{n+1}]$. Then we encode ξ, once and for all, by a finite collection Φ of 4-tuples

$$(\mathbf{u}, n(\mathbf{u}), \{\mathbf{v}\}, \{M(\mathbf{v})\}),$$

where

(1) $\mathbf{u}$ ranges over the support of ξ,

(2) $n(\mathbf{u}) \in \mathbb{C}$ is the coefficient of $\mathbf{u}$ in ξ,

(3) $\{\mathbf{v}\}$ is the set of submodular symbols appearing in the boundary of $\mathbf{u}$, and

(4) $\{M(\mathbf{v})\}$ is a set of lifts for $\{\mathbf{v}\}$.

Moreover, the lifts in (4) are chosen to satisfy the following Γ-equivariance condition. Suppose that for $\mathbf{u}, \mathbf{u}' \in \operatorname{supp} \xi$ we have $\mathbf{v} \in \operatorname{supp}(\partial \mathbf{u})$ and $\mathbf{v}' \in \operatorname{supp}(\partial \mathbf{u}')$ satisfying $\mathbf{v} = \gamma \cdot \mathbf{v}'$ for some $\gamma \in \Gamma$. Then we require $M(\mathbf{v}) = \gamma M(\mathbf{v}')$. This is possible since ξ is a cycle modulo Γ, although there is one complication since Γ has torsion: it can happen that some submodular symbol $\mathbf{v}$ of a 1-sharbly $\mathbf{u}$ is identified to *itself* by an element of Γ. This means that in constructing $\{M(\mathbf{v})\}$ for $\mathbf{u}$, we must somehow choose more than one lift for $\mathbf{v}$. To deal with this, let $M(\mathbf{v})$ be any lift of $\mathbf{v}$, and let $\Gamma(\mathbf{v}) \subset \Gamma$ be the stabilizer of $\mathbf{v}$. Then in ξ, we replace $q(\mathbf{u})\mathbf{u}$ by

$$\frac{1}{\#\Gamma(\mathbf{v})} \sum_{\gamma \in \Gamma(\mathbf{v})} q(\mathbf{u}) \mathbf{u}_\gamma,$$

where $\mathbf{u}_\gamma$ has the same data as $\mathbf{u}$, except[11] that we give $\mathbf{v}$ the lift $\gamma M(\mathbf{v})$.

Next we compute and store the 1-sharbly transformation laws generalizing (A.5.5)–(A.5.7). As a part of this we fix triangulations of certain cross polytopes as in (A.5.7).

We are now ready to begin the actual reduction algorithm. We take a Hecke operator $T(l, k)$ and build the coset representatives Ω as in (A.5.1). For each $h \in \Omega$ and each 1-sharbly $\mathbf{u}$ in the support of ξ, we obtain a nonreduced 1-sharbly $\mathbf{u}_h := h \cdot \mathbf{u}$. Here h acts on all the data attached to $\mathbf{u}$ in the list Φ. In particular, we replace each lift $M(\mathbf{v})$ with $h \cdot M(\mathbf{v})$, where the dot means matrix multiplication.

Now we check the submodular symbols of $\mathbf{u}_h$ and choose reducing points for the nonunimodular symbols. This is where the lifts come in handy. Recall that reduction points must be chosen Γ-equivariantly over the entire cycle. Instead of explicitly keeping track of the identifications between modular symbols, we do the following trick:

[11] In fact, we can be slightly more clever than this and only introduce denominators that are powers of 2.

(1) Construct the *Hermite normal form* $M_{\mathrm{her}}(\mathbf{v})$ of the lift $M(\mathbf{v})$ (see [**Coh93**, §2.4] and Exercise 7.5). Record the transformation matrix $U \in \mathrm{GL}_4(\mathbb{Z})$ such that $UM(\mathbf{v}) = M_{\mathrm{her}}(\mathbf{v})$.

(2) Choose a reducing point u for $M_{\mathrm{her}}(\mathbf{v})$.

(3) Then the reducing point for $M(\mathbf{v})$ is $U^{-1}u$.

This guarantees Γ-equivariance: if $\mathbf{v}$, $\mathbf{v}'$ are submodular symbols of ξ with $\gamma\cdot\mathbf{v} = \mathbf{v}'$ and with reducing points u, u', we have $\gamma u = u'$. The reason is that the Hermite normal form $M_{\mathrm{her}}(\mathbf{v})$ is a *uniquely determined* representative of the $\mathrm{GL}_4(\mathbb{Z})$-orbit of $M(\mathbf{v})$ [**Coh93**]. Hence if $\gamma M(\mathbf{v}) = M(\mathbf{v}')$, then $M_{\mathrm{her}}(\mathbf{v}) = M_{\mathrm{her}}(\mathbf{v}')$.

After computing all reducing points, we apply the appropriate transformation law. The result will be a chain of 1-sharblies, each of which has (conjecturally) smaller norm than the original 1-sharbly $\mathbf{u}$. We output these 1-sharblies if they are reduced; otherwise they are fed into the reduction algorithm again. Eventually we obtain a reduced 1-sharbly cycle ξ' homologous to the original cycle ξ.

The final step of the algorithm is to rewrite ξ' as a cocycle on W. This is easy to do since the relevant cells of W are in bijection with the reduced 1-sharblies. There are some nuisances in keeping orientations straight, but the computation is not difficult. We refer to [**AGM02**] for details.

A.5.13. We now give some examples, taken from [**AGM02**], of Hecke eigenclasses in $H^5(\Gamma_0(N);\mathbb{C})$ for various levels N. Instead of giving a table of eigenvalues, we give the *Hecke polynomials*. If β is an eigenclass with $T(l,k)(\beta) = a(l,k)\beta$, then we define

$$H(\beta, l) = \sum_k (-1)^k l^{k(k-1)/2} a(l,k) X^k \in \mathbb{C}[X].$$

For almost all l, after putting $X = l^{-s}$ where s is a complex variable, the function $H(\beta, s)$ is the inverse of the local factor at l of the automorphic representation attached to β.

Example A.28. Suppose $N = 11$. Then the cohomology $H^5(\Gamma_0(11);\mathbb{C})$ is 2-dimensional. There are two Hecke eigenclasses u_1, u_2, each with rational Hecke eigenvalues.

u_1	T_2	$(1-4X)(1-8X)(1+2X+2X^2)$
	T_3	$(1-9X)(1-27X)(1+X+3X^2)$
	T_5	$(1-25X)(1-125X)(1-X+5X^2)$
	T_7	$(1-49X)(1-343X)(1+2X+7X^2)$
u_2	T_2	$(1-X)(1-2X)(1+8X+32X^2)$
	T_3	$(1-X)(1-3X)(1+9X+243X^2)$
	T_5	$(1-X)(1-5X)(1-25X+3125X^2)$
	T_7	$(1-X)(1-7X)(1+98X+16807X^2)$

Example A.29. Suppose $N = 19$. Then the cohomology $H^5(\Gamma_0(19);\mathbb{C})$ is 3-dimensional. There are three Hecke eigenclasses u_1, u_2, u_3, each with rational Hecke eigenvalues.

u_1	T_2	$(1-4X)(1-8X)(1+2X^2)$
	T_3	$(1-9X)(1-27X)(1+2X+3X^2)$
	T_5	$(1-25X)(1-125X)(1-3X+5X^2)$
u_2	T_2	$(1-X)(1-2X)(1+32X^2)$
	T_3	$(1-X)(1-3X)(1+18X+243X^2)$
	T_5	$(1-X)(1-5X)(1-75X+3125X^2)$
u_3	T_2	$(1-2X)(1-4X)(1+3X+8X^2)$
	T_3	$(1-3X)(1-9X)(1+5X+27X^2)$
	T_5	$(1-5X)(1-25X)(1+12X+125X^2)$

In these examples, the cohomology is completely accounted for by the Eisenstein summand of (A.2.8). In fact, let $\Gamma_0'(N) \subset \mathrm{SL}_2(\mathbb{Z})$ be the usual Hecke congruence subgroup of matrices upper-triangular modulo N. Then the cohomology classes above actually come from classes in $H^1(\Gamma_0'(N))$, that is from holomorphic modular forms of level N.

For $N = 11$, the space of weight two cusp forms $S_2(11)$ is 1-dimensional. This cusp form f lifts in two different ways to $H^5(\Gamma_0(11);\mathbb{C})$, which can be seen from the quadratic part of the Hecke polynomials for the u_i. Indeed, for u_i the quadratic part is exactly the inverse of the local factor for the L-function attached to f, after the substitution $X = l^{-s}$. For u_2, we see that the lift is also twisted by the square of the cyclotomic character. (In fact the linear terms of the Hecke polynomials come from powers of the cyclotomic character.)

For $N = 19$, the space of weight two cusp forms $S_2(19)$ is again 1-dimensional. The classes u_1 and u_2 are lifts of this form, exactly as for $N = 11$. The class u_3, on the other hand, comes from $S_4(19)$, the space of weight 4 cusp forms on $\Gamma_0'(19)$. In fact, $\dim S_4(19) = 4$, with one Hecke eigenform defined over $\mathbb{Q}$ and another defined over a totally real cubic extension of $\mathbb{Q}$. Only the rational weight four eigenform contributes to $H^5(\Gamma_0(19);\mathbb{C})$. One can show that whether or not a weight four cuspidal eigenform f contributes to the cohomology of $\Gamma_0(N)$ depends only on the sign of the functional equation of $L(f, s)$ [**Wes**]. This phenomenon is typical of what one encounters when studying Eisenstein cohomology.

In addition to the lifts of weight 2 and weight 4 cusp forms, for other levels one finds lifts of Eisenstein series of weights 2 and 4 and lifts of cuspidal cohomology classes from subgroups of $\mathrm{SL}_3(\mathbb{Z})$. For some levels one finds cuspidal classes that appear to be lifts from the group of symplectic similitudes GSp(4). More details can be found in [**AGM02, AGM**].

A.5.14. Here are some notes on the reduction algorithm and its implementation:

- Some additional care must be taken when selecting reducing points for the submodular symbols of $\mathbf{u}$. In particular, in practice one should choose w for $\mathbf{v}$ such that $\sum \|\mathbf{v}_i(w)\|$ is minimized. Similar remarks apply when choosing a subdivision of the crosspolytopes in Section A.5.10.
- In practice, the reduction algorithm has *always* terminated with a reduced 1-sharbly cycle ξ' homologous to ξ. However, at the moment we cannot prove that this will always happen.
- Experimentally, the efficiency of the reduction step appears to be comparable to that of Theorem A.22. In other words the depth of the "reduction tree" associated to a given 1-sharbly $\mathbf{u}$ seems to be bounded by a polynomial in $\log\log \|\mathbf{u}\|$. Hence computing the Hecke action using this algorithm is extremely efficient.

  On the other hand, computing Hecke operators on SL_4 is still a much bigger computation—relative to the level—than on SL_2 and SL_3. For example, the size of the full retract W modulo $\Gamma_0(p)$ is roughly $O(p^6)$, which grows rapidly with p. The portion of the retract corresponding to H^5 is much smaller, around $p^3/10$, but this still grows quite quickly. This makes computing with $p > 100$ out of reach at the moment.

  The number of Hecke cosets grows rapidly as well, e.g., the number of coset representatives of $T(l, 2)$ is $l^4 + l^3 + 2l^2 + l + 1$. Hence it is only feasible to compute Hecke operators for small l; for large levels only $l = 2$ is possible.

  Here are some numbers to give an idea of the size of these computations. For level 73, the rank of H^5 is 20. There are 39504 cells of codimension 1 and 4128 top-dimensional cells in W modulo $\Gamma_0(73)$. The computational techniques in [**AGM02**] used at this level (a Lanczos scheme over a large finite field) tend to produce sharbly cycles supported on almost all the cells. Computing $T(2, 1)$

requires a reduction tree of depth 1 and produces as many as 26 reduced 1-sharblies for each of the 15 nonreduced Hecke images. Thus one cycle produces a cycle supported on as many as 15406560 1-sharblies, all of which must be converted to an appropriate cell of W modulo Γ. Also this is just what needs to be done for *one* cycle; do not forget that the rank of H^5 is 20.

In practice the numbers are slightly better, since the reduction step produces fewer 1-sharblies on average and since the support of the initial cycle has size less than 39504. Nevertheless the orders of magnitude are correct.

- Using lifts is a convenient way to encode the global Γ-identifications in the cycle ξ, since it means we do not have to maintain a big data structure keeping track of the identifications on $\partial\xi$. However, there is a certain expense in computing the Hermite normal form. This is balanced by the benefit that working with the data Φ associated to ξ allows us to reduce the supporting 1-sharblies $\mathbf{u}$ *independently.* This means we can cheaply parallelize our computation: each 1-sharbly, encoded as a 4-tuple $(\mathbf{u}, n(\mathbf{u}), \{\mathbf{v}\}, \{M(\mathbf{v})\})$, can be handled by a separate computer. The results of all these individual computations can then be collated at the end, when producing a W-cocycle.

A.6. Complements and Open Problems

A.6.1. We conclude this appendix by giving some complements and describing some possible directions for future work, both theoretical and computational. Since a full explanation of the material in this section would involve many more pages, we will be brief and will provide many references.

A.6.2. Perfect Quadratic Forms over Number Fields and Retracts. Since Voronoĭ's pioneering work [**Vor08**], it has been the goal of many to extend his results from $\mathbb{Q}$ to a general algebraic number field F. Recently Coulangeon [**Cou01**], building on work of Icaza and Baeza [**Ica97, BI97**], has found a good notion of *perfection* for quadratic forms over number fields[12]. One of the key ideas in [**Cou01**] is that the correct notion of equivalence between Humbert forms involves not only the action of $\mathrm{GL}_n(\mathscr{O}_F)$, where $\mathscr{O}_F$ is the ring of integers of F, but also the action of a certain continuous group U related to the units $\mathscr{O}_F^\times$. One of Coulangeon's basic results is that there are finitely many equivalence classes of perfect Humbert forms modulo these actions.

[12] Such forms are called *Humbert forms* in the literature.

On the other hand, Ash's original construction of retracts [**Ash77**] introduces a geometric notion of perfection. Namely he generalizes the Voronoĭ polyhedron Π and defines a quadratic form to be perfect if it naturally indexes a facet of Π. What is the connection between these two notions? Can one use Coulangeon's results to construct cell complexes to be used in cohomology computations? One tempting possibility is to try to use the group U to collapse the Voronoĭ cells of [**Ash77**] into a cell decomposition of the symmetric space associated to $\mathrm{SL}_n(F)$.

A.6.3. The Modular Complex. In his study of multiple ζ-values, Goncharov has recently defined the *modular complex* M^* [**Gon97, Gon98**]. This is an n-step complex of $\mathrm{GL}_n(\mathbb{Z})$-modules closely related both to the properties of multiple polylogarithms evaluated at μ_N, the Nth roots of unity, and to the action of $G_{\mathbb{Q}}$ on $\pi_{1,N} = \pi_1^l(\mathbb{P}^1 \smallsetminus \{0, \infty, \mu_N\})$, the pro-$l$ completion of the algebraic fundamental group of $\mathbb{P}^1 \smallsetminus \{0, \infty, \mu_N\}$.

Remarkably, the modular complex is very closely related to the Voronoĭ decomposition $\mathscr{V}$. In fact, one can succinctly describe the modular complex by saying that it is the chain complex of the cells coming from the top-dimensional Voronoĭ cone of type A_n. This is all of the Voronoĭ decomposition for $n = 2, 3$, and Goncharov showed that the modular complex is quasi-isomorphic to the full Voronoĭ complex for $n = 4$. Hence there is a precise relationship among multiple polylogarithms, the Galois action on $\pi_{1,N}$, and the cohomology of level N congruence subgroups of $\mathrm{SL}_n(\mathbb{Z})$.

The question then arises, how much of the cohomology of congruence subgroups is captured by the modular complex for all n? Table A.3.2 indicates that asymptotically very little of the Voronoĭ decomposition comes from the A_n cone, but this says nothing about the cohomology. The first interesting case to consider is $n = 5$.

A.6.4. Retracts for Other Groups. The most general construction of retracts W known [**Ash84**] applies only to *linear* symmetric spaces. The most familiar example of such a space is $\mathrm{SL}_n(\mathbb{R})/\mathrm{SO}(n)$; other examples are the symmetric spaces associated to SL_n over number fields and division algebras.

Now let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic group, and let $X = G/K$ be the associated symmetric space. What can one say about cell complexes that can be used to compute $H^*(\Gamma; \mathscr{M})$? The theorem of Borel–Serre mentioned in Section A.3.3 implies the vanishing of $H^k(\Gamma; \mathscr{M})$ for $k > \nu := \dim X - q$, where q is the $\mathbb{Q}$-*rank* of Γ. For example, for the split form of SL_n, the $\mathbb{Q}$-rank is $n - 1$. For the split symplectic group Sp_{2n}, the $\mathbb{Q}$-rank is n. Moreover, this bound is sharp: there will be coefficient modules $\mathscr{M}$ for which $H^{\nu}(\Gamma; \mathscr{M}) \neq 0$. Hence any minimal cell complex used to compute the cohomology of Γ should have dimension ν.

Ideally one would like to see such a complex realized as a subspace of X and would like to be able to treat all finite index subgroups of Γ simultaneously. This leads to the following question: is there a Γ-equivariant deformation retraction of X onto a regular cell complex W of dimension ν?

For $\mathbf{G} = \mathrm{Sp}_4$, McConnell and MacPherson showed that the answer is yes. Their construction begins by realizing the symplectic symmetric space X_{Sp} as a subspace of the special linear symmetric space X_{SL}. They then construct subsets of X_{Sp} by intersecting the Voronoĭ cells in X_{SL} with X_{Sp}. Through explicit computations in coordinates they prove that these intersections are cells and give a cell decomposition of X_{Sp}. By taking an appropriate dual complex (as suggested by Figures A.3.2 and A.3.3 and as done in [**Ash77**]), they construct the desired cell complex W.

Other progress has been recently made by Bullock [**Bul00**], Bullock and Connell [**BC06**], and Yasaki [**Yas05b, Yas05a**] in the case of groups of $\mathbb{Q}$-rank 1. In particular, Yasaki uses the *tilings* of Saper [**Sap97**] to construct an explicit retract for the unitary group $\mathrm{SU}(2,1)$ over the Gaussian integers. His method also works for Hilbert modular groups, although further refinement may be needed to produce a regular cell complex. Can one generalize these techniques to construct retracts for groups of arbitrary $\mathbb{Q}$-rank? Is there an analogue of the Voronoĭ decomposition for these retracts (i.e., a dual cell decomposition of the symmetric space)? If so, can one generalize ideas in Sections A.4–A.5 and use that generalization to compute the action of the Hecke operators on the cohomology?

A.6.5. Deeper Cohomology Groups. The algorithm in Section A.5 can be used to compute the Hecke action on $H^{\nu-1}(\Gamma)$. For $n > 4$, this group no longer contains cuspidal cohomology classes. Can one generalize this algorithm to compute the Hecke action on deeper cohomology groups? The first practical case is $n = 5$. Here $\nu = 10$, and the highest degree in which cuspidal cohomology can live is 8. This case is also interesting since the cohomology of full level has been studied [**EVGS02**].

Here are some indications of what one can expect. The general strategy is the same: for a k-sharbly ξ representing a class in $H^{\nu-k}(\Gamma)$, begin by Γ-equivariantly choosing reducing points for the nonunimodular submodular symbols of ξ. This data can be packaged into a new k-sharbly cycle as in Section A.5.7ff, but the crosspolytopes must be replaced with *hypersimplices*. By definition, the hypersimplex $\Delta(n,k)$ is the convex hull in $\mathbb{R}^n$ of the points $\{\sum_{i\in I} e_i\}$, where I ranges over all order k subsets of $\{1,\dots,n\}$ and $e_1,\dots,e_n$ denotes the standard basis of $\mathbb{R}^n$.

The simplest example is $n = 2$, $k = 2$. From the point of view of cohomology, this is even less interesting than $n = 2$, $k = 1$, since now we are computing the Hecke action on $H^{-1}(\Gamma)$! Nevertheless, the geometry here illustrates what one can expect in general.

Each 2-sharbly in the support of ξ can be written as $[v_1, v_2, v_3, v_4]$ and determines six submodular symbols, of the form $[v_i, v_j]$, $i \neq j$. Assume for simplicity that all these submodular symbols are nonunimodular. Let w_{ij} be the reducing point for $[v_i, v_j]$. Then use the ten points v_i, w_{ij} to label the vertices of the hypersimplex $\Delta(5, 2)$ as in Figure A.6.1 (note that $\Delta(5, 2)$ is 4-dimensional).

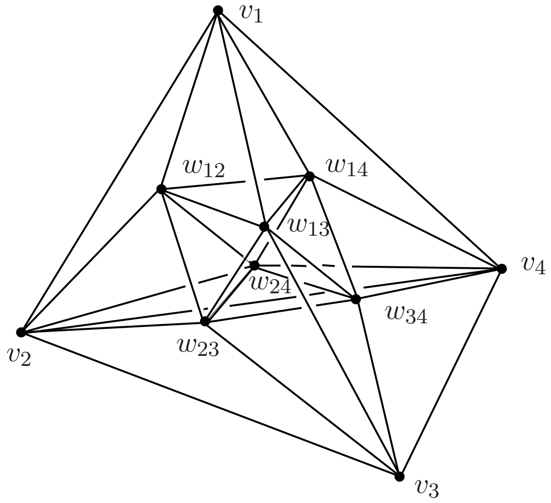


Figure A.6.1.

The boundary of this hypersimplex gives the analogue of (A.5.4). Which 2-sharblies will appear in ξ'? The boundary $\partial\Delta(5, 2)$ is a union of five tetrahedra and five octahedra. The outer tetrahedron will not appear in ξ', since that is the analogue of the left side of (A.5.4). The four octahedra sharing a triangular face with the outer tetrahedron also will not appear, since they disappear when considering ξ' modulo Γ. The remaining four tetrahedra and the central octahedron survive to ξ' and constitute the right side of the analogue of (A.5.4). Note that we must choose a simplicial subdivision of the central octahedron to write the result as a 2-sharbly cycle and that this must be done with care since it introduces a new submodular symbol.

If some submodular symbols are unimodular, then again one must consider iterated cones on hypersimplices, just as in Section A.5.10. The analogues of these steps become more complicated, since there are now many simplicial subdivisions of a hypersimplex[13]. There is one final complication: in general we cannot use reduced k-sharblies alone to represent cohomology classes. Thus one must terminate the algorithm when $\|\xi\|$ is less than some predetermined bound.

A.6.6. Other Linear Groups. Let F be a number field, and let $\mathbf{G} = \mathbf{R}_{F/\mathbb{Q}}(\mathrm{SL}_n)$ (Example A.2). Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. Can one compute the action of the Hecke operators on $H^*(\Gamma)$?

There are two completely different approaches to this problem. The first involves the generalization of the modular symbols method. One can define the analogue of the sharbly complex, and can try to extend the techniques of Sections A.4–A.5.

This technique has been extensively used when F is *imaginary quadratic* and $n = 2$. We have $X = \mathrm{SL}_2(\mathbb{C})/\,\mathrm{SU}(2)$, which is isomorphic to 3-dimensional hyperbolic space $\mathfrak{h}_3$. The arithmetic groups $\Gamma \subset \mathrm{SL}_2(\mathscr{O}_F)$ are known as *Bianchi groups*. The retracts and cohomology of these groups have been well studied; as a representative sample of works we mention [**Men79, EGM98, Vog85, GS81**].

Such groups have $\mathbb{Q}$-rank 1 and thus have cohomological dimension 2. One can show that the cuspidal classes live in degrees 1 and 2. This means that we can use modular symbols to investigate the Hecke action on cuspidal cohomology. This was done by Cremona [**Cre84**] for *euclidean* fields F. In that case Theorem A.22 works with no trouble (the euclidean algorithm is needed to construct reducing points). For noneuclidean fields further work has been done by Whitley [**Whi90**], Cremona and Whitely [**CW94**] (both for principal ideal domains), Bygott [**Byg99**] (for $F = \mathbb{Q}(\sqrt{-5})$ and any field with class group an elementary abelian 2-group), and Lingham [**Lin05**] (any field with odd class number). Putting all these ideas together allows one to generalize the modular symbols method to *any* imaginary quadratic field [**Cre**].

For F imaginary quadratic and $n > 2$, very little has been studied. The only related work to the best of our knowledge is that of Staffeldt [**Sta79**]. He determined the structure of the Voronoĭ polyhedron in detail for $\mathbf{R}_{F/\mathbb{Q}}(\mathrm{SL}_3)$, where $F = \mathbb{Q}(\sqrt{-1})$. We have $\dim X = 8$ and $\nu = 6$. The cuspidal cohomology appears in degrees $3, 4, 5$, so one could try to use the techniques of Section A.5 to investigate it.

Similar remarks apply to F *real quadratic* and $n = 2$. The symmetric space $X \simeq \mathfrak{h} \times \mathfrak{h}$ has dimension 4 and the $\mathbb{Q}$-rank is 1, which means $\nu = 3$. Unfortunately the cuspidal cohomology appears only in degree 2, which

[13] Indeed, computing all simplicial subdivisions of $\Delta(n, k)$ is a difficult problem in convex geometry.

means modular symbols cannot see it. On the other hand, 1-sharblies can see it, and so one can try to use ideas in Section A.5 here to compute the Hecke operators. The data needed to build the retract W already (essentially) appears in the literature for certain fields; see for example [**Ong86**].

The second approach shifts the emphasis from modular symbols and the sharbly complex to the Voronoĭ fan and its cones. For this approach we must assume that the group Γ is associated to a *self-adjoint homogeneous cone* over $\mathbb{Q}$. (cf. [**Ash77**]). This class of groups includes arithmetic subgroups of $\mathbf{R}_{F/\mathbb{Q}}(\mathrm{SL}_n)$, where F is a totally real or CM field. Such groups have all the nice structures in Section A.3.2. For example, we have a cone C with a G-action. We also have an analogue of the Voronoĭ polyhedron Π. There is a natural compactification $\tilde{C}$ of C obtained by adjoining certain self-adjoint homogeneous cones of lower rank. The quotient $\Gamma\backslash\tilde{C}$ is singular in general, but it can still be used to compute $H^*(\Gamma;\mathbb{C})$. The polyhedron Π can be used to construct a fan $\mathscr{V}$ that gives a Γ-equivariant decomposition of all of $\tilde{C}$. But the most important structure we have is the Voronoĭ reduction algorithm: given any point $x \in \tilde{C}$, we can determine the unique Voronoĭ cone containing x.

Here is how this setup can be used to compute the Hecke action. Full details are in [**Gun99, GM03**]. We define two chain complexes $\mathbf{C}_*^V$ and $\mathbf{C}_*^R$. The latter is essentially the chain complex generated by all simplicial rational polyhedral cones in $\tilde{C}$; the former is the subcomplex generated by the Voronoĭ cones. These are the analogues of the sharbly complex and the chain complex associated to the retract W, and one can show that either can be used to compute $H^*(\Gamma;\mathbb{C})$. Take a cycle $\xi \in \mathbf{C}_*^V$ representing a cohomology class in $H^*(\Gamma;\mathbb{C})$ and act on it by a Hecke operator T. We have $T(\xi) \in \mathbf{C}_*^R$, and we must push $T(\xi)$ back to $\mathbf{C}_*^V$.

To do this, we use the linear structure on $\tilde{C}$ to subdivide $T(\xi)$ very finely into a chain ξ'. For each 1-cone τ in $\operatorname{supp}\xi'$, we choose a 1-cone $\rho_\tau \in \tilde{C} \smallsetminus C$ and assemble them using the combinatorics of ξ' into a polyhedral chain ξ'' homologous to ξ'. Under certain conditions involved in the construction of ξ', this chain ξ'' will lie in $\mathbf{C}_*^V$.

We illustrate this process for the split group SL_2; more details can be found in [**Gun99**]. We work modulo homotheties, so that the three-dimensional cone $\tilde{C}$ becomes the extended upper half plane $\mathfrak{h}^* := \mathfrak{h}\cup\mathbb{Q}\cup\{\infty\}$, with $\partial\tilde{C}$ passing to the cusps $\mathfrak{h}^* \smallsetminus \mathfrak{h}$. As usual top-dimensional Voronoĭ cones become the triangles of the Farey tessellation, and the cones ρ_τ become cusps. Given any $x \in \mathfrak{h}$, let $R(x)$ be the set of cusps of the unique triangle or edge containing x (this can be computed using the Voronoĭ reduction algorithm). Extend R to a function on $\mathfrak{h}^*$ by putting $R(u) = \{u\}$ for any cusp u.

In $\mathfrak{h}$, the support of $T(\xi)$ becomes a geodesic μ between two cusps u, u', in other words the support of a modular symbol $[u, u']$ (Figure A.6.2). Subdivide μ by choosing points $x_0, \ldots, x_n$ such that $x_0 = u$, $x_n = u'$, and $R(x_i) \cap R(x_{i+1}) \neq \varnothing$. (This is easily done, for example by repeatedly barycentrically subdividing μ.) For each $i < n$ choose a cusp $q_i \in R(x_i) \cap R(x_{i+1})$, and put $q_n = u'$. Then we have a relation in H^1:

$$[u, u'] = [q_0, q_1] + \cdots + [q_{n-1}, q_n]. \tag{A.6.1}$$

Moreover, each $[q_i, q_{i+1}]$ is unimodular, since q_i and q_{i+1} are both vertices of a triangle containing x_{i+1}. Upon lifting (A.6.1) back to $\mathbf{C}_*^R$, the cusps q_i become the 1-cones ρ_τ and give us a relation $T(\xi) = \xi'' \in \mathbf{C}_*^V$.

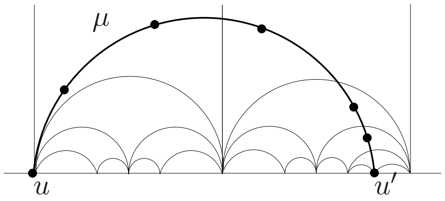


Figure A.6.2. A subdivision of μ; the solid dots are the x_i. Since the x_i lie in the same or adjacent Voronoĭ cells, we can assign cusps to them to construct a homology to a cycle in $\mathbf{C}_*^V$.

A.6.7. The Sharbly Complex for General Groups. In [**Gun00b**] we generalized Theorem A.22 (without the complexity statement) to the symplectic group Sp_{2n}. Using this algorithm and the symplectic retract [**MM93, MM89**], one can compute the action of the Hecke operators on the top-degree cohomology of subgroups of $\mathrm{Sp}_4(\mathbb{Z})$.

More recently, Toth has investigated modular symbols for other groups. He showed that the unimodular symbols generate the top-degree cohomology groups for Γ an arithmetic subgroup of a split classical group or a split group of type E_6 or E_7 [**Tot05**]. His technique of proof is completely different from that of [**Gun00b**]. In particular he does not give an analogue of the Manin trick. Can one extract an algorithm from Toth's proof that can be used to explicitly compute the action of the Hecke operators on cohomology?

The proof of the main result of [**Gun00b**] uses a description of the relations among the modular symbols. These relations were motivated by the structure of the cell complex in [**MM93, MM89**]. The modular symbols and these relations are analogues of the groups S_0 and S_1 in the sharbly complex. Can one extend these combinatorial constructions to form a *symplectic sharbly complex*? What about for general groups $\mathbf{G}$?

Already for Sp_4, resolution of this question would have immediate arithmetic applications. Indeed, Harder has a beautiful conjecture about certain congruences between holomorphic modular forms and Siegel modular forms of full level [**Hara**]. Examples of these congruences were checked numerically in [**Hara**] using techniques of [**FvdG**] to compute the Hecke action.

However, to investigate higher levels, one needs a different technique. The relevant cohomology classes live in $H^{\nu-1}(\Gamma;\mathscr{M})$, so one only needs to understand the first three terms of the complex $S_0 \leftarrow S_1 \leftarrow S_2$. We understand S_0, S_1 from [**Gun00b**]; the key is understanding S_2, which should encode relations among elements of S_1. If one could do this and then could generalize the techniques of [**Gun00a**], one would have a way to investigate Harder's conjecture.

A.6.8. Generalized Modular Symbols. We conclude this appendix by discussing a geometric approach to modular symbols. This complements the algebraic approaches presented in this book and leads to many new interesting phenomena and problems.

Suppose $\mathbf{H}$ and $\mathbf{G}$ are connected semisimple algebraic groups over $\mathbb{Q}$ with an injective map $f\colon \mathbf{H} \to \mathbf{G}$. Let K_H be a maximal compact subgroup of $H = \mathbf{H}(\mathbb{R})$, and suppose $K \subset G$ is a maximal compact subgroup containing $f(K_H)$. Let $X = G/K$ and $Y = H/K_H$.

Now let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be a torsion-free arithmetic subgroup. Let $\Gamma_H = f^{-1}(\Gamma)$. We get a map $\Gamma_H\backslash Y \to \Gamma\backslash X$, and we denote the image by $S(H,\Gamma)$. Any compactly supported cohomology class $\xi \in H_c^{\dim Y}(\Gamma\backslash X;\mathbb{C})$ can be pulled back via f to $\Gamma_H\backslash Y$ and integrated to obtain a complex number. Hence $S(H,\Gamma)$ defines a linear form on $H_c^{\dim Y}(\Gamma\backslash X;\mathbb{C})$. By Poincaré duality, this linear form determines a class $[S(H,\Gamma)] \in H^{\dim X-\dim Y}(\Gamma\backslash X;\mathbb{C})$, called a *generalized modular symbol*. Such classes have been considered by many authors, for example [**AB90, SV03, Har05, AGR93**].

As an example, we can take $\mathbf{G}$ to be the split form of SL_2, and we can take $f\colon \mathbf{H} \to \mathbf{G}$ to be the inclusion of connected component of the diagonal subgroup. Hence $H \simeq \mathbb{R}_{>0}$. In this case K_H is trivial. The image of Y in X is the ideal geodesic from 0 to ∞. One way to vary f is by taking an $\mathrm{SL}_2(\mathbb{Q})$-translate of this geodesic, which gives a geodesic between two cusps. Hence we can obtain the support of any modular symbol this way. This example generalizes to SL_n to yield the modular symbols in Section A.4. Here $H \simeq (\mathbb{R} > 0)^{n-1}$. Note that $\dim Y = n-1$, so the cohomology classes we have constructed live in the top degree $H^{\nu}(\Gamma\backslash X;\mathbb{C})$.

Another family of examples is provided by taking $\mathbf{H}$ to be a Levi factor of a parabolic subgroup; these are the modular symbols studied in [**AB90**].

There are many natural questions to study for such objects. Here are two:

- Under what conditions on $\mathbf{G}, \mathbf{H}, \Gamma$ is $[S(H,\Gamma)]$ nonzero? This question is connected to relations between periods of automorphic forms and functoriality lifting. There are a variety of partial results known; see for example [**SV03, AGR93**].
- We know the usual modular symbols span the top-degree cohomology for any arithmetic group Γ. Fix a class of generalized modular symbols by fixing the pair $\mathbf{G}, \mathbf{H}$ and fixing some class of maps f. How much of the cohomology can one span for a general arithmetic group $\Gamma \subset \mathbf{G}(\mathbb{Q})$?

  A simple example is given by the Ash–Borel construction for $\mathbf{G} = \mathrm{SL}_3$ and $\mathbf{H}$ a Levi factor of a rational parabolic subgroup $\mathbf{P}$ of type $(2,1)$. In this case $H \simeq \mathrm{SL}_2(\mathbb{R}) \times \mathbb{R}_{>0}$ and sits inside G via

$$g \begin{pmatrix} \alpha^{-1}M & 0 \\ 0 & \alpha \end{pmatrix} g^{-1}, \quad M \in \mathrm{SL}_2(\mathbb{R}), \quad \alpha \in \mathbb{R}_{>0}, \quad g \in \mathrm{SL}_3(\mathbb{Q}).$$

  For $\Gamma \subset \mathrm{SL}_3(\mathbb{Z})$ these symbols define a subspace

$$S_{(2,1)} \subset H^2(\Gamma\backslash X; \mathbb{C}).$$

  Are there Γ for which $S_{(2,1)}$ equals the full cohomology space? For general Γ how much is captured? Is there a nice combinatorial way to write down the relations among these classes? Can one cook up a generalization of Theorem A.22 for these classes and use it to compute Hecke eigenvalues?

Bibliography

[AB90] A. Ash and A. Borel, *Generalized modular symbols*, Cohomology of arithmetic groups and automorphic forms (Luminy-Marseille, 1989), Springer, Berlin, 1990, pp. 57–75.

[ADT04] Nadia Ben Atti and Gema M. Díaz-Toca, `http://hlombardi.free.fr/publis/ABMAvar.html` (2004).

[AG00] Avner Ash and Robert Gross, *Generalized non-abelian reciprocity laws: a context for Wiles' proof*, Bull. London Math. Soc. **32** (2000), no. 4, 385–397. MR 1760802 (2001h:11142)

[Aga00] A. Agashe, *The Birch and Swinnerton-Dyer formula for modular abelian varieties of analytic rank* 0, Ph.D. thesis, University of California, Berkeley (2000).

[AGG84] Avner Ash, Daniel Grayson, and Philip Green, *Computations of cuspidal cohomology of congruence subgroups of* $\mathrm{SL}(3, \mathbf{Z})$, J. Number Theory **19** (1984), no. 3, 412–436. MR 769792 (86g:11032)

[AGM] Avner Ash, Paul E. Gunnells, and Mark McConnell, *Cohomology of congruence subgroups of* $\mathrm{SL}_4(\mathbb{Z})$ *II*, in preparation.

[AGM02] ———, *Cohomology of congruence subgroups of* $\mathrm{SL}_4(\mathbb{Z})$, J. Number Theory **94** (2002), no. 1, 181–212. MR 1904968 (2003f:11072)

[AGR93] Avner Ash, David Ginzburg, and Steven Rallis, *Vanishing periods of cusp forms over modular symbols*, Math. Ann. **296** (1993), no. 4, 709–723. MR 1233493 (94f:11044)

[Ahl78] Lars V. Ahlfors, *Complex analysis*, third ed., McGraw-Hill Book Co., New York, 1978, An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics. MR 510197 (80c:30001)

[AL70] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[AO01] Scott Ahlgren and Ken Ono, *Addition and counting: the arithmetic of partitions*, Notices Amer. Math. Soc. **48** (2001), no. 9, 978–984. MR 1854533 (2002e:11136)

[AR79] Avner Ash and Lee Rudolph, *The modular symbol and continued fractions in higher dimensions*, Invent. Math. **55** (1979), no. 3, 241–250. MR 553998 (82g:12011)

[Art79] James Arthur, *Eisenstein series and the trace formula*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 253–274. MR 546601 (81b:10020)

[Ash77] Avner Ash, *Deformation retracts with lowest possible dimension of arithmetic quotients of self-adjoint homogeneous cones*, Math. Ann. **225** (1977), no. 1, 69–76. MR 0427490 (55 #522)

[Ash80] ______, *Cohomology of congruence subgroups* $\mathrm{SL}(n, \mathbb{Z})$, Math. Ann. **249** (1980), no. 1, 55–73. MR 82f:22010

[Ash84] ______, *Small-dimensional classifying spaces for arithmetic subgroups of general linear groups*, Duke Math. J. **51** (1984), no. 2, 459–468. MR 747876 (85k:22027)

[Ash86] ______, *A note on minimal modular symbols*, Proc. Amer. Math. Soc. **96** (1986), no. 3, 394–396. MR 822426 (87e:22024)

[Ash94] ______, *Unstable cohomology of* $\mathrm{SL}(n, \mathcal{O})$, J. Algebra **167** (1994), no. 2, 330–342. MR 1283290 (95g:20050)

[Bar57] E. S. Barnes, *The perfect and extreme senary forms*, Canad. J. Math. **9** (1957), 235–242. MR 0086834 (19,251e)

[Bar94] A. Barvinok, *A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed*, Math. Oper. Res. **19** (1994), no. 4, 769–779.

[Bas96] Jacques Basmaji, *Ein Algorithmus zur Berechnung von Hecke-Operatoren und Anwendungen auf modulare Kurven,* `http://modular.math.washington.edu/scans/papers/basmaji/`, 1996.

[BC06] S. S. Bullock and C. Connell, *Equivariant retracts of geometrically finite discrete groups acting on negatively pinched Hadamard manifolds*, in preparation, 2006.

[BCDT01] C. Breuil, B. Conrad, Fred Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[BCS92] J. P. Buhler, R. E. Crandall, and R. W. Sompolski, *Irregular primes to one million*, Math. Comp. **59** (1992), no. 200, 717–722. MR 1134717 (93a:11106)

[BHKS06] K. Belebas, M. Van Hoeij, J. Klüners, and A. Steel, *Factoring polynomials over global fields*, preprint at `http://www.math.fsu.edu/~hoeij/papers.html` (2006).

[BI97] R. Baeza and M. I. Icaza, *On Humbert-Minkowski's constant for a number field*, Proc. Amer. Math. Soc. **125** (1997), no. 11, 3195–3202. MR 1403112 (97m:11092)

[Bir71] B. J. Birch, *Elliptic curves over* **Q***: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, pp. 396–400.

[BK90] S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, The Grothendieck Festschrift, Vol. I, Birkhäuser Boston, Boston, MA, 1990, pp. 333–400.

[BMS06] Yann Bugeaud, Maurice Mignotte, and Samir Siksek, *Classical and modular approaches to exponential Diophantine equations. II. The Lebesgue-Nagell equation*, Compos. Math. **142** (2006), no. 1, 31–62. MR 2196761

[Bro94] Kenneth S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics, vol. 87, Springer-Verlag, New York, 1994, corrected reprint of the 1982 original. MR 1324339 (96a:20072)

[BS73] A. Borel and J.-P. Serre, *Corners and arithmetic groups*, Comment. Math. Helv. **48** (1973), 436–491, avec un appendice: Arrondissement des variétés à coins, par A. Douady et L. Hérault. MR 0387495 (52 #8337)

[BS02] K. Buzzard and W. A. Stein, *A mod five approach to modularity of icosahedral Galois representations*, Pacific J. Math. **203** (2002), no. 2, 265–282. MR 2003c:11052

[BT82] Raoul Bott and Loring W. Tu, *Differential forms in algebraic topology*, Graduate Texts in Mathematics, vol. 82, Springer-Verlag, New York, 1982. MR 658304 (83i:57016)

[Bul00] S. S. Bullock, *Well-rounded retracts of rank one symmetric spaces*, preprint, 2000.

[Bum84] Daniel Bump, *Automorphic forms on* $\mathrm{GL}(3, \mathbf{R})$, Lecture Notes in Mathematics, vol. 1083, Springer-Verlag, Berlin, 1984. MR 765698 (86g:11028)

[Bum97] ______, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics, vol. 55, Cambridge University Press, Cambridge, 1997. MR 1431508 (97k:11080)

[Buz96] Kevin Buzzard, *On the eigenvalues of the Hecke operator* T_2, J. Number Theory **57** (1996), no. 1, 130–132. MR 96m:11033

[BW00] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, second ed., Mathematical Surveys and Monographs, vol. 67, American Mathematical Society, Providence, RI, 2000. MR 1721403 (2000j:22015)

[Byg99] J. Bygott, *Modular forms and modular symbols over imaginary quadratic fields*, Ph.D. thesis, Exeter University, 1999.

[Car59a] L. Carlitz, *Arithmetic properties of generalized Bernoulli numbers*, J. Reine Angew. Math. **202** (1959), 174–182. MR 0109132 (22 #20)

[Car59b] ______, *Some arithmetic properties of generalized Bernoulli numbers*, Bull. Amer. Math. Soc. **65** (1959), 68–69. MR 0104630 (21 #3383)

[CDT99] Brian Conrad, Fred Diamond, and Richard Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), no. 2, 521–567. MR 1639612 (99i:11037)

[CF67] George E. Cooke and Ross L. Finney, *Homology of cell complexes*, Based on lectures by Norman E. Steenrod, Princeton University Press, Princeton, N.J., 1967. MR 0219059 (36 #2142)

[Che05] Imin Chen, *A Diophantine equation associated to* $X_0(5)$, LMS J. Comput. Math. **8** (2005), 116–121 (electronic). MR 2153792 (2006b:11052)

[CL04] J. Cremona and M. P. Lingham, *Finding all elliptic curves with good reduction outside a given set of primes*, in progress (2004).

[CO77] H. Cohen and J. Oesterlé, *Dimensions des espaces de formes modulaires*, 69–78. Lecture Notes in Math., Vol. 627. MR 57 #12396

[Coh93] H. Cohen, *A course in computational algebraic number theory*, Springer-Verlag, Berlin, 1993. MR 94i:11105

[Cou01] Renaud Coulangeon, *Voronoï theory over algebraic number fields*, Réseaux euclidiens, designs sphériques et formes modulaires, Monogr. Enseign. Math., vol. 37, Enseignement Math., Geneva, 2001, pp. 147–162. MR 1878749 (2002m:11064)

[Cre] J. E. Cremona, personal communication.

[Cre84] ______, *Hyperbolic tessellations, modular symbols, and elliptic curves over complex quadratic fields*, Compositio Math. **51** (1984), no. 3, 275–324.

[Cre92] ______, *Modular symbols for $\Gamma_1(N)$ and elliptic curves with everywhere good reduction*, Math. Proc. Cambridge Philos. Soc. **111** (1992), no. 2, 199–218.

[Cre97a] ______, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, `http://www.maths.nott.ac.uk/personal/jec/book/`.

[Cre97b] ______, *Computing periods of cusp forms and modular elliptic curves*, Experiment. Math. **6** (1997), no. 2, 97–107.

[Cre06] ______, Proceedings of the 7th International Symposium (ANTS-VII) (2006).

[CS88] J. H. Conway and N. J. A. Sloane, *Low-dimensional lattices. III. Perfect forms*, Proc. Roy. Soc. London Ser. A **418** (1988), no. 1854, 43–80. MR 953277 (90a:11073)

[CW94] J. E. Cremona and E. Whitley, *Periods of cusp forms and elliptic curves over imaginary quadratic fields*, Math. Comp. **62** (1994), no. 205, 407–429.

[CWZ01] Janos A. Csirik, Joseph L. Wetherell, and Michael E. Zieve, *On the genera of $X_0(N)$*, `http://www.csirik.net/papers.html` (2001).

[Dar97] H. Darmon, *Faltings plus epsilon, Wiles plus epsilon, and the generalized Fermat equation*, C. R. Math. Rep. Acad. Sci. Canada **19** (1997), no. 1, 3–14. MR 1479291 (98h:11034a)

[Dem04] L. Dembélé, *Quaternionic Manin symbols, Brandt matrices and Hilbert modular forms*, preprint, 2004.

[Dem05] L. Dembélé, *Explicit computations of Hilbert modular forms on $\mathbb{Q}(\sqrt{5})$*, Experiment. Math. **14** (2005), no. 4, 457–466. MR 2193808

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[Dia96] F. Diamond, *On deformation rings and Hecke rings*, Ann. of Math. (2) **144** (1996), no. 1, 137–166. MR 1405946 (97d:11172)

[Dix82] John D. Dixon, *Exact solution of linear equations using p-adic expansions*, Numer. Math. **40** (1982), no. 1, 137–141. MR 681819 (83m:65025)

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149.

[DP04] H. Darmon and R. Pollack, *The efficient calculation of Stark-Heegner points via overconvergent modular symbols.*

[DS05] Fred Diamond and Jerry Shurman, *A first course in modular forms*, Graduate Texts in Mathematics, vol. 228, Springer-Verlag, New York, 2005.

[DVS05] M. Dutour, F. Vallentin, and A. Schürmann, *Classification of perfect forms in dimension* 8, talk at Oberwolfach meeting *Sphere packings: Exceptional structures and relations to other fields*, November 2005.

[Ebe02] Wolfgang Ebeling, *Lattices and codes*, revised ed., Advanced Lectures in Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 2002, a course partially based on lectures by F. Hirzebruch.

[ECdJ$^+$06] Bas Edixhoven, Jean-Marc Couveignes, Robin de Jong, Franz Merkl, and Johan Bosman, *On the computation of coefficients of modular form*, `http://www.arxiv.org/abs/math.NT/0605244` (2006).

[EGM98] J. Elstrodt, F. Grunewald, and J. Mennicke, *Groups acting on hyperbolic space*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998, Harmonic analysis and number theory. MR 1483315 (98g:11058)

[Eil47] Samuel Eilenberg, *Homology of spaces with operators. I*, Trans. Amer. Math. Soc. **61** (1947), 378–417; errata, 62, 548 (1947). MR 0021313 (9,52b)

[Elk98] Noam D. Elkies, *Elliptic and modular curves over finite fields and related computational issues*, Computational perspectives on number theory (Chicago, IL, 1995), AMS/IP Stud. Adv. Math., vol. 7, Amer. Math. Soc., Providence, RI, 1998, pp. 21–76. MR 1486831 (99a:11078)

[EVGS02] Philippe Elbaz-Vincent, Herbert Gangl, and Christophe Soulé, *Quelques calculs de la cohomologie de* $\mathrm{GL}_N(\mathbb{Z})$ *et de la* K*-théorie de* $\mathbb{Z}$, C. R. Math. Acad. Sci. Paris **335** (2002), no. 4, 321–324. MR 1931508 (2003h:19002)

[FH91] William Fulton and Joe Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer-Verlag, New York, 1991, A first course, Readings in Mathematics. MR 1153249 (93a:20069)

[FJ02] D.W. Farmer and K. James, *The irreducibility of some level 1 Hecke polynomials*, Math. Comp. **71** (2002), no. 239, 1263–1270 (electronic). MR 2003e:11046

[FL] D.W. Farmer and Stefan Lemurell, *Maass forms and their L-functions*, AIM 2005-15, arXiv:math.NT/0506102.

[FM99] G. Frey and M. Müller, *Arithmetic of modular curves and applications*, Algorithmic algebra and number theory (Heidelberg, 1997), Springer, Berlin, 1999, pp. 11–48.

[Fra98] J. Franke, *Harmonic analysis in weighted* L_2*-spaces*, Ann. Sci. École Norm. Sup. (4) **31** (1998), no. 2, 181–279.

[FT93] A. Fröhlich and M. J. Taylor, *Algebraic number theory*, Cambridge University Press, Cambridge, 1993.

[FvdG] C. Faber and G. van der Geer, *Sur la cohomologie des Systémes Locaux sur les Espaces des Modules des Courbes de Genus 2 and des Surfaces Abéliennes*, arXiv:math.AG/0305094.

[Gel75] Stephen S. Gelbart, *Automorphic forms on adèle groups*, Princeton University Press, Princeton, N.J., 1975, Annals of Mathematics Studies, No. 83. MR 0379375 (52 #280)

[GH81] M.J. Greenberg and J.R. Harper, *Algebraic topology*, Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., 1981, A first course. MR 83b:55001

[GLQ04] Josep González, Joan-Carles Lario, and Jordi Quer, *Arithmetic of $\mathbb{Q}$-curves*, Modular curves and abelian varieties, Progr. Math., vol. 224, Birkhäuser, Basel, 2004, pp. 125–139. MR 2058647 (2005c:11068)

[GM03] P. E. Gunnells and M. McConnell, *Hecke operators and $\mathbb{Q}$-groups associated to self-adjoint homogeneous cones*, J. Number Theory **100** (2003), no. 1, 46–71.

[Gol05] Dorian Goldfeld, *Automorphic forms and L-functions on the general linear group*, to appear, 2005.

[Gon97] A. B. Goncharov, *The double logarithm and Manin's complex for modular curves*, Math. Res. Lett. **4** (1997), no. 5, 617–636.

[Gon98] ———, *Multiple polylogarithms, cyclotomy and modular complexes*, Math. Res. Lett. **5** (1998), no. 4, 497–516.

[Gor93] D. Gordon, *Discrete logarithms in* $\mathrm{GF}(p)$ *using the number field sieve*, SIAM J. Discrete Math. **6** (1993), no. 1, 124–138. MR 94d:11104

[Gor04] ———, *Discrete logarithm problem,* `http://www.win.tue.nl/~henkvt/content.html`.

[GP05] Benedict H. Gross and David Pollack, *On the Euler characteristic of the discrete spectrum*, J. Number Theory **110** (2005), no. 1, 136–163. MR 2114678 (2005k:11100)

[Gre83] Ralph Greenberg, *On the Birch and Swinnerton-Dyer conjecture*, Invent. Math. **72** (1983), no. 2, 241–265. MR 700770 (85c:11052)

[Gri05] G. Grigorov, *Kato's Euler System and the Main Conjecture*, Harvard Ph.D. Thesis (2005).

[Gro98] Benedict H. Gross, *On the Satake isomorphism*, Galois representations in arithmetic algebraic geometry (Durham, 1996), London Math. Soc. Lecture Note Ser., vol. 254, Cambridge Univ. Press, Cambridge, 1998, pp. 223–237. MR 1696481 (2000e:22008)

[GS81] F. Grunewald and J. Schwermer, *A nonvanishing theorem for the cuspidal cohomology of SL_2 over imaginary quadratic integers*, Math. Ann. **258** (1981), 183–200.

[GS02] Mark Giesbrecht and Arne Storjohann, *Computing rational forms of integer matrices*, J. Symbolic Comput. **34** (2002), no. 3, 157–172. MR 1935075 (2003j:15016)

[Gun99] P. E. Gunnells, *Modular symbols for $\mathbb{Q}$-rank one groups and Voronoĭ reduction*, J. Number Theory **75** (1999), no. 2, 198–219.

[Gun00a] ———, *Computing Hecke eigenvalues below the cohomological dimension*, Experiment. Math. **9** (2000), no. 3, 351–367. MR 1 795 307

[Gun00b] ———, *Symplectic modular symbols*, Duke Math. J. **102** (2000), no. 2, 329–350.

[Hara] G. Harder, *Congruences between modular forms of genus 1 and of genus 2*, Arbeitstagung.

[Harb] ———, *Kohomologie arithmetischer Gruppen*, lecture notes, Universität Bonn, 1987–1988.

[Har87] ———, *Eisenstein cohomology of arithmetic groups. The case* GL_2, Invent. Math. **89** (1987), no. 1, 37–118. MR 892187 (89b:22018)

[Har91] ______, *Eisenstein cohomology of arithmetic groups and its applications to number theory*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 779–790. MR 1159264 (93b:11057)

[Har05] ______, *Modular symbols and special values of automorphic L-functions*, preprint, 2005.

[HC68] Harish-Chandra, *Automorphic forms on semisimple Lie groups*, Notes by J. G. M. Mars. Lecture Notes in Mathematics, No. 62, Springer-Verlag, Berlin, 1968. MR 0232893 (38 #1216)

[Hel01] Sigurdur Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Graduate Studies in Mathematics, vol. 34, American Mathematical Society, Providence, RI, 2001, corrected reprint of the 1978 original. MR 1834454 (2002b:53081)

[Hij74] H. Hijikata, *Explicit formula of the traces of Hecke operators for* $\Gamma_0(N)$, J. Math. Soc. Japan **26** (1974), no. 1, 56–82.

[Hsu96] Tim Hsu, *Identifying congruence subgroups of the modular group*, Proc. Amer. Math. Soc. **124** (1996), no. 5, 1351–1359. MR 1343700 (96k:20100)

[HT01] Michael Harris and Richard Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, Princeton, NJ, 2001, with an appendix by Vladimir G. Berkovich. MR 1876802 (2002m:11050)

[Hum80] James E. Humphreys, *Arithmetic groups*, Lecture Notes in Mathematics, vol. 789, Springer, Berlin, 1980. MR 584623 (82j:10041)

[Ica97] M. I. Icaza, *Hermite constant and extreme forms for algebraic number fields*, J. London Math. Soc. (2) **55** (1997), no. 1, 11–22. MR 1423282 (97j:11034)

[Jaq91] David-Olivier Jaquet, *Classification des réseaux dans* $\mathbf{R}^7$ *(via la notion de formes parfaites)*, Astérisque (1991), no. 198-200, 7–8, 177–185 (1992), Journées Arithmétiques, 1989 (Luminy, 1989). MR 1144322 (93g:11071)

[JBS03] A. Jorza, J. Balakrishna, and W. Stein, *The Smallest Conductor for an Elliptic Curve of Rank Four is Composite,* `http://modular.math.washington.edu/rank4/`.

[JC93] David-Olivier Jaquet-Chiffelle, *Énumération complète des classes de formes parfaites en dimension* 7, Ann. Inst. Fourier (Grenoble) **43** (1993), no. 1, 21–55. MR 1209694 (94d:11048)

[Kan00] Masanobu Kaneko, *The Akiyama-Tanigawa algorithm for Bernoulli numbers*, J. Integer Seq. **3** (2000), no. 2, Article 00.2.9, 6 pp. (electronic). MR 1800883 (2001k:11026)

[Kel06] Bernd C. Kellner, *Bernoulli numbers*, `http://www.bernoulli.org` (2006).

[Kna92] A. W. Knapp, *Elliptic curves*, Princeton University Press, Princeton, NJ, 1992.

[Knu] Donald E. Knuth, *The art of computer programming. Vol. 2*, third ed., Addison-Wesley Publishing Co., Reading, Mass., Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[Kob84] N. Koblitz, *Introduction to elliptic curves and modular forms*, Graduate Texts in Mathematics, vol. 97, Springer-Verlag, New York, 1984. MR 86c:11040

[Kri90] Aloys Krieg, *Hecke algebras*, Mem. Amer. Math. Soc. **87** (1990), no. 435, x+158. MR 1027069 (90m:16024)

[Laf02] Laurent Lafforgue, *Chtoucas de Drinfeld et correspondance de Langlands*, Invent. Math. **147** (2002), no. 1, 1–241. MR 1875184 (2002m:11039)

[Lan66] R. P. Langlands, *Eisenstein series*, Algebraic Groups and Discontinuous Subgroups (Proc. Sympos. Pure Math., Boulder, Colo., 1965), Amer. Math. Soc., Providence, R.I., 1966, pp. 235–252. MR 0249539 (40 #2784)

[Lan76] Robert P. Langlands, *On the functional equations satisfied by Eisenstein series*, Springer-Verlag, Berlin, 1976, Lecture Notes in Mathematics, Vol. 544. MR 0579181 (58 #28319)

[Lan95] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, with appendixes by D. Zagier and W. Feit, corrected reprint of the 1976 original.

[Lem01] Dominic Lemelin, *Mazur-tate type conjectures for elliptic curves defined over quadratic imaginary fields.*

[Leo58] Heinrich-Wolfgang Leopoldt, *Eine Verallgemeinerung der Bernoullischen Zahlen*, Abh. Math. Sem. Univ. Hamburg **22** (1958), 131–140. MR 0092812 (19,1161e)

[Li75] W-C. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.

[Lin05] M. Lingham, *Modular forms and elliptic curves over imaginary quadratic fields*, Ph.D. thesis, Nottingham, 2005.

[LLL82] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), no. 4, 515–534. MR 682664 (84a:12002)

[LS76] Ronnie Lee and R. H. Szczarba, *On the homology and cohomology of congruence subgroups*, Invent. Math. **33** (1976), no. 1, 15–53. MR 0422498 (54 #10485)

[LS90] J.-P. Labesse and J. Schwermer (eds.), *Cohomology of arithmetic groups and automorphic forms*, Lecture Notes in Mathematics, vol. 1447, Berlin, Springer-Verlag, 1990. MR 1082959 (91h:11033)

[LS02] Joan-C. Lario and René Schoof, *Some computations with Hecke rings and deformation rings*, Experiment. Math. **11** (2002), no. 2, 303–311, with an appendix by Amod Agashe and William Stein. MR 1959271 (2004b:11072)

[LS04] Jian-Shu Li and Joachim Schwermer, *On the Eisenstein cohomology of arithmetic groups*, Duke Math. J. **123** (2004), no. 1, 141–169. MR 2060025 (2005h:11108)

[Lub94] A. Lubotzky, *Discrete groups, expanding graphs and invariant measures*, Progress in Mathematics, vol. 125, Birkhäuser Verlag, Basel, 1994, with an appendix by Jonathan D. Rogawski.

[Man72] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66. MR 47 #3396

[Mar01] François Martin, *Périodes de formes modulaires de poids 1.*

[Mar03] Jacques Martinet, *Perfect lattices in Euclidean spaces*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 327, Springer-Verlag, Berlin, 2003. MR 1957723 (2003m:11099)

[Mar05] Greg Martin, *Dimensions of the spaces of cusp forms and newforms on $\Gamma_0(N)$ and $\Gamma_1(N)$*, J. Number Theory **112** (2005), no. 2, 298–331. MR 2141534 (2005m:11069)

[Maz73] B. Mazur, *Courbes elliptiques et symboles modulaires*, Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 414, Springer, Berlin, 1973, pp. 277–294. Lecture Notes in Math., Vol. 317. MR 55 #2930

[McC91] M. McConnell, *Classical projective geometry and arithmetic groups*, Math. Ann. **290** (1991), no. 3, 441–462. MR 92k:22020

[Men79] Eduardo R. Mendoza, *Cohomology of* PGL_2 *over imaginary quadratic integers*, Bonner Mathematische Schriften [Bonn Mathematical Publications], 128, Universität Bonn Mathematisches Institut, Bonn, 1979, Dissertation, Rheinische Friedrich-Wilhelms-Universität, Bonn, 1979. MR 611515 (82g:22012)

[Mer94] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations, Springer, 1994, pp. 59–94.

[Mer99] ______, *Arithmetic of elliptic curves and Diophantine equations*, J. Théor. Nombres Bordeaux **11** (1999), no. 1, 173–200, Les XXèmes Journées Arithmétiques (Limoges, 1997). MR 1730439 (2000j:11084)

[Mes86] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242.

[Miy89] T. Miyake, *Modular forms*, Springer-Verlag, Berlin, 1989, translated from the Japanese by Yoshitaka Maeda.

[MM89] R. MacPherson and M. McConnell, *Classical projective geometry and modular varieties*, Algebraic analysis, geometry, and number theory (Baltimore, MD, 1988), Johns Hopkins Univ. Press, Baltimore, MD, 1989, pp. 237–290. MR 98k:14076

[MM93] ______, *Explicit reduction theory for Siegel modular threefolds*, Invent. Math. **111** (1993), no. 3, 575–625. MR 94a:32052

[MTT86] B. Mazur, J. Tate, and J. Teitelbaum, *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer*, Invent. Math. **84** (1986), no. 1, 1–48.

[MW94] Colette Mœglin and Jean-Loup Waldspurger, *Décomposition spectrale et séries d'Eisenstein*, Progress in Mathematics, vol. 113, Birkhäuser Verlag, Basel, 1994, Une paraphrase de l'Écriture [A paraphrase of Scripture]. MR 1261867 (95d:11067)

[Nec94] V. I. Nechaev, *On the complexity of a deterministic algorithm for a discrete logarithm*, Mat. Zametki **55** (1994), no. 2, 91–101, 189. MR 96a:11145

[Ong86] Heidrun E. Ong, *Perfect quadratic forms over real-quadratic number fields*, Geom. Dedicata **20** (1986), no. 1, 51–77. MR 823160 (87f:11023)

[PR94] Vladimir Platonov and Andrei Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., Boston, MA, 1994, translated from the 1991 Russian original by Rachel Rowen. MR 1278263 (95b:11039)

[Que06] J. Quer, *Dimensions of spaces of modular forms for $\Gamma_H(N)$*, Preprint.

[Rib92] K. A. Ribet, *Abelian varieties over* $\mathbf{Q}$ *and modular forms*, Algebra and topology 1992 (Taejŏn), Korea Adv. Inst. Sci. Tech., Taejŏn, 1992, pp. 53–79. MR 94g:11042

[Ros86] M. Rosen, *Abelian varieties over* **C**, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 79–101.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 143–232. MR 2002h:11047

[Sap97] Leslie Saper, *Tilings and finite energy retractions of locally symmetric spaces*, Comment. Math. Helv. **72** (1997), no. 2, 167–202. MR 1470087 (99a:22019)

[Sar03] Peter Sarnak, *Spectra of hyperbolic surfaces*, Bull. Amer. Math. Soc. (N.S.) **40** (2003), no. 4, 441–478 (electronic). MR 1997348 (2004f:11107)

[SC03] Samir Siksek and John E. Cremona, *On the Diophantine equation* $x^2+7=y^m$, Acta Arith. **109** (2003), no. 2, 143–149. MR 1980642 (2004c:11109)

[Sch86] Joachim Schwermer, *Holomorphy of Eisenstein series at special points and cohomology of arithmetic subgroups of* $\mathrm{SL}_n(\mathbf{Q})$, J. Reine Angew. Math. **364** (1986), 193–220. MR 817646 (87h:11048)

[Sch90] A. J. Scholl, *Motives for modular forms*, Invent. Math. **100** (1990), no. 2, 419–430.

[Sch95] R. Schoof, *Counting points on elliptic curves over finite fields*, J. Théor. Nombres Bordeaux **7** (1995), no. 1, 219–254, Les Dix-huitièmes Journées Arithmétiques (Bordeaux, 1993). MR 1413578 (97i:11070)

[Ser73] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[Ser87] ———, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. **54** (1987), no. 1, 179–230.

[Shi59] G. Shimura, *Sur les intégrales attachées aux formes automorphes*, J. Math. Soc. Japan **11** (1959), 291–311.

[Shi94] ———, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, reprint of the 1971 original, Kan Memorial Lectures, 1.

[Sho80a] V. V. Shokurov, *Shimura integrals of cusp forms*, Izv. Akad. Nauk SSSR Ser. Mat. **44** (1980), no. 3, 670–718, 720. MR 582162 (82b:10029)

[Sho80b] ———, *A study of the homology of Kuga varieties*, Izv. Akad. Nauk SSSR Ser. Mat. **44** (1980), no. 2, 443–464, 480. MR 571104 (82f:14023)

[Sho97] Victor Shoup, *Lower bounds for discrete logarithms and related problems*, Advances in cryptology—EUROCRYPT '97 (Konstanz), Lecture Notes in Comput. Sci., vol. 1233, Springer, Berlin, 1997, pp. 256–266. MR 98j:94023

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, corrected reprint of the 1986 original.

[Sou75] Christophe Soulé, *Cohomologie de* $SL_3(Z)$, C. R. Acad. Sci. Paris Sér. A-B **280** (1975), no. 5, Ai, A251–A254. MR 0396849 (53 #709)

[Sta79] R. E. Staffeldt, *Reduction theory and* K_3 *of the Gaussian integers*, Duke Math. J. **46** (1979), no. 4, 773–798. MR 552526 (80m:22014)

[Ste] Allan Steel, *Advanced matrix algorithms*, Seminar Talk at Harvard University.

[Ste97] ———, *A new algorithm for the computation of canonical forms of matrices over fields*, J. Symbolic Comput. **24** (1997), no. 3-4, 409–432, Computational algebra and number theory (London, 1993). MR 1484489 (98m:65070)

[Ste99a] Norman Steenrod, *The topology of fibre bundles*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1999, reprint of the 1957 edition, Princeton Paperbacks. MR 1688579 (2000a:55001)

[Ste99b] W. A. Stein, `HECKE`*: The Modular Symbols Calculator*, software (available online) (1999).

[Ste00] ———, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley (2000).

[Ste06] ———, `SAGE`*: Software for Algebra and Geometry Experimentation,* `http://sage.scipy.org/sage`.

[Str69] Volker Strassen, *Gaussian elimination is not optimal*, Numerische Mathematik **13** (1969), 354–356.

[Stu87] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin, 1987, pp. 275–280.

[SV01] W. A. Stein and H. A. Verrill, *Cuspidal modular symbols are transportable*, LMS J. Comput. Math. **4** (2001), 170–181 (electronic). MR 1 901 355

[SV03] B. Speh and T. N. Venkataramana, *Construction of some generalised modular symbols*, preprint, 2003.

[SW02] William A. Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, pp. 267–275. MR 2041090 (2005h:11113)

[SW05] Jude Socrates and David Whitehouse, *Unramified Hilbert modular forms, with examples relating to elliptic curves*, Pacific J. Math. **219** (2005), no. 2, 333–364. MR 2175121

[Tat75] J. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1975, pp. 33–52. Lecture Notes in Math., Vol. 476. MR 52 #13850

[Tho89] J. G. Thompson, *Hecke operators and noncongruence subgroups*, Group theory (Singapore, 1987), de Gruyter, Berlin, 1989, including a letter from J.-P. Serre, pp. 215–224. MR 981844 (90a:20105)

[Tot05] A. Toth, *On the Steinberg module of Chevalley groups*, Manuscripta Math. **116** (2005), no. 3, 277–295.

[TW95] R. Taylor and A. J. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[vdG] Gerard van der Geer, *Siegel Modular Forms*, arXiv:math.AG/0605346.

[vGvdKTV97] Bert van Geemen, Wilberd van der Kallen, Jaap Top, and Alain Verberkmoes, *Hecke eigenforms in the cohomology of congruence subgroups of* $\mathrm{SL}(3, \mathbf{Z})$, Experiment. Math. **6** (1997), no. 2, 163–174. MR 1474576 (99a:11059)

[Vig77] Marie-France Vignéras, *Séries thêta des formes quadratiques indéfinies*, Séminaire Delange-Pisot-Poitou, 17e année (1975/76), Théorie des nombres: Fasc. 1, Exp. No. 20, Secrétariat Math., Paris, 1977, p. 3. MR 0480352 (58 #521)

[Vog85] K. Vogtmann, *Rational homology of Bianchi groups*, Math. Ann. **272** (1985), no. 3, 399–419.

[Vog97] David A. Vogan, Jr., *Cohomology and group representations*, Representation theory and automorphic forms (Edinburgh, 1996), Proc. Sympos. Pure Math., vol. 61, Amer. Math. Soc., Providence, RI, 1997, pp. 219–243. MR 1476500 (98k:22064)

[Vor08] G. Voronoĭ, *Nouvelles applications des paramétres continus à la théorie des formes quadratiques, I. Sur quelques propriétés des formes quadratiques positives parfaites*, J. Reine Angew. Math. **133** (1908), 97–178.

[VZ84] David A. Vogan, Jr. and Gregg J. Zuckerman, *Unitary representations with nonzero cohomology*, Compositio Math. **53** (1984), no. 1, 51–90. MR 762307 (86k:22040)

[Wan82] Kai Wang, *A proof of an identity of the Dirichlet L-function*, Bull. Inst. Math. Acad. Sinica **10** (1982), no. 3, 317–321. MR 679019 (84c:10040)

[Wan95] Xiang Dong Wang, 2-*dimensional simple factors of* $J_0(N)$, Manuscripta Math. **87** (1995), no. 2, 179–197. MR 1334940 (96h:11059)

[Wes] U. Weselman, personal communication.

[Whi90] E. Whitley, *Modular symbols and elliptic curves over imaginary quadratic number fields*, Ph.D. thesis, Exeter University, 1990.

[Wie05] Gabor Wiese, *Modular Forms of Weight One Over Finite Fields*, Ph.D. thesis (2005).

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551. MR 1333035 (96d:11071)

[Wil00] ——— , *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[Yas05a] D. Yasaki, *On the cohomology of* SU(2, 1) *over the Gaussian integers*, preprint, 2005.

[Yas05b] ——— , *On the existence of spines for* **Q**-*rank 1 groups*, preprint, 2005.

Index

Symbol Index

Algorithm Index

Definition Index

SAGE Index

General Index 148, 150

30 Explicit Heegner points: Kolyvagins conjecture and non-trivial elements in the Shafarevich-Tate group, with D. Jetchev and K. Lauter

Explicit Heegner points: Kolyvagin's conjecture and non-trivial elements in the Shafarevich-Tate group

Dimitar Jetchev, Kristin Lauter and William Stein

Abstract

Kolyvagin used Heegner points to associate a system of cohomology classes to an elliptic curve over $\mathbb{Q}$ and conjectured that the system contains a nontrivial class. His conjecture has profound implications on the structure of Selmer groups. We provide new computational and theoretical evidence for Kolyvagin's conjecture. More precisely, we apply results of Zhang and others to deduce that Kolyvagin classes are computable, then explicitly study Heegner points over ring class fields and Kolyvagin's conjecture for specific elliptic curves of rank two. We explain how Kolyvagin's conjecture implies that if the analytic rank of an elliptic curve is at least two then the $\mathbb{Z}_p$-corank of the corresponding Selmer group is at least two as well. We also use explicitly computed Heegner points to produce non-trivial classes in the Shafarevich-Tate group.

1. Introduction

Let $E_{/F}$ be an elliptic curve over a number field F. The analytic rank $r_{\mathrm{an}}(E/F)$ of E is the order of vanishing of the L-function $L(E_{/F}, s)$ at $s = 1$. The Mordell-Weil rank $r_{\mathrm{MW}}(E/F)$ is the rank of the Mordell-Weil group $E(F)$. The conjecture of Birch and Swinnerton-Dyer asserts that $r_{\mathrm{an}}(E/F) = r_{\mathrm{MW}}(E/F)$.

Kolyvagin constructed explicit cohomology classes from Heegner points over certain abelian extensions of quadratic imaginary fields and used these classes to bound the size of the Selmer groups for elliptic curves over $\mathbb{Q}$ of analytic rank at most one (see [Kol90], [Kol91b] and [Gro91]). His results, together with the Gross-Zagier formula (see [GZ86]), imply the following theorem:

Theorem 1.1 (Gross-Zagier, Kolyvagin). *Let $E_{/\mathbb{Q}}$ be an elliptic curve which satisfies $r_{\mathrm{an}}(E/\mathbb{Q}) \leqslant 1$. Then the Shafarevich-Tate group $Ш(E/\mathbb{Q})$ is finite and $r_{\mathrm{an}}(E/\mathbb{Q}) = r_{\mathrm{MW}}(E/\mathbb{Q})$.*

Unfortunately, very little is known about the Birch and Swinnerton-Dyer conjecture for elliptic curves $E_{/\mathbb{Q}}$ with $r_{\mathrm{an}}(E/\mathbb{Q}) \geqslant 2$. Still, it implies the following conjecture:

Conjecture 1. If $r_{\mathrm{an}}(E/\mathbb{Q}) \geqslant 2$ then $r_{\mathrm{MW}}(E/\mathbb{Q}) \geqslant 2$.

As far as we know, nothing has been proved towards the above assertion. A weaker conjecture can be formulated in the language of Selmer coranks. The Selmer corank $r_p(E/F)$ of $E_{/F}$ is the $\mathbb{Z}_p$-corank of the Selmer group $\mathrm{Sel}_{p^\infty}(E/F)$. Using Kummer theory, one shows that $r_p(E/\mathbb{Q}) \geqslant r_{\mathrm{MW}}(E/\mathbb{Q})$ with an equality occuring if and only if the p-primary part of the Shafarevich-Tate group $Ш(E/\mathbb{Q})$ is finite. Thus, one obtains the following weaker conjecture:

2000 Mathematics Subject Classification 11G05
Keywords: Elliptic curves, Heegner points, Selmer groups
The first author was supported by summer internships at Microsoft Research as well as by graduate fellowships from Microsoft Research and University of California at Berkeley. The third author was partially supported by NSF grant 0555776.

Conjecture 2. If $r_{\mathrm{an}}(E/\mathbb{Q}) \geqslant 2$ then $r_p(E/\mathbb{Q}) \geqslant 2$.

For elliptic curves E of arbitrary analytic rank, Kolyvagin was able to explain the exact structure of the Selmer group $\mathrm{Sel}_{p^\infty}(E/\mathbb{Q})$ in terms of Heegner points and the associated cohomology classes under a conjecture about the non-triviality of these classes (see [Kol91a, Conj.A]). Unfortunately, Kolyvagin's conjecture appears to be extremely difficult to prove. Until the present paper, there has been no example of an elliptic curve over $\mathbb{Q}$ of rank at least two for which the conjecture has been verified.

In this paper, we present a complete algorithm to compute Kolyvagin's cohomology classes by explicitly computing the corresponding Heegner points over ring class fields. We use this algorithm to verify Kolyvagin's conjecture for the first time for elliptic curves of analytic rank two. We also explain (see Corollary 3.3) how Kolyvagin's conjecture implies Conjecture 2. In addition, we use methods of Cornut (see [Cor02]) to provide theoretical evidence for Kolyvagin's conjecture. As a separate application of the explicit computation of Heegner points, we construct nontrivial cohomology classes in the Shafarevich-Tate group $\text{Ш}(E/K)$ of elliptic curves E over certain quadratic imaginary fields. One of the main contributions of this paper is that by establishing certain height bounds, we prove that there exists an algorithm which *provably* computes the correct Heegner points over ring class fields.

The paper is organized as follows. Section 2 introduces Heegner points over ring class fields and Kolyvagin cohomology classes. We explain the methods of computation and illustrate them with several examples. In Section 3 we state Kolyvagin's conjecture, discuss Kolyvagin's work on Selmer groups and establish Conjecture 2 as a corollary. Moreover, we present a proof of the theoretical evidence following closely Cornut's arguments. Section 3.6 contains the essential examples for which we manage to explicitly verify the conjecture. Finally, in Section 4 we apply our computational techniques to produce explicit non-trivial elements in the Shafarevich-Tate groups for specific elliptic curves. Finally, the appendix establishes certain bounds on the logarithmic heights of the Heegner points over ring class fields.

2. Heegner points over ring class fields

We discuss Heegner points over ring class fields in Section 2.1 and describe a method for computing them in Section 2.2. Height estimates for these points are given in the appendix. We illustrate the method with some examples in Section 2.3. The standard references are [Gro91], [Kol90] and [McC91].

2.1 Heegner points over ring class fields

Let E be an elliptic curve over $\mathbb{Q}$ of conductor N and let $K = \mathbb{Q}(\sqrt{-D})$ for some fundamental discriminant $D > 0$, $D \neq 3, 4$, such that all prime factors of N are split in K. We refer to such a discriminant as a *Heegner discriminant* for $E/\mathbb{Q}$. Let $\mathcal{O}_K$ be the ring of integers of K. It follows that $N\mathcal{O}_K = \mathcal{N}\bar{\mathcal{N}}$ for an ideal $\mathcal{N}$ of $\mathcal{O}_K$ with $\mathcal{O}_K/\mathcal{N} \simeq \mathbb{Z}/N\mathbb{Z}$.

By the modularity theorem (see [BCDT01]), there exists an optimal (having minimal degree) modular parameterization $\varphi : X_0(N) \to E$. Let $\mathcal{N}^{-1}$ be the fractional ideal of $\mathcal{O}_K$ for which $\mathcal{N}\mathcal{N}^{-1} = \mathcal{O}_K$. We view $\mathcal{O}_K$ and $\mathcal{N}$ as $\mathbb{Z}$-lattices of rank two in $\mathbb{C}$ and observe that $\mathbb{C}/\mathcal{O}_K \to \mathbb{C}/\mathcal{N}^{-1}$ is a cyclic isogeny of degree N between the elliptic curves $\mathbb{C}/\mathcal{O}_K$ and $\mathbb{C}/\mathcal{N}^{-1}$. This isogeny corresponds to a complex point $x_1 \in X_0(N)(\mathbb{C})$. According to the theory of complex multiplication [Sil94, Ch.II], the point x_1 is defined over the Hilbert class field H_K of K.

More generally, for an integer c, let $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$ be the order of conductor c in $\mathcal{O}_K$ and let $\mathcal{N}_c = \mathcal{N} \cap \mathcal{O}_c$, which is an invertible ideal of $\mathcal{O}_c$. Then $\mathcal{O}_c/\mathcal{N}_c \simeq \mathbb{Z}/N\mathbb{Z}$ and the map $\mathbb{C}/\mathcal{O}_c \to \mathbb{C}/\mathcal{N}_c^{-1}$

is a cyclic isogeny of degree N. Thus, it defines a point $x_c \in X_0(N)(\mathbb{C})$. By the theory of complex multiplication, this point is defined over the ring class field $K[c]$ of conductor c over K (that is, the unique abelian extension of K corresponding to the norm subgroup $\widehat{\mathcal{O}_c}^{\times} K^\times \subset \widehat{K}^\times$; e.g., if $c = 1$ then $K[1] = H_K$).

We use the parameterization $\varphi : X_0(N) \to E$ to obtain points

$$y_c = \varphi(x_c) \in E(K[c]).$$

Let $y_K = \mathrm{Tr}_{H_K/K}(y_1)$. We refer to y_K as the *Heegner point* for the discriminant D, even though it is only well defined up to sign and torsion (if $\mathcal{N}'$ is another ideal with $\mathcal{O}/\mathcal{N}' \simeq \mathbb{Z}/N\mathbb{Z}$ then the new Heegner point differs from y_K by at most a sign change and a rational torsion point).

2.2 Explicit computation of the points y_c

Significant work has been done on explicit calculations of Heegner points on elliptic curves (see [Coh07], [Del02], [Elk94], [Wat04]). Yet, all of these only compute the points y_1 and y_K. In [EJL06] explicit computations of the points y_c were considered in several examples and some difficulties were outlined. However, there has been no algorithm which provably computes the points y_c. One of the main contributions of this paper is the description of such an algorithm.

To compute the point $y_c = [\mathbb{C}/\mathcal{O}_c \to \mathbb{C}/\mathcal{N}_c^{-1}] \in E(K[c])$ we let $f \in S_2(\Gamma_0(N))$ be the newform corresponding to the elliptic curve E and Λ be the complex lattice (defined up to homothety), such that $E \cong \mathbb{C}/\Lambda$. Let $\mathfrak{h}^\times = \mathfrak{h} \cup \mathbb{P}^1(\mathbb{Q}) \cup \{i\infty\}$, where $\mathfrak{h} = \{z \in \mathbb{C} : \Im(z) > 0\}$ is the upper-half plane equipped with the action of $\Gamma_0(N)$ by linear fractional transformations. The modular parametrization $\varphi : X_0(N) \to E$ is then given by the function $\varphi : \mathfrak{h}^\times \to \mathbb{C}/\Lambda$

$$\varphi(\tau) = \int_\tau^{i\infty} f(z)dz = \sum_{n \geqslant 1} \frac{a_n}{n} e^{2\pi i n \tau}, \tag{1}$$

where $f = \sum_{n=1}^{\infty} a_n q^n$ is the Fourier expansion of the modular form f.

We first compute ideal class representatives $\mathfrak{a}_1, \mathfrak{a}_2, \dots, \mathfrak{a}_{h_c}$ for the Picard group $\mathrm{Pic}(\mathcal{O}_c) \cong \mathrm{Gal}(K[c]/K)$, where $h_c = \#\,\mathrm{Pic}(\mathcal{O}_c)$. Let $\sigma_i \in \mathrm{Gal}(K[c]/K)$ be the image of the ideal class of $\mathfrak{a}_i$ under the Artin map. We use the ideal $\mathfrak{a}_i$ to compute a complex number (a *quadratic surd*)$\tau_i \in \mathfrak{h}$ representing the CM point $\sigma_i(x_c)$ for each $i = 1, \dots, h_c$ (since $X_0(N) = \Gamma_0(N)\backslash\mathfrak{h}^\times$). Explicitly, the Galois conjugates of x_c are

$$\sigma_i(x_c) = [\mathbb{C}/\mathfrak{a}_i^{-1} \to \mathbb{C}/\mathfrak{a}_i^{-1}\mathcal{N}_c^{-1}],\ i = 1, \dots, h_c.$$

Next, we use (1) to approximate $\varphi(\sigma_i(x_c))$ as an element of $\mathbb{C}/\Lambda$ by truncating the infinite series up to sufficiently many terms whose number is determined precisely by the results of the appendix. Finally, the image of $\varphi(\tau_i) + \Lambda$ under the Weierstrass $\wp$-function gives us an approximation of the x-coordinate of the point y_c on the Weierstrass model of the elliptic curve E. On the other hand, this coordinate is $K[c]$-rational. Thus, if we compute the map (1) with sufficiently many terms and up to high enough floating point accuracy, we must be able to recognize the correct x-coordinate of y_c on the Weierstrass model as an element of $K[c]$.

To implement the last step, we use the upper bound established in the appendix on the logarithmic height of the Heegner point y_c. The bound on the logarithmic height comes from a bound on the canonical height combined with bounds on the height difference (see the appendix for complete details). Once we have a height bound, we estimate the floating point accuracy required for the computation. Finally, we estimate the number of terms of (1) necessary to compute the point y_c up to the corresponding accuracy (see [Coh07, p.591] for more details).

Remark 1. In practice, there are two ways to implement the above algorithm. The first approach is to compute an approximation x_i of the x-coordinates of $y_c^{\sigma_i}$ for every $i = 1, \dots, c$ and form the polynomial $F(z) = \prod_{i=1}^{h_c}(z - x_i)$. The coefficients of this polynomial are very close to the rational coefficients of the minimal polynomial of the actual x-coordinate of y_c. Thus, one can try to recognize the coefficients of $F(z)$ by using the continued fractions method. The second approach is to search for the τ_i with the largest imaginary part (which will make the convergence of the corresponding series (1) defining the modular parametrization fast) and then try to search for an algebraic dependence of degree $[K[c] : K]$ using standard algorithms implemented in PARI/GP. Indeed, computing a conjugate with a smaller imaginary part might be significantly harder since the infinite series in (1) will converge slower and one will need more terms to compute the image up to the required accuracy.

Remark 2. We did *not* actually implement an algorithm for computing bounds on heights of Heegner points as described in the appendix of this paper. Thus, the computations in the specific examples below are not provably correct, though we did many consistency checks and we are convinced that our computational observations are correct. The primary goal of the examples and practical implementation of our algorithm is to provide tools and data for improving our theoretical understanding of Kolyvagin's conjecture, and not making the computations below provably correct does not detract from either of these goals.

2.3 Examples

We compute the Heegner points y_c for specific elliptic curves and choices of quadratic imaginary fields.

53a1: Let $E_{/\mathbb{Q}}$ be the elliptic curve with label **53a1** in Cremona's database (see [Cre]). Explicitly, E is the curve $y^2 + xy + y = x^3 - x^2$. Let $D = 43$ and $c = 5$. The conductor of E is 53 which is split in $K = \mathbb{Q}(\sqrt{-D})$, so D is a Heegner discriminant for E. The modular form associated to E is $f_E(q) = q - q^2 - 3q^3 - q^4 + 3q^6 - 4q^7 + 3q^8 + 6q^9 + \cdots$. One applies the methods from Section 2.2 to compute the minimal polynomial of the x-coordinate of y_5 for the above model

$$F(x) = x^6 - 12x^5 + 1980x^4 - 5855x^3 + 6930x^2 - 3852x + 864.$$

Since $F(x)$ is an irreducible polynomial over K, it generates the ring class field $K[5]/K$, i.e., $K[5] = K[\alpha] \cong K[x]/\langle F(x)\rangle$, where α is one of the roots. To find the y-coordinate of y_5 we substitute α into the equation of E and factor the resulting quadratic polynomial over $K[5]$ to obtain that the point y_5 is equal to

$$\big(\alpha, -4/315\alpha^5 + 43/315\alpha^4 - 7897/315\alpha^3 + 2167/35\alpha^2 - 372/7\alpha + 544/35\big) \in E(K[5]).$$

389a1: The elliptic curve with label **389a1** is $y^2 + y = x^3 + x^2 - 2x$ and the associated modular form is $f_E(q) = q - 2q^2 - 2q^3 + 2q^4 - 3q^5 + 4q^6 - 5q^7 + q^9 + 6q^{10} + \cdots$. Let $D = 7$ (which is a Heegner discriminant for E) and $c = 5$. As above, we compute the minimal polynomial of the x-coordinate of y_5:

$$F(x) = x^6 + \frac{10}{7}x^5 - \frac{867}{49}x^4 - \frac{76}{245}x^3 + \frac{3148}{35}x^2 - \frac{25944}{245}x + \frac{48771}{1225}.$$

If α is a root of $F(x)$ then $y_5 = (\alpha, \beta)$ where

$$\beta = \frac{280}{7761}\sqrt{-7}\alpha^5 + \frac{1030}{7761}\sqrt{-7}\alpha^4 - \frac{12305}{36218}\sqrt{-7}\alpha^3 - \frac{10099}{15522}\sqrt{-7}\alpha^2 + \frac{70565}{54327}\sqrt{-7}\alpha + \frac{-18109 - 33814\sqrt{-7}}{36218}.$$

709a1: The elliptic curve **709a1** with equation $y^2 + y = x^3 - x^2 - 2x$ has an associated modular form $f_E(q) = q - 2q^2 - q^3 + 2q^4 - 3q^5 + 2q^6 - 4q^7 - 2q^9 + \cdots$. Let $D = 7$ (a Heegner discriminant for E) and $c = 5$. The minimal polynomial of the x-coordinate of y_5 is

$$F(x) = \frac{1}{5^2 \cdot 7^2 \cdot 19^2}\left(442225x^6 - 161350x^5 - 2082625x^4 - 387380x^3 + 2627410x^2 + 18136030x + 339921\right),$$

and if α is a root of x then $y_5 = (\alpha, \beta)$ for

$$\beta = \frac{341145}{62822}\sqrt{-7}\alpha^5 - \frac{138045}{31411}\sqrt{-7}\alpha^4 - \frac{31161685}{1319262}\sqrt{-7}\alpha^3 + \frac{7109897}{1319262}\sqrt{-7}\alpha^2 + \\ + \frac{39756589}{1319262}\sqrt{-7}\alpha + \frac{-219877 + 4423733\sqrt{-7}}{439754}.$$

718b1: The curve **718b1** has equation $y^2 + xy + y = x^3 - 5x$ with associated modular form $f_E(q) = q - q^2 - 2q^3 + q^4 - 3q^5 + 2q^6 - 5q^7 - q^8 + q^9 + 3q^{10} + \dots$. Again, for $D = 7$ and $c = 5$ we find $F(x) = \frac{1}{3^4 \cdot 5^2}\left(2025x^6 + 12400x^5 + 32200x^4 + 78960x^3 + 289120x^2 + 622560x + 472896\right)$ and $y_5 = (\alpha, \beta)$ with

$$\beta = \frac{16335}{12271}\sqrt{-7}\alpha^5 + \frac{206525}{36813}\sqrt{-7}\alpha^4 + \frac{54995}{5259}\sqrt{-7}\alpha^3 + \frac{390532}{12271}\sqrt{-7}\alpha^2 + \\ + \frac{-36813 + 9538687\sqrt{-7}}{73626}\alpha + \frac{-12271 + 4018835\sqrt{-7}}{24542}.$$

3. Kolyvagin's conjecture: consequences and evidence

We recall Kolyvagin's construction of the cohomology classes in Section 3.2 and state Kolyvagin's conjecture in Section 3.3. Section 3.4 is devoted to the proof of the promised consequence regarding the $\mathbb{Z}_p$-corank of the Selmer group of an elliptic curve with large analytic rank. In Section 3.5 we provide Cornut's arguments for the theoretical evidence for Kolyvagin's conjecture and finally, in Section 3.6 we verify Kolyvagin's conjecture for particular elliptic curves. Throughout the entire section we assume that $E_{/\mathbb{Q}}$ is an elliptic curve of conductor N, D is a Heegner discriminant for E and $p \nmid ND$ is a prime such that the mod p Galois representation $\overline{\rho}_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[p])$ is surjective.

3.1 Preliminaries

Most of this section follows the exposition in [Gro91], [McC91] and [Kol91c].

1. Kolyvagin primes. We refer to a prime number ℓ as a *Kolyvagin prime* if ℓ is inert in K and p divides both a_ℓ and $\ell + 1$). For a Kolyvagin prime ℓ let

$$M(\ell) = \mathrm{ord}_p(\gcd(a_\ell, \ell + 1)).$$

We denote by Λ^r the set of all square-free products of exactly r Kolyvagin primes and let $\Lambda = \bigcup_r \Lambda^r$. For any $c \in \Lambda$, let $M(c) = \min_{\ell \mid c} M(\ell)$. Finally, let

$$\Lambda^r_m = \{c \in \Lambda^r : M(c) \geqslant m\}$$

and let $\Lambda_m = \bigcup_r \Lambda^r_m$.

2. Kolyvagin derivative operators. Let $\mathcal{G}_c = \mathrm{Gal}(K[c]/K)$ and $G_c = \mathrm{Gal}(K[c]/K[1])$. For each $\ell \in \Lambda^1$, the group G_ℓ is cyclic of order $\ell + 1$. Indeed,

$$G_\ell \simeq (\mathcal{O}_K/\ell\mathcal{O}_K)^\times/(\mathbb{Z}/\ell\mathbb{Z})^\times \simeq \mathbb{F}_\lambda^\times/\mathbb{F}_\ell^\times.$$

Moreover, $G_c \cong \prod_{\ell | c} G_\ell$ (since $\mathrm{Gal}(K[c]/K[c/\ell]) \cong G_\ell$). Next, fix a generator σ_ℓ of G_ℓ for each $\ell \in \Lambda^1$. Define $D_\ell = \sum_{i=1}^{\ell} i\sigma_\ell^i \in \mathbb{Z}[G_\ell]$ and let

$$D_c = \prod_{\ell | c} D_\ell \in \mathbb{Z}[G_c].$$

Note that $(\sigma_\ell - 1)D_\ell = 1 + \ell - \mathrm{Tr}_{K[\ell]/K[1]}$.

We refer to D_c as the *Kolyvagin derivative operators.* Finally, let S be a set of coset representatives for the subgroup $G_c \subseteq \mathcal{G}_c$. Define

$$P_c = \sum_{s \in S} s D_c y_c \in E(K[c]).$$

The points P_c are derived from the points y_c, so we will refer to them as *derived Heegner points.*

3. The function $m : \Lambda \to \mathbb{Z}$ and the sequence $\{m_r\}_{r \geqslant 0}$. For any $c \in \Lambda$ let $m'(c)$ be the largest positive integer such that $P_c \in p^{m'(c)}E(K[c])$ (if P_c is torsion then $m'(c) = \infty$). Define a function $m : \Lambda \to \mathbb{Z}$ by

$$m(c) = \begin{cases} m'(c) & \text{if } m'(c) \leqslant M(c), \\ \infty & \text{otherwise.} \end{cases}$$

Finally, let $m_r = \min_{c \in \Lambda^r} m(c)$.

Proposition 3.1. *The sequence $\{m_r\}_{r \geqslant 0}$ is non-increasing, i.e., $m_r \geqslant m_{r+1}$ for every $r \geqslant 0$.*

Proof. This is proved in [Kol91c, Thm.C]. □

3.2 Kolyvagin cohomology classes

Kolyvagin uses the points P_c to construct classes $\kappa_{c,m} \in \mathrm{H}^1(K, E[p^m])$ for any $c \in \Lambda_m$. For the details of the construction, we refer to [Gro91, pp.241-242]) and [McC91, §4]. The class $\kappa_{c,m}$ is explicit, in the sense that it is represented by the 1-cocycle

$$\sigma \mapsto \sigma\left(\frac{P_c}{p^m}\right) - \frac{P_c}{p^m} - \frac{(\sigma - 1)P_c}{p^m}, \tag{2}$$

where $\dfrac{(\sigma-1)P_c}{p^m}$ is the unique p^m-division point of $(\sigma - 1)P_c$ in $E(K[c])$ (see [McC91, Lem. 4.1]). The class $\kappa_{c,m}$ is non-trivial if and only if $P_c \notin p^m E(K[c])$ (which is equivalent to $m > m(c)$).

Finally, let $-\varepsilon$ be the sign of the functional equation corresponding to E. For each $c \in \Lambda_m$, let $\varepsilon(c) = \varepsilon \cdot (-1)^{f_c}$ where $f_c = \#\{\ell : \ell \mid c\}$ (e.g., $f_1 = 0$). It follows from [Gro91, Prop.5.4(ii)] that $\kappa_{c,m}$ lies in the $\varepsilon(c)$-eigenspace for the action of complex conjugation on $\mathrm{H}^1(K, E[p^m])$.

3.3 Statement of the conjecture

We are interested in $m_\infty = \min_{c \in \Lambda} m(c) = \lim_{r \to \infty} m_r$. In the case when the Heegner point $P_1 = y_K$ has infinite order in $E(K)$, the Gross-Zagier formula (see [GZ86]) implies that $r_{\mathrm{an}}(E/K) = 1$, so (by the results of Kolyvagin) $r_{\mathrm{MW}}(E/K) = 1$. This means that $m_0 = \mathrm{ord}_p([E(K) : \mathbb{Z}y_K]) < \infty$. In particular, $m_\infty < \infty$ which is equivalent to the system of cohomology classes

$$T = \{\kappa_{c,m} : m \leqslant M(c)\}$$

containing at least one non-zero class. A much more interesting and subtle is the case of an elliptic curves E over K of analytic rank at least two. In this case, Kolyvagin conjectured (see [Kol91a, Conj.C]) that T contains a non-trivial class as well.

Conjecture 3 (Kolyvagin's conjecture). We have $m_\infty < \infty$, i.e., T contains at least one class $\kappa_{c,m} \neq 0$.

Remark 3. Although Kolyvagin's conjecture is obvious in the case of elliptic curves of analytic rank one over K, the number m_∞ is still interesting. Indeed, the p-part of the Birch and Swinnerton-Dyer conjectural formula for $E_{/K}$ is equivalent to $m_\infty = \mathrm{ord}_p \left(\prod_{q \mid N} c_q \right)$, where c_q is the Tamagawa number of $E_{/\mathbb{Q}}$ at q. See [Jet07] for some new results related to this question which imply (in many cases) the exact upper bound on the order of the p-primary part of the Shafarevich-Tate group as predicted by the conjectural formula.

3.4 A consequence on the structure of Selmer groups

Let $r_p^{\pm}(E/K) = \mathrm{corank}_{\mathbb{Z}_p} \mathrm{Sel}_{p^\infty}(E/K)^{\pm}$. Kolyvagin (see [Kol91a]) proved the following:

Theorem 3.2 (Kolyvagin). *Assume Conjecture 3 and let f be the smallest nonnegative integer for which $m_f < \infty$. Then*

$$\mathrm{Sel}_{p^\infty}(E/K)^{\varepsilon(-1)^{f+1}} \cong (\mathbb{Q}_p/\mathbb{Z}_p)^{f+1} \oplus \textit{(a finite group)}$$

and

$$\mathrm{Sel}_{p^\infty}(E/K)^{\varepsilon(-1)^{f}} \cong (\mathbb{Q}_p/\mathbb{Z}_p)^{r} \oplus \textit{(a finite group)}$$

where $r \leqslant f$ and $f - r$ is even. In other words, $r_p^{\varepsilon(-1)^f}(E/K) = r$ and $r_p^{\varepsilon(-1)^{f+1}}(E/K) = f + 1$.

The above structure theorem of Kolyvagin has the following consequence which strongly supports Conjecture 2.

Corollary 3.3. *Assume Conjecture 3. Then (i) If $r_{\mathrm{an}}(E/\mathbb{Q})$ is even and nonzero then*

$$r_p(E/\mathbb{Q}) \geqslant 2.$$

(ii) If $r_{\mathrm{an}}(E/\mathbb{Q})$ is odd and strictly larger than one then

$$r_p(E/\mathbb{Q}) \geqslant 3.$$

Proof. (i) By [BFH90] or [MM97], one can choose a quadratic imaginary field $K = \mathbb{Q}(\sqrt{-D})$ with Heegner discriminant D, such that $L'(E^D_{/\mathbb{Q}}, 1) \neq 0$, where E^D is the twist of E by the quadratic character associated to K (note that D is a Heegner discriminant, so the sign of the functional equation of E^D is always odd since E has even sign). Hence, by Theorem 1.1, the Selmer group $\mathrm{Sel}_{p^\infty}(E^D/\mathbb{Q})$ has $\mathbb{Z}_p$-corank one, i.e., $r_p^-(E/K) = r_p(E^D/\mathbb{Q}) = 1$. We want to prove $r_p^+(E/K) = r_p(E/\mathbb{Q}) \geqslant 2$. Assume the contrary, i.e., $r_p^+(E/K) = r_p(E/\mathbb{Q}) \leqslant 1$. This means (by Theorem 3.2) that $r = f = 0$ (here, r and f are as in Theorem 3.2). Therefore, $m_0 < \infty$ which means that the Heegner point y_K has infinite order in $E(K)$ and hence (by the Gross-Zagier formula), $L'(E_{/K}, 1) \neq 0$. But this is a contradiction since

$$L'(E_{/K}, s) = L'(E_{/\mathbb{Q}}, s) L(E^D_{/\mathbb{Q}}, s) + L(E_{/\mathbb{Q}}, s) L'(E^D_{/\mathbb{Q}}, s),$$

which vanishes at $s = 1$ since $L(E_{/\mathbb{Q}}, 1) = L'(E_{/\mathbb{Q}}, 1) = 0$. Thus, $r_p(E/\mathbb{Q}) = r_p^+(E/K) \geqslant 2$.

(ii) It follows from the work of Waldspurger (see also [BFH90, pp.543-44]) that one can choose a quadratic imaginary field $K = \mathbb{Q}(\sqrt{-D})$ with a Heegner discriminant D, such that $L(E^D_{/\mathbb{Q}}, 1) \neq 0$ (this uses the fact that $r_{\mathrm{an}}(E/\mathbb{Q})$ is odd). Therefore, by Theorem 1.1, $r_p(E^D/\mathbb{Q}) = 0$, i.e., $r_p^-(E/K) = 0$. By Theorem 3.2, $r = 0$ and f is even. If $f \geqslant 2$ we are done since $r_p(E/\mathbb{Q}) = r_p^+(E/K) = f + 1 \geqslant 3$. If $f = 0$, we use the same argument as in (i) to arrive at a contradiction. Therefore, $r_p(E/\mathbb{Q}) = r_p^+(E/K) \geqslant 3$. □

Remark 4. The parity conjecture proved by Nekovář (see [Nek01]) implies that

$$r_p(E/\mathbb{Q}) \equiv r_{\mathrm{an}}(E/\mathbb{Q}) \mod 2.$$

Yet, Nekovář's result does not imply in any obvious way the statements of the above proposition.

3.5 Cornut's theoretical evidence for Kolyvagin's conjecture

The following evidence for Conjecture 3 was proven by Cornut.

Proposition 3.4. *For all but finitely many $c \in \Lambda$ there exists a set R of liftings for the elements of $\mathrm{Gal}(K[1]/K)$ into $\mathrm{Gal}(K^{\mathrm{ab}}/K)$, such that if $P_c = D_0 D_c y_c$ is the derived Heegner point defined in terms of this choice of liftings (i.e, if $D_0 = \sum_{\sigma \in R} \sigma$), then P_c is non-torsion.*

Remark 5. For a non-torsion point P_c, let $m'(c)$ be the function defined in Section 3.1. Proposition 3.4 provides little evidence towards Kolyvagin's conjecture. The reason is that even if one gets non-torsion points P_c, it might still happen that for each such c we have $m'(c) > M(c)$ (i.e., $m(c) = \infty$) in which case all classes $\kappa_{c,m}$ with $m \leqslant M(c)$ will be trivial.

Let $K[\infty] = \bigcup_{c \in \Lambda} K[c]$. The proof of the proposition depends on the following two lemmas:

Lemma 3.5. *The group $E(K[\infty])_{\mathrm{tors}}$ is finite.*

Proof. Let q be any prime which is a prime of good reduction for E, which is inert in K and which is different from the primes in Λ^1 (there are infinitely many such primes according to Čebotarev density theorem). Let $\mathfrak{q}$ be the unique prime of K over q. It follows from class field theory that the prime $\mathfrak{q}$ splits completely in $K[\infty]$ since it splits completely in each of the finite extensions $K[c]$. Thus, the completion of $K[\infty]$ at any prime which lies over $\mathfrak{q}$ is isomorphic to $K_{\mathfrak{q}}$ and therefore, $E(K[\infty])_{\mathrm{tors}} \hookrightarrow E(K_{\mathfrak{q}})_{\mathrm{tors}}$. The last group is finite since it is isomorphic to an extension of $\mathbb{Z}_q^2$ by a finite group (see [Mil86, Lem.I.3.3] or [Tat67, p.168-169]). Therefore, $E(K[\infty]_{\mathrm{tors}})$ is finite. □

Let $|E(K[\infty])_{\mathrm{tors}}| = M < \infty$ and let $d(c) = [K[c] : K[1]] = \prod_{\ell | c} (\ell + 1)$ for any $c \in \Lambda$. Let m_E be the modular degree of E, i.e., the degree of the fixed optimal modular parametrization $\varphi : X_0(N) \to E$.

Lemma 3.6. *Suppose that $c \in \Lambda$ satisfies $d(c) > m_E M$. There exists a set of lifting R of $\mathrm{Gal}(K[1]/K)$ into $\mathrm{Gal}(K[c]/K)$, such that $D_0 y_c \notin E(K[c])_{\mathrm{tors}}$, where $D_0 = \sum_{\sigma \in R} \sigma$.*

Proof. The $\mathrm{Gal}(K[c]/K[1])$-orbit of the point $x_c \in X_0(N)(K[c])$ consists of $d(c)$ distinct points (since $K[c] = K(j(\mathcal{O}_c)))$, so there are at least $d(c)/m_E$ distinct points in the orbit $\mathrm{Gal}(K[c]/K[1])y_c$. Choose a set of representatives R of $\mathrm{Gal}(K[c]/K)/\mathrm{Gal}(K[c]/K[1])$ which contains the identity element $1 \in \mathrm{Gal}(K[c]/K)$. For $\tau \in \mathrm{Gal}(K[c]/K[1])$ define

$$R_\tau = (R - \{1\}) \cup \{\tau\}.$$

Let $S = \sum_{\sigma \in R} \sigma y_c$ and $S_\tau = \sum_{\sigma \in R_\tau} \sigma y_c$. Then

$$S_\tau - S = \tau y_c - y_c,$$

which takes at least $d(c)/m_E > M$ distinct values. Therefore, there exists an automorphism $\tau \in \mathrm{Gal}(K[c]/K[1])$, for which $S_\tau \notin E(K[c])_{\mathrm{tors}}$, which proves the lemma. □

Proof of Proposition 3.4. Suppose that $c \in \Lambda$ satisfies the statement of Lemma 3.6 for some choice of liftings R and the corresponding $D_0 = \sum_{\sigma \in R}$, i.e., $D_0 y_c \notin E(K[c])_{\mathrm{tors}}$. For any ring class character $\chi : \mathrm{Gal}(K[c]/K) \to \mathbb{C}^\times$, let $e_\chi \in \mathbb{C}[\mathrm{Gal}(K[c]/K)]$ be the eidempotent projector corresponding to χ. Explicitly,

$$e_\chi = \frac{1}{\# \mathrm{Gal}(K[c]/K)} \sum_{\sigma \in \mathrm{Gal}(K[c]/K)} \chi^{-1}(\sigma)\sigma \in \mathbb{C}[\mathrm{Gal}(K[c]/K)].$$

Consider $V = E(K[c]) \otimes \mathbb{C}$ as a complex representation of $\mathrm{Gal}(K[c]/K)$. The representation V decomposes as

$$V = \bigoplus_{\chi : \mathrm{Gal}(K[c]/K) \to \mathbb{C}^\times} V_\chi,$$

where V_χ is the one-dimensional subspace on which $\mathrm{Gal}(K[c]/K[1])$ acts via the character χ. Since the vector $D_0 y_c \otimes 1 \in V$ is non-zero, there exists a ring class character χ, such that $e_\chi(D_0 y_c \otimes 1) \neq 0$.

Next, we consider the point $D_0 D_c y_c \in E(K[c])$ and claim that $D_0 D_c y_c \otimes 1 \in E(K[c]) \otimes \mathbb{C}$ is non-zero, which is sufficient to conclude that $P_c = D_0 D_c y_c \notin E(K[c])_{\mathrm{tors}}$. We will prove that $e_\chi(D_0 D_c y_c \otimes 1) \neq 0$. Indeed,

$$e_\chi(D_0 D_c y_c \otimes 1) = e_\chi D_c (D_0 y_c \otimes 1) = \prod_{\ell \mid c} \left(\sum_{i=1}^{\ell} i \sigma_\ell^i \right) e_\chi (D_0 y_c \otimes 1) =$$

$$= \prod_{\ell \mid c} \left(\sum_{i=1}^{\ell} i \chi(\sigma_\ell)^i \right) e_\chi (D_0 y_c \otimes 1),$$

the last equality holding since $\tau e_\chi = \chi(\tau) e_\chi$ in $\mathbb{C}[\mathrm{Gal}(K[c]/K)]$ for all $\tau \in \mathrm{Gal}(K[c]/K)$. Thus, it remains to compute $\sum_{i=1}^{\ell} i\chi(\sigma_\ell)^i$ for every $\ell \mid c$. It is not hard to show that

$$\sum_{i=1}^{\ell} i\chi(\sigma_\ell)^i = \begin{cases} \frac{\ell+1}{\chi(\sigma_\ell)-1} & \text{if } \chi(\sigma_\ell) \neq 1 \\ \frac{\ell(\ell+1)}{2} & \text{if } \chi(\sigma_\ell) = 1. \end{cases}$$

Thus, $e_\chi(D_0 D_c y_c \otimes 1) \neq 0$ which means that $P_c = D_0 D_c y_c \notin E(K[c])_{\mathrm{tors}}$ for any c satisfying $D_0 y_c \notin E(K[c])_{\mathrm{tors}}$. To complete the proof, notice that for all, but finitely many $c \in \Lambda$, the hypothesis of Lemma 3.6 will be satisfied. □

3.6 Computational evidence for Kolyvagin's conjecture

Consider the example $E =$ **389a1** with equation $y^2 + y = x^3 + x^2 - 2x$. As in Section 2.3, let $D = 7$, $\ell = 5$, and $p = 3$. Using the algorithm of [GJP$^+$05, §2.1] we verify that the mod p Galois representation $\overline{\rho}_{E,p}$ is surjective. Next, we observe that $\ell = 5$ is a Kolyvagin prime for E, p and D. Let $c = 5$ and consider the class $\kappa_{5,1} \in \mathrm{H}^1(K, E[3])$. We claim that $\kappa_{5,1} \neq 0$, which will verify Kolyvagin's conjecture.

Proposition 3.7. *The class $\kappa_{5,1} \neq 0$. In other words, Kolyvagin's conjecture holds for $E =$* **389a1**, *$D = 7$ and $p = 3$.*

Before proving the proposition, we recall some standard facts about division polynomials (see, e.g., [Sil92, Ex.3.7]). For an elliptic curve given in Weierstrass form over any field of characteristic different from 2 and 3, $y^2 = x^3 + Ax + B$, one defines a sequence of polynomials $\psi_m \in \mathbb{Z}[A, B, x, y]$

inductively as follows:

$$\begin{aligned}
&\psi_1 = 1,\ \psi_2 = 2y,\\
&\psi_3 = 3x^4 + 6Ax^2 + 12Bx - A^2,\\
&\psi_4 = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3),\\
&\psi_{2m+1} = \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3 \ \text{ for } m \geqslant 2,\\
&2y\psi_{2m} = \psi_m(\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2) \ \text{ for } m \geqslant 3.
\end{aligned}$$

Define also polynomials ϕ_m and ω_m by

$$\phi_m = x\psi_m^2 - \psi_{m+1}\psi_{m-1},\ 4y\omega_m = \psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2.$$

After replacing y^2 by $x^3 + Ax + B$, the polynomials ϕ_m and ψ_m^2 can be viewed as polynomials in x with leading terms x^{m^2} and $m^2x^{m^2-1}$, respectively. Finally, multiplication-by-m is given by

$$mP = \left(\frac{\phi_m(P)}{\psi_m(P)^2}, \frac{\omega_m(P)}{\psi_m(P)^3}\right).$$

Proof of Proposition 3.7. We already computed the Heegner point y_5 on the model $y^2 + y = x^3 + x^2 - 2x$ in Section 2.3. The Weierstrass model for E is $y^2 = x^3 - 7/3x + 107/108$, so $A = -7/3$ and $B = 107/108$. We now compute the point $P_5 = \sum_{i=1}^{5} i\sigma^i(y_5) \in E(K[5])$ on the Weierstrass model, where σ is a generator of $\mathrm{Gal}(K[5]/K)$. To show that $\kappa_{5,1} \neq 0$ we need to check that there is no point $Q = (x, y)$, such that $3Q = P_5$. For the verification of this fact, we use the division polynomial ψ_3 and the polynomial ϕ_3. Indeed, it follows from the recursive definitions that

$$\begin{aligned}
\phi_3(x) = {} & x^9 - 12Ax^7 - 168Bx^6 + (30A^2 + 72B)x^5 - 168ABx^4 +\\
& + (36A^3 + 144AB - 96B^2)x^3 + 72A^2Bx^2 +\\
& + (9A^4 - 24A^2B + 96AB^2 + 144B^2)x + 8A^3B + 64B^3.
\end{aligned}$$

Consider the polynomial $g(x) = \phi_3(x) - X(P_5)\psi_3(x)^2$, where $X(P_5)$ is the x-coordinate of the point P_5 on the Weierstrass model. We factor $g(x)$ (which has degree 9) over the number field $K[5]$ and check that it is irreducible. In particular, there is no root of $g(x)$ in $K[5]$, i.e., there is no $Q \in E(K[5])$, such that $3Q = P_5$. Thus, $\kappa_{5,1} \neq 0$. □

Remark 6. Using exactly the same method as above, we verify Kolyvagin's conjecture for the other two elliptic curves of rank two from Section 2.3. For both $E =$ **709a1** and $E =$ **718b1** we use $D = 7$, $p = 3$ and $\ell = 5$ (which are valid parameters), and verify that $\kappa_{5,1} \neq 0$ in the two cases. For completeness, we provide all the data of each computation in the three examples in the files `389a1.txt`, `709a1.txt` and `718a1.txt`.

4. Non-trivial elements in the Shafarevich-Tate group

Throughout the entire section, let $E_{/\mathbb{Q}}$ be a non-CM elliptic curve, $K = \mathbb{Q}(\sqrt{-D})$, where D is a Heegner discriminant for E such that the Heegner point y_K has infinite order in $E(K)$ (which, by the Gross-Zagier formula and Kolyvagin's result, means that $E(K)$ has Mordell-Weil rank one) and let p be a prime, such that $p \nmid DN$ and the mod p Galois representation $\overline{\rho}_{E,p}$ is surjective.

4.1 Non-triviality in Kolyvagin classes.

Under the above assumptions, the next proposition provides a criterion which guarantees that an explicit class in the Shafarevich-Tate group $\text{Ш}(E/K)$ is non-zero.

Proposition 4.1. *Let $c \in \Lambda_m$. Assume that the following hypotheses are satisfied:*

i) *[Selmer hypothesis]: The class* $\kappa_{c,m} \in \mathrm{H}^1(K, E[p^m])$ *is an element of the Selmer group* $\mathrm{Sel}_{p^m}(E/K)$.

ii) *[Non-divisibility]: The derived Heegner point* P_c *is not divisible by* p^m *in* $E(K[c])$*, i.e.,* $P_c \notin p^m E(K[c])$.

iii) *[Parity]: The number* $f_c = \#\{\ell : \ell \mid c\}$ *is odd.*

Then the image $\kappa'_{c,m} \in \mathrm{H}^1(K, E)[p^m]$ *of* $\kappa_{c,m}$ *is a non-zero element of* $\text{Ш}(E/K)[p^m]$.

Proof. The first hypothesis implies that the image $\kappa'_{c,m}$ of $\kappa_{c,m}$ in $\mathrm{H}^1(K, E)[p^m]$ is an element of the Shafarevich-Tate group $\text{Ш}(E/K)$. The second one implies that $\kappa_{c,m} \neq 0$. To show that $\kappa'_{c,m} \neq 0$ we use the exact sequence

$$0 \to E(K)/p^m E(K) \to \mathrm{Sel}_{p^m}(E/K) \to \text{Ш}(E/K)[p^m] \to 0$$

which splits under the action of complex conjugation as

$$0 \to (E(K)/p^m E(K))^{\pm} \to \mathrm{Sel}_{p^m}(E/K)^{\pm} \to \text{Ш}(E/K)^{\pm}[p^m] \to 0.$$

According to [Gro91, Prop.5.4(2)], the class $\kappa_{c,m}$ lies in the ε_c-eigenspace of the Selmer group $\mathrm{Sel}_{p^m}(E/K)$ for the action of complex conjugation, where $\varepsilon_c = \varepsilon(-1)^{f_c} = -1$ (f_c is odd by the third hypothesis and $\varepsilon = 1$ since $-\varepsilon$ is the sign of the functional equation for $E_{/K}$ which is -1 by Gross-Zagier). On the other hand, the Heegner point $y_K = P_1$ lies in the ε_1-eigenspace of complex conjugation (again, by [Gro91, Prop.5.4(2)]) where $\varepsilon_1 = \varepsilon(-1)^{f_1} = 1$. Since $E(K)$ has rank one, the group $E(K)^-$ is torsion and since $E(K)[p] = 0$, we obtain that $(E(K)/p^m E(K))^- = 0$. Therefore,

$$\mathrm{Sel}_{p^m}(E/K)^- \cong \text{Ш}(E/K)^-[p^m],$$

which implies $\kappa'_{c,m} \neq 0$. □

4.2 The example $E =$ 53a1.

The Weierstrass equation for the curve $E =$**53a1** is $y^2 = x^3 + 405x + 16038$ and E has rank one over $\mathbb{Q}$. The Fourier coefficient $a_5(f) \equiv 5 + 1 \equiv 0 \mod 3$, so $\ell = 5$ is a Kolyvagin prime for E, the discriminant $D = 43$ and the prime $p = 3$. Kolyvagin's construction exhibits a class $\kappa_{5,1} \in \mathrm{H}^1(K, E[3])$. We will prove the following proposition:

Proposition 4.2. *The cohomology class* $\kappa_{5,1} \in \mathrm{H}^1(K, E[3])$ *lies in the Selmer group* $\mathrm{Sel}_3(E/K)$ *and its image* $\kappa'_{5,1}$ *in the Shafarevich-Tate group* $\text{Ш}(E/K)$ *is a nonzero* 3*-torsion element.*

Remark 7. Since E/K has analytic rank one, Kolyvagin's conjecture is automatic (since $m_0 < \infty$ by Gross-Zagier's formula) and one knows (see [McC91, Thm. 5.8]) that there exist Kolyvagin classes $\kappa'_{c,m}$ which generate $\text{Ш}(E/K)[p^\infty]$. Yet, this result is not explicit in the sense that one does not know any particular Kolyvagin class which is non-trivial. The above proposition exhibits an explicit non-zero cohomology class in the p-primary part of the Shafarevich-Tate group $\text{Ш}(E/K)$.

Proof. Using the data computed in Section 2.3 for this curve, we apply the Kolyvagin derivative to compute the point P_5. In order to do this, one needs a generator of the Galois group $\mathrm{Gal}(K[5]/K)$. Such a generator is determined by the image of α, which will be another root of $f(x)$ in $K[5]$. We check that the automorphism σ defined by

$$\begin{aligned} \alpha \mapsto\ & \frac{1}{1601320}(47343 + 54795\sqrt{-43})\alpha^5 + \frac{1}{2401980}(-614771 - 936861\sqrt{-43})\alpha^4 + \\ & + \frac{1}{600495}(34507457 + 40541607\sqrt{-43})\alpha^3 + \frac{1}{4803960}(102487877 - 767102463\sqrt{-43})\alpha^2 + \\ & + \frac{1}{400330}(-61171198 + 52833377\sqrt{-43})\alpha + \frac{1}{200165}(18971815 - 7453713\sqrt{-43}) \end{aligned}$$

is a generator (we found this automorphism by factoring the defining polynomial of the number field over the number field $K[5]$). Thus, we can compute $P_5 = \sum_{i=1}^{5} i\sigma^i(y_5)$.

Note that we are computing the point on the Weierstrass model of E rather than on the original model. The cohomology class $\kappa_{5,1}$ is trivial if and only if $P_5 \in 3E(K[5])$. To show that $P_5 \notin 3E(K[5])$, we repeat the argument from Proposition 3.7 and verify (using any factorization algorithm for polynomials over number fields) that the polynomial $g(x) = \phi_3(x) - X(P_5)\psi_3(x)^2$ has no linear factors over $K[5]$ (here, $X(P_5)$ is the x-coordinate of P_5). This means that there is no point $Q = (x, y) \in E(K[5])$, such that $3Q = P_5$, i.e., $\kappa_{5,1} \neq 0$. Finally, using Proposition 4.1 we conclude that the class $\kappa'_{5,1} \in Ш(E/K)[3]$ is non-trivial. □

Remark 8. For completeness, all the computational data is provided (with the appropriate explanations) in the file `53a1.txt`. We verified the irreducibility of $g(x)$ using MAGMA and PARI/GP independently.

Appendix A. Upper bounds on the logarithmic heights of the Heegner points y_c

We explain how to compute an upper bound on the logarithmic height $h(y_c)$. The method first relates the canonical height $\widehat{h}(y_c)$ to special values of the first derivatives of certain automorphic L-functions via Zhang's generalization of the Gross-Zagier formula. Then we either compute the special values up to arbitrary precision using a well-known algorithm (recently implemented by Dokchitser) or use effective asymptotic upper bounds (convexity bounds) on the special values and Cauchy's integral formula. Finally, using some known bounds on the difference between the canonical and the logarithmic heights, we obtain explicit upper bounds on the logarithmic height $h(y_c)$. We provide a summary of the asymptotic bounds in Section A.4 and refer the reader to [Jet] for complete details.

A.1 The automorphic L-functions $L(f, \chi, s)$ and $L(\pi_{f\otimes\theta_\chi}, s)$

Let $d_c = c^2 D$ and let $f = \sum_{n\geqslant 1} a_n q^n$ be the new eigenform of level N and weight two corresponding to E. Let $\chi : \mathrm{Gal}(K[c]/K) \to \mathbb{C}^\times$ be a ring class character.

1. The theta series θ_χ. Recall that ideal classes for $\mathrm{Pic}(\mathcal{O}_c)$ correspond to primitive, reduced binary quadratic forms of discriminants d_c. To each ideal class $\mathcal{A}$ we consider the corresponding binary quadratic form $Q_\mathcal{A}$ and the theta series $\theta_{Q_\mathcal{A}}$ associated to it via

$$\theta_{Q_\mathcal{A}} = \sum_M e^{2\pi i z Q_\mathcal{A}(M)}$$

which is a modular form for $\Gamma_0(d_c)$ of weight one with character ε (the quadratic character of K) according to Weil's converse theorem (see [Shi71] for details). This allows us to define a cusp form

$$\theta_\chi = \sum_{\mathcal{A}\in \mathrm{Pic}(\mathcal{O}_c)} \chi^{-1}(\mathcal{A})\theta_{Q_\mathcal{A}} \in S_1(\Gamma_0(d_c), \varepsilon).$$

Here, we view χ^{-1} as a character of $\mathrm{Pic}(\mathcal{O}_c)$ via the isomorphism $\mathrm{Pic}(\mathcal{O}_c) \cong \mathrm{Gal}(K[c]/K)$. Let $\theta_\chi = \sum_{m\geqslant 0} b_m q^m$ be the Fourier expansion. By $L(f, \chi, s)$ we will always mean the Rankin L-function $L(f \otimes \theta_\chi, s)$ (equivalently, the L-function associated to the automorphic representation $\pi = f \otimes \theta_\chi$ of GL_4).

2. The functional equation of $L(f, \chi, s)$. We recall some basic facts about the Rankin L-series $L(f \otimes \theta_\chi, s)$ following [Gro84, §III]. Since $(N, D) = 1$, the *conductor* of $L(f \otimes \theta_\chi, s)$ is $Q = N^2 d_c^2$.

The Euler factor at infinity (the gamma factor) is $L_\infty(f \otimes \theta_\chi, s) = \Gamma_{\mathbb{C}}(s)^2$. If we set

$$\Lambda(f \otimes \theta_\chi, s) = Q^{s/2} L_\infty(f \otimes \theta_\chi, s) L(f \otimes \theta_\chi, s)$$

then the function Λ has a holomorphic continuation to the entire complex plane and satisfies the functional equation

$$\Lambda(f \otimes \theta_\chi, s) = -\Lambda(f \otimes \theta_\chi, 2 - s).$$

In particular, the order of vanishing of $L(f \otimes \theta_\chi, s)$ at $s = 1$ is non-negative and odd, i.e., $L(f \otimes \theta_\chi, 1) = 0$.

3. The automorphic L-function $L(\pi_{f\otimes\theta_\chi}, s)$. In order to center the critical line at $\mathrm{Re}(s) = \frac{1}{2}$ instead of $\mathrm{Re}(s) = 1$ (which is consistent with Langlands convention), we consider the L-function $L(\pi_{f\otimes\theta_\chi}, s)$ corresponding to the automorphic representation attached to $f \otimes \theta_\chi$ for GL_4. This function satisfies

$$L(\pi_{f\otimes\theta_\chi}, s) = L\left(f \otimes \theta_\chi, s + \frac{1}{2}\right)$$

The function $L(\pi_{f\otimes\theta_\chi}, s)$ then satisfies a functional equation relating the values at s and $1 - s$. For a general automorphic L-function $L(\pi, s)$, we consider the corresponding Dirichlet series and Euler product

$$L(\pi, s) = \sum_{n \geqslant 1} \frac{\lambda_\pi(n)}{n^s} = \prod_p (1 - \alpha_{\pi,1}(p)p^{-s})^{-1} \ldots (1 - \alpha_{\pi,d}(p)p^{-s})^{-1},$$

which are absolutely convergent for $\mathrm{Re}(s) > 1$.

A.2 Zhang's formula

For a character χ of $\mathrm{Gal}(K[c]/K)$, let

$$e_\chi = \frac{1}{\#\,\mathrm{Gal}(K[c]/K)} \sum_{\sigma \in \mathrm{Gal}(K[c]/K)} \chi^{-1}(\sigma)\sigma \in \mathbb{C}[\mathrm{Gal}(K[c]/K)]$$

be the associated eidempotent (see also Section 3.5). The canonical height $\widehat{h}(e_\chi y_c)$ is related via the generalized Gross-Zagier formula of Zhang to a special value of the derivative of the Rankin-Selberg L-function $L(f, \chi, s)$ at $s = 1$ (see [Zha01, Thm.1.2.1]). More precisely,

Theorem A.1 (Zhang). *If $(\,,)$ denotes the Petersson inner product on $S_2(\Gamma_0(N))$ then*

$$L'(f, \chi, 1) = \frac{4}{\sqrt{D}}(f, f)\widehat{h}(e_\chi y_c).$$

Since $\langle e_{\chi'} y_c, e_{\chi''} y_c \rangle = 0$ whenever $\chi' \neq \chi''$ (here, $\langle\,,\rangle$ denotes the Néron-Tate height pairing for E) and since $\widehat{h}(x) = \langle x, x \rangle$ then

$$\widehat{h}(y_c) = \widehat{h}\left(\sum_\chi e_\chi y_c\right) = \sum_\chi \widehat{h}(e_\chi y_c). \tag{3}$$

Thus, we will have an upper bound on the canonical height $\widehat{h}(y_c)$ if we obtain upper bounds on the special values $L'(f, \chi, 1)$ for every character χ of $\mathrm{Gal}(K[c]/K)$.

A.3 Computing special values of derivatives of automorphic L-functions

For simplicity, let $\gamma(s) = L_\infty(f \otimes \theta_\chi, s + 1/2)$ be the gamma factor of the L-function $L(\pi, s)$. This means that if $\Lambda(\pi, s) = Q^{s/2}\gamma(s)L(\pi, s)$ then $\Lambda(\pi, s)$ satisfies the functional equation $\Lambda(\pi, s) = \Lambda(\pi, 1 - s)$. We will describe a classical algorithm to compute the value of $L^{(k)}(\pi, s)$ at $s = s_0$ up to arbitrary precision. The algorithm and its implementation is discussed in a greater generality in [Dok04]. The main idea is to express $\Lambda(\pi, s)$ as an infinite series with rapid convergence which is usually done in the following sequence of steps:

i) Consider the inverse Mellin transform of the gamma factor $\gamma(s)$, i.e., the function $\phi(t)$ which satisfies
$$\gamma(s) = \int_0^\infty \phi(t) t^s \frac{dt}{t}.$$
One can show (see [Dok04, §3]) that $\phi(t)$ decays exponentially for large t. Hence, the sum
$$\Theta(t) = \sum_{n=1}^\infty \lambda_\pi(n) \phi\left(\frac{nt}{\sqrt{Q}}\right)$$
converges exponentially fast. The function $\phi(t)$ can be computed numerically as explained in [Dok04, §3-5].

ii) The Mellin transform of $\Theta(t)$ is exactly the function $\Lambda(\pi, s)$. Indeed,
$$\int_0^\infty \Theta(t) t^s \frac{dt}{t} = \int_0^\infty \sum_{n=1}^\infty \lambda_\pi(n) \phi\left(\frac{nt}{\sqrt{Q}}\right) t^s \frac{dt}{t} = \sum_{n=1}^\infty \lambda_\pi(n) \int_0^\infty \phi\left(\frac{nt}{\sqrt{Q}}\right) t^s \frac{dt}{t} =$$
$$= \sum_{n=1}^\infty \lambda_\pi(n) \left(\frac{\sqrt{Q}}{n}\right)^s \int_0^\infty \phi(t') t'^s \frac{dt'}{t'} = Q^{s/2} \gamma(s) L(\pi, s) = \Lambda(\pi, s).$$

iii) Next, we obtain a functional equation for $\Theta(t)$ which relates $\Theta(t)$ to $\Theta(1/t)$. Indeed, since $\Lambda(\pi, s)$ is holomorphic, Mellin's inversion formula implies that
$$\Theta(t) = \int_{c-i\infty}^{c+i\infty} \Lambda(\pi, s) t^{-s} ds, \ \forall c.$$
Therefore,
$$\Theta(1/t) = \int_{c-i\infty}^{c+i\infty} \Lambda(\pi, s)(1/t)^{-s} ds = -t \int_{c-i\infty}^{c+i\infty} \Lambda(\pi, 1-s) t^{-(1-s)} ds =$$
$$= -t \int_{c-i\infty}^{c+i\infty} \Lambda(\pi, s') t^{-s'} ds' = -t\Theta(t).$$
Thus, $\Theta(t)$ satisfies the functional equation $\Theta(1/t) = -t\Theta(t)$.

iv) Next, we consider the incomplete Mellin transform
$$G_s(t) = t^{-s} \int_t^\infty \phi(x) x^s \frac{dx}{x}, \ t > 0$$
of $\phi(t)$. The function $G_s(t)$ satisfies $\lim_{t \to 0} t^s G_s(t) = \gamma(s)$ and it decays exponentially. Moreover, it can be computed numerically (see [Dok04, §4-5]).

v) Finally, we use the functional equation for $\Theta(t)$ to obtain
$$\Lambda(\pi, s) = \int_0^\infty \Theta(t) t^s \frac{dt}{t} = \int_0^1 \Theta(t) t^s \frac{dt}{t} + \int_1^\infty \Theta(t) t^s \frac{dt}{t} =$$
$$= \int_1^\infty \Theta(1/t') t'^{-s} \frac{dt'}{t'} + \int_1^\infty \Theta(t) t^s \frac{dt}{t} =$$
$$= -\int_1^\infty \Theta(t') t'^{1-s} \frac{dt'}{t'} + \int_1^\infty \Theta(t) t^s \frac{dt}{t}.$$

vi) Finally, we compute

$$\int_1^\infty \Theta(t)t^s\frac{dt}{t} = \int_1^\infty \sum_{n=1}^\infty \lambda_\pi(n)\phi\left(\frac{nt}{\sqrt{Q}}\right)t^s\frac{dt}{t} = \sum_{n=1}^\infty \lambda_\pi(n)\int_1^\infty \phi\left(\frac{nt}{\sqrt{Q}}\right)t^s\frac{dt}{t} =$$

$$= \sum_{n=1}^\infty \lambda_\pi(n)\int_{\frac{n}{\sqrt{Q}}}^\infty \phi\left(t'\right)\left(\frac{\sqrt{Q}t'}{n}\right)^s = \sum_{n=1}^\infty \lambda_\pi(n)G_s\left(\frac{n}{\sqrt{Q}}\right).$$

Thus,

$$\Lambda(\pi, s) = \sum_{n=1}^\infty \lambda_\pi(n)G_s\left(\frac{n}{\sqrt{Q}}\right) - \sum_{n=1}^\infty \lambda_\pi(n)G_{1-s}\left(\frac{n}{\sqrt{Q}}\right)$$

is the desired expansion. From here, we obtain a formula for the k-th derivative

$$\frac{\partial^k}{\partial s^k}\Lambda(\pi, s) = \sum_{n=1}^\infty \lambda_\pi(n)\frac{\partial^k}{\partial s^k}G_s\left(\frac{n}{\sqrt{Q}}\right) - \sum_{n=1}^\infty \lambda_\pi(n)\frac{\partial^k}{\partial s^k}G_{1-s}\left(\frac{n}{\sqrt{Q}}\right).$$

The computation of the derivatives of $G_s(x)$ is explained in [Dok04, §3-5].

A.4 Asymptotic estimates on the canonical heights $\widehat{h}(y_c)$

In this section we provide an asymptotic bound on the canonical height $\widehat{h}(y_c)$ by using convexity bounds on the special values of the automorphic L-functions $L(\pi, s)$ defined in Section A.1. We only outline the basic techniques used to prove the asymptotic bounds and refer the reader to [Jet] for the complete details. Asymptotic bounds on heights of Heegner points are obtained in [RV], but these bounds are of significantly different type than ours. In our case, we fix the elliptic curve E and let the fundamental discriminant D and the conductor c of the ring class field both vary. We obtain the following:

Proposition A.2. *Fix the elliptic curve E and let the fundamental discriminant D and the conductor c vary. For any $\varepsilon > 0$ the following asymptotic bound holds*

$$\widehat{h}(y_c) \ll_{\varepsilon,f} h_D D^\varepsilon c^{2+\varepsilon},$$

where h_D is the class number of the quadratic imaginary field $K = \mathbb{Q}(\sqrt{-D})$. Moreover, the implied constant depends only on ε and the cusp form f.

One proves the proposition by combining the formula of Zhang with convexity bounds on special values of automorphic L-functions. The latter are conveniently expressed in terms of a quantity known as the *analytic conductor* associated to the automorphic representation π (see [Mic02, p.12]). It is a function $Q_\pi(t)$ over the real line, which is defined as

$$Q_\pi(t) = Q \cdot \prod_{i=1}^d (1 + |it - \mu_{\pi,i}|), \ \forall t \in \mathbb{R},$$

where $\mu_{\pi,i}$ are obtained from the gamma factor

$$L_\infty(\pi, s) = \prod_{i=1}^d \Gamma_\mathbb{R}(s - \mu_{\pi,i}), \ \Gamma_\mathbb{R}(s) = \pi^{-s/2}\Gamma(s/2).$$

In our situation for $\pi = \pi_{f\otimes\theta_\chi}$, $d = 4$ and $\mu_{\pi,1} = \mu_{\pi,2} = 0$, $\mu_{\pi,3} = \mu_{\pi,4} = 1$ (see [Mic02, §1.1.1] and [Ser70, §3] for discussions of local factors at archimedian places). Moreover, we let $Q_\pi = Q_\pi(0)$.

The main idea is to prove that for a fixed f, $|L'(\pi_{f\otimes\theta_\chi}, 1/2)| \ll_{\varepsilon,f} Q_{\pi_{f\otimes\theta_\chi}}^{1/4+\varepsilon}$, where the implied constant only depends on f and ε (and is independent of χ and the discriminant D). To establish the bound, we first prove an asymptotic bound for the L-function $L(\pi_{f\otimes\theta_\chi}, s)$ on the vertical line $\mathrm{Re}(s) =$

$1+\varepsilon$ by either using the Ramanujan-Petersson conjecture or a method of Iwaniec (see [Mic02, p.26]). This gives us the estimate $|L(\pi_{f\otimes\theta_\chi}, 1+\varepsilon+it)| \ll_{\varepsilon,f} Q_{\pi_{f\otimes\theta_\chi}}(t)^{\varepsilon}$. Then, by the functional equation for $L(\pi_{f\otimes\theta_\chi}, s)$ and Stirling's approximation formula, we deduce an upper bound for the L-function on the vertical line $\mathrm{Re}(s) = -\varepsilon$, i.e., $|L(\pi_{f\otimes\theta_\chi}, -\varepsilon+t)| \ll_{\varepsilon,f} Q_{\pi_{f\otimes\theta_\chi}}(t)^{1/2+\varepsilon}$. Next, we apply Phragmen-Lindelöf's convexity principle (see [IK04, Thm.5.53]) to obtain the bound $|L(\pi_{f\otimes\theta_\chi}, 1/2+it)| \ll_{\varepsilon,f} Q_{\pi}(t)^{1/4+\varepsilon}$ (also known as *convexity bound*). Finally, by applying Cauchy's integral formula for a small circle centered at $s = 1/2$, we obtain the asymptotic estimate $|L'(\pi_{f\otimes\theta_\chi}, 1/2)| \ll_{\varepsilon,f} Q_{\pi_{f\otimes\theta_\chi}}^{1/4+\varepsilon}$. Since $Q = N^2 d_c^2 = N^2 D^2 c^4$ in our situation and since $[K[c] : K] = h_D \prod_{\ell|c}(\ell+1)$, Zhang's formula (Theorem A.1) and equation (3) imply that for any $\varepsilon > 0$,

$$\widehat{h}(y_c) \ll_{\varepsilon,f} h_D D^{\varepsilon} c^{2+\varepsilon}.$$

Remark 9. In the above situation (the Rankin-Selberg L-function of two cusp forms of levels N and $d_c = c^2 D$), one can even prove a subconvexity bound $|L'(\pi_{f\otimes\theta_\chi}, 1/2)| \ll_f D^{1/2-1/1057} c^{1-2/1057}$, where the implied constant depends only on f and is independent of χ (see [Mic04, Thm.2]). Yet, the proof relies on much more involved analytic number theory techniques than the convexity principle, so we do not discuss it here.

A.5 Height difference bounds and the main estimates

To estimate $h(y_c)$ we need a bound on the difference between the canonical and the logarithmic heights. Such a bound has been established in [Sil90] and [CPS06] and is effective.

Let F be a number field. For any non-archimedian place v of K, let $E^0(F_v)$ denote the points of $E(F_v)$ which specialize to the identity component of the Néron model of E over the ring of integers $\mathcal{O}_v$ of F_v. Moreover, let $n_v = [F_v : \mathbb{Q}_v]$ and let M_F^∞ denote the set of all archimedian places of F. A slightly weakened (but easier to compute) bounds on the height difference are provided by the following result of [CPS06, Thm.2]

Theorem A.3 (Cremona-Prickett-Siksek). *Let $P \in E(F)$ and suppose that $P \in E^0(F_v)$ for every non-archimedian place v of F. Then*

$$\frac{1}{3[F:\mathbb{Q}]} \sum_{v\in M_F^\infty} n_v \log \delta_v \leqslant h(P) - \widehat{h}(P) \leqslant \frac{1}{3[F:\mathbb{Q}]} \sum_{v\in M_F^\infty} n_v \log \varepsilon_v,$$

where ε_v and δ_v are defined in [CPS06, §2].

Remark 10. All of the points y_c in our particular examples satisfies the condition $y_c \in E^0(K[c]_v)$ for all non-archimedian places v of $K[c]$. Indeed, according to [GZ86, §III.3] (see also [Jet07, Cor.3.2]) the point y_c lies in $E^0(K[c]_v)$ up to a rational torsion point. Since $E(\mathbb{Q})_{\mathrm{tors}}$ is trivial for all the curves that we are considering, the above proposition is applicable. In general, one does not need this assumption in order to compute height difference bounds (see [CPS06, Thm.1] for the general case).

Remark 11. A method for computing ε_v and δ_v up to arbitrary precision for real and complex archimedian places is provided in [CPS06, §7-9].

Acknowledgements

We are indebted to Christophe Cornut for sharing his thoughts on the Kolyvagin's conjecture. We would like to thank Jan Nekovář for his helpful discussions on height bounds, Stephen D. Miller and Mark Watkins for pointing out analytic methods for estimating special values of derivatives of automorphic L-functions, Henri Darmon and David Jao for useful conversations on Heegner points computations, and Ken Ribet for reading the preliminary draft and for useful conversations.

References

BCDT01 C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

BFH90 D. Bump, S. Friedberg, and J. Hoffstein, *Eisenstein series on the metaplectic group and nonvanishing theorems for automorphic L-functions and their derivatives*, Ann. of Math. (2) **131** (1990), no. 1, 53–127.

Coh07 H. Cohen, *Number theory II: Analytic and modern methods*, Graduate Texts in Mathematics, vol. 240, Springer, 2007.

Cor02 C. Cornut, *Non-trivialité des points de Heegner*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 12, 1039–1042.

CPS06 J. E. Cremona, M. Prickett, and Samir Siksek, *Height difference bounds for elliptic curves over number fields*, J. Number Theory **116** (2006), no. 1, 42–68.

Cre J. E. Cremona, *Tables of Elliptic Curves,* `http://www.maths.nott.ac.uk/personal/jec/ftp/data/`.

Del02 C. Delauney, *Formes modulaires et invariants de courbes elliptiques définies sur* **Q**, Université Bordeaux I, PhD thesis (2002).

Dok04 T. Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149.

EJL06 K. Eisentraeger, D. Jetchev, and K. Lauter, *On the computation of the Cassels pairing for certain Kolyvagin classes in the Shafarevich-Tate group*, preprint (2006).

Elk94 N. D. Elkies, *Heegner point computations*, Algorithmic number theory (Ithaca, NY, 1994), Springer, Berlin, 1994, pp. 122–133.

GJP$^+$05 G. Grigorov, A. Jorza, S. Patrikis, C. Patrascu, and W. Stein, *Verification of the Birch and Swinnerton-Dyer Conjecture for Specific Elliptic Curves*, (Submitted) `http://wstein.org/papers/bsdalg/` (2005).

Gro84 B. Gross, *Heegner points on $X_0(N)$*, Modular forms (Durham, 1983), Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., Horwood, Chichester, 1984, pp. 87–105.

Gro91 B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 235–256.

GZ86 B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

IK04 H. Iwaniec and E. Kowalski, *Analytic Number Theory*, vol. 53, American Mathematical Society, 2004.

Jet D. Jetchev, *Asymptotic heights of Heegner points over ring class fields*, in preparation.

Jet07 ———, *Global divisibility of Heegner points and Tamagawa numbers*, preprint (2007).

Kol90 V. A. Kolyvagin, *Euler systems*, The Grothendieck Festschrift, Vol. II, Birkhäuser Boston, Boston, MA, 1990, pp. 435–483.

Kol91a ———, *On the structure of Selmer groups*, Math. Ann. **291** (1991), no. 2, 253–259. MR 93e:11073

Kol91b V. A. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 429–436.

Kol91c ———, *On the structure of Shafarevich-Tate groups*, Algebraic geometry (Chicago, IL, 1989), Springer, Berlin, 1991, pp. 94–121.

McC91 W. G. McCallum, *Kolyvagin's work on Shafarevich-Tate groups*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 295–316.

Mic02 P. Michel, *Analytic number theory and families of L-functions*, `www.math.univ-montp2.fr/~michel/fichierdvi/Parkcitylectures.pdf` (2002).

Mic04 ———, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. (2) **160** (2004), no. 1, 185–236.

Mil86 J. S. Milne, *Arithmetic duality theorems*, Academic Press Inc., Boston, Mass., 1986.

MM97 M. R. Murty and V. K. Murty, *Non-vanishing of L-functions and applications*, Birkhäuser Verlag, Basel, 1997.

Nek01 Jan Nekovář, *On the parity of ranks of Selmer groups. II*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), no. 2, 99–104.

RV G. Ricotta and T. Vidick, *Hauteurs asymptotique des points de Heegner*, preprint.

Ser70 J.-P. Serre, *Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures)*, Séminaire Delange-Pisot-Poitou. Théorie des nombres 11, no. 2, Exposé No. 19 (1969-70).

Shi71 G. Shimura, *On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields*, Nagoya Math. J. **43** (1971), 199–208.

Sil90 Joseph H. Silverman, *The difference between the Weil height and the canonical height on elliptic curves*, Math. Comp. **55** (1990), no. 192, 723–743.

Sil92 J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

Sil94 ______, *Advanced topics in the arithmetic of elliptic curves*, Springer-Verlag, New York, 1994.

Tat67 J. T. Tate, *$p-$ divisible groups.*, Proc. Conf. Local Fields (Driebergen, 1966), Springer, Berlin, 1967, pp. 158–183.

Wat04 Mark Watkins, *Some remarks on Heegner point computations*, Preprint, 2004.

Zha01 S. W. Zhang, *Gross-Zagier formula for GL_2*, Asian J. Math. **5** (2001), no. 2, 183–290.

Dimitar Jetchev jetchev@math.berkeley.edu
Department of Mathematics, University of California at Berkeley, Berkeley, CA 94720

Kristin Lauter klauter@microsoft.com
Microsoft Research, One Microsoft Way, Redmond, WA 98052

William Stein wstein@math.washington.edu
Department of Mathematics, University of Washington, Seattle, WA 98195

31 Book – The Birch and Swinnerton-Dyer Conjecture, a Computational Approach

The Birch and Swinnerton-Dyer Conjecture, a Computational Approach

William A. Stein

Department of Mathematics, University of Washington
E-mail address: wstein@math.washington.edu

1991 *Mathematics Subject Classification.* Primary 11;
Secondary 11-04

Key words and phrases. abelian varieties, computation, elliptic curves, Hecke operators, modular curves, modular forms, modular symbols, Manin symbols, number theory

Abstract.

Contents

Preface

This is an introductory graduate-level textbook about the Birch and Swinnerton-Dyer conjecture and modern approaches to the arithmetic of elliptic curves.

Other very relevant books: Darmon's *Rational Points on Modular Elliptic Curves*.

Acknowledgements. I would like to acknowledge the partial support of NSF Grant DMS 05-55776.

Notation and Conventions.

Chapter 1

The BSD Rank Conjecture

This chapter explains the conjecture that Birch and Swinnerton-Dyer made about ranks of elliptic curves (the BSD rank conjecture).

1.1. Statement of the BSD Rank Conjecture

An excellent reference for this section is Andrew Wiles's Clay Math Institute paper [**Wil00**]. The reader is also strongly encouraged to look Birch's original paper [**Bir71**] to get a better sense of the excitement surrounding this conjecture, as exemplified in the following quote:

> "I want to describe some computations undertaken by myself and Swinnerton-Dyer on EDSAC by which we have calculated the zeta-functions of certain elliptic curves. As a result of these computations we have found an analogue for an elliptic curve of the Tamagawa number of an algebraic group; and conjectures (due to ourselves, due to Tate, and due to others) have proliferated."

An *elliptic curve* E over a field K is the projective closure of the zero locus of a nonsingular affine curve

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.1.1}$$

where $a_1, a_2, a_3, a_4, a_6 \in K$. There is a simple algebraic condition on the a_i that ensures that (1.1.1) defines a nonsingular curve (see, e.g., [**Sil92**]).

An elliptic curve E has genus 1, and the set of points on E has a natural structure of *abelian group*, with identity element the one extra projective point at ∞. Again, there are simple algebraic formulas that, given two points P and Q on an elliptic curve, produce a third point $P+Q$ on the elliptic curve. Moreover, if P and Q both have coordinates in K, then so does $P+Q$. The *Mordell-Weil group*

$$E(K) = \{ \text{ points on } E \text{ with coordinates in } K \ \}$$

of E over K plays a central role in this book.

In the 1920s, Mordell proved that if $K=\mathbb{Q}$, then $E(\mathbb{Q})$ is finitely generated, and soon after Weil proved that $E(K)$ is finitely generated for any number field K, so

$$E(K) \approx \mathbb{Z}^r \oplus T, \tag{1.1.2}$$

where T is a finite group. Perhaps the chief invariant of an elliptic curve E over a number field K is the *rank*, which is the number r in (1.1.2).

Fix an elliptic curve E over $\mathbb{Q}$. For all but finitely many prime numbers p, the equation (1.1.1) reduces modulo p to define an elliptic curve over the finite field $\mathbb{F}_p$. The primes that must be excluded are exactly the primes that divide the discriminant Δ of (1.1.1).

As above, the set of points $E(\mathbb{F}_p)$ is an abelian group. This group is finite, because it is contained in the set $\mathbb{P}^2(\mathbb{F}_p)$ of rational points in the projective plane. Moreover, since it is the set of points on a (genus 1) curve, a theorem of Hasse implies that

$$|p+1-\#E(\mathbb{F}_p)| \leq 2\sqrt{p}.$$

The error terms

$$a_p = p+1-\#E(\mathbb{F}_p)$$

play a central role in almost everything in this book. We next gather together the error terms into a single "generating function":

$$\tilde{L}(E,s) = \prod_{p\nmid\Delta}\left(\frac{1}{1-a_pp^{-s}+p^{1-2s}}\right).$$

The function $\tilde{L}(E,s)$ defines a complex analytic function on some right half plane $\mathrm{Re}(s) > \frac{3}{2}$.

A deep theorem of Wiles et al. [**Wil95, BCDT01**], which many consider the crowning achievement of 1990s number theory, implies that $\tilde{L}(E,s)$ can be analytically continued to an analytic function on all $\mathbb{C}$. This implies that $\tilde{L}(E,s)$ has a Taylor series expansion about $s=1$:

$$\tilde{L}(E,s) = c_0 + c_1(s-1) + c_2(s-1)^2 + \cdots$$

Define the *analytic rank* r_{an} of E to be the order of vanishing of $\tilde{L}(E,s)$ as $s=1$, so

$$\tilde{L}(E,s) = c_{r_{\mathrm{an}}}(s-1)^{r_{\mathrm{an}}} + \cdots .$$

The definitions of the analytic and Mordell-Weil ranks could not be more different – one is completely analytic and the other is purely algebraic.

Conjecture 1.1 (Birch and Swinnerton-Dyer Rank Conjecture)**.** *Let E be an elliptic curve over $\mathbb{Q}$. Then the algebraic and analytic ranks of E are the same.*

This problem is extremely difficult. The conjecture was made in the 1960s, and hundreds of people have thought about it for over 4 decades. The work of Wiles et al. on modularity in late 1999, combined with earlier work of Gross, Zagier, and Kolyvagin, and many others proves the following partial result toward the conjecture.

Theorem 1.2. *Suppose E is an elliptic curve over $\mathbb{Q}$ and that $r_{\mathrm{an}} \leq 1$. Then the algebraic and analytic ranks of E are the same.*

In 2000, Conjecture 1.1 was declared a million dollar millenium prize problem by the Clay Mathematics Institute, which motivated even more work, conferences, etc., on the conjecture. Since then, to the best of my knowledge, not a single new result directly about Conjecture 1.1 has been proved[1]. The class of curves for which we know the conjecture is still the set of curves over $\mathbb{Q}$ with $r_{\mathrm{an}} \leq 1$, along with a finite set of individual curves on which further computer calculations have been performed (by Cremona, Watkins, myself, and others).

> *"A new idea is needed."*
> – Nick Katz on BSD, at a 2001 Arizona Winter School

And another quote from Bertolini-Darmon (2001):

> "The following question stands as the ultimate challenge concerning the Birch and Swinnerton-Dyer conjecture for elliptic curves over $\mathbb{Q}$: *Provide evidence for the Birch and Swinnerton-Dyer conjecture in cases where* $\mathrm{ord}_{s=1} L(E,s) > 1$."

1.2. The BSD Rank Conjecture Implies that $E(\mathbb{Q})$ is Computable

Proposition 1.3. *Let E be an elliptic curve over $\mathbb{Q}$. If Conjecture 1.1 is true, then there is an algorithm to compute the rank of E.*

Proof. By naively searching for points in $E(\mathbb{Q})$ we obtain a lower bound on r, which is closer and closer to the true rank r, the longer we run the search. At some point this lower bound will equal r, but without using further information we do not know when that will occur.

[1] Much interesting new work has been done on related conjectures and problems.

As explained, e.g., in [**Cre97**] (see also [**Dok04**]), we can for any k compute $L^{(k)}(E,1)$ to any desired precision. Such computations yield upper bounds on r_{an}. In particular, if we compute $L^{(k)}(E,1)$ and it is nonzero (to the precision of our computation), then $r_{\mathrm{an}} \leq k$. Eventually this method will also converge to give an upper bound on r_{an}, though again without further information we do not know when our computed upper bound on r_{an} equals to the true value of r_{an}.

Since we are assuming that Conjecture 1.1 is true, we know that $r = r_{\mathrm{an}}$, hence at some point the lower bound on r computed using point searches will equal the upper bound on r_{an} computed using the L-series. At this point, by Conjecture 1.1, we know the true value of r. □

Next we show that given the rank r, the full group $E(\mathbb{Q})$ is computable. The issue is that what we did above might have only computed a subgroup of finite index. The argument below follows [**Cre97**, §3.5] closely.

The *naive height* $h(P)$ of a point $P = (x,y) \in E(\mathbb{Q})$ is

$$h(P) = \log(\max(\mathrm{numer}(x), \mathrm{denom}(x))).$$

The *Néron-Tate canonical height* of P is

$$\hat{h}(P) = \lim_{n\to\infty} \frac{h(2^n P)}{4^n}.$$

Note that if P has finite order then $\hat{h}(P) = 0$. Also, a standard result is that the *height pairing*

$$\langle P, Q\rangle = \frac{1}{2}\left(\hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q)\right)$$

defines a nondegenerate real-valued quadratic form on $E(\mathbb{Q})/_{\mathrm{tor}}$ with discrete image.

Lemma 1.4. *Let $B > 0$ be a positive real number such that*

$$S = \{P \in E(\mathbb{Q}) \,:\, \hat{h}(P) \leq B\}$$

contains a set of generators for $E(\mathbb{Q})/2E(\mathbb{Q})$. Then S generates $E(\mathbb{Q})$.

Proof. Let A be the subgroup of $E(\mathbb{Q})/_{\mathrm{tor}}$ generated by the points in S. Suppose for the sake of contradiction that A is a proper subgroup. Then there is $Q \in E(\mathbb{Q}) \setminus A$ with $\hat{h}(Q)$ minimal, since $\hat{h}$ takes a discrete set of values. Since S contains generators for $E(\mathbb{Q})/2E(\mathbb{Q})$, there is an element $P \in S$ that is congruent to Q modulo $2E(\mathbb{Q})$, i.e., so that

$$Q = P + 2R,$$

for some $R \in E(\mathbb{Q})$. We have $R \notin A$ (since otherwise Q would be in A), so $\hat{h}(R) \geq \hat{h}(Q)$ by minimality. Finally, since $\hat{h}$ is quadratic and nonnegative, we have

$$\begin{aligned}\hat{h}(P) &= \frac{1}{2}\left(\hat{h}(Q+P) + \hat{h}(Q-P) - \hat{h}(Q)\right)\\ &\geq \frac{1}{2}\hat{h}(2R) - \hat{h}(Q)\\ &= 2\hat{h}(R) - \hat{h}(Q) \geq \hat{h}(Q) > B.\end{aligned}$$

(Here we use that $\hat{h}(P) = \langle P, P\rangle$ and use properties of a bilinear form.) □

Proposition 1.5. *Let E be an elliptic curve over $\mathbb{Q}$. If Conjecture 1.1 is true, then there is an algorithm to compute $E(\mathbb{Q})$.*

Proof. By Proposition 1.3 we can compute the rank r of $E(\mathbb{Q})$. Note that we can also trivially compute the subgroup $E(\mathbb{Q})[2]$ of elements of order 2 in $E(\mathbb{Q})$, since if E is given by $y^2 = x^3+ax+b$, then this subgroup is generated by points (α, β), where α is a rational root of $x^3 + ax + b$. Thus we can compute $s = \dim_{\mathbb{F}_2} E(\mathbb{Q})/2E(\mathbb{Q})$, since it is equal to $r + \dim E(\mathbb{Q})[2]$.

Run any search for points in $E(\mathbb{Q})$ and use that $\hat{h}$ is a nondegenerate quadratic form to find independent points $P_1, \ldots, P_r$ of infinite order. It is easy to check whether a point P is twice another point (just solve a relatively simple algebraic equation). Run through all subsets of the points P_i, and if any subset of the P_i sums to $2Q$ for some point $Q \in E(\mathbb{Q})$, then we replace one of the P_i by Q and decrease the index of our subgroup in $E(\mathbb{Q})$ by a factor of 2. Because $E(\mathbb{Q})$ is a finitely generated group, after a finite number of steps (and including the 2-torsion points found above) we obtain independent points $P_1, \ldots, P_s$ that generate $E(\mathbb{Q})/2E(\mathbb{Q})$.

Let C the the explicit bound of Cremona-Pricket-Siksek on the difference between the naive and canonical height (i.e., for any $P \in E(\mathbb{Q})$, we have $|h(P) - \hat{h}(P)| < C$). Let

$$B = \max\{\hat{h}(P_1), \ldots, \hat{h}(P_s)\}.$$

Then by a point search up to naive height $B + C$, we compute a set that contains the set S in Lemma 1.4. This set then contains generators for $E(\mathbb{Q})$, hence we have computed $E(\mathbb{Q})$.

□

1.3. The Complex L-series $L(E, s)$

In Section 1.1 we defined a function $\tilde{L}(E, s)$, which encoded information about $E(\mathbb{F}_p)$ for all but finitely many primes p. In this section we define the function $L(E, s)$, which includes information about all primes, and the function $\Lambda(E, s)$ that also includes information "at infinity".

Let E be an elliptic curve over $\mathbb{Q}$ defined by *a minimal Weierstrass equation*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{1.3.1}$$

A minimal Weierstrass equation in one for which the a_i are all integers and the discriminant $\Delta \in \mathbb{Z}$ is minimal amongs all discriminants of Weierstrass equations for E (again, see [**Sil92**] for the definition of the discriminant of a Weierstrass equation, and also for an explicit description of the allowed transformations of a Weierstrass equation).

For each prime number $p \nmid \Delta$, the equation (1.3.1) reduces modulo p to define an elliptic $E_{\mathbb{F}_p}$ over the finite field $\mathbb{F}_p$. Let

$$a_p = p + 1 - \#E(\mathbb{F}_p).$$

For each prime $p \mid \Delta$, we use the following recipe to define a_p. If the *singular* curve $E_{\mathbb{F}_p}$ has a cuspidal singularity, e.g., is $y^2 = x^3$, then let $a_p = 0$. If it has a a nodal singularity, e.g., like $y^2 = x^3 + x^2$, let $a_p = 1$ if the slope of the tangent line at the singular point is in $\mathbb{F}_p$ and let $a_p = -1$ if the slope is not in $\mathbb{F}_p$. Summarizing:

$$a_p = \begin{cases} 0 & \text{if the reduction is cuspidal (``additive''),} \\ 1 & \text{if the reduction is nodal and tangent line is } \mathbb{F}_p\text{-rational (``split multiplicative'')} \\ -1 & \text{if the reduction is nodal and tangent line is not } \mathbb{F}_p\text{-rational (``non-split multiplicative'} \end{cases}$$

Even in the cases when $p \mid \Delta$, we still have

$$a_p = p + 1 - \#E(\mathbb{F}_p).$$

When E has additive reduction, the nonsingular points form a group isomorphic to $(\mathbb{F}_p, +)$, and there is one singular point, hence $p+1$ points, so

$$a_p = p + 1 - (p + 1))) = 0.$$

When E has split multiplicative reduction, there is 1 singular point plus the number of elements of a group isomorphic to $(\mathbb{F}_p^*, \times)$, so $1 + (p - 1) = p$ points, and

$$a_p = p + 1 - p = 1.$$

When E has non-split multiplicative reduction, there is 1 singular point plus the number of elements of a group isomorphic $(\mathbb{F}_{p^2}^*/\mathbb{F}_p^*, \times)$, i.e., $p+2$ points, and

$$a_p = p + 1 - (p + 2) = -1.$$

The definition of the full L-function of E is then

$$L(E, s) = \prod_{p \mid \Delta} \frac{1}{1 - a_p p^{-s}} \cdot \prod_{p \nmid \Delta} \frac{1}{1 - a_p p^{-s} + p \cdot p^{-2s}}. = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

If in addition we add in a few more analytic factors to the L-function we obtain a function $\Lambda(E, s)$ that satisfies a remarkably simple functional equation. Let

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$$

be the Γ-*function* (e.g., $\Gamma(n) = (n-1)!$), which defines a meromorphic function on $\mathbb{C}$, with poles at the non-positive integers.

Theorem 1.6 (Hecke, Wiles et al.). *There is a unique positive integer $N = N_E$ and sign $\varepsilon = \varepsilon_E \in \{\pm 1\}$ such that the function*

$$\Lambda(E, s) = N^{s/2} \cdot (2\pi)^{-s} \cdot \Gamma(s) \cdot L(E, s)$$

extends to a complex analytic function on all $\mathbb{C}$ that satisfies the functional equation

$$\Lambda(E, 2-s) = \varepsilon \cdot \Lambda(E, s), \tag{1.3.2}$$

for all $s \in \mathbb{C}$.

Proof. Wiles et al. prove that $L(E, s)$ is the L-series attached to a modular form (see Section **??** below), and Hecke proved that the L-series of a modular form analytically continues and satisfies the given functional equation. □

The integer $N = N_E$ is called the *conductor* of E and $\varepsilon = \varepsilon_E$ is called the *sign in the functional equation* for E or the *root number* of E. One can prove that the primes that divide N are the same as the primes that divide Δ. Moreover, for $p \geq 5$, we have that

$$\mathrm{ord}_p(N) = \begin{cases} 0, & \text{if } p \nmid \Delta, \\ 1, & \text{if } E \text{ has multiplicative reduction at } p, \text{ and} \\ 2, & \text{if } E \text{ has additive reduction at } p. \end{cases}$$

There is a geometric algorithm called Tate's algorithm that computes N in all cases and ε.

Example 1.7. Consider the elliptic curve E defined by

$$y^2 + y = x^3 + 50x + 31.$$

The above Weierstrass equation is minimal and has discriminant

$$-1 \cdot 5^6 \cdot 7^2 \cdot 11.$$

```
sage: e = EllipticCurve('1925d'); e
Elliptic Curve defined by y^2 + y = x^3 + 50*x + 31 over Rational Field
sage: e.is_minimal()
True
sage: factor(e.discriminant())
-1 * 5^6 * 7^2 * 11
```

At 5 the curve has additive reduction so $a_5 = 0$. At 7 the curve has split multiplicative reduction so $a_7 = 1$. At 11 the curve has nonsplit multiplicative reduction, so $a_{11} = -1$. Counting points for $p = 2, 3$, we find that

$$L(E, s) = \frac{1}{1^{-s}} + \frac{3}{3^{-s}} + \frac{-2}{4^{-s}} + \frac{1}{7^{-s}} + \frac{6}{9^{-s}} + \frac{-1}{11^{-s}} + \frac{-6}{12^{-s}} + \cdots$$

```
sage: [e.ap(p) for p in primes(14)]
[0, 3, 0, 1, -1, 4]
```

Corollary 1.8. *Let E be an elliptic curve over $\mathbb{Q}$, let $\varepsilon \in \{1, -1\}$ be the sign in the functional equation* (1.3.2)*, and let $r_{E,\mathrm{an}} = \mathrm{ord}_{s=1} L(E, s)$. Then*

$$\varepsilon = (-1)^{r_{E,\mathrm{an}}}.$$

Proof. Because $\Gamma(1) = 1$, we have $\mathrm{ord}_{s=1} L(E, s) = \mathrm{ord}_{s=1} \Lambda(E, s)$. It thus sufficies to prove the corollary with $L(E, s)$ replaced by $\Lambda(E, s)$. Note that $r = r_{E,\mathrm{an}}$ is the minimal integer $r \geq 0$ such that $\Lambda^{(r)}(E, 1) \neq 0$. By repeated differentiation, we see that for any integer $k \geq 0$, we have

$$(-1)^k \Lambda^{(k)}(E, 2 - s) = \varepsilon \cdot \Lambda^{(k)}(s). \tag{1.3.3}$$

Setting $s = 1$ and $k = r$, and using that $\Lambda^{(r)}(E, 1) \neq 0$, shows that $(-1)^r = \varepsilon$, as claimed. □

Conjecture 1.9 (The Parity Conjecture)**.** *Let E be an elliptic curve over $\mathbb{Q}$, let $r_{E,\mathrm{an}}$ be the analytic rank and $r_{E,\mathrm{alg}}$ be the algebraic rank. Then*

$$r_{E,\mathrm{alg}} \equiv r_{E,\mathrm{an}} \pmod 2.$$

Jan Nekovar has done a huge amount of work toward Conjecture 1.9; in particular, he proves it under the (as yet unproved) hypothesis that $Ш(E)$ is finite (see Section 2.2 below).

1.4. Computing $L(E, s)$

In this section we briefly describe one way to evaluate $L(E, s)$, for s real. See [**Dok04**] for a more sophisticated analysis of computing $L(E, s)$ and its Taylor expansion for any complex number s.

Theorem 1.10 (Lavrik). *We have the following rapidly-converging series expression for $L(E,s)$, for any complex number s:*

$$L(E,s) = N^{-s/2} \cdot (2\pi)^s \cdot \Gamma(s)^{-1} \cdot \sum_{n=1}^{\infty} a_n \cdot (F_n(s-1) - \varepsilon F_n(1-s))$$

where

$$F_n(t) = \Gamma\left(t+1, \frac{2\pi n}{\sqrt{N}}\right) \cdot \left(\frac{\sqrt{N}}{2\pi n}\right)^{t+1},$$

and

$$\Gamma(z,\alpha) = \int_{\alpha}^{\infty} t^{z-1} e^{-t} dt$$

is the incomplete Γ-function.

Theorem 1.10 above is a special case of a more general theorem that gives rapidly converging series that allow computation of any Dirichlet series $\sum a_n n^s$ that meromorphically continues to the whole complex plane and satisfies an appropriate functional equation. For more details, see [**Coh00**, §10.3], especially Exercise 24 on page 521 of [**Coh00**].

1.4.1. Approximating the Rank. Fix an elliptic curve E over $\mathbb{Q}$. The usual method to *approximate* the rank is to find a series that rapidly converges to $L^{(r)}(E,1)$ for $r = 0,1,2,3,\ldots$, then compute $L(E,1)$, $L'(E,1)$, $L^{(2)}(E,1)$, etc., until one appears to be nonzero. Note that half of the $L^{(k)}(E,1)$ are automatically 0 because of equation (1.3.3). For more details, see [**Cre97**, §2.13] and [**Dok04**].

In this section, we describe a slightly different method, which only uses Theorem 1.10 and the definition of the derivative.

Proposition 1.11. *Write*

$$L(E,s) = c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots .$$

with $c_r \neq 0$. Then

$$\lim_{s\to 1} (s-1) \cdot \frac{L'(E,s)}{L(E,s)} = r.$$

Proof. Setting $L(s) = L(E,s)$, we have

$$\begin{aligned}\lim_{s\to 1} (s-1) \cdot \frac{L'(s)}{L(s)} &= \lim_{s\to 1} (s-1) \cdot \frac{rc_r(s-1)^{r-1} + (r+1)c_{r+1}(s-1)^r + \cdots}{c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots} \\ &= r \cdot \lim_{s\to 1} \frac{c_r(s-1)^r + \frac{(r+1)}{r}c_{r+1}(s-1)^{r+1} + \cdots}{c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots} \\ &= r.\end{aligned}$$

□

Thus the rank r is the limit as $s \to 1$ of a certain (smooth) function. We know this limit is an integer. But, for example, for the rank 4 curve

$$y^2 + xy = x^3 - x^2 - 79x + 289 \tag{1.4.1}$$

of conductor 234446 nobody has succeeded in proving that this integer limit is 4. (We can prove that the limit is either 2 or 4 by using the functionality equation (1.3.2) to show that the order of vanishing is even, then verifying by computation that $L^{(4)}(E,1) = 214.65233\ldots \neq 0$.)

Using the definition of derivative, we approximate $(s-1)\frac{L'(s)}{L(s)}$ as follows. For $|s-1|$ small, we have

$$\begin{aligned}(s-1)\frac{L'(s)}{L(s)} &= \frac{s-1}{L(s)} \cdot \lim_{h\to 0} \frac{L(s+h)-L(s)}{h} \\ &\approx \frac{s-1}{L(s)} \cdot \frac{L(s+(s-1)^2)-L(s)}{(s-1)^2} \\ &= \frac{L(s^2-s+1)-L(s)}{(s-1)L(s)}\end{aligned}$$

In fact, we have

$$\lim_{s\to 1}(s-1)\cdot\frac{L'(s)}{L(s)} = \lim_{s\to 1}\frac{L(s^2-s+1)-L(s)}{(s-1)L(s)}.$$

We can use this formula in SAGE to "approximate" r. First we start with a curve of rank 2.

```
sage: e = EllipticCurve('389a'); e.rank()
2
sage: L = e.Lseries_dokchitser()
sage: def r(e,s): L1=L(s); L2=L(s^2-s+1); return (L2-L1)/((s-1)*L1)
sage: r(e,1.01)
2.00413534247395
sage: r(e,1.001)
2.00043133754756
sage: r(e,1.00001)
2.00000433133371
```

Next consider the curve $y^2 + xy = x^3 - x^2 - 79x + 289$ of rank 4:

```
sage: e =  EllipticCurve([1, -1, 0, -79, 289])
sage: e.rank()
4
sage: L = e.Lseries_dokchitser(100)
sage: def r(e,s): L1=L(s); L2=L(s^2-s+1); return (L2-L1)/((s-1)*L1)
sage: R = RealField(100)
sage: r(e,R('1.01'))
4.0212949184444018810727106489
sage: r(e,R('1.001'))
4.0022223745190806421850637523
sage: r(e,R('1.00001'))
4.0000223250026401574120263050
sage: r(e,R('1.000001'))
4.0000022325922257758141597819
```

It certainly looks like $\lim_{s\to 1} r(s) = 4$. We know that $\lim_{s\to 1} r(s) \in \mathbb{Z}$, and if only there were a good way to bound the error we could conclude that the limit is 4. But this has stumped people for years, and probably it is nearly impossible without a deep result that somehow interprets $L''(E, 1)$ in a completely different way.

1.5. The p-adic $\mathcal{L}$-series

Fix[2] an elliptic curve E defined over $\mathbb{Q}$. We say a prime p is a prime of *good ordinary reduction* for E if $p \nmid N_E$ and $a_p \not\equiv 0 \pmod{p}$. The Hasse bound, i.e., that $|a_p| < 2\sqrt{p}$ on implies that if $p \geq 5$ then ordinary at p is the same as $a_p \neq 0$.

In this section, we define for each odd prime number p of good ordinary reduction for E a p-adic L-function $L_p(E, T)$. This is a p-adic analogue of the complex L-function $L(E, s)$ about which there are similar analogue of the BSD conjecture.

1.5.1. Hensel's lemma and the Teichmuller lift. The following standard lemma is proved by Newton iteration.

Lemma 1.12 (Hensel)**.** *If $f \in \mathbb{Z}_p[x]$ is a polynomial and $\beta \in \mathbb{Z}/p\mathbb{Z}$ is a multiplicity one root of $\overline{f}$, then there is a unique lift of β to a root of f.*

For example, consider the polynomial $f(x) = x^{p-1} - 1$. By Fermat's little theorem, it has $p-1$ distinct roots in $\mathbb{Z}/p\mathbb{Z}$, so by Lemma 1.12 there are $p-1$ roots of $f(x)$ in $\mathbb{Z}_p$, i.e., all the $p-1$st roots of unity are elements of $\mathbb{Z}_p$. The *Teichmuller lift* is the map that sends any $\beta \in (\mathbb{Z}/p\mathbb{Z})^*$ to the unique $(p-1)$st root of unity in $\mathbb{Z}_p^*$ that reduces to it.

The *Teichmuller character* is the homomorphism

$$\tau : \mathbb{Z}_p^* \to \mathbb{Z}_p^*$$

obtained by first reducing modulo p, then sending an element to its Teichmuller lift. The 1-*unit projection* character is the homomorphism

$$\langle \bullet \rangle : \mathbb{Z}_p^* \to 1 + p\mathbb{Z}_p$$

given by

$$\langle x \rangle = \frac{x}{\tau(x)}.$$

1.5.2. Modular Symbol and Measures. Let

$$f_E(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z} \in S_2(\Gamma_0(N))$$

be the modular form associated to E, which is a holomorphic function on the extended upper half plane $\mathfrak{h} \cup \mathbb{Q} \cup \{\infty\}$. Let

$$\Omega_E = \int_{E(\mathbb{R})} \frac{dx}{2y + \underline{a}_1 x + \underline{a}_3} \in \mathbb{R}$$

[2] This section is based on correspondence with Robert Pollack and Koopa Koo.

be the real period associated to a minimal Weierstrass equation

$$y^2 + \underline{a}_1 xy + \underline{a}_3 y = x^3 + \underline{a}_2 x^2 + \underline{a}_4 x + \underline{a}_6$$

for E.

The *plus modular symbol map* associated to the elliptic curve E is the map $\mathbb{Q} \to \mathbb{Q}$ given by sending $r \in \mathbb{Q}$ to

$$[r] = [r]_E = \frac{2\pi i}{\Omega_E}\left(\int_r^{i\infty} f_E(z)dz + \int_{-r}^{i\infty} f_E(z)dz\right).$$

Question 1.13. Let E vary over all elliptic curve over $\mathbb{Q}$ and r over all rational numbers. Is the set of denominators of the rational numbers $[r]_E$ bounded? Thoughts: For a given curve E, the denominators are bounded by the order of the image in $E(\overline{\mathbb{Q}})$ of the cuspidal subgroup of $J_0(N)(\overline{\mathbb{Q}})$. It is likely one can show that if a prime ℓ divides the order of the image of this subgroup, then E admits a rational ℓ-isogeny. Mazur's theorem would then prove that the set of such ℓ is bounded, which would imply a "yes" answer to this question. Also, for any particular curve E, one can compute the cuspidal subgroup precisely, and hence bound the denominators of $[r]_E$.

Let a_p be the pth Fourier coefficient of E and note that the polynomial

$$x^2 - a_p x + p \equiv x(x - a_p) \pmod{p}$$

has distinct roots because p is an ordinary prime. Let α be the root of $x^2 - a_p x + p$ with $|\alpha|_p = 1$, i.e., the lift of the root a_p modulo p, which exists by Lemma 1.12.

Define a *measure on* $\mathbb{Z}_p^*$ by

$$\mu_E(a + p^n\mathbb{Z}_p) = \frac{1}{\alpha^n}\left[\frac{a}{p^n}\right] - \frac{1}{\alpha^{n+1}}\left[\frac{a}{p^{n-1}}\right].$$

That μ_E is a measure follows from the formula for the action of Hecke operators on modular symbols and that f_E is a Hecke eigenform. We will not prove this here[3].

1.5.3. The p-Adic L-function. Define the p-adic L-function as a function on characters

$$\chi \in \mathrm{Hom}(\mathbb{Z}_p^*, \mathbb{C}_p^*)$$

as follows. Send a character χ to

$$L_p(E, \chi) = \int_{\mathbb{Z}_p^*} \chi \, d\mu_E.$$

We will later make the integral on the right more precise, as a limit of Riemann sums (see Section 1.6).

[3] Add proof or good reference.

Remark 1.14. For any Dirichlet character $\chi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, let $L(E, \chi, s)$ be the entire L-function defined by the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{\chi(n)a_n}{n^s}.$$

The standard *interpolation property* of L_p is that for any primitive Dirichlet character χ of conductor p^n (for any n), we have[4]

$$L_p(E, \chi) = \begin{cases} p^n \cdot g(\chi) \cdot L(E, \bar{\chi}, 1)/\Omega_E & \text{for } \chi \neq 1, \\ (1 - \alpha^{-1})^2 L(E, 1)/\Omega_E & \text{if } \chi = 1, \end{cases} \tag{1.5.1}$$

where $g(\chi)$ is the Gauss sum:

$$g(\chi) = \sum_{a \mod p^n} \chi(a) e^{\frac{2\pi i a}{p^n}}.$$

Note, in particular, that $L(E, 1) \neq 0$ if and only if $L_p(E, 1) \neq 0$.

In order to obtain a Taylor series attached to L_p, we view L_p as a p-adic analytic function on the open disk

$$D = \{u \in \mathbb{C}_p : |u - 1|_p < 1\},$$

as follows. We have that $\gamma = 1 + p$ is a topological generator for $1 + p\mathbb{Z}_p$. For any $u \in D$, let $\psi_u : 1 + p\mathbb{Z}_p \to \mathbb{C}_p^*$ be the character given by sending γ to u and extending by using the group law and continuity. Extend ψ_u to a character $\chi_u : \mathbb{Z}_p^* \to \mathbb{C}_p^*$ by letting $\chi_u(x) = \psi_u(\langle x \rangle)$. Finally, overloading notation, let

$$L_p(E, u) = L_p(E, \chi_u).$$

Theorem 1.15. *The function $L_p(E, u)$ is a p-adic analytic function on D with Taylor series about $u = 1$ in the variable T*

$$\mathcal{L}_p(E, T) \in \mathbb{Q}_p[[T]].$$

that converges on $\{z \in \mathbb{C}_p : |z|_p < 1\}$. (Note that $L_p(E, u) = \mathcal{L}_p(E, u - 1)$.)

It is $\mathcal{L}_p(E, T)$ that we will compute explicitly.

Conjecture 1.16 (Mazur, Tate, Teitelbaum)**.**

$$\operatorname{ord}_T \mathcal{L}_p(E, T) = \operatorname{rank} E(\mathbb{Q}).$$

Proposition 1.17. *Conjecture 1.16 is true if* $\operatorname{ord}_T \mathcal{L}_p(E, T) \leq 1$.

[4] I copied this from Bertolini-Darmon, and I don't trust it exactly yet, especially because the line from Bertolini-Darmon for $\chi = 1$ was wrong.

Sketch of Proof. By Remark 1.14, we have $\mathrm{ord}_T(\mathcal{L}_p(E,T)) = 0$ if and only if

$$r_{E,\mathrm{an}} = \mathrm{ord}_{s=1} L(E,s) = 0.$$

Since the BSD rank conjecture (Conjecture 1.1) is a theorem when $r_{E,\mathrm{an}} = 0$, Conjecture 1.16 is also known under the hypothesis that $\mathrm{ord}_T(\mathcal{L}_p(E,T)) = 0$.

Recall that the BSD rank conjecture is also a theorem when $r_{E,\mathrm{an}} = 1$. It turns out that the same is true of Conjecture 1.16 above. If $\mathrm{ord}_T(\mathcal{L}_p(E,T)) = 1$, then a theorem of Perrin-Riou implies that a certain Heegner point has nonzero p-adic height, hence is non-torsion, so by the Gross-Zagier theorem $r_{E,\mathrm{an}} = 1$. Kolyvagin's theorem then implies that $\mathrm{rank}\, E(\mathbb{Q}) = 1$. □

Remark 1.18. Mazur, Tate, and Teitelbaum also define an analogue of $\mathcal{L}_p(E,T)$ for primes of bad multiplicative reduction and make a conjecture. A prime p is *supersingular* for E if $a_p \equiv 0 \pmod{p}$; it is a theorem of Elkies [**Elk87**] that for any elliptic curve E there are infinitely many supersingular primes p. Perrin-Riou, Pollack, Greenberg and others have studied $\mathcal{L}_p(E,T)$ at good supersingular primes. More works needs to be done on finding a definition of $\mathcal{L}_p(E,T)$ when p is a prime of bad additive reduction for E.

Remark 1.19. A theorem of Rohrlich implies that there is some character as in (1.5.1) such that $L(E,\chi,1) \neq 0$, so $\mathcal{L}_p(E,T)$ is not identically zero. Thus $\mathrm{ord}_T \mathcal{L}_p(T) < \infty$.

1.6. Computing $\mathcal{L}_p(E,T)$

Fix notation as in Section 1.5. In particular, E is an elliptic curve over $\mathbb{Q}$, p is an odd prime of good ordinary reduction for E, and α is the root of $x^2 - a_p x + p$ with $|\alpha|_p = 1$.

For each integer $n \geq 1$, define a polynomial

$$P_n(T) = \sum_{a=1}^{p-1} \left(\sum_{j=0}^{p^{n-1}-1} \mu_E \left(\tau(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot (1+T)^j \right) \in \mathbb{Q}_p[T].$$

Recall that $\tau(a) \in \mathbb{Z}_p^*$ is the Teichmuller lift of a.

Proposition 1.20. *We have that the p-adic limit of these polynomials is the p-adic L-series:*

$$\lim_{n\to\infty} P_n(T) = \mathcal{L}_p(E,T).$$

This convergence is coefficient-by-coefficient, in the sense that if $P_n(T) = \sum_j a_{n,j} T^j$ and $\mathcal{L}_p(E,T) = \sum_j a_j T^j$, then

$$\lim_{n\to\infty} a_{n,j} = a_j.$$

We now give a proof of this convergence and in doing so obtain an upper bound for $|a_j - a_{n,j}|$.

For any choice ζ_r of p^r-th root of unity in $\mathbb{C}_p$, let χ_r be the $\mathbb{C}_p$-valued character of $\mathbb{Z}_p^\times$ of order p^r which factors through $1 + p\mathbb{Z}_p$ and sends $1 + p$ to ζ_r. Note that the conductor of χ_r is p^{r+1}.

Lemma 1.21. *Let ζ_r be a p^r-th root of unity with $1 \leq r \leq n-1$, and let χ_r be the corresponding character of order p^{r+1}, as above. Then*

$$P_n(\zeta_r - 1) = \int_{\mathbb{Z}_p^\times} \chi_r \; d\mu_E$$

In particular, note that the right hand side does not depend on n.

Proof. Writing $\chi = \chi_r$, we have

$$\begin{aligned} P_n(\zeta_r - 1) &= \sum_{a=1}^{p-1} \sum_{j=0}^{p^{n-1}-1} \mu_E \left(\tau(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot \zeta_r^j \\ &= \sum_{a=1}^{p-1} \sum_{j=0}^{p^{n-1}-1} \mu_E \left(\tau(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot \chi\left((1+p)^j\right) \\ &= \sum_{b \in (\mathbb{Z}/p^n\mathbb{Z})^*} \mu_E\left(b + p^n\mathbb{Z}_p\right) \cdot \chi(b) \\ &= \int_{\mathbb{Z}_p^\times} \chi \; d\mu_E. \end{aligned}$$

In the second to the last equality, we use that

$$(\mathbb{Z}/p^n\mathbb{Z})^* \cong (\mathbb{Z}/p\mathbb{Z})^* \times (1 + p(\mathbb{Z}/p^n\mathbb{Z}))^*$$

to sum over lifts of $b \in (\mathbb{Z}/p^n\mathbb{Z})^*$ of the form $\tau(a)(1+p)^j$, i.e., a Teichmuller lift times a power of $(1+p)^j$. In the last equality, we use that χ has conductor p^n, so is constant on the residue classes modulo p^n, i.e., the last equality is just the Riemann sums definition of the given integral.

□

For each positive integer n, let $w_n(T) = (1+T)^{p^n} - 1$.

Corollary 1.22. *We have that*

$$w_{n-1}(T) \text{ divides } P_{n+1}(T) - P_n(T).$$

Proof. By Lemma 1.21, $P_{n+1}(T)$ and $P_n(T)$ agree on $\zeta_j - 1$ for $0 \leq j \leq n-1$ and any choice ζ_j of p^j-th root of unity, so their difference vanishes on every root of the polynomial $w_{n-1}(T) = (1+T)^{p^{n-1}} - 1$. The claimed divisibility follows, since $w_{n-1}(T)$ has distinct roots. □

Lemma 1.23. *Let $f(T)=\sum_j b_j T^j$ and $g(T)=\sum_j c_j T^j$ be in $\mathcal{O}[T]$ with $\mathcal{O}$ a finite extension of $\mathbb{Z}_p$. If $f(T)$ divides $g(T)$, then*

$$\operatorname{ord}_p(c_j) \geq \min_{0\leq i\leq j} \operatorname{ord}_p(b_i).$$

Proof. We have $f(T)k(T)=g(T)$. The lemma follows by using the definition of polynomial multiplication and the non-archimedean property of ord_p on each coefficient of $g(T)$. □

As above, let $a_{n,j}$ be the jth coefficient of the polynomial $P_n(T)$. Let

$$c_n = \max(0, -\min_j \operatorname{ord}_p(a_{n,j}))$$

so that $p^{c_n}P_n(T)\in\mathbb{Z}_p[T]$, i.e., c_n is the smallest power of p that clears the denominator. Note that c_n is an integer since $a_{n,j}\in\mathbb{Q}$. *Probably if $E[p]$ is irreducible then $c_n=0$ – see Question 1.13.* Also, for any $j>0$, let

$$e_{n,j} = \min_{1\leq i\leq j} \operatorname{ord}_p \binom{p^n}{i}.$$

be the min of the valuations of the coefficients of $w_n(T)$, as in Lemma 1.23.

Proposition 1.24. *For all $n\geq 0$, we have $a_{n+1,0}=a_{n,0}$, and for $j>0$,*

$$\operatorname{ord}_p(a_{n+1,j}-a_{n,j}) \geq e_{n-1,j} - \max(c_n, c_{n+1}).$$

Proof. Let $c=\max(c_n,c_{n+1})$. The divisibility of Corollary 1.8 implies that there is a polynomial $h(T)\in\mathbb{Z}_p[T]$ with

$$w_{n-1}(T)\cdot p^c h(T) = p^c P_{n+1}(T) - p^c P_n(T)$$

and thus (by Gauss' lemma) $p^c h(T)\in\mathbb{Z}_p[T]$ since the right hand side of the equation is integral and $w_{n-1}(T)$ is a primitive polynomial. Applying Lemma 1.23 and renormalizing by p^c gives the result. □

For j fixed, $e_{n-1,j}-\max(c_{n+1},c_n)$ goes to infinity as n grows since the c_k are uniformly bounded (they are bounded by the power of p that divides the order of the cuspidal subgroup of E). Thus, $\{a_{n,j}\}$ is a Cauchy and Proposition 1.24 implies that that

$$\operatorname{ord}_p(a_j - a_{n,j}) \geq e_{n-1,j} - \max(c_{n+1}, c_n).$$

Remark 1.25. Recall that presently there is not a single example where we can provably show that $\operatorname{ord}_{s=1} L(E,s)\geq 4$. Amazingly $\operatorname{ord}_T \mathcal{L}_p(E,T)$ is "computable in practice" because Kato has proved, using his Euler system in K_2, that $\operatorname{rank} E(\mathbb{Q}) \leq \operatorname{ord}_T \mathcal{L}_p(E,T)$ by proving a divisibility predicted by Iwasawa Theory. Thus computing elements of $E(\mathbb{Q})$ gives a provable lower bound, and approximating $\mathcal{L}_p(E,T)$ using Riemann sums gives a provable upper bound – in practice these meet.

Chapter 2

The Birch and Swinnerton-Dyer Formula

2.1. Galois Cohomology

Galois cohomology is the basic language used for much research into algebraic aspects of the BSD conjecture. It was introduced by Lang and Tate in 1958 in [**LT58**]. This section contains a survey of the basic facts we will need in order to define Shafarevich-Tate groups, discuss descent, and construct Kolyvagin's cohomology classes.

The best basic reference on Galois cohomology is chapters VII and X of Serre's *Local Fields* [**Ser79**] or the (very similar!) article by Atiyah and Wall in Cassels-Frohlich [**Cp86**, Ch. IV]. See also the article by Gruenberg in [**Cp86**, Ch. V] for an introduction to profinite groups such as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Since this section is only a survey, you should read one of the above two references in detail, if you haven't already. You might also want to read Chapter 1 of [**CS00**] by Coates and Sujatha, which contains an excellent summary of more advanced topics in Galois cohomology, and Serre's book *Galois Cohomology* [**Ser97**] discusses many general advanced topics in depth. The original article [**LT58**] is also well worth reading.

2.1.1. Group Cohomology. If G is a multiplicative group, the *group ring* $\mathbb{Z}[G]$ is the ring of all finite formal sums of elements of G, with multiplication defined using distributivity and extending linearly. Let A be an additive group. We say that A is a *G-module* if A is equipped with a module structure over the group ring $\mathbb{Z}[G]$.

Let A^G be the submodule of elements of A that are fixed by G. Notice that if $A \to B$ is a homomorphism of G-modules, then restriction defines a homomorphism $A^G \to B^G$, so $A \mapsto A^G$ is a *functor*. In fact, it is a *left-exact* functor:

Proposition 2.1. *If $0 \to A \to B \to C$ is an exact sequence of G modules, then $0 \to A^G \to B^G \to C^G$ is also exact.*

Definition 2.2 (Group Cohomology)**.** The *group cohomology $H^n(G, A)$* is by definition the *right derived functors* of the left exact functor $A \to A^G$. These are the unique, up to canonical equivalence, functors H^n such that

- The sequence

$$0 \to A^G \to B^G \to C^G \xrightarrow{\delta} H^1(G, A) \to \cdots \to H^n(G, A) \to H^n(G, B) \to H^n(G, C) \xrightarrow{\delta} H^{n+1}(G, A) \to$$

  is exact.
- If A is *coinduced*, i.e., $A = \mathrm{Hom}(\mathbb{Z}[G], X)$ for X an abelian group, then

$$\mathrm{H}^n(G, A) = 0 \text{ for all } n \geq 1.$$

Remark 2.3. For those familiar with the Ext functor, we have

$$\mathrm{H}^n(G, A) = \mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, A).$$

We construct $\mathrm{H}^n(G, A)$ explicitly as follows. Consider $\mathbb{Z}$ as a G-module, equipped with the trivial G-action. Consider the following free resolution of $\mathbb{Z}$. Let P_i be the free $\mathbb{Z}$-module with basis the set of $i+1$ tuples $(g_0, \ldots, g_i) \in G^{i+1}$, and with G acting on P_i componentwise:

$$s(g_0, \ldots, g_i) = (sg_0, \ldots, sg_i).$$

The homomorphism $d : P_i \to P_{i+1}$ is given by

$$d(g_0, \ldots, g_i) = \sum_{j=0}^{i} (-1)^j (g_0, \ldots, g_{j-1}, g_{j+1}, \ldots g_i),$$

and $P_0 \to \mathbb{Z}$ is given by sending every element (g_0) to $1 \in \mathbb{Z}$.

The cohomology groups $\mathrm{H}^i(G, A)$ are then the cohomology groups of the complex $K_i = \mathrm{Hom}_{\mathbb{Z}[G]}(P_i, A)$. We identify an element of K_i with a function $f : G^{i+1} \to A$ such that the condition

$$f(sg_0, \ldots, sg_i) = sf(g_0, \ldots, g_i)$$

holds. Notice that such an $f \in K_i$ is uniquely determined by the function (of i inputs)

$$\varphi(g_1, \ldots, g_i) = f(1, g_1, g_1 g_2, \ldots, g_1 \cdots g_i).$$

The boundary map $d : K_i \to K_{i+1}$ on such functions $\varphi \in K_i$ is then given explicitly by the formula

$$\begin{aligned}(d\varphi)(g_1, \ldots, g_{i+1}) = g_1\varphi(g_2, \ldots, g_{i+1}) + \sum_{j=1}^{i}(-1)^j\varphi(g_2, \ldots, g_jg_{j+1}, \ldots, g_{i+1})\\ + (-1)^{i+1}\varphi(g_1, \ldots, g_i).\end{aligned}$$

The group of *n-cocycles* is the group of $\varphi \in K_n$, as above are functions of n variables such that $d\varphi = 0$. The subgroup of *n-coboundaries* is the image of K_{n+1} under d. Explicitly, the cohomology group $H^n(G, A)$ is the quotient of the group group of n-cocycles modulo the subgroup of n-coboundaries.

When $n = 1$, the 1-cocycles are the maps $G \to A$ such that

$$\varphi(gg') = g\varphi(g') + \varphi(g),$$

and φ is a coboundary if there exists $a \in A$ such that $\varphi(g) = ga - a$ for all $g \in G$. Notice that if G acts trivially on A, then

$$\mathrm{H}^1(G, A) = \mathrm{Hom}(G, A).$$

2.1.2. The inf-res Sequence. Suppose G is a group and H is a normal subgroup of G, and A is a G-module. Then for any $n \geq 0$, there are natural homomorphisms

$$\mathrm{res} : \mathrm{H}^n(G, A) \to \mathrm{H}^n(H, A)$$

and

$$\mathrm{inf} : \mathrm{H}^n(G/H, A^H) \to \mathrm{H}^n(G, A)$$

Require that we view n-cocycles as certain maps on the n-fold product of the group. On cocycles, the map res is obtained by simply restricting a cocycle, which is a map $G^i \to A$, to a map $H^i \to A$. The second map inf is obtained by precomposing a cocycle $(G/H)^i \to A^H$ with the natural map $G^i \to (G/H)^i$.

Proposition 2.4. *The* inf-res sequence

$$0 \to \mathrm{H}^1(G/H, A^H) \xrightarrow{\mathrm{inf}} \mathrm{H}^1(G, A) \xrightarrow{\mathrm{res}} \mathrm{H}^1(H, A)$$

is exact.

Proof. See [**Ser79**, §VII.6]. □

2.1.3. Galois Cohomology. Let K be a field and L a finite *Galois extension* of K, so the set of field automorphisms of L that fix K equals the dimension of L viewed as a K-vector space.

For any $\mathrm{Gal}(L/K)$-module A and any $n \geq 0$, let

$$\mathrm{H}^n(L/K, A) = \mathrm{H}^n(\mathrm{Gal}(L/K), A).$$

If $M/L/K$ is a tower of Galois extensions of K and suppose $\mathrm{Gal}(M/K)$ acts on A. Then inf defines a map

$$\mathrm{H}^n(L/K, A^L) \to \mathrm{H}^n(M/K, A). \tag{2.1.1}$$

Let K^{sep} denote a separable closure of K and suppose A is a (continuous) $\mathrm{Gal}(K^{\mathrm{sep}}/K)$-module. (Note – if K has characteristic 0, then a separable closure is the same thing as an algebraic closure.) For any subfield $L \subset K^{\mathrm{sep}}$ that contains K, let $A(L) = A^L$. Let

$$\mathrm{H}^n(K, A) = \varinjlim_{L/K \text{ finite Galois}} \mathrm{H}^n(L/K, A(L)),$$

where the direct limit is with respect to the maps (2.1.1). We can think of this direct limit as simply the union of all the groups, where we identify two elements if they are eventually equal under some map (2.1.1).

One can prove (see [**Cp86**, Ch. V]) that changing the choice of separable closure K^{sep} only changes $\mathrm{H}^n(K, A)$ by unique isomorphism, i.e., the construction is essentially independent of the choice of seperable closure.

2.2. The Shafarevich-Tate Group

In this section we discuss Galois cohomology of elliptic curves, introduce the Kummmer sequence, define the Selmer group, the Shafarevich-Tate group and dicuss descent and the Mordell-Weil theorem.

2.2.1. The Elliptic Curve Kummer Sequence. Let E be an elliptic curve over a number field K. Consider the abelian group $E(\overline{\mathbb{Q}})$ of all points on E defined over a fixed choice $\overline{\mathbb{Q}}$ of algebraic closure of $\mathbb{Q}$. Then A is a module over $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$, and we may consider the Galois cohomology groups

$$\mathrm{H}^n(K, E), \qquad \text{for } n = 0, 1, 2, \ldots$$

which are of great interest in the study of elliptic curves, especially for $n = 0, 1$.

If L is a finite Galois extension of K, then the inf-res sequence, written in terms of Galois chomology, is

$$0 \to \mathrm{H}^1(L/K, E(L)) \to \mathrm{H}^1(K, E) \to \mathrm{H}^1(L, E).$$

For any positive integer n consider the homomorphism

$$[n] : E(\overline{\mathbb{Q}}) \to E(\overline{\mathbb{Q}}).$$

This is a surjective homomorphism of abelian groups, so we have an exact sequence

$$0 \to E[n] \to E \xrightarrow{[n]} E \to 0.$$

The associated long exact sequence of Galois cohomology is

$$0 \to E(K)[n] \to E(K) \xrightarrow{[n]} E(K) \to \mathrm{H}^1(K, E[n]) \to \mathrm{H}^1(K, E) \xrightarrow{[n]} \mathrm{H}^1(K, E) \to \cdots.$$

An interesting way to rewrite the begining part of this sequence is as

$$0 \to E(K)/nE(K) \to \mathrm{H}^1(K, E[n]) \to H^1(K, E)[n] \to 0. \tag{2.2.1}$$

The sequence (2.2.1) is called the *Kummer sequence* associated to the elliptic curve.

2.2.2. The Global-to-Local Restriction Maps. Let $\wp$ be a prime ideal of the ring $\mathcal{O}_K$ of integers of the number field K, and let $K_\wp$ be the completion of K with respect to $\wp$. Thus $K_\wp$ is a finite extension the field $\mathbb{Q}_p$ of p-adic numbers.

More explicitly, if $K = \mathbb{Q}(\alpha)$, with α a root of the irreducible polynomial $f(x)$, then the prime ideals $\wp$ correspond to the irreducible factors of $f(x)$ in $\mathbb{Z}_p[x]$. The fields $K_\wp$ then correspond to adjoing roots of each of these irreducible factors of $f(x)$ in $\mathbb{Z}_p[x]$. Note that for most p, a generalization of Hensel's lemma (see Section 1.5.1) asserts that we can factor $f(x)$ by factoring $f(x)$ modulo p and iteratively lifting the factorization.

We have a natural map $\mathrm{Gal}(\overline{\mathbb{Q}}_p/K_\wp) \to \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ got by restriction; implicit in this is a *choice* of embedding of $\overline{\mathbb{Q}}$ in $\overline{\mathbb{Q}}_p$ that sends K into K_v. We may thus view $\mathrm{Gal}(\overline{\mathbb{Q}}_p/K_\wp)$ as a subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$.

Let A be any $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ module. Then this restriction map induces a restriction map on Galois cohomology

$$\mathrm{res}_\wp : \mathrm{H}^1(K, A) \to \mathrm{H}^1(K_\wp, A).$$

Recall that in terms of 1-cocycles this sends a set-theoretic map (a crossed-homomorphism) $f : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to A$ to a map $\mathrm{res}_\wp(f) : \mathrm{Gal}(\overline{\mathbb{Q}}_p/K_\wp) \to A$.

Likewise there is a restriction map for each real Archimedian prime v, i.e., for each embedding $K \to \mathbb{R}$ we have a map

$$\mathrm{res}_v : \mathrm{H}^1(K, A) \to \mathrm{H}^1(\mathbb{R}, A).$$

Exercise 2.5. Let $A = E(\mathbb{C})$ be the group of points on an elliptic curve over $\mathbb{R}$. Prove that $\mathrm{H}^1(\mathbb{R}, E) = \mathrm{H}^1(\mathbb{C}/\mathbb{R}, E(\mathbb{C}))$ is a group of order 1 or 2.

Exercise 2.6. Prove that for any Galois moduloe A and for all primes $\wp$ the kernel of $\mathrm{res}_\wp$ does not depend on the choice of embedding of $\overline{\mathbb{Q}}$ in $\overline{\mathbb{Q}}_p$. (See [**Cp86**, Ch. V]).

2.2.3. The Selmer Group.

Let E be an elliptic curve over a number field K. Let v be either a prime $\wp$ of K or a real Archimedian place (i.e., embedding $K \to \mathbb{R}$). As in Section 2.2.1 we also obtain a local Kummer sequence

$$0 \to E(K_v)/nE(K_v) \to \mathrm{H}^1(K_v, E[n]) \to H^1(K_v, E)[n] \to 0.$$

Putting these together for all v we obtain a commutative diagram:
(2.2.2)

$$\begin{tikzcd}
0 \arrow[r] & E(K)/nE(K) \arrow[r] \arrow[d] & \mathrm{H}^1(K, E[n]) \arrow[r] \arrow[d] & H^1(K, E)[n] \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \prod_v E(K_v)/nE(K_v) \arrow[r] & \prod_v \mathrm{H}^1(K_v, E[n]) \arrow[r] & \prod_v H^1(K_v, E)[n] \arrow[r] & 0.
\end{tikzcd}$$

Definition 2.7. The *n-Selmer group* of an elliptic curve E over a number field K is

$$\mathrm{Sel}^{(n)}(E/K) = \ker\left(\mathrm{H}^1(K, E[n]) \to \prod_v \mathrm{H}^1(K_v, E)[n]\right).$$

2.2.4. The Shafarevich-Tate Group and the Mordell-Weil Theorem.

Definition 2.8 (Shafarevich-Tate Group)**.** The *Shafarevich-Tate group* of an elliptic curve E over a number field K is

$$\text{Ш}(E/K) = \ker\left(\mathrm{H}^1(K,E) \to \prod_v \mathrm{H}^1(K_v,E)\right).$$

For any positive integer n, we may thus add in a row to (2.2.2):

$$\begin{tikzcd}
0 \arrow[r] & E(K)/nE(K) \arrow[r] \arrow[d, equal] & \mathrm{Sel}^{(n)}(E/K) \arrow[r] \arrow[d, hook] & \text{Ш}(E/K)[n] \arrow[r] \arrow[d, hook] & 0 \\
0 \arrow[r] & E(K)/nE(K) \arrow[r] \arrow[d] & \mathrm{H}^1(K,E[n]) \arrow[r] \arrow[d] & H^1(K,E)[n] \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \prod_v E(K_v)/nE(K_v) \arrow[r] & \prod_v \mathrm{H}^1(K_v,E[n]) \arrow[r] & \prod_v H^1(K_v,E)[n] \arrow[r] & 0.
\end{tikzcd}$$

The *n-descent sequence* for E is the short exact sequence

$$0 \to E(K)/nE(K) \to \mathrm{Sel}^{(n)}(E/K) \to \text{Ш}(E/K)[n] \to 0. \tag{2.2.3}$$

Theorem 2.9. *For every integer n the group* $\mathrm{Sel}^{(n)}(E/K)$ *is finite.*

Sketch of Proof. Let $K(E[n])$ denote the finite Galois extension of K obtained by adjoining to K all x and y coordinates of elements of $E(\overline{\mathbb{Q}})$ of order dividing n. The inf-res sequence for $K(E[n])/K$ is

$$0 \to \mathrm{H}^1(K(E[n])/K, E[n]) \to \mathrm{H}^1(K,E[n]) \to \mathrm{H}^1(K(E[n]),E[n]). \tag{2.2.4}$$

Because $\mathrm{Gal}(K(E[n])/K)$ and $E[n]$ are both finite groups, the cohomology group $\mathrm{H}^1(K(E[n])/K, E[n])$ is also finite.

Since $\mathrm{Sel}^{(n)}(E/K) \subset \mathrm{H}^1(K,E[n])$, restriction defines a map

$$\mathrm{Sel}^{(n)}(E/K) \to \mathrm{Sel}^{(n)}(E/K[n]). \tag{2.2.5}$$

The kernel of (2.2.5) is finite since it is contained in the first term of (2.2.4), which is finite. It thus suffices to prove that $\mathrm{Sel}^{(n)}(E/K[n])$ is finite.

But

$$\mathrm{Sel}^{(n)}(E/K[n]) \subset \mathrm{H}^1(K[n],E[n]) \cong \mathrm{Hom}(\mathrm{Gal}(\overline{\mathbb{Q}}/K[n]),E[n])).$$

So each element of $\mathrm{Sel}^{(n)}(E/K[n])$ determines (and is determined by) a homomorphism $\mathrm{Gal}(\overline{\mathbb{Q}}/K[n]) \to (\mathbb{Z}/n\mathbb{Z})^2$. That that the fixed field of such a homomorphism is a Galois extension of $K[n]$ with Galois group contained in $(\mathbb{Z}/n\mathbb{Z})^2$.

To complete the proof, one uses the theory of elliptic curves over local fields to show that there is a finite set S of primes such that any such homomorphism corresponding to an element of the Selmer group corresponds to an extension of $K[n]$ ramified only at primes in S. Then the two main theorems of algebraic number theory — that class groups are finite and unit groups are finitely generated — together imply that there are only finitely many such extensions of $K[n]$.

□

Exercise 2.10. Prove the that $E[n]$ is a finite Galois extension of K.

Theorem 2.11 (Mordell-Weil)**.** *The group $E(\mathbb{Q})$ is finitely generated.*

Proof. The exact sequence (2.2.3) with $n = 2$ and Theorem 2.9 imply that $E(\mathbb{Q})/2E(\mathbb{Q})$ is a finite group. Recall Lemma 1.4 which asserted that if B is a positive real number such that

$$S = \{P \in E(\mathbb{Q}) \,:\, \hat{h}(P) \leq B\}$$

contains a set of generators for $E(\mathbb{Q})/2E(\mathbb{Q})$, then S generates $E(\mathbb{Q})$. Since $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite, it makes sense to define B to be the maximum of the heights of arbitrary lifts of all the elements of $E(\mathbb{Q})/2E(\mathbb{Q})$. Then the corresponding set S generates $E(\mathbb{Q})$. A basic fact about heights is that the set of points of bounded height is finite, i.e., S is finite, so $E(\mathbb{Q})$ is finitely generated. □

2.2.5. Some Conjectures and Theorems about the Shafarevich-Tate Group.

Conjecture 2.12 (Shafarevich-Tate)**.** *Let E be an elliptic curve over a number field K. Then the group $Ш(E/K)$ is finite.*

Theorem 2.13 (Rubin)**.** *If E is a CM elliptic curve over $\mathbb{Q}$ with $L(E, 1) \neq 0$, then $Ш(E/\mathbb{Q})$ is finite. (He proved more than just this.)*

Thus Rubin's theorem proves that the Shafarevich-Tate group of the CM elliptic curve $y^2 + y = x^3 - 7$ of conductor 27 is finite.

Theorem 2.14 (Kolyvagin et al.)**.** *If E is an elliptic curve over $\mathbb{Q}$ with $\mathrm{ord}_{s=1} L(E, s) \leq 1$, then $Ш(E/\mathbb{Q})$ is finite.*

Kolyvagin's theorem is proved in a completely different way than Rubin's theorem. It combines the Gross-Zagier theorem, the modularity theorem that there is a map $X_0(N) \to E$, a nonvanishing result about the special values $L(E^D, 1)$ of quadratic twists of E, and a highly original explicit study of the structure of the images of certain points on $X_0(N)(\overline{\mathbb{Q}})$ in $E(\overline{\mathbb{Q}})$.

Theorem 2.15 (Cassels)**.** *Let E be an elliptic curve over a number field K. There is an alternating pairing on $Ш(E/K)$, which is nondegenerate on the quotient of $Ш(E/K)$ by its maximal divisible subgroup. Moreover, if $Ш(E/K)$ is finite then $\#Ш(E/K)$ is a perfect square.*

For an abelian group A and a prime p, let $A(p)$ denote the subgroup of elements of p power order in A.

The following problem remains open. It helps illustrate our ignorance about Conjecture 2.12 in any cases beyond those mentioned above.

Problem 2.16. Show that there is an elliptic curve E over $\mathbb{Q}$ with rank ≥ 2 such that $Ш(E/\mathbb{Q})(p)$ is finite for infinitely many primes p.

2.3. The Birch and Swinnerton-Dyer Formula

> "The subject of this lecture is rather a special one. I want to describe some computations undertaken by myself and Swinnerton-Dyer on EDSAC, by which we have calculated the zeta-functions of certain elliptic curves. As a result of these computations we have found an analogue for an elliptic curve of the Tamagawa number of an algebraic group; and conjectures have proliferated. [...] I would like to stress that though the associated theory is both abstract and technically complicated, the objects about which I intend to talk are usually simply defined and often machine computable; experimentally we have detected certain relations between different invariants, but we have been unable to approach proofs of these relations, which must lie very deep."
>
> – Bryan Birch

Conjecture 2.17 (Birch and Swinnerton-Dyer)**.** *Let E be an elliptic curve over $\mathbb{Q}$ of rank r. Then $r = \mathrm{ord}_{s=1} L(E,s)$ and*

$$\frac{L^{(r)}(E,1)}{r!} = \frac{\Omega_E \cdot \mathrm{Reg}(E) \cdot \#Ш(E/\mathbb{Q}) \cdot \prod_p c_p}{\#E(\mathbb{Q})_{\mathrm{tor}}^2}. \tag{2.3.1}$$

Let

$$y^2 + \underline{a}_1 xy + \underline{a}_3 y = x^3 + \underline{a}_2 x^2 + \underline{a}_4 x + \underline{a}_6 \tag{2.3.2}$$

be a minimal Weierstrass equation for E.

Recall from Section 1.5.2 that the *real period* Ω_E is the integral

$$\Omega_E = \int_{E(\mathbb{R})} \frac{dx}{2y + \underline{a}_1 x + \underline{a}_3}.$$

See [**Cre97**, §3.7] for an explanation about how to use the Gauss arithmetic-geometry mean to efficiently compute Ω_E.

To define the *regulator* $\mathrm{Reg}(E)$ let $P_1, \ldots, P_n$ be a basis for $E(\mathbb{Q})$ modulo torsion and recall the Néron-Tate canonical height pairing $\langle\, , \rangle$ from

Section 1.2. The real number $\mathrm{Reg}(E)$ is the absolute value of the determinant of the $n \times n$ matrix whose (i, j) entry is $\langle P_i, P_j \rangle$. See [**Cre97**, §3.4] for a discussion of how to compute $\mathrm{Reg}(E)$.

We defined the group $Ш(E/\mathbb{Q})$ in Section 2.2.4. In general it is not known to be finite, which led to Tate's famous assertion that the above conjecture "relates the value of a function at a point at which it is not known to be defined[1] to the order of a group that is not known to be finite." The paper [**GJP$^+$05**] discusses methods for computing $\#Ш(E/\mathbb{Q})$ in practice, though no general algorithm for computing $\#Ш(E/\mathbb{Q})$ is known. In fact, in general even if we assume truth of the BSD rank conjecture (Conjecture 1.1) and assume that $Ш(E/\mathbb{Q})$ is finite, there is still no known way to compute $\#Ш(E/\mathbb{Q})$, i.e., there is no analogue of Proposition 1.3. Given finiteness of $Ш(E/\mathbb{Q})$ we can compute the p-part $Ш(E/\mathbb{Q})(p)$ of $Ш(E/\mathbb{Q})$ for any prime p, but we don't know when to stop considering new primes p. (Note that when $r_{E,\mathrm{an}} \leq 1$, Kolyvagin's work provides an explicit upper bound on $\#Ш(E/\mathbb{Q})$, so in that case $Ш(E/\mathbb{Q})$ is computable.)

The *Tamagawa numbers* c_p are 1 for all primes $p \nmid \Delta_E$, where Δ_E is the discriminant of (2.3.2). When $p \mid \Delta_E$, the number c_p is a more refined measure of the structure of the E locally at p. If p is a prime of *additive reduction* (see Section 1.3), then one can prove that $c_p \leq 4$. The other alternatives are that p is a prime of split or nonsplit multiplicative reduction. If p is a *nonsplit prime*, then

$$c_p = \begin{cases} 1 & \text{if } \mathrm{ord}_p(\Delta) \text{ is odd} \\ 2 & \text{otherwise} \end{cases}$$

If p is a prime of *split multiplicative* reduction then

$$c_p = \mathrm{ord}_p(\Delta)$$

can be arbitrarily large. The above discussion completely determines c_p except when p is an additive prime – see [**Cre97**, §3.2] for a discussion of how to compute c_p in general.

For those that are very familiar with elliptic curves over local fields,

$$c_p = [E(\mathbb{Q}_p) : E^0(\mathbb{Q}_p)],$$

where $E^0(\mathbb{Q}_p)$ is the subgroup of $E(\mathbb{Q}_p)$ of points that have nonsingular reduction modulo p.

For those with more geometric background, we offer the following conceptual definition of c_p. Let $\mathcal{E}$ be the *Néron model* of E. This is the unique, up to unique isomorphism, smooth commutative (but not proper!) group

[1] When E is defined over $\mathbb{Q}$ it is now known that $L(E, s)$ is defined overwhere.

scheme over $\mathbb{Z}$ that has generic fiber E and satisfies the Néron mapping property:

> for any smooth group scheme X over $\mathbb{Z}$ the natural map
> $$\mathrm{Hom}(X, \mathcal{E}) \to \mathrm{Hom}(X_{\mathbb{Q}}, E)$$
> is an isomorphism.

In particular, note that $\mathcal{E}(\mathbb{Z}) \cong E(\mathbb{Q})$. For each prime p, the reduction $\mathcal{E}_{\mathbb{F}_p}$ of the Néron model modulo p is a smooth commutative group scheme over $\mathbb{F}_p$ (smoothness is a property of morphisms that is closed under base extension). Let $\mathcal{E}^0_{\mathbb{F}_p}$ be the identity component of the group scheme $\mathcal{E}_{\mathbb{F}_p}$, i.e., the connected component of $\mathcal{E}^0_{\mathbb{F}_p}$ that contains the 0 section. The *component group* of E at p is the quotient group scheme

$$\Phi_{E,p} = \mathcal{E}_{\mathbb{F}_p}/\mathcal{E}^0_{\mathbb{F}_p},$$

which is a finite étale group scheme over $\mathbb{F}_p$. Finally

$$c_p = \#\Phi_{E,p}(\mathbb{F}_p).$$

2.4. Examples: The Birch and Swinnerton-Dyer Formula

In each example below we use SAGE to compute the conjectural order of $\text{Ш}(E/\mathbb{Q})$ and find that it appears to be the square of an integer as predicted by Theorem 2.15.

2.4.1. Example: A Curve of Rank 0. Consider the elliptic curve E with Cremona label 11a, which is one the 3 curves of smallest conductor. We now compute each of the quantities in Conjecture 2.17. First we define the curve E in SAGE and compute its rank:

```
sage: E = EllipticCurve('11a'); E
Elliptic Curve defined by y^2 + y = x^3 - x^2 - 10*x - 20
over Rational Field
sage: E.rank()
0
```

Next we compute the number $L(E, 1)$ to double precision (as an element of the real double field RDF):

```
sage: L = RDF(E.Lseries(1)); L
0.253841860856
```

We next compute the real period:

```
sage: Om = RDF(E.omega()); Om
1.26920930428
```

To compute $\prod c_p$ we factor the discriminant of E. It turns at that only 11 divides the discriminant, and since the reduction at 11 is split multiplicative the Tamagawa number is $5 = \mathrm{ord}_{11}(\Delta_E)$.

```
sage: factor(discriminant(E))
-1 * 11^5
sage: c11 = E.tamagawa_number(11); c11
5
```

Next we compute the regulator, which is 1 since E rank 0.

```
sage: Reg = RDF(E.regulator()); Reg
1.0
```

The torsion subgroup has order 5.

```
sage: T = E.torsion_order(); T
5
```

Putting everything together in (2.3.1) and solving for the conjectural order of $Ш(E/\mathbb{Q})$, we see that Conjecture 2.17 for E is equivalent to the assertion that $Ш(E/\mathbb{Q})$ has order 1.

```
sage: Sha_conj = L * T^2 / (Om * Reg * c11); Sha_conj
1.0
```

2.4.2. Example: A Rank 0 curve with nontrivial Sha. Consider the curve E with label 681b. This curve has rank 0, and we compute the conjectural order of $\#Ш(E/\mathbb{Q})$ as in the previous section:

```
sage: E = EllipticCurve('681b'); E
Elliptic Curve defined by y^2 + x*y  = x^3 + x^2 - 1154*x - 15345
over Rational Field
sage: E.rank()
0
sage: L = RDF(E.Lseries(1)); L
1.84481520613
sage: Om = RDF(E.omega()); Om
0.81991786939
```

There are two primes of bad reduction this time.

```
sage: factor(681)
3 * 227
sage: factor(discriminant(E))
3^10 * 227^2
sage: c3 = E.tamagawa_number(3); c227 = E.tamagawa_number(227)
sage: c3, c227
(2, 2)
sage: Reg = RDF(E.regulator()); Reg
1.0
sage: T = E.torsion_order(); T
4
```

In this case it turns out that $\#\text{Ш}(E/\mathbb{Q})$ is conjecturally 9.

```
sage: Sha_conj = L * T^2 / (Om * Reg * c3*c227); Sha_conj
9.0
```

2.4.3. Example: A Curve of Rank 1. Let E be the elliptic curve with label 37a, which is the curve of rank 1 with smallest conductor. We define E and compute its rank, which is 1.

```
sage: E = EllipticCurve('37a'); E
Elliptic Curve defined by y^2 + y = x^3 - x over
Rational Field
sage: E.rank()
1
```

We next compute the value $L'(E,1)$. The corresponding function in SAGE takes a bound on the number of terms of the L-series to use, and returns an approximate to $L'(E,1)$ along with a bound on the error (coming from the tail end of the series).

```
sage: L, error = E.Lseries_deriv_at1(200); L, error
(0.305999773834879, 2.10219814818300e-90)
sage: L = RDF(L); L
0.305999773835
```

We compute Ω_E and the Tamagawa number, regulator, and torsion as above.

```
sage: Om = RDF(E.omega()); Om
5.98691729246
sage: factor(discriminant(E))
37
sage: c37 = 1
sage: Reg = RDF(E.regulator()); Reg
0.05111140824
sage: T = E.torsion_order(); T
1
```

Finally, we solve and find that the conjectural order of $Ш(E/\mathbb{Q})$ is 1.

```
sage: Sha_conj = L * T^2 / (Om * Reg * c37); Sha_conj
1.0
```

2.4.4. Example: A curve of rank 2. Let E be the elliptic curve 389a of rank 2, which is the curve of rank 2 with smallest conductor.

```
sage: E = EllipticCurve(’389a’); E
Elliptic Curve defined by y^2 + y = x^3 + x^2 - 2*x
over Rational Field
sage: E.rank()
2
```

Because the curve has rank 2, we use Dokchitser's L-function package to approximate $L^{(2)}(E,1)$ to high precision:

```
sage: Lser = E.Lseries_dokchitser()
sage: L = RDF(abs(Lser.derivative(1,2))); L
1.51863300058
```

We compute the regulator, Tamagawa numbers, and torsion as usual:

```
sage: Om = RDF(E.omega()); Om
4.98042512171
sage: factor(discriminant(E))
389
sage: c389 = 1
sage: Reg = RDF(E.regulator()); Reg
0.152460177943
sage: T = E.torsion_order(); T
1
```

Finally we solve for the conjectural order of $\#\text{Ш}(E/\mathbb{Q})$.

```
sage: Sha_conj = (L/2) * T^2 / (Om * Reg * c389)
sage: Sha_conj
1.0
```

We pause to emphasize that just getting something that looks like an integer by computing

$$\frac{L^{(r)}(E,1)}{r!} \cdot \#E(\mathbb{Q})^2_{\text{tor}}/(\Omega_E \cdot \text{Reg}(E) \cdot \prod c_P) \tag{2.4.1}$$

is already excellent evidence for Conjecture 2.17. There is also a subtle and deep open problem here:

Open Problem 2.18. Let E be the elliptic curve 389a above. Prove that the quantity (2.4.1) is a rational number.

For curves E of analytic rank 0 it is easy to prove using modular symbols that the conjectural order of $\text{Ш}(E/\mathbb{Q})$ is a rational number. For curves with analytic rank 1, this rationality follows from the very deep Gross-Zagier theorem. For curves of analytic rank ≥ 2 there is not a single example in which the conjectural order of $\text{Ш}(E/\mathbb{Q})$ is known to be a rational number.

2.4.5. Example: A Rank 3 curve. The curve E with label 5077a has rank 3. This is the curve with smallest conductor and rank 3.

```
sage: E = EllipticCurve('5077a'); E
Elliptic Curve defined by y^2 + y = x^3 - 7*x + 6
over Rational Field
sage: E.rank()
3
```

We compute $L(E,s)$ using Dokchitser's algorithm. Note that the order of vanishing appears to be 3.

```
sage: E.root_number()
-1
sage: Lser = E.Lseries_dokchitser()
sage: Lser.derivative(1,1)
-5.63436295355925e-22
sage: Lser.derivative(1,2)
2.08600476044634e-21
sage: L = RDF(abs(Lser.derivative(1,3))); L
10.3910994007
```

That the order of vanishing is really 3 follows from the Gross-Zagier theorem, which asserts that $L'(E,1)$ is a nonzero multiple of the Néron-Tate canonical height of a certain point on E called a Heegner point. One can explicitly construct this point[2] on E and find that it is torsion, hence has height 0, so $L'(E,1) = 0$. That $L''(E,1) = 0$ then follows from the functional equation (see Section 1.3). Finally we compute the other BSD invariants:

```
sage: Om = RDF(E.omega()); Om
4.15168798309
sage: factor(discriminant(E))
5077
sage: c5077 = 1
sage: Reg = RDF(E.regulator()); Reg
0.417143558758
sage: T = E.torsion_order(); T
1
```

Putting everything together we see that the conjectural order of $\text{Ш}(E/\mathbb{Q})$ is 1.

[2] This is not yet implemented in SAGE; if it were, there would be an example right here.

```
sage: Sha_conj = (L/6) * T^2 / (Om * Reg * c5077)
sage: Sha_conj
1.0
```

Note that just as was the case with the curve 389a above, we do not know that the above conjectural order of $Ш(E/\mathbb{Q})$ is a rational number, since there are no know theoretical results that relate any of the three real numbers $L^{(3)}(E,1)$, $\mathrm{Reg}(E/\mathbb{Q})$, and $\Omega_{E/\mathbb{Q}}$.

2.4.6. Example: A Rank 4 curve. Let E be the curve of rank 4 with label 234446b. It is likely that this is the curve with smallest conductor and rank 4 (a big calculation of the author et al. shows that there are no rank 4 curves with smaller *prime* conductor).

```
sage: E = EllipticCurve([1, -1, 0, -79, 289]); E
Elliptic Curve defined by y^2 + x*y  = x^3 - x^2 - 79*x + 289
over Rational Field
sage: E.rank()
4
```

We next compute $L(E,1)$, $L'(E,1)$, $L^{(2)}(E,1)$, $L^{(3)}(E,1)$, and $L^{(4)}(E,1)$. All these special values *look* like they are 0, except for $L^{(4)}(E,1)$ which is about 214, hence clearly nonzero. One can prove that $L(E,1) = 0$ (e.g., using denominator bounds coming from modular symbols), hence since the root number is $+1$, we have either $r_{E,\mathrm{an}} = 2$ or $r_{E,\mathrm{an}} = 4$, and of course suspect (but cannot prove yet) that $r_{E,\mathrm{an}} = 4$.

```
sage: E.root_number()
1
sage: Lser = E.Lseries_dokchitser()
sage: Lser(1)
1.43930352980778e-18
sage: Lser.derivative(1,1)
-4.59277879927938e-24
sage: Lser.derivative(1,2)
-8.85707917856308e-22
sage: Lser.derivative(1,3)
1.01437455701212e-20
sage: L = RDF(abs(Lser.derivative(1,4))); L
214.652337502
```

As above, we compute the other BSD invariants of E.

```
sage: Om = RDF(E.omega()); Om
2.97267184726
sage: factor(discriminant(E))
2^2 * 117223
sage: c2 = 2
sage: c117223 = 1
sage: Reg = RDF(E.regulator()); Reg
1.50434488828
sage: T = E.torsion_order(); T
1
```

Finally, putting everything together, we see that the conjectural order of Ш$(E/\mathbb{Q})$ is 1.

```
sage: Sha_conj = (L/24) * T^2 / (Om * Reg * c2 * c117223)
sage: Sha_conj
1.0
```

Again we emphasize that we do not even know that the conjectural order computed above is a rational number.

It seems almost a miracle that $L^{(4)}(E,1) = 214.65\ldots$, $\Omega_E = 2.97\ldots$, and $\mathrm{Reg}(E) = 1.50\ldots$ have anything to do with each other, but indeed they do:

```
sage: L/24, 2*Om*Reg
(8.9438473959, 8.9438473959)
```

That these two numbers are the same to several decimal places is a fact, independent of any conjectures.

2.5. The p-adic BSD Conjectural Formula

Let E be an elliptic curve over $\mathbb{Q}$ and let p be a prime of good ordinary reduction for E.

In Chapter 1 (see Theorem 1.15) we defined a p-adic L-series

$$\mathcal{L}_p(E,T) \in \mathbb{Q}_p[[T]].$$

Conjecture 1.16 asserted that $\operatorname{ord}_T \mathcal{L}_p(E,T) = \operatorname{rank} E(\mathbb{Q})$. Just as is the cases for $L(E,s)$, there is a conjectural formula for the leading coefficient of the power series $\mathcal{L}_p(E,T)$. This formula is due to Mazur, Tate, and Teitelbaum [**MTT86**].

First, suppose $\operatorname{ord}_T \mathcal{L}_p(E,T) = 0$, i.e., $\mathcal{L}_p(E,0) \neq 0$. Recall that the interpolation property (1.5.1) for $\mathcal{L}_p(E,T)$ implies that

$$\mathcal{L}_p(E,0) = \varepsilon_p \cdot L(E,1)/\Omega_E,$$

where

$$\varepsilon_p = (1-\alpha^{-1})^2, \tag{2.5.1}$$

and $\alpha \in \mathbb{Z}_p$ is the unit root of $x^2 - a_p x + p = 0$. Thus the usual BSD conjecture predicts that if the rank is 1, then

$$\mathcal{L}_p(E,0) = \varepsilon_p \cdot \frac{\prod_\ell c_\ell \cdot \#\text{Ш}(E/\mathbb{Q}) \cdot \operatorname{Reg}(E)}{\#E(\mathbb{Q})_{\text{tor}}^2} \tag{2.5.2}$$

Notice in (2.5.2) that since $E(\mathbb{Q})$ has rank 0, we have $\operatorname{Reg}(E) = 1$, so there is no issue with the left hand side being a p-adic number and the right hand side not making sense. It would be natural to try to generalize (2.5.2) to higher order of vanishing as follows. Let $\mathcal{L}_p^*(E,0)$ denote the leading coefficient of the power series $\mathcal{L}_p(E,T)$. Then

$$\mathcal{L}_p^*(E,0)\text{“}=\text{”}\varepsilon_p \cdot \frac{\prod_\ell c_\ell \cdot \#\text{Ш}(E/\mathbb{Q}) \cdot \operatorname{Reg}(E)}{\#E(\mathbb{Q})_{\text{tor}}^2} \qquad \text{(nonsense!!)}. \tag{2.5.3}$$

Unfortunately (2.5.2) is total nonsense when the rank is bigger than 0. The problem is that $\operatorname{Reg}(E) \in \mathbb{R}$ is a real number, whereas ε_p and $\mathcal{L}_p^*(E,0)$ are both p-adic numbers.

The key *new idea* needed to make a conjecture is to replace the real-number regulator $\operatorname{Reg}(E)$ with a p-adic regulator $\operatorname{Reg}_p(E) \in \mathbb{Q}_p$. This new regulator is defined in a way analogous to the classical regulator, but where many classical complex analytic objects are replaced by p-adic analogues. Moreover, the p-adic regulator was, until recently (see [**MST06**]), much more difficult to compute than the classical real regulator. We will define the p-adic number $\operatorname{Reg}_p(E) \in \mathbb{Q}_p$ in the next section.

Conjecture 2.19 (Mazur, Tate, and Teitelbaum)**.** *Let E be an elliptic curve over $\mathbb{Q}$ and let p be a prime of good ordinary reduction for E. Then the rank of E equals $\mathrm{ord}_T(\mathcal{L}_p(E,T))$ and*

$$\mathcal{L}_p^*(E,0) = \varepsilon_p \cdot \frac{\prod_\ell c_\ell \cdot \#\text{Ш}(E/\mathbb{Q}) \cdot \mathrm{Reg}_p(E)}{\#E(\mathbb{Q})_{\mathrm{tor}}^2}, \tag{2.5.4}$$

where ε_p is as in (2.5.1)*, and the p-adic regulator $\mathrm{Reg}_p(E) \in \mathbb{Q}_p$ will be defined below.*

Remark 2.20. There are analogous conjectures in many other cases, e.g., good supersingular, bad multiplicative, etc. See [**SW07**] for more details.

2.5.1. Example: A Curve of Rank 2. We only consider primes p of good ordinary reduction for a given curve E in this section. If E is an elliptic curve with analytic rank 0, then the p-adic and classical BSD conjecture are the same, so there is nothing new to illustrate. We will thus consider only curves of rank ≥ 1 in this section.

We consider the elliptic curve 446d1 of rank 2 at the prime $p = 5$.

```
sage: E = EllipticCurve('446d1'); p = 5; E
Elliptic Curve defined by y^2 + x*y  = x^3 - x^2 - 4*x + 4
over Rational Field
```

Next we verify that the rank is 2, that p is a good ordinary prime, and that there are 10 points on E modulo p (so E is *ananomolous at* p, i.e., $p \mid \#E(\mathbb{F}_p)$).

```
sage: E.rank()
2
sage: E.is_ordinary(p)
True
sage: E.Np(p)
10
```

Next we compute the p-adic L-series of E at p. We add $O(T^7)$ so that the displayed series doesn't take several lines.

```
sage: Lp  = E.padic_lseries(p)
sage: LpT = Lp.series(4)
sage: LpT = LpT.add_bigoh(7); LpT
(5 + 5^2 + O(5^3))*T^2 + (2*5 + 3*5^2 + O(5^3))*T^3
          + (4*5^2 + O(5^3))*T^4 + (4*5 + O(5^2))*T^5
          + (1 + 2*5 + O(5^3))*T^6 + O(T^7)
```

We compute the p-adic modular form E_2 evaluated on our elliptic curve with differential ω to precision $O(p^8)$. This is the key difficult input to the computation of the p-adic regulator $\mathrm{Reg}_p(E)$.

```
sage: E.padic_E2(p, prec=8)
3*5 + 4*5^2 + 5^3 + 5^4 + 5^5 + 2*5^6 + 4*5^7 + O(5^8)
```

We compute the normalized p-adic regulator, normalized to the choice of $1+p$ as a topological generator of $1+p\,\mathbb{Z}_p$.

```
sage: Regp = E.padic_regulator(p, 10)
sage: R = Regp.parent()
sage: kg = log(R(1+p))
sage: reg = Regp * p^2 / log(R(1+p))^2
sage: reg*kg^2
2*5 + 2*5^2 + 5^4 + 4*5^5 + 2*5^7 + O(5^8)
```

We compute the Tamagawa numbers and torsion subgroup.

```
sage: E.tamagawa_numbers()
[2, 1]
sage: E.torsion_order()
1
```

We compute $\mathcal{L}_p^*(E,0)$, which is the leading term of the p-adic L-function. It is not a unit, so we call the prime p an *irregular* prime.

```
sage: Lpstar = LpT[2]; Lpstar
5 + 5^2 + O(5^3)
```

Finally, putting everything together we compute the conjectural p-adic order of $\#\text{Ш}(E/\mathbb{Q})$. In particular, we see that conjecturally $\#\text{Ш}(E/\mathbb{Q})(5)$ is trivial.

```
sage: eps = (1-1/Lp.alpha(20))^2
sage: Lpstar / (eps*reg*(2*1)) * (1)^2
1 + O(5^2)
```

2.5.2. The p-adic Regulator. Fix an elliptic curve E defined over $\mathbb{Q}$ and a prime p of good ordinary reduction for E. In this section we define the p-adic regulator $\mathrm{Reg}_p(E)$. See [**MTT86**], [**MST06**] and [**SW07**] and the references listed there for a more general discussion of p-adic heights, especially for bad or supersingular primes, and for elliptic curves over number fields. See also forthcoming work of David Harvey for highly optimized computation of p-adic regulators.

The p-adic logarithm $\log_p : \mathbb{Q}_p^* \to (\mathbb{Q}_p, +)$ is the unique group homomorphism with $\log_p(p) = 0$ that extends the homomorphism $\log_p : 1+p\mathbb{Z}_p \to \mathbb{Q}_p$ defined by the usual power series of $\log(x)$ about 1. Explicitly, if $x \in \mathbb{Q}_p^*$, then

$$\log_p(x) = \frac{1}{p-1} \cdot \log_p(u^{p-1}),$$

where $u = p^{-\operatorname{ord}_p(x)} \cdot x$ is the unit part of x, and the usual series for log converges at u^{p-1}.

Example 2.21. For example, in SAGE we compute the logs of a couple of non-unit elements of $\mathbb{Q}_5$ as follows:

```
sage: K = Qp(5,8); K
5-adic Field with capped relative precision 8
sage: a = K(-5^2*17); a
3*5^2 + 5^3 + 4*5^4 + 4*5^5 + 4*5^6 + 4*5^7 + 4*5^8 + 4*5^9 + O(5^10)
sage: u = a.unit_part()
3 + 5 + 4*5^2 + 4*5^3 + 4*5^4 + 4*5^5 + 4*5^6 + 4*5^7 + O(5^8)
sage: b = K(1235/5); b
2 + 4*5 + 4*5^2 + 5^3 + O(5^8)
sage: log(a)
5 + 3*5^2 + 3*5^3 + 4*5^4 + 4*5^5 + 5^6 + O(5^8)
sage: log(a*b) - log(a) - log(b)
O(5^8)
```

Note that we can recover b:

```
sage: c = a^b; c
2*5^494 + 4*5^496 + 2*5^497 + 5^499 + 3*5^500 + 5^501 + O(5^502)
sage: log(c)/log(a)
2 + 4*5 + 4*5^2 + 5^3 + O(5^7)
```

Let $\mathcal{E}$ denote the Néron model of E over $\mathbb{Z}$. Let $P \in E(\mathbb{Q})$ be a non-torsion point that reduces to $0 \in E(\mathbb{F}_p)$ and to the connected component of $\mathcal{E}_{\mathbb{F}_\ell}$ at all primes ℓ of bad reduction for E. For example, given any point $Q \in E(\mathbb{Q})$ one can construct such a P by multiplying it by the least common multiple of the Tamagawa numbers of E.

Exercise 2.22. Show that any nonzero point $P = (x(P), y(P)) \in E(\mathbb{Q})$ can be written uniquely in the form $(a/d^2, b/d^3)$, where $a, b, d \in \mathbb{Z}$, $\gcd(a, d) = \gcd(b, d) = 1$, and $d > 0$. (Hint: Use that $\mathbb{Z}$ is a unique factorization domain.)

The function $d(P)$ assigns to P this square root d of the denominator of the x-coordinate $x(P)$.

Example 2.23. We compute a point on a curve, and observe that the denominator of the x coordinate is a perfect square.

```
sage: E = EllipticCurve('446d1')
sage: P = 3*E.gen(0); P
(32/49 : -510/343 : 1)
```

Let

$$x(t) = \frac{1}{t^2} + \cdots \in \mathbb{Z}_p((t)) \tag{2.5.5}$$

be the formal power series that expresses x in terms of the local parameter $t = -x/y$ at infinity. Similarly, let $y(t) = -x(t)/t$ be the corresponding series for y. If we do the change of variables $t = -x/y$ and $w = -1/y$, so $x = t/w$ and $y = -1/w$, then the Weierstrass equation for E becomes

$$s = t^3 + a_1 st + a_2 wt^2 + a_3 w^2 + a_4 w^2 t + a_6 w^3 = F(w, t).$$

Repeatedly substituting this equation into itself recursively yields a power series expansion for $w = -1/y$ in terms of t, hence for both x and y.

Remark 2.24. The *formal group* of E is a power series

$$F(t_1, t_2) \in R = \mathbb{Z}[a_1, \dots, a_6][[t_1, t_2]].$$

defined as follows. Since $x(t)$ and $y(t)$ satisfy the equation of E, the points $P_1 = (x(t_1), y(t_1))$ and $P_2 = (x(t_2), y(t_2))$ are in $E(R)$. As explained explicitly in [**Sil92**, §IV.1], their sum is

$$Q = P_1 + P_2 = (x(F), y(F)) \in E(R)$$

for some $F = F(t_1, t_2) \in R$.

Example 2.25. We compute the above change of variables in SAGE:

```
sage: var('a1 a2 a3 a4 a6')
sage: E = EllipticCurve([a1,a2,a3,a4,a6]); E
Elliptic Curve defined by
     y^2 + a1*x*y + a3*y = x^3 + a2*x^2 + a4*x + a6
over Symbolic Ring
sage: eqn = SR(E); eqn
(y^2 + a1*x*y + a3*y) == (x^3 + a2*x^2 + a4*x + a6)
sage: F = eqn.lhs() - eqn.rhs(); F
y^2 + a1*x*y + a3*y - x^3 - a2*x^2 - a4*x - a6
sage: G = w^3*F(x=t/s, y=-1/w); G.expand()
-t^3 - a2*w*t^2 - a4*w^2*t - a1*w*t - a6*w^3 - a3*w^2 + w
```

Example 2.26. We use SAGE to compute the formal power series $x(t)$ and $y(t)$ for the rank 1 elliptic curve 37a.

```
sage: E = EllipticCurve('37a'); E
Elliptic Curve defined by y^2 + y = x^3 - x over Rational Field
sage: F = E.formal_group(); F
Formal Group associated to the Elliptic Curve defined by
y^2 + y = x^3 - x over Rational Field
sage: x = F.x(prec=8); x
t^-2 - t + t^2 - t^4 + 2*t^5 - t^6 - 2*t^7 + O(t^8)
sage: y = F.y(prec=8); y
-t^-3 + 1 - t + t^3 - 2*t^4 + t^5 + 2*t^6 - 6*t^7 + O(t^8)
```

Notice that the power series satisfy the equation of the curve.

```
sage: y^2 + y == x^3 - x
True
```

Recall that $\omega_E = \frac{dx}{2y+\underline{a}_1x+\underline{a}_3}$ is the differential on a fixed choice of Weierstrass equation for E. Let

$$\omega(t) = \frac{dx}{2y + \underline{a}_1 x + \underline{a}_3} \in \mathbb{Q}((t))dt$$

be the formal invariant holomorphic differential on E.

Example 2.27. Continuing the above example, we compute the formal differential on E:

```
sage: F.differential(prec=8)
1 + 2*t^3 - 2*t^4 + 6*t^6 - 12*t^7 + O(t^8)
```

We can also compute $\omega(t)$ directly from the definition:

```
sage: x.derivative()/(2*y+1)
1 + 2*t^3 - 2*t^4 + 6*t^6 - 12*t^7 + 6*t^8 + 20*t^9 + O(t^10)
```

The following theorem, which is proved in [**MT91**], uniquely determines a power series $\sigma \in t\mathbb{Z}_p[[t]]$ and constant $c \in \mathbb{Z}_p$.

Theorem 2.28 (Mazur-Tate). *There is exactly one odd function $\sigma(t) = t + \cdots \in t\mathbb{Z}_p[[t]]$ and constant $c \in \mathbb{Z}_p$ that together satisfy the differential equation*

$$x(t) + c = -\frac{d}{\omega}\left(\frac{1}{\sigma}\frac{d\sigma}{\omega}\right), \tag{2.5.6}$$

where ω is the invariant differential $dx/(2y + a_1x + a_3)$ associated with our chosen Weierstrass equation for E.

The above theorem produces a (very inefficient) algorithm to compute c and $\sigma(t)$. Just view c as a formal indeterminate and compute $\sigma(t) \in \mathbb{Q}[c][[t]]$, then obtain constraints on c using that the coefficients of σ must be in $\mathbb{Z}_p$. These determine c to some precision, which increases as we compute $\sigma(t)$ to higher precision. Until recently this was the only known way to compute c and $\sigma(t)$ – fortunately the method of [**MST06**] is much faster in general.

Definition 2.29 (Canonical p-adic Height)**.** Let E be an elliptic curve over $\mathbb{Q}$ with good ordinary reduction at the odd prime p. Let $\log_p$, d, and $\sigma(t)$ be as above and suppose $P \in E(\mathbb{Q})$ and that nP is a nonzero multiple of P such that nP reduces to the identity component of the Néron model of E at each prime of bad reduction. Then the *p-adic canonical height* of P is

$$h_p(P) = \frac{1}{n^2} \cdot \frac{1}{p} \cdot \log_p\left(\frac{\sigma(P)}{d(P)}\right).$$

Definition 2.30 (p-adic Regulator)**.** The *p-adic regulator* of E is the discriminant (well defined up to sign) of the bilinear $\mathbb{Q}_p$-valued pairing

$$(P,Q)_p = h_p(P) + h_p(Q) - h_p(P+Q).$$

Conjecture 2.31 (Schneider)**.** *The p-adic regulator $\mathrm{Reg}_p(E)$ is nonzero.*

Theorem 2.32 (Kato, Schneider, et al.)**.** *Let E be an elliptic curve over $\mathbb{Q}$ with good ordinary reduction at the odd prime p and assume that the p-adic Galois representation $\rho_{E,p}$ is surjective. If*

$$\mathrm{ord}_T(\mathcal{L}_p(E,T)) \leq \mathrm{rank}\, E(\mathbb{Q}),$$

then $\#\text{Ш}(E/\mathbb{Q})(p)$ is finite. Moreover, if $\mathrm{Reg}_p(E)$ is nonzero, then

$$\mathrm{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)) \leq \mathrm{ord}_p\left(\frac{\mathcal{L}_p^*(E,0)}{\prod c_\ell \cdot \mathrm{Reg}_p(E)}\right).$$

Chapter 3

Heegner Points and Kolyvagin's Euler System

3.1. CM Elliptic Curves

In this section we state, and in some cases sketch proofs of, some basic facts about CM elliptic curves.

If E is an elliptic curve over a field K we let $\mathrm{End}(E/K)$ be the ring of all endomorphisms of E that are defined over K.

Definition 3.1 (CM Elliptic Curve)**.** An elliptic curve E over a subfield of $\mathbb{C}$ has *complex multiplication* if $\mathrm{End}(E/\mathbb{C}) \neq \mathbb{Z}$.

Remark 3.2. If E is an elliptic curve over $\mathbb{Q}$, then $\mathrm{End}(E/\mathbb{Q}) = \mathbb{Z}$. This is true even if E has complex multiplication, in which case the complex multiplication must be defined over a bigger field than $\mathbb{Q}$. The reason $\mathrm{End}(E/\mathbb{Q}) = \mathbb{Z}$ is because $\mathrm{End}(E/\mathbb{Q})$ acts faithfully on the 1-dimensional $\mathbb{Q}$-vector space of invariant holomorphic differentials on E over $\mathbb{Q}$ and $\mathrm{End}(E/\mathbb{Q})$ is finitely generated as a $\mathbb{Z}$-module.

A *complex lattice* $\Lambda \subset \mathbb{C}$ is a subgroup abstractly isomorphic to $\mathbb{Z} \times \mathbb{Z}$ such that $\mathbb{R}\Lambda = \mathbb{C}$. Using the Weirestrass $\wp$-function associated to the lattice Λ, one proves that there is a group isomorphism

$$\mathbb{C}/\Lambda \cong E_\Lambda(\mathbb{C}),$$

where E_Λ is an elliptic curve over $\mathbb{C}$. Conversely, if E is any elliptic curve over $\mathbb{C}$, then there is a lattice Λ such that $E = E_\Lambda$. Explicitly, if ω_E is an invariant differential we may take Λ to be the lattice of all periods $\int_\gamma \omega_E \in \mathbb{C}$, where γ runs through the integral homology $\mathrm{H}_1(E(\mathbb{C}), \mathbb{Z})$.

Proposition 3.3. *Let Λ_1 and Λ_2 be complex lattices. Then*

$$\mathrm{Hom}(\mathbb{C}/\Lambda_1, \mathbb{C}/\Lambda_2) \cong \{\alpha \in \mathbb{C} : \alpha\Lambda_1 \subset \Lambda_2\},$$

where the homomorphisms on the left side are as elliptic curves over $\mathbb{C}$. Moreover, the complex number $\alpha \in \mathbb{C}$ corresponds to the homomorphism $[\alpha]$ induced by multiplication by α, and the kernel of $[\alpha]$ is isomorphic to $\Lambda_2/(\alpha\Lambda_1)$.

Corollary 3.4. *If α is any nonzero complex number and Λ is a lattice, then $\mathbb{C}/\Lambda \cong \mathbb{C}/(\alpha\Lambda)$.*

Proof. Since multiplication by α sends Λ into $\alpha\Lambda$, Proposition 3.3 implies that α defines a homomorphism with 0 kernel, hence an isomorphism. □

Now suppose $E/\mathbb{C}$ is a CM elliptic curve, and let Λ be a lattice such that $E \cong E_\Lambda$. Then

$$\mathrm{End}(E/\mathbb{C}) \cong \{\alpha \in \mathbb{C} : \alpha\Lambda \subset \Lambda\}.$$

Proposition 3.5. *Let $E = E_\Lambda$ be a CM elliptic curve. Then there is a complex number ω and a quadratic imaginary field such K that*

$$\omega\Lambda \subset \mathcal{O}_K,$$

where $\mathcal{O}_K$ is the ring of integers of K. Moreover, $\mathrm{End}(E/\mathbb{C})$ is an order (=subring of rank 2) of $\mathcal{O}_K$.

Proof. Write $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$. By Corollary 3.4, we have $E_\Lambda \cong E_{\omega_1^{-1}\Lambda}$, so we may assume that $\omega_1 = 1$, i.e., that $\Lambda = \mathbb{Z} + \beta\mathbb{Z}$ for some $\beta \in \mathbb{C}$. To complete the proof, we will show that $\omega\Lambda \subset \mathcal{O}_K$ for some quadratic imaginary field K and complex number ω.

By our hypothesis that E is CM there is a complex number $\alpha \notin \mathbb{Z}$ such that $\alpha\Lambda \subset \Lambda$. Fixing a basis for Λ, we see that α acts on Λ via a 2×2 integral matrix, so satisfies a quadratic equation. Thus α is an algebraic integer of degree 2. In particular, there are integers a, b, c, d such that

$$\alpha 1 = a + b\beta, \qquad \text{and} \qquad \alpha\beta = c + d\beta.$$

Since $\alpha \notin \mathbb{Z}$, the first equation above implies that $\beta \in \mathbb{Q}(\alpha)$, so since $\beta \notin \mathbb{Q}$, $\mathbb{Q}(\beta) = \mathbb{Q}(\alpha)$. Note that $\beta \notin \mathbb{R}$ since Λ is a lattice with basis 1 and β, so $K = \mathbb{Q}(\beta)$ is a quadratic imaginary field. Thus the ring $\mathrm{End}(E/\mathbb{C})$ generated by all such α is an order in the ring $\mathcal{O}_K$ of integers of an imaginary quadratic field. Finally, since $\beta \in K$, there is a complex number ω such that $\omega(\mathbb{Z} + \mathbb{Z}\beta) \subset \mathcal{O}_K$, where ω is chosen so that $\omega\beta \in \mathcal{O}_K$. □

3.1.1. The Set of CM Elliptic Curves with Given CM.

Definition 3.6 (Fractional Ideal)**.** A *fractional ideal* $\mathfrak{a}$ of a number field K is an $\mathcal{O}_K$-submodule of K that is isomorphic to $\mathbb{Z}^{[K:\mathbb{Q}]}$ as an abelian group. In particular, $\mathfrak{a}$ is nonzero.

If $\mathfrak{a}$ is a fractional ideal, the *inverse* $\mathfrak{a}^{-1}$ of $\mathfrak{a}$, which is the set of $x \in K$ such that $x\mathfrak{a} \subset \mathcal{O}_K$, is also a fractional ideal. Moreover, $\mathfrak{a}\mathfrak{a}^{-1} = \mathcal{O}_K$.

Fix a quadratic imaginary field K. Let $\mathrm{Ell}(\mathcal{O}_K)$ be the set of $\mathbb{C}$-isomorphism classes of elliptic curves $E/\mathbb{C}$ with $\mathrm{End}(E) \cong \mathcal{O}_K$. By the above results we may also view $\mathrm{Ell}(\mathcal{O}_K)$ as the set of lattices Λ with $\mathrm{End}(E_\Lambda) \cong \mathcal{O}_K$.

If $\mathfrak{a}$ is a fractional $\mathcal{O}_K$ ideal, then $\mathfrak{a} \subset K \subset \mathbb{C}$ is a lattice in $\mathbb{C}$. For the elliptic curve $E_\mathfrak{a}$ we have

$$\mathrm{End}(E_\mathfrak{a}) = \mathcal{O}_K,$$

because $\mathfrak{a}$ is an $\mathcal{O}_K$-module by definition. Since rescaling a lattice produces an isomorphic elliptic curve, for any nonzero $c \in K$ the fractional ideals $\mathfrak{a}$ and $c\mathfrak{a}$ define the same elements of $\mathrm{Ell}(\mathcal{O}_K)$.

The *class group* $\mathrm{Cl}(\mathcal{O}_K)$ is the group of fractional ideals modulo principal fractional ideals. If $\mathfrak{a}$ is a fractional $\mathcal{O}_K$ ideal, denote by $\overline{\mathfrak{a}}$ its ideal class in the class group $\mathrm{Cl}(\mathcal{O}_K)$ of K. We have a natural map

$$\mathrm{Cl}(\mathcal{O}_K) \to \mathrm{Ell}(\mathcal{O}_K),$$

which sends $\overline{\mathfrak{a}}$ to $E_\mathfrak{a}$.

Theorem 3.7. *Fix a quadratic imaginary field K, and let Λ be a lattice in $\mathbb{C}$ such that $E_\Lambda \in \mathrm{Ell}(\mathcal{O}_K)$. Let $\mathfrak{a}$ and $\mathfrak{b}$ be nonzero fractional $\mathcal{O}_K$-ideals. Then*

1. *$\mathfrak{a}\Lambda$ is a lattice in $\mathbb{C}$,*
2. *We have $\mathrm{End}(E_{\mathfrak{a}\Lambda}) \cong \mathcal{O}_K$.*
3. *We have $E_{\mathfrak{a}\Lambda} \cong E_{\mathfrak{b}\Lambda}$ if and only if $\overline{\mathfrak{a}} = \overline{\mathfrak{b}}$.*

Thus there is a well-defined action of $\mathrm{Cl}(\mathcal{O}_K)$ on $\mathrm{Ell}(\mathcal{O}_K)$ given by

$$\overline{\mathfrak{a}}E_\Lambda = E_{\overline{\mathfrak{a}}^{-1}\Lambda}.$$

Theorem 3.8. *The action of $\mathrm{Cl}(\mathcal{O}_K)$ on $\mathrm{Ell}(\mathcal{O}_K)$ is simply transitive.*

Example 3.9. Let $K = \mathbb{Q}(\sqrt{-23})$. Then the class number h_K is 3. An elliptic curve with CM by $\mathcal{O}_K$ is $\mathbb{C}/(\mathbb{Z} + (1 + \sqrt{-23})/2\mathbb{Z})$, and one can obtain the other two elements of $\mathrm{Ell}(\mathcal{O}_K)$ by multiplying the lattice $\mathbb{Z} + (1 + \sqrt{-23})/2\mathbb{Z}$ by two representative ideal classes for $\mathrm{Cl}(\mathcal{O}_K)$.

3.1.2. Class Field Theory.

Class field theory makes sense for arbitrary number fields, but for simplicity in this section and because it is all that is needed for our application to the BSD conjecture, we assume henceforth that K is a totally imaginary number field, i.e., one with no real embeddings.

Let L/K be a finite abelian extension of number fields, and let $\mathfrak{a}$ be any unramified prime ideal in $\mathcal{O}_K$. Let $\mathfrak{b}$ be an prime of $\mathcal{O}_L$ over $\mathfrak{a}$ and consider the extension $k_{\mathfrak{b}} = \mathcal{O}_L/\mathfrak{b}$ of the finite field $k_{\mathfrak{a}} = \mathcal{O}_K/\mathfrak{a}$. There is an element $\overline{\sigma} \in \mathrm{Gal}(k_{\mathfrak{b}}/k_{\mathfrak{a}})$ that acts via qth powering on $k_{\mathfrak{b}}$, where $q = \#k_{\mathfrak{a}}$. A basic fact one proves in algebraic number theory is that there is an element $\sigma \in \mathrm{Gal}(L/K)$ that acts as $\overline{\sigma}$ on $\mathcal{O}_L/\mathfrak{b}$; moreover, replacing $\mathfrak{b}$ by a different ideal over $\mathfrak{a}$ just changes σ by conjugation. Since $\mathrm{Gal}(L/K)$ is abelian it follows that σ is uniquely determined by $\mathfrak{a}$. The association $\mathfrak{a} \mapsto \sigma = [\mathfrak{a}, L/K]$ is called the *Artin reciprocity map.*

Exercise 3.10. Prove that if an unramified prime $\mathfrak{p}$ of K splits completely in an abelian exension L/K, then $[\mathfrak{p}, L/K] = 1$.

Let $\mathfrak{c}$ be an integral ideal divisible by all primes of K that ramify in L, and let $I(\mathfrak{c})$ be the group of fractional ideals that are coprime to $\mathfrak{c}$. Then the reciprocity map extends to a map

$$I(\mathfrak{c}) \to \mathrm{Gal}(L/K) \qquad a \mapsto [\mathfrak{a}, L/K]$$

Let

$$P(\mathfrak{c}) = \{(\alpha) : \alpha \in K^*, \quad \alpha \equiv 1 \pmod{\mathfrak{c}}\}.$$

Here $\alpha \equiv 1 \pmod{\mathfrak{c}}$ means that $\mathrm{ord}_{\mathfrak{p}}(\alpha - 1) \geq \mathrm{ord}_p(\mathfrak{c})$ for each prime divisor $\mathfrak{p} \mid \mathfrak{c}$.

Definition 3.11 (Conductor of Extension)**.** The *conductor* of an abelian extension L/K is the largest (nonzero) integral ideal $\mathfrak{c} = \mathfrak{c}_{L/K}$ of $\mathcal{O}_K$ such that $[(\alpha), L/K] = 1$ for all $\alpha \in K^*$ such that $\alpha \equiv 1 \pmod{\mathfrak{c}}$.

Proposition 3.12. *The conductor of L/K exists.*

If $\mathfrak{c} = \mathfrak{c}_{L/K}$ is the conductor of L/K then Artin reciprocity induces a group homomorphism

$$I(\mathfrak{c})/P(\mathfrak{c}) \to \mathrm{Gal}(L/K).$$

Definition 3.13 (Ray Class Field)**.** Let $\mathfrak{c}$ be a nonzero integral ideal of $\mathcal{O}_K$. A *ray class field* associated to $\mathfrak{c}$ is a finite abelian extension $K_{\mathfrak{c}}$ of K such that whenever L/K is an abelian extension such that $\mathfrak{c}_{L/K} \mid \mathfrak{c}$, then $L \subset K_{\mathfrak{c}}$.

Theorem 3.14 (Existence Theorem of Class Field Theory)**.** *Given any nonzero integral ideal $\mathfrak{c}$ of $\mathcal{O}_K$ there exists a unique ray class field $K_{\mathfrak{c}}$ associated to $\mathfrak{c}$, and the conductor of $K_{\mathfrak{c}}$ divides $\mathfrak{c}$.*

Theorem 3.15 (Reciprocity Law of Class Field Theory)**.** *Let L/K be a finite abelian extension.*

(1) *The Artin map is a surjective homomorphism $I(\mathfrak{c}_{L/K}) \to \mathrm{Gal}(L/K)$.*

(2) *The kernel of the Artin map is $N_{L/K}(I_L)\cdot P(\mathfrak{c}_{L/K})$, where $N_{L/K}(I_L)$ is the group of norms from L to K of the fractional ideals of L.*

Definition 3.16 (Hilbert Class Field)**.** The *Hilbert class field* of a number field K is the maximal unramified abelian extension of K.

In particular, since the Hilbert class field is unramified over K, we have:

Theorem 3.17. *Let K be a number field and let H be the Hilbert class field of K. The Artin reciprocity map induces an isomorphism*

$$\mathrm{Cl}(\mathcal{O}_K) \xrightarrow{\cong} \mathrm{Gal}(H/K).$$

3.1.3. The Field of Definition of CM Elliptic Curves.

Theorem 3.18. *Let F be an elliptic curve over $\mathbb{C}$ with CM by $\mathcal{O}_K$, where K is a quadratic imaginary field. Let H be the Hilbert Class Field of K.*

(1) *There is an elliptic curve E defined over K such that $F \cong E_{\mathbb{C}}$.*

(2) *The $\mathrm{Gal}(H/K)$-conjugates of E are representative elements for $\mathrm{Ell}(\mathcal{O}_K)$.*

(3) *If $\sigma \in \mathrm{Gal}(H/K)$ corresponds via Artin reciprocity to $\overline{\mathfrak{a}} \in \mathrm{Cl}(\mathcal{O}_K)$, then*
$$E^{\sigma} = \overline{\mathfrak{a}}E.$$

Theorem 3.18 generalizes in a natural way to the more general situation in which $\mathcal{O}_K$ is replaced by an order $\mathcal{O}_f = \mathbb{Z}+f\mathcal{O}_K \subset \mathcal{O}_K$. Then the Hilbert class field is replaced by the ray class field K_f, which is a finite abelian extension of H that is unramified outside f (see Definition 3.13 above). There is an elliptic curve E defined over K_f whose endomorphism ring is $\mathcal{O}_f$, and the set of $\mathrm{Gal}(K_f/K)$-conjugates of E forms a set of representatives for $\mathrm{Ell}(\mathcal{O}_f)$. Moreover, the group $I(\mathfrak{c}_{L/K})/(N \cdot P(\mathfrak{c}_{L/K}))$ of Theorem 3.15 acts simply transitively on $\mathrm{Ell}(\mathcal{O}_f)$, and the action of $\mathrm{Gal}(K_f/K)$ on the set of conjugates of E is consistent with the Artin reciprocity map.

3.2. Heegner Points

Let E be an elliptic curve defined over $\mathbb{Q}$ with conductor N, and fix a modular parametrization $\pi_E : X_0(N) \to E$.

Let K be a quadratic imaginary field such that the primes dividing N are all unramified and split in K. For simplicity, we will also assume that $K \neq \mathbb{Q}(i), \mathbb{Q}(\sqrt{-3})$. Let $\mathcal{N}$ be an integral ideal of $\mathcal{O}_K$ such that $\mathcal{O}_K/\mathcal{N} \cong \mathbb{Z}/N\mathbb{Z}$. Then $\mathbb{C}/\mathcal{O}_K$ and $\mathbb{C}/\mathcal{N}^{-1}$ define two elliptic curves over $\mathbb{C}$, and since $\mathcal{O}_K \subset \mathcal{N}^{-1}$, there is a natural map

$$\mathbb{C}/\mathcal{O}_K \to \mathbb{C}/\mathcal{N}^{-1}. \tag{3.2.1}$$

By Proposition 3.3 the kernel of this map is

$$\mathcal{N}^{-1}/\mathcal{O}_K \cong \mathcal{O}_K/\mathcal{N} \cong \mathbb{Z}/N\mathbb{Z}.$$

Exercise 3.19. Prove that there is an isomorphism $\mathcal{N}^{-1}/\mathcal{O}_K \cong \mathcal{O}_K/\mathcal{N}$ of finite abelian group.

The modular curve $X_0(N)$ parametrizes isomorphism classes of pairs (F, ϕ), where ϕ is an isogeny with kernel cyclic of order N. Thus $\mathbb{C}/\mathcal{O}_K$ and the isogeny (3.2.1) define an element $x_1 \in X_0(N)(\mathbb{C})$. The discussion of Section 3.1.3 along with properties of modular curves proves the following proposition.

Proposition 3.20. *We have*

$$x_1 \in X_0(N)(H),$$

where H is the Hilbert class field of K.

Definition 3.21 (Heegner point)**.** The *Heegner point* associated to K is

$$y_K = \mathrm{Tr}_{H/K}(\pi_E(x_1)) \in E(K).$$

More generally, for any integer n, let $\mathcal{O}_n = \mathbb{Z} + n\mathcal{O}_K$ be the order in $\mathcal{O}_K$ of index n. Then $\mathcal{N}_n = \mathcal{N} \cap \mathcal{O}_n$ satisfies $\mathcal{O}_n/\mathcal{N}_n \cong \mathbb{Z}/N\mathbb{Z}$, and the pair

$$(\mathbb{C}/\mathcal{O}_n,\ \mathbb{C}/\mathcal{O}_n \to \mathbb{C}/\mathcal{N}_n^{-1})$$

defines a point $x_n \in X_0(N)(K_n)$, where K_n is the ray class field of conductor n over K.

Definition 3.22 (Heegner point of conductor n)**.** The Heegner point of conductor n is

$$y_n = \pi_E(x_n) \in E(K_n).$$

3.3. Computing Heegner Points

[[This section will be my take on what's in Cohen's book and Watkins paper, hopefully generalized to compute Heegner points over ring class fields (?).]]

3.4. Kolyvagin's Euler System

3.4.1. Kolyvagin's Cohomology Classes. In this section we define Kolyvagin's cohomology classes. Later we will explain the properties that these classes have, and eventually use them to sketch a proof of finiteness of Shafarevich-Tate groups of certain elliptic curves.

We will use, when possible, similar notation to the notation Kolyvagin uses in his papers (e.g., [**Kol91**]). If A is an abelian group let $A/M = A/(MA)$. Kolyvagin writes A_M for the M-torsion subgroup, but we will instead write $A[M]$ for this group.

Let E be an elliptic curve over $\mathbb{Q}$ with no constraint on the rank of E. Fix a modular parametrization $\pi : X_0(N) \to E$, where N is the conductor of E. Let K be a quadratic imaginary field with discriminant D that satisfies the Heegner hypothesis for E, so each prime dividing N splits in K, and assume for simplicity that $D \neq -3, -4$.

Let $\mathcal{O}_K$ be the ring of integer of K. Since K satisfies the Heegner hypothesis, there is an ideal $\mathcal{N}$ in $\mathcal{O}_K$ such that $\mathcal{O}_K/\mathcal{N}$ is cyclic of order N. For any positive integer λ, let K_λ be the ray class field of K associated to the conductor λ (see Definition 3.13). Recall that K_λ is an abelian extension of K that is unramified outside λ, whose existence is guaranteed by class field theory. Let $\mathcal{O}_\lambda = \mathbb{Z} + \lambda\mathcal{O}_K$ be the order in $\mathcal{O}_K$ of conductor λ, and let $\mathcal{N}_\lambda = \mathcal{N} \cap \mathcal{O}_\lambda$. Let

$$z_\lambda = [(\mathbb{C}/\mathcal{O}_\lambda, \mathcal{N}_\lambda^{-1}/\mathcal{O}_\lambda)] = X_0(N)(K_\lambda)$$

be the Heegner point associated to λ. Also, let

$$y_\lambda = \pi(z_\lambda) \in E(K_\lambda)$$

be the image of the Heegner point on the curve E.

Let $R = \mathrm{End}(E/\mathbb{C})$, and let $B(E)$ be the set of primes $\ell \geq 3$ in $\mathbb{Z}$ that do not divide the discriminant of R and are such that the image of the representation

$$\rho_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(\mathrm{Tate}_\ell(E))$$

contains $\mathrm{Aut}_R(\mathrm{Tate}_\ell(E))$, where $\mathrm{Aut}_R(\mathrm{Tate}_\ell(E))$ is the set of automorphisms that commute with the action of R on $\mathrm{Tate}_\ell(E)$. Note that if $\ell \geq 5$ the condition that $\rho_{E,\ell}$ is surjective is equivalent to the simpler condition that

$$\overline{\rho}_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_R(E[\ell])$$

is surjective. The set $B(E)$ contains all but finitely many primes, by theorems of Serre [**Ser72**], Mazur [**Maz78**], and CM theory, and one can compute $B(E)$.

```
sage: E = EllipticCurve('11a')
sage: E.non_surjective()
[(5, '5-torsion')]
sage: E = EllipticCurve('389a')
sage: E.non_surjective()
[]
```

Fix a prime $\ell \in B(E)$. We next introduce some very useful notation. Let Λ^1 denote the set of all primes $p \in \mathbb{Z}$ such that $p \nmid N$, p remains prime in $\mathcal{O}_K$, and for which

$$n(p) = \mathrm{ord}_\ell(\gcd(p+1, a_p)) \geq 1.$$

For any positive integer r, let Λ^r denote the set of all products of r distinct primes in Λ^1; by definition $\Lambda^0 = \{1\}$. Finally, let

$$\Lambda = \bigcup_{r \geq 0} \Lambda^r.$$

For any $r > 0$ and $\lambda \in \Lambda^r$, let

$$n(\lambda) = \min_{p \mid \lambda} n(p)$$

be the "worst" of all the powers of p that divide $\gcd(p+1, a_p)$. If $\lambda = 1$, set $n(\lambda) = +\infty$.

Fix an element $\lambda \in \Lambda$, with $\lambda \neq 1$, and consider the ℓ-power

$$M = M_\lambda = \ell^{n(\lambda)}.$$

Recall from Section 2.2.1 that we associate to the short exact sequence

$$0 \to E[M] \to E \xrightarrow{[M]} E \to 0$$

an exact sequence

$$0 \to E(K)/M \to \mathrm{H}^1(K, E[M]) \to H^1(K, E)[M] \to 0.$$

Our immediate goal is to construct an *interesting* cohomology class

$$c_\lambda \in \mathrm{H}^1(K, E[M]).$$

If L/K is any Galois extension, we have (see Section 2.1.2 for most of this) an exact sequence

(3.4.1)

$$0 \to \mathrm{H}^1(L/K, E[M](L)) \to \mathrm{H}^1(K, E[M]) \to \mathrm{H}^1(L, E[M])^{\mathrm{Gal}(L/K)} \to 0.$$

Lemma 3.23. *We have $E[M](K_\lambda) = 0$.*

Proof. For simplicity we prove the statement only in the non-CM case. The integer M is a power of a prime ℓ, so it suffices to show that $E[\ell](K_\lambda) = 0$. Since $\ell \in B(E)$ the Galois representation

$$\overline{\rho}_{E,\ell} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_\ell)$$

is surjective. The group $\mathrm{GL}_2(\mathbb{F}_\ell)$ acts transitively on $(\mathbb{F}_\ell)^2$, so the $G_{\mathbb{Q}}$ orbit of any nonzero point in $E[\ell](\overline{\mathbb{Q}})$ is equal to the set of all nonzero points in $E[\ell](\overline{\mathbb{Q}})$. By class field theory, the extension K_λ of $\mathbb{Q}$ is Galois, so if $E[\ell](K_\lambda)$ is nonzero, then it is equal to $E[\ell](\overline{\mathbb{Q}})$. Using properties of the Weil pairing, we see that the field generated by the coordinates of the elements of $E[\ell](\overline{\mathbb{Q}})$ contains the cyclotomic field $\mathbb{Q}(\zeta_\ell)$, which is a field totally ramified at ℓ. But $K \cap \mathbb{Q}(\zeta_\ell) = \mathbb{Q}$, since $\mathrm{disc}(K) \neq -3, -4$, and K_λ is ramified only at primes in Λ^1 and $\ell \notin \Lambda^1$. We conclude that $K_\lambda \cap \mathbb{Q}(\zeta_\ell) = \mathbb{Q}$, so we must have $E[\ell](K_\lambda) = 0$. (Compare [**Gro91**, Lem. 4.3].) □

Thus (3.4.1) with $L = K_\lambda$ becomes

$$\mathrm{H}^1(K, E[M]) \xrightarrow{\cong} \mathrm{H}^1(K_\lambda, E[M])^{G_\lambda} \tag{3.4.2}$$

where $G_\lambda = \mathrm{Gal}(K_\lambda/K)$. Putting this together, we obtain the following commutative diagram with exact rows and columns:

$$\begin{tikzcd}
0 \arrow[r] & (E(K_\lambda)/M)^{G_\lambda} \arrow[r] & \mathrm{H}^1(K_\lambda, E[M])^{G_\lambda} \arrow[r] & \mathrm{H}^1(K_\lambda, E)[M]^{G_\lambda} & \\
0 \arrow[r] & E(K)/M \arrow[r] \arrow[u, hook] & \mathrm{H}^1(K, E[M]) \arrow[r] \arrow[u, "\cong", "\mathrm{res}"'] & \mathrm{H}^1(K, E)[M] \arrow[r] \arrow[u, "\mathrm{res}"] & 0 \\
 & & & \mathrm{H}^1(K_\lambda/K, E)[M] \arrow[u, hook, "\mathrm{inf}"] &
\end{tikzcd}$$

Thus to construct $c_\lambda \in \mathrm{H}^1(K, E[M])$, it suffices to construct a class $c'_\lambda \in \mathrm{H}^1(K_\lambda, E[M])$ that is invariant under the action of G_λ. We will do this by constructing an element of $E(K_\lambda)$ and using the inclusion

$$E(K_\lambda)/M \hookrightarrow \mathrm{H}^1(K_\lambda, E[M]). \tag{3.4.3}$$

In particular, we will construct an element of the group $E(K_\lambda)/M$ that is invariant under the action of G_λ.

Recall that $y_\lambda \in E(K_\lambda)$. Unfortunately, there is no reason that the class

$$[y_\lambda] \in E(K_\lambda)/M$$

should be invariant under the action of G_λ. To deal with this problem, Kolyvagin introduced a *new and original idea* which we now explain.

Let $H = K_1$ be the Hilbert class field of K. Write $\lambda = p_1 \cdots p_r$, and for each $p = p_i$ let $G_p = \mathrm{Gal}(K_p/K)$ where K_p is the ray class field associated to p. Class field theory implies that the natural map

$$\mathrm{Gal}(K_\lambda/K_1) \cong\!\rightarrow G_{p_1} \times G_{p_2} \times \cdots \times G_{p_r}$$

is an isomorphism. Moreover, each group G_{p_i} is cyclic of order $p_i + 1$. For each $p = p_i$, let σ_p be a fixed choice of generator of G_p, and let

$$\mathrm{Tr}_p = \sum_{\sigma \in G_p} \sigma \in \mathbb{Z}[G_p].$$

Finally, let $D_p \in \mathbb{Z}[G_p]$ be any solution of the equation

$$(\sigma_p - 1) \cdot D_p = p + 1 - \mathrm{Tr}_p\,. \tag{3.4.4}$$

For example, Kolyvagin always takes

$$D_p = \sum_{i=1}^{p} i\sigma_p^i = -\sum_{i=1}^{p+1} (\sigma_p^i - 1)/(\sigma_p - 1).$$

Notice that the choice of D_p is well defined up to addition of elements in $\mathbb{Z}\,\mathrm{Tr}_p$. Let

$$D_\lambda = \prod D_p = D_{p_1} \cdot D_{p_2} \cdot \cdots \cdot D_{p_r} \in \mathbb{Z}[G_\lambda].$$

Finally, let S be a set of coset representatives for $\mathrm{Gal}(K_\lambda/K_1)$ in $G_\lambda = \mathrm{Gal}(K_\lambda/K)$, and let

$$J_\lambda = \sum_{\sigma \in S} \sigma \in \mathbb{Z}[G_\lambda].$$

Let

$$P_\lambda = J_\lambda D_\lambda y_\lambda \in E(K_\lambda).$$

Note that if $\lambda = 1$, then $K_\lambda = K_1$, so

$$P_1 = J_1 y_\lambda = \mathrm{Tr}_{K_1/K}(y_\lambda) = y_K \in E(K).$$

Before proving that we can use P_λ to define a cohomology class in $\mathrm{H}^1(K, E[M])$, we state two crucial facts about the structure of the Heegner points y_λ.

Proposition 3.24. *Write $\lambda = p\lambda'$, and let $a_p = a_p(E) = p + 1 - \#E(\mathbb{F}_p)$.*

(1) *We have*

$$\mathrm{Tr}_p(y_\lambda) = a_p y_{\lambda'}$$

in $E(K_{\lambda'})$.

(2) *Each prime factor $\wp_\lambda$ of p in K_λ divides a unique prime $\wp_{\lambda'}$ of $K_{\lambda'}$, and we have a congruence*

$$y_\lambda \equiv \mathrm{Frob}(\wp_{\lambda'})(y_{\lambda'}) \pmod{\wp_\lambda}.$$

Proof. See [**Gro91**, Prop. 3.7]. The proof uses a description of the action of Hecke operators on modular curves. □

Proposition 3.25. *The class* $[P_\lambda]$ *of* P_λ *in* $E(K_\lambda)/M$ *is fixed by* G_λ.

Proof. We follow the proof of [**Gro91**, Prop. 3.6]. It suffices to show that $[D_\lambda y_\lambda]$ is fixed by σ_p for each prime $p \mid \lambda$, since the σ_p generate $\mathrm{Gal}(K_\lambda/K_1)$, the elements of the set S of coset representatives fix the image of J_λ, and G_λ is generated by the σ_p and S. Thus we will prove that

$$(\sigma_p - 1)D_\lambda y_\lambda \in ME(K_\lambda)$$

for each $p \mid \lambda$.

Write $\lambda = pm$. By (3.4.4), we have in $\mathbb{Z}[G_\lambda]$ that

$$(\sigma_p - 1)D_\lambda = (\sigma_p - 1)D_p D_m = (p + 1 - \mathrm{Tr}_p)D_m,$$

so using Proposition 3.24 we have

$$\begin{aligned}(\sigma_p - 1)D_\lambda y_\lambda &= (p + 1 - \mathrm{Tr}_p)D_m y_\lambda\\ &= (p+1)D_m y_\lambda - D_m \,\mathrm{Tr}_p(y_\lambda)\\ &= (p+1)D_m y_\lambda - a_p D_m y_{\lambda'}\end{aligned}$$

Since $p \in \Lambda^1$ and $M = \ell^{n(p)}$ and $n(p) = \min(\mathrm{ord}_\ell(p+1), \mathrm{ord}_\ell(a_p))$, we have $M \mid p+1$ and $M \mid a_p$. Thus $(p+1)D_m y_\lambda \in ME(K_\lambda)$ and $a_p y_{\lambda'} \in ME(K_\lambda)$, which proves the proposition. □

We have now constructed an element of $E(K_\lambda)/M$ that is fixed by G_λ. Via (3.4.3) this defines an element $c'_\lambda \in \mathrm{H}^1(K_\lambda, E[M])$. But then using (3.4.2) we obtain our sought after class $c_\lambda \in \mathrm{H}^1(K, E[M])$.

We will also be interested in the image d_λ of c_λ in $\mathrm{H}^1(K, E)[M]$.

Proposition 3.26. *If* v *is archimedean or* $v \nmid \lambda$, *then*

$$\mathrm{res}_v(d_\lambda) = 0.$$

Proof. If v is archimedean we are done, since $K_v = \mathbb{C}$ is algebraically closed. Otherwise, the class d_λ splits over K_λ and K_λ is unramified at v, so

$$\mathrm{res}_v(d_\lambda) \in \mathrm{H}^1(K_v^{\mathrm{unr}}/K_v, E).$$

But the latter group is isomorphic to the component group of E at v, and a theorem of Gross-Zagier implies that the Heegner point maps to the identity component. (See [**Gro91**, Prop. 6.2] for more details.) □

Proposition 3.27. *Write $\lambda = pm$ and let $\wp = p\mathcal{O}_K$ be the unique prime ideal of K dividing p. Let v be a place of K_m that divides $\wp$. Then the order of $\mathrm{res}_\wp(d_\lambda)$ is the same as the order of*

$$[P_m] \in E(K_\wp)/ME(K_\wp),$$

where $K_\wp$ denotes the completion of K at $\wp$. (Note that $\wp$ splits completely in K_m/K by class field theory, since $\wp = p\mathcal{O}_K$ is principal and coprime to m, so $P_m \in E(K_\wp)$.)

Proof. See [**Gro91**, Prop. 6.2] for the case $M = \ell$. The argument involves standard properties of Galois cohomology of elliptic curves, some diagram chasing, reduction modulo a prime, and use of formal groups. □

Next we consider a consequence of Proposition 3.27 when y_K is not a torsion point. Note that y_K nontorsion implies that $y_K \notin ME(K)$ for all but finitely many M. Moreover, the Gross-Zagier theorem implies that y_K is nontorsion if and only if $\mathrm{ord}_{s=1} L(E, s) \leq 1$.

Proposition 3.28. *Suppose that $y_K \in E(K)$ is not divisible by M. Then there are infinitely many $p \in \Lambda^1$ such that $d_p \in \mathrm{H}^1(K, E)[M]$ is nonzero.*

Proof. This follows from Proposition 3.27 with $m = 1$ and the Chebotarev density theorem. See e.g., [**Ste02**, §4.1] for a proof. □

Remark 3.29. See, e.g., [**Ste02**] for an application of this idea to a problem raised by Lang and Tate in [**LT58**].

Theorem 3.30 (Kolyvagin)**.** *Suppose E is a modular elliptic curve over $\mathbb{Q}$ and K is a quadratic imaginary field that satisfies the Heegner hypothesis for E and is such that $y_K \in E(K)$ is nontorsion. Then $E(K)$ has rank 1 and*

$$\#\text{Ш}(E/K) \mid b \cdot [E(K) : \mathbb{Z}y_K]^2,$$

where b is a positive integer divisible only by primes $\ell \in B(E)$ (i.e., for which the ℓ-adic representation is not as surjective as possible).

Proof. See the entire paper [**Gro91**]. Kolyvagin proves this theorem by bounding $\mathrm{Sel}^{(M)}(E/K)$ for various M using Proposition 3.28 in conjunction with a careful study of various pairings coming from Galois cohomology, the Weil pairing, Tate local daulity, etc. Since

$$0 \to E(K)/ME(K) \to \mathrm{Sel}^{(M)}(E/K) \to \text{Ш}(E/K),$$

a bound on the Selmer group translates into a bound on $E(K)$ and $\text{Ш}(E/K)$. □

After Kolyvagin proved his theorem, independently Murty-Murty, Bump-Friedberg-Hoffstein, Waldspurger, each proved that infinitely many such quadratic imaginary K always exists so long as E has analytic rank 0 or 1. Also, Taylor and Wiles proved that every E over $\mathbb{Q}$ is modular. Thus we have the following theorem:

Theorem 3.31. *Suppose E is an elliptic curve over $\mathbb{Q}$ with*

$$r_{E,\mathrm{an}} = \mathrm{ord}_{s=1} L(E,s) \le 1.$$

Then $E(\mathbb{Q})$ has rank $r_{E,\mathrm{an}}$, the group $Ш(E/\mathbb{Q})$ is finite, and there is an explicit computable upper bound on $\#Ш(E/\mathbb{Q})$.

The author has computed the upper bound of the theorem for all elliptic curves with conductor up to 1000 and $r_{E,\mathrm{an}} \le 1$.

3.4.2. Kolyvagin's Conjectures. What about curves E with $r_{E,\mathrm{an}} \ge 2$? Suppose that E is an elliptic curve over $\mathbb{Q}$ with $r_{E,\mathrm{an}} \ge 2$. In the short paper [**Kol91**], Kolyvagin states an amazing structure theorems for Selmer groups assuming the following unproved conjecture, which is the appropriate generalization of the condition that P_1 has infinite order.

Conjecture 3.32 (Kolyvagin [**Kol91**])**.** *Let E be any elliptic curve over $\mathbb{Q}$ and fix a prime $\ell \in B(E)$ and a prime power $M = \ell^n$ of ℓ. Then there is at least one cohomology class $c_\lambda \in H^1(K, E[M])$ that is nonzero.*

So far nobody has been able to show that Conjecture 3.32 is satisfied by every elliptic curve E over $\mathbb{Q}$, though several people are currently working hard on this problem (including Vatsal and Cornut). Proposition 3.28 above implies that Conjecture 3.32 is true for elliptic curves with $r_{E,\mathrm{an}} \le 1$.

Kolyvagin also goes on in [**Kol91**] to give a *conjectural construction* of a subgroup

$$V \subset E(K)/E(K)_{\mathrm{tor}}$$

for which $\mathrm{rank}(E(\mathbb{Q})) = \mathrm{rank}(V)$. Let ℓ be an arbitrary prime, i.e., so we do not necessarily assume $\ell \in B(E)$. One can construct cohomology class $c_\lambda \in \mathrm{H}^1(K, E[M])$, so long as $\lambda \in \Lambda^{n+k_0}$, where $\ell^{k_0/2}E(\mathbf{K})[\ell^\infty] = 0$, and $\mathbf{K}$ is the compositum of all class field K_λ for $\lambda \in \Lambda$. For any $n \ge 1$, $k \ge k_0$, and $r \ge 0$, let

$$V^r_{n,k} \subset \varinjlim_m \mathrm{H}^1(K, E[\ell^m])/E(K)_{\mathrm{tor}}$$

be the subgroup generated by the images of the classes $\tau_\lambda = \tau_{\lambda,n} \in \mathrm{H}^1(K, E[\ell^n])$ where λ runs through Λ^r_{n+k}.

Conjecture 3.33 (Kolyvagin)**.** *Let E be any elliptic curve over $\mathbb{Q}$. Then for all prime numbers ℓ, there exists an integer r such that for all $k \ge k_0$ there is an n such that $V^r_{n,k} \ne 0$.*

Recall that

$$n(p) = \mathrm{ord}_\ell(\gcd(p+1, a_p)) \geq 1$$

and

$$n(\lambda) = \min_{p|\lambda} n(p).$$

Let $m'(\lambda)$ be the maximal nonnegative integer such that $P_\lambda \in \ell^{m'(\lambda)} E(K_\lambda)$. Let $m(\lambda) = m'(\lambda)$ if $m'(\lambda) < n(\lambda)$, and $m(\lambda) = \infty$ otherwise. For any $r \geq 0$, let

$$m_r = \min\{m(\lambda) : \lambda \in \Lambda^r\},$$

and let f be the minimal r such that m_r is finite.

Proposition 3.34. *We have $f = 0$ if and only if y_K has infinite order.*

Let $SD = \ell^n S$, where

$$S = \varinjlim_{n} \mathrm{Sel}^{(\ell^n)}(K, E[\ell^n]).$$

If A is a $\mathbb{Z}[1,\sigma]$-module and $\varepsilon = (-1)^{r_{E,\mathrm{an}}-1}$. then

$$A^v = \{b \in A : \sigma(b) = (-1)^{v+1}\varepsilon b\}$$

Assuming his conjectures, Kolyvagin deduces that for every prime number ℓ there exists integers k_1 and k_2 such that for any integer $k \geq k_1$ we have

$$\ell^{k_2} SD^{(f+1)}[M] \subset V^f_{n,k} \subset SD^{(f+1)}[M].$$

Here the exponent of $f+1$ means the $+1$ or -1 eigenspace for the conjugation action.

Conjecture 3.35 (Kolyvagin)**.** *Let E be any elliptic curve over $\mathbb{Q}$ and ℓ any prime. There exists $v \in \{0,1\}$ and a subgroup*

$$V \subset (E(K)/E(K)_{\mathrm{tors}})^{(v)}$$

such that

$$1 \leq \mathrm{rank}(V) \equiv v \pmod{2}.$$

Let $a = \mathrm{rank}(V) - 1$. Then for all sufficiently large k and all n, one has that

$$V^a_{n,k} \equiv V \mod \ell^n(E(K)/E(K)_{\mathrm{tor}}).$$

Assuming the above conjecture for all primes ℓ, the group V is uniquely determined by the congruence condition in the second part of the conjecture. Also, Kolyvagin proves that if the above conjecture is true, then the rank of $E^v(\mathbb{Q})$ equals the rank of V, and that $Ш(E^v/\mathbb{Q})[\ell^\infty]$ is finite. (Here E^v is E or its quadratic twist.)

When P_1 has infinite order, the conjecture is true with $v = 1$ and $V = \mathbb{Z}P_1$. (I think here E has $r_{E,\mathrm{an}} = 0$.)

3.5. The Gross-Zagier Theorem

Chapter 4

Computational Verification of the Conjecture

Bibliography

[BCDT01] C. Breuil, B. Conrad, Fred Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[Bir71] B. J. Birch, *Elliptic curves over* **Q***: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, pp. 396–400.

[Coh00] Henri Cohen, *Advanced topics in computational number theory*, Graduate Texts in Mathematics, vol. 193, Springer-Verlag, New York, 2000. MR MR1728313 (2000k:11144)

[Cp86] J. W. S. Cassels and A. Fröhlich (eds.), *Algebraic number theory*, London, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1986, Reprint of the 1967 original.

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997,
`http://www.maths.nott.ac.uk/personal/jec/book/`.

[CS00] J. Coates and R. Sujatha, *Galois cohomology of elliptic curves*, Tata Institute of Fundamental Research Lectures on Mathematics, 88, Published by Narosa Publishing House, New Delhi, 2000. MR MR1759312 (2001b:11046)

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149.

[Elk87] Noam D. Elkies, *The existence of infinitely many supersingular primes for every elliptic curve over* **Q**, Invent. Math. **89** (1987), no. 3, 561–567. MR MR903384 (88i:11034)

[GJP$^+$05] G. Grigorov, A. Jorza, S. Patrikis, C. Patrascu, and W. Stein, *Verification of the Birch and Swinnerton-Dyer Conjecture for Specific Elliptic Curves*, (Submitted)
`http://www.wstein.org/papers/bsdalg/` (2005).

[Gro91] B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 235–256.

[Kol91] V. A. Kolyvagin, *On the structure of Selmer groups*, Math. Ann. **291** (1991), no. 2, 253–259. MR 93e:11073

[LT58] S. Lang and J. Tate, *Principal homogeneous spaces over abelian varieties*, Amer. J. Math. **80** (1958), 659–684.

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[MST06] Barry Mazur, William Stein, and John Tate, *Computation of p-adic heights and log convergence*, Doc. Math. (2006), no. Extra Vol., 577–614 (electronic). MR MR2290599

[MT91] B. Mazur and J. Tate, *The p-adic sigma function*, Duke Math. J. **62** (1991), no. 3, 663–688. MR 93d:11059

[MTT86] B. Mazur, J. Tate, and J. Teitelbaum, *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer*, Invent. Math. **84** (1986), no. 1, 1–48.

[Ser72] J-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331.

[Ser79] ———, *Local fields*, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

[Ser97] ———, *Galois cohomology*, Springer-Verlag, Berlin, 1997, Translated from the French by Patrick Ion and revised by the author.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, corrected reprint of the 1986 original.

[Ste02] W. A. Stein, *There are genus one curves over* **Q** *of every odd index*, J. Reine Angew. Math. **547** (2002), 139–147. MR 2003c:11059

[SW07] William Stein and Chris Wuthrich, *Computations About Tate-Shafarevich Groups Uusing Iwasawa Theory*, In preparation (2007).

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551. MR 1333035 (96d:11071)

[Wil00] ———, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

32 Open Source Mathematical Software, with David Joyner

Open Source Mathematical Software

Mathematical software has greatly contributed to mathematical research, enabling exciting advances in mathematics and providing extensive data for conjectures. Perhaps three of the most well-known applications of computation to mathematical research are the resolution of the *four-color conjecture* by Appel and Haken in 1976 (though it is now reproven with less need for computer verification by N. Robertson, D. P. Sanders, P. D. Seymour and R. Thomas), Thomas Hales's proof of *Kepler's conjecture*, and the formulation of the *Birch and Swinnerton-Dyer conjecture*, which grew out of extensive numerical computation.

Open source software, such as TEX, Mozilla Firefox, and Linux has had a profound effect on computing during the last decade, and we hope that open source mathematical software will have a similar positive impact on mathematics.

> I think we need a symbolic standard to make computer manipulations easier to document and verify. And with all due respect to the free market, perhaps we should not be dependent on commercial software here. An open source project could, perhaps, find better answers to the obvious problems such as availability, bugs, backward compatibility, platform independence, standard libraries, etc. One can learn from the success of TEX and more specialized software like Macaulay2. I do hope that funding agencies are looking into this.
>
> —*Andrei Okounkov, 2006 Fields medalist*
> (see "Interviews with three Fields medalists" *Notices of the AMS*, **54**(3) (2007), 405–410).

The term *open source* is defined at `http://www.opensource.org/`, but basically it means anyone (including commercial companies or the defense department) should be able to inspect open source software, modify it, and share it with others.

One key difference between mathematical theorems and software is that theorems require little maintenance, whereas *mathematical software requires substantial and potentially expensive maintenance* (bug fixes, updates when algorithms or languages change, etc.). Mathematical research usually generates no direct revenue for researchers, and likewise open source mathematical software is free to share and extend, so it rarely generates revenue. Volunteer effort, donations, and financial support from the NSF and other organizations is thus critical to the success of open source mathematical software.

There is a proof in the article by Campbell et al. in *The Atlas of Finite Groups—Ten Years On* (1998) that describes how many separate software packages were "easily used" to deduce various mathematical facts—no code is given, and some of the programs are proprietary software that runs only on hardware many years out of date. Such proofs may become increasingly common in mathematics if something isn't done to reverse this trend.

Suppose Jane is a well-known mathematician who announces she has proved a theorem. We probably will believe her, but she knows that she will be required to produce a proof if requested. However, suppose now Jane says a theorem is true based partly on the results of software. The closest we can reasonably hope to get to a rigorous proof (without new ideas) is the open inspection and ability to use all the computer code on which the result depends. If the program is proprietary, this is not possible. We have every right to be distrustful, not only due to a vague distrust of computers but because even the best programmers regularly make mistakes.

If one reads the proof of Jane's theorem in hopes of extending her ideas or applying them in a new context, it is limiting to not have access to the inner workings of the software on which Jane's result builds. For example, consider the following quote from the Mathematica tutorial[1]:

> Particularly in more advanced applications of Mathematica, it may sometimes seem worthwhile to try to analyze internal algorithms in order to predict which way of doing a given computation will be the most efficient. [...] But most often the analyses will not be worthwhile. For the internals of Mathematica are quite complicated, and even given a basic description of the algorithm used for a particular purpose, it is usually extremely difficult to reach a reliable conclusion about how the detailed implementation of this algorithm will actually behave in particular circumstances.

No journal would make a statement like the above about the proofs of the theorems they publish. Increasingly, proprietary software and the algorithms used are an essential part of mathematical proofs. To quote J. Neubüser, "*with this situation two of the most basic rules of conduct in mathematics are violated: In mathematics information is passed on free of charge and everything is laid open for checking.*"

Full disclosure: The second author started a new mathematics software system in 2005 called SAGE (see `www.sagemath.org`), which combines Python, GAP, Singular, PARI, Maxima, SciPy, etc. with several hundred thousand lines of new code. SAGE receives contributions from many mathematicians worldwide that synthesize the latest algorithms from a broad range of topics into a comprehensive toolkit for mathematical research.

—*David Joyner*
U. S. Naval Academy, Annapolis
`wdj@usna.edu`
—*William Stein*
University of Washington, Seattle
`wstein@u.washington.edu`

[1] `http://reference.wolfram.com/mathematica/tutorial/WhyYouDoNotUsuallyNeedToKnowAboutInternals.html`

33 On the generation of the coefficient field of a newform by a single Hecke eigenvalue, with K. Koo and G. Wiese

On the generation of the coefficient field of a newform by a single Hecke eigenvalue

Koopa Tak-Lun Koo* and William Stein† and Gabor Wiese‡

November 20, 2007

Abstract

Let f be a non-CM newform of weight $k \geq 2$ without nontrivial inner twists. In this article we study the set of primes p such that the eigenvalue $a_p(f)$ of the Hecke operator T_p acting on f generates the field of coefficients of f. We show that this set has density 1, and prove a natural analogue for newforms having inner twists. We also present some new data on reducibility of Hecke polynomials, which suggest questions for further investigation.

Mathematics Subject Classification (2000): 11F30 (primary); 11F11, 11F25, 11F80, 11R45 (secondary).

1 Introduction

The main aim of this paper is to prove the following theorem.

Theorem 1. *Let f be a newform (i.e., a new normalized cuspidal Hecke eigenform) of weight $k \geq 2$, level N and Dirichlet character χ which does not have complex multiplication (CM, see [R80, p. 48]). Let $E_f = \mathbf{Q}(a_n(f) : (n, N) = 1)$ be the field of coefficients of f and $F_f = \mathbf{Q}\left(\frac{a_n(f)^2}{\chi(n)} : (n, N) = 1\right)$.*

*Department of Mathematics, University of Washington, Seattle, Box 354350 WA 98195, USA; e-mail: koopakoo@gmail.com

†Department of Mathematics, University of Washington, Seattle, Box 354350 WA 98195, USA; e-mail: wstein@math.washington.edu

‡Institut für Experimentelle Mathematik, Universität Duisburg-Essen, Ellernstraße 29, 45326 Essen, Germany; e-mail: gabor@pratum.net

(a) The set

$$\left\{p \text{ prime} : \mathbf{Q}\left(\frac{a_p(f)^2}{\chi(p)}\right) = F_f\right\}$$

has density 1.

(b) If f does not have any nontrivial inner twists, then the set

$$\{p \text{ prime} : \mathbf{Q}(a_p(f)) = E_f\}$$

has density 1.

A twist of f by a Dirichlet character ϵ is said to be *inner* if there exists a (necessarily unique) field automorphism $\sigma_\epsilon : E_f \to E_f$ such that

$$a_p(f \otimes \epsilon) = a_p(f)\epsilon(p) = \sigma_\epsilon(a_p(f))$$

for almost all primes p. If N is square free, $k = 2$ and the Dirichlet character χ of f is the trivial character, then there are no nontrivial inner twists of f. For a discussion of inner twists we refer the reader to [R80, §3] and [R85, §3].

In the presence of nontrivial inner twists, the conclusion of Part (b) of the theorem never holds. To see this, we let ϵ be a nontrivial inner twist with associated field automorphism σ_ϵ. The set of primes p such that $\epsilon(p) = 1$ has a positive density and for any such p we have $\sigma_\epsilon(a_p(f)) = a_p(f)$. Therefore, $a_p(f) \in E_f^{\langle\sigma\rangle} \subsetneq E_f$ for a set of primes p of positive density.

In the literature there are related but weaker results in the context of Maeda's conjecture, i.e., they concern the case of level 1 and assume that $S_k(1)$ consists of a single Galois orbit of newforms (see, e.g., [JO98] and [BM03]). We now show how Part (b) of Theorem 1 extends the principal results of these two papers.

Let f be a newform of level N, weight $k \geq 2$ and trivial Dirichlet character $\chi = 1$ which neither has CM nor nontrivial inner twists. This is true when $N = 1$. Let $\mathbb{T}$ be the $\mathbf{Q}$-algebra generated by all T_n with $n \geq 1$ inside $\mathrm{End}(S_k(N, 1))$ and let $\mathfrak{P}$ be the kernel of the $\mathbf{Q}$-algebra homomorphism $\mathbb{T} \xrightarrow{T_n \mapsto a_n(f)} E_f$. As $\mathbb{T}$ is reduced, the map $\mathbb{T}_{\mathfrak{P}} \xrightarrow{T_n \mapsto a_n(f)} E_f$ is a ring isomorphism with $\mathbb{T}_{\mathfrak{P}}$ the localization of $\mathbb{T}$ at $\mathfrak{P}$. Non canonically $\mathbb{T}_{\mathfrak{P}}$ is also isomorphic as a $\mathbb{T}_{\mathfrak{P}}$-module (equivalently as an E_f-vector space) to its $\mathbf{Q}$-linear dual, which can be identified with the localization at $\mathfrak{P}$ of the $\mathbf{Q}$-vector space $S_k(N, 1; \mathbf{Q})$ of cusp forms in $S_k(N, 1)$ with q-expansion in $\mathbf{Q}[[q]]$. Hence, $\mathbf{Q}(a_p(f)) = E_f$ precisely means that the characteristic polynomial $P_p \in \mathbf{Q}[X]$ of T_p acting on the localization at $\mathfrak{P}$ of $S_k(N, 1; \mathbf{Q})$

is irreducible. Part (b) of Theorem 1 hence shows that the set of primes p such that P_p is irreducible has density 1.

This extends Theorem 1 of [JO98] and Theorem 1.1 of [BM03]. Both theorems restrict to the case $N = 1$ and assume that there is a unique Galois orbit of newforms, i.e., a unique $\mathfrak{P}$, so that no localization is needed. Theorem 1 of [JO98] says that

$$\#\{p < X \text{ prime } : P_p \text{ is irreducible in } \mathbf{Q}[X]\} \gg \frac{X}{\log X}$$

and Theorem 1.1 of [BM03] states that there is $\delta > 0$ such that

$$\#\{p < X \text{ prime } : P_p \text{ is reducible in } \mathbf{Q}[X]\} \ll \frac{X}{(\log X)^{1+\delta}}.$$

Acknowledgements. The authors would like to thank the MSRI, where part of this research was done, for its hospitality. The first author would like to thank his advisor Ralph Greenberg for suggesting the problem. The second author acknowledges partial support from the National Science Foundation grant No. 0555776, and also used [SAGE] for some calculations related to this paper. All three authors thank Jordi Quer for useful discussions.

2 Group theoretic input

Lemma 1. *Let q be a prime power and ϵ a generator of the cyclic group $\mathbb{F}_q^\times$.*

(a) The conjugacy classes c in $\mathrm{GL}_2(\mathbb{F}_q)$ have the following four kinds of representatives:

$$S_a = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad T_a = \begin{pmatrix} a & 0 \\ 1 & a \end{pmatrix}, \quad U_{a,b} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad V_{x,y} = \begin{pmatrix} x & \epsilon y \\ y & x \end{pmatrix}$$

where $a \neq b$, and $y \neq 0$.

(b) The number of elements in each of these conjugacy classes are: $1, q^2 - 1, q^2 + q$, and $q^2 - q$, respectively.

Proof. See Fulton-Harris [FH91], page 68. □

We use the notation $[g]_G$ for the conjugacy class of g in G.

Proposition 1. *Let q be a prime power and r a positive integer. Let further $R \subseteq \widetilde{R} \subseteq \mathbb{F}_{q^r}^\times$ be subgroups. Put $\sqrt{\widetilde{R}} = \{s \in \mathbb{F}_{q^r}^\times \,:\, s^2 \in \widetilde{R}\}$. Set*

$$H = \{g \in \mathrm{GL}_2(\mathbb{F}_q) \,:\, \det(g) \in R\}$$

and let

$$G \subseteq \{g \in \mathrm{GL}_2(\mathbb{F}_{q^r}) \,:\, \det(g) \in \widetilde{R}\}$$

be any subgroup such that H is a normal subgroup of G. Then the following statements hold.

(a) The group $G/(G \cap \mathbb{F}_{q^r}^\times)$ (with $\mathbb{F}_{q^r}^\times$ identified with scalar matrices) is either equal to $\mathrm{PSL}_2(\mathbb{F}_q)$ or to $\mathrm{PGL}_2(\mathbb{F}_q)$. More precisely, if we let $\{s_1, \dots, s_n\}$ be a system of representatives for $\sqrt{\widetilde{R}}/R$, then for all $g \in G$ there is i such that $g \begin{pmatrix} s_i^{-1} & 0 \\ 0 & s_i^{-1} \end{pmatrix} \in G \cap \mathrm{GL}_2(\mathbb{F}_q)$ and $\left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right) \in G$.

(b) Let $g \in G$ such that $g \begin{pmatrix} s_i^{-1} & 0 \\ 0 & s_i^{-1} \end{pmatrix} \in G \cap \mathrm{GL}_2(\mathbb{F}_q)$ and $\left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right) \in G$. Then

$$[g]_G = [g \begin{pmatrix} s_i^{-1} & 0 \\ 0 & s_i^{-1} \end{pmatrix}]_{G \cap \mathrm{GL}_2(\mathbb{F}_q)} \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right).$$

(c) Let $P(X) = X^2 - aX + b \in \mathbb{F}_{q^r}[X]$ be a polynomial. Then the inequality

$$\sum_C |C| \;\leq\; 2|\widetilde{R}/R|(q^2 + q)$$

holds, where the sum runs over the conjugacy classes C of G with characteristic polynomial equal to $P(X)$.

Proof. (a) The classification of the finite subgroups of $\mathrm{PGL}_2(\overline{\mathbb{F}}_q)$ yields that the group $G/(G \cap \mathbb{F}_{q^r}^\times)$ is either $\mathrm{PGL}_2(\mathbb{F}_{q^u})$ or $\mathrm{PSL}_2(\mathbb{F}_{q^u})$ for some $u \mid r$. This, however, can only occur with $u = 1$, as $\mathrm{PSL}_2(\mathbb{F}_{q^u})$ is simple. The rest is only a reformulation.

(b) This follows from (a), since scalar matrices are central.

(c) From (b) we get the inclusion

$$\bigsqcup_C C \subseteq \bigsqcup_{i=1}^n \bigsqcup_D D \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right),$$

where C runs over the conjugacy classes of G with characteristic polynomial equal to $P(X)$ and D runs over the conjugacy classes of $G \cap \mathrm{GL}_2(\mathbb{F}_q)$ with characteristic polynomial equal to $X^2 - as_i^{-1}X + bs_i^{-2}$ (such a conjugacy class is empty if the polynomial is not in $\mathbb{F}_q[X]$). The group $G \cap \mathrm{GL}_2(\mathbb{F}_q)$ is normal in $\mathrm{GL}_2(\mathbb{F}_q)$, as it contains $\mathrm{SL}_2(\mathbb{F}_q)$. Hence, any conjugacy class of $\mathrm{GL}_2(\mathbb{F}_q)$ either has an empty intersection with $G \cap \mathrm{GL}_2(\mathbb{F}_q)$ or is a disjoint union of conjugacy classes of $G \cap \mathrm{GL}_2(\mathbb{F}_q)$. Consequently, by Lemma 1, the disjoint union $\bigsqcup_D D \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right)$ is equal to one of

(i) $[U_{a,b}]_{\mathrm{GL}_2(\mathbb{F}_q)} \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right)$,

(ii) $[V_{x,y}]_{\mathrm{GL}_2(\mathbb{F}_q)} \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right)$ or

(iii) $[S_a]_{\mathrm{GL}_2(\mathbb{F}_q)} \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right) \sqcup [T_a]_{\mathrm{GL}_2(\mathbb{F}_q)} \left(\begin{smallmatrix} s_i & 0 \\ 0 & s_i \end{smallmatrix}\right)$.

Still by Lemma 1, the first set contains q^2+q, the second set q^2-q and the third one q^2 elements. Hence, the set $\bigsqcup_C C$ contains at most $2|\widetilde{R}/R|(q^2+q)$ elements. □

3 Proof

The proof of Theorem 1 relies on the following important theorem by Ribet, which, roughly speaking, says that the image of the mod ℓ Galois representation attached to a fixed newform is as big as it can be for almost all primes ℓ.

Theorem 2 (Ribet). *Let f be a Hecke eigenform of weight $k \geq 2$, level N and Dirichlet character $\chi : (\mathbf{Z}/N\mathbf{Z})^\times \to \mathbf{C}^\times$. Suppose that f does not have CM. Let E_f and F_f be as in Theorem 1 and denote by $\mathcal{O}_{E_f}$ and $\mathcal{O}_{F_f}$ the corresponding rings of integers.*

There exists an abelian extension $K/\mathbf{Q}$ such that for almost all prime numbers ℓ the following statement holds:

> *Let $\widetilde{\mathcal{L}}$ be a prime ideal of $\mathcal{O}_{E_f}$ dividing ℓ. Put $\mathcal{L} = \widetilde{\mathcal{L}} \cap \mathcal{O}_{F_f}$ and $\mathcal{O}_{F_f}/\mathcal{L} \cong \mathbb{F}$. Consider the residual Galois representation*
>
> $$\overline{\rho}_{f,\widetilde{\mathcal{L}}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O}_{E_f}/\widetilde{\mathcal{L}})$$
>
> *attached to f. Then the image $\overline{\rho}_{f,\widetilde{\mathcal{L}}}(\mathrm{Gal}(\overline{\mathbf{Q}}/K))$ is equal to*
>
> $$\{g \in \mathrm{GL}_2(\mathbb{F}) \,:\, \det(g) \in \mathbb{F}_\ell^{\times(k-1)}\}.$$

Proof. It suffices to take Ribet [R85, Thm. 3.1] mod $\widetilde{\mathcal{L}}$. Note that F_f is the field E_f^Γ. To see this, one checks immediately that $F_f \subseteq E_f^\Gamma$ with Γ the group of the field automorphisms associated with the inner twists as in [R80, §3]. On the other hand, let σ be a field embedding $E_f \to \mathbf{C}$ which is the identity on F_f, i.e., on $\frac{a_n(f)^2}{\chi(n)}$ for all n with $(n, N) = 1$. Then $\frac{\sigma(a_n(f))^2}{a_n(f)^2} = \frac{\sigma(\chi(n))}{\chi(n)}$ is a root of unity, and, thus, so is $\epsilon(n) = \frac{\sigma(a_n(f))}{a_n(f)}$. This defines a Dirichlet character ϵ by which f has an inner twist. Hence, $\sigma \in \Gamma$ and $F_f = E_f^\Gamma$.

Ribet does not say explicitly that $K/\mathbf{Q}$ is abelian, but this follows since it is a composite of abelian extensions of K, which are each cut out by a character. □

Remark 1. *The field F_f defined in Theorem 1 is invariant under twisting. More precisely, let ϵ be any Dirichlet character and consider the twisted modular form $f \otimes \epsilon$, the Dirichlet character of which is $\chi\epsilon^2$. Then the Fourier coefficients satisfy $a_n(f \otimes \epsilon) = a_n(f)\epsilon(n)$ and, thus, $\frac{a_n(f\otimes\epsilon)^2}{\chi(n)\epsilon(n)^2} = \frac{a_n(f)^2}{\chi(n)}$.*

Remark 2. *If f in Theorem 1 does not have any nontrivial inner twists, then $K = \mathbf{Q}$ and $F_f = E_f$, since $F_f = E_f^\Gamma$ with Γ the group of field automorphisms associated with the inner twists (see the proof of Theorem 2).*

Theorem 3. *Let f be a non-CM newform of weight $k \geq 2$, level N and Dirichlet character χ. Let F_f be as in Theorem 1 and let $L \subset F_f$ be any proper subfield. Then the set*

$$\left\{ p \text{ prime} : \frac{a_p(f)^2}{\chi(p)} \in L \right\}$$

has density zero.

Proof. Let $L \subsetneq F_f$ be a proper subfield and $\mathcal{O}_L$ its integer ring. We define the set

$$S := \{\mathcal{L} \subset \mathcal{O}_{F_f} \text{ prime ideal} : [\mathcal{O}_{F_f}/\mathcal{L} : \mathcal{O}_L/(L \cap \mathcal{L})] \geq 2\}.$$

Notice that this set is infinite. For, if it were finite, then all but finitely many primes would split completely in the extension F_f/L, which is not the case by Chebotarev's density theorem.

Let $\mathcal{L} \in S$ be any prime, ℓ its residue characteristic and $\widetilde{\mathcal{L}}$ a prime of $\mathcal{O}_{E_f}$ lying over $\mathcal{L}$. Put $\mathbb{F}_q = \mathcal{O}_L/(L \cap \mathcal{L})$, $\mathbb{F}_{q^r} = \mathcal{O}_{F_f}/\mathcal{L}$ and $\mathbb{F}_{q^{rs}} = \mathcal{O}_{E_f}/\widetilde{\mathcal{L}}$. We have $r \geq 2$. Let W be the subgroup of $\mathbb{F}_{q^{rs}}^\times$ consisting of the values of χ modulo $\widetilde{\mathcal{L}}$; its size $|W|$ is less than or equal to $|(\mathbf{Z}/N\mathbf{Z})^\times|$. Let $R = \mathbb{F}_\ell^{\times(k-1)}$ be the subgroup of $(k-1)$st powers of elements in the multiplicative group $\mathbb{F}_\ell^\times$ and let $\widetilde{R} = \langle R, W\rangle \subset \mathbb{F}_{q^{rs}}^{\times}$. The size of $\widetilde{R}$ is less than or equal to $|R| \cdot |W|$. Let $H = \{g \in \mathrm{GL}_2(\mathbb{F}_{q^r}) \ : \ \det(g) \in R\}$ and $G = \mathrm{Gal}(\overline{\mathbf{Q}}^{\ker \overline{\rho}_{f,\widetilde{\mathcal{L}}}}/\mathbf{Q})$. By Galois theory, G can be identified with the image of the residual representation $\overline{\rho}_{f,\widetilde{\mathcal{L}}}$, and we shall make this identification from now on. By Theorem 2 we have the inclusion of groups

$$H \subseteq G \subseteq \{g \in \mathrm{GL}_2(\mathbb{F}_{q^{rs}}) : \det(g) \in \widetilde{R}\}$$

with H being normal in G.

If C is a conjugacy class of G, by Chebotarev's density theorem the density of

$$\{p\,\text{prime} : \ [\overline{\rho}_{f,\widetilde{\mathcal{L}}}(\mathrm{Frob}_p)]_G = C\}$$

equals $|C|/|G|$. We consider the set

$$M_{\mathcal{L}} := \bigsqcup_C \{p\,\text{prime} : \ [\overline{\rho}_{f,\widetilde{\mathcal{L}}}(\mathrm{Frob}_p)]_G = C\} \supseteq \left\{ p\,\text{prime} : \ \overline{\left(\frac{a_p(f)^2}{\chi(p)}\right)} \in \mathbb{F}_q \right\},$$

where the reduction modulo $\mathcal{L}$ of an element $x \in \mathcal{O}_{F_f}$ is denoted by $\overline{x}$ and C runs over the conjugacy classes of G with characteristic polynomials equal to some $X^2 - aX + b \in \mathbb{F}_{q^{rs}}[X]$ such that

$$a^2 \in \{t \in \mathbb{F}_{q^{rs}} \ : \ \exists u \in \mathbb{F}_q \ \exists w \in W : t = uw\}$$

and automatically $b \in \widetilde{R}$. The set $M_{\mathcal{L}}$ has the density $\delta(M_{\mathcal{L}}) = \sum_C \frac{|C|}{|G|}$ with C as before. There are at most $2q|W|^2 \cdot |R|$ such polynomials. We are now precisely in the situation to apply Prop. 1, Part (c), which yields the inequality

$$\delta(M_{\mathcal{L}}) \leq \frac{4|W|^3 q(q^{2r} + q^r)}{(q^{3r} - q^r)} = O\left(\frac{1}{q^{r-1}}\right) \leq O\left(\frac{1}{q}\right),$$

where for the denominator we used $|G| \geq |H| = |R| \cdot |\mathrm{SL}_2(\mathbb{F}_{q^r})|$.

Since q is unbounded for $\mathcal{L} \in S$, the intersection $M := \bigcap_{\mathcal{L} \in S} M_{\mathcal{L}}$ is a set having a density and this density is 0. The inclusion

$$\left\{ p\,\text{prime} : \ \frac{a_p(f)^2}{\chi(p)} \in L \right\} \subseteq M$$

finishes the proof. □

Proof of Theorem 1. To obtain (a), it suffices to apply Theorem 3 to each of the finitely many sub-extension of F_f. (b) follows from (a) by Remark 2 and the fact that χ must take values in $\{\pm 1\}$, as otherwise E_f would be a CM-field and complex conjugation would give a nontrivial inner twist. □

4 Reducibility of Hecke polynomials: questions

Motivated by a conjecture of Maeda, there has been some speculation that for every integer k and prime number p, the characteristic polynomial of T_p acting on $S_k(1)$ is irreducible. See, for example, [FJ02], which verifies this for all $k < 2000$ and $p < 2000$. The most general such speculation might be the following question: *if f is a non-CM newform of level $N \geq 1$ and weight $k \geq 2$ such that some $a_p(f)$ generates the field $E_f = \mathbf{Q}(a_n(f) : n \geq 1)$, do all but finitely many prime-indexed Fourier coefficients $a_p(f)$ have irreducible characteristic polynomial?* The answer in general is no. An example is given by the newform in level 63 and weight 2 that has an inner twist by $\left(\frac{\cdot}{4}\right)$. Also for non-CM newforms of weight 2 without nontrivial inner twists such that $[E_f : \mathbf{Q}] = 2$, we think that the answer is likely no.

Let $f \in S_k(\Gamma_0(N))$ be a newform of weight k and level N. The *degree* of f is the degree of the field E_f, and we say that f is a *reducible newform* if the characteristic polynomial of $a_p(f)$ is reducible for infinitely many primes p.

For each even weight $k \leq 12$ and degree $d = 2, 3, 4$, we used [SAGE] to find newforms f of weight k and degree d. For each of these forms, we computed the *reducible primes* $p < 1000$, i.e., the primes such that the characteristic polynomial of $a_p(f)$ is reducible. The result of this computation is given in Table 1. Table 2 contains the number of reducible primes $p < 10000$ for the first 20 newforms of degree 2 and weight 2. This data inspires the following question.

Question 1. *If $f \in S_2(\Gamma_0(N))$ is a newform of degree 2, is f necessarily reducible? That is, are there infinitely many primes p such that $a_p(f) \in \mathbf{Z}$, or equivalently, such that the characteristic polynomial of $a_p(f)$ is reducible?*

Tables 4–6 contain additional data about the first few newforms of given degree and weight, which may suggest other similar questions. In particular, Table 4 contains data for all primes up to 10^6 for the first degree 2 form f with $L(f, 1) \neq 0$, and for the first degree 2 form g with $L(g, 1) = 0$. We find that there are 386 primes $< 10^6$ with $a_p(f) \in \mathbf{Z}$ (i.e., has reducible characteristic polynomial), and 309 with $a_p(g) \in \mathbf{Z}$.

Question 2. *If $f \in S_2(\Gamma_0(N))$ is a newform of degree 2, can the asymptotic behaviour of the function*

$$N(x) := \#\{p\,\text{prime} : \, p < x, a_p(f) \in \mathbf{Z}\}$$

be described as a function of x?

The authors intend to investigate these questions in a subsequent paper.

Table 1: Counting Reducible Characteristic Polynomials

k	d	N	reducible $p < 1000$
2	2	23	13, 19, 23, 29, 43, 109, 223, 229, 271, 463, 673, 677, 883, 991
2	3	41	17, 41
2	4	47	47
4	2	11	11
4	3	17	17
4	4	23	23
6	2	7	7
6	3	11	11
6	4	17	17
8	2	5	5
8	3	17	17
8	4	11	11
10	2	5	5
10	3	7	7
10	4	13	13
12	2	5	5
12	3	7	7
12	4	21	3, 7

Table 2: First 20 Newforms of Degree 2 and Weight 2

k	d	N	$\#\{$reducible $p < 10000\}$
2	2	23	47
2	2	29	42
2	2	31	78
2	2	35	48
2	2	39	71
2	2	43	43
2	2	51	64
2	2	55	95
2	2	62	77
2	2	63	622 (inner twist by $\left(\frac{\cdot}{4}\right)$)
2	2	65	43
2	2	65	90
2	2	67	51
2	2	67	19
2	2	68	53
2	2	69	47
2	2	73	43
2	2	73	55
2	2	74	52
2	2	74	21

Table 3: Newforms 23a and 67b: values of $\psi(x) = \#\{\text{reducible } p < x \cdot 10^5\}$

k	d	N	r_{an}	1	2	3	4	5	6	7	8	9	10
2	2	23	0	127	180	210	243	277	308	331	345	360	386
2	2	67	1	111	159	195	218	240	257	276	288	301	309

Table 4: First 5 Newforms of Degrees 3, 4 and Weight 2

k	d	N	reducible $p < 10000$
2	3	41	17, 41
2	3	53	13, 53
2	3	61	61, 2087
2	3	71	23, 31, 71, 479, 647, 1013, 3181
2	3	71	13, 71, 509, 3613

k	d	N	reducible $p < 10000$
2	4	47	47
2	4	95	5, 19
2	4	97	97
2	4	109	109, 4513
2	4	111	3, 37

Table 5: First 5 Newforms of Degrees 2, 3 and Weight 4

k	d	N	reducible $p < 1000$
4	2	11	11
4	2	13	13
4	2	21	3, 7
4	2	27	3, 7, 13, 19, 31, 37, 43, 61, 67, 73, 79, 97, 103, 109, 127, 139, 151, 157, 163, 181, 193, 199, 211, 223, 229,241, 271, 277, 283, 307, 313, 331, 337, 349, 367, 373, 379, 397, 409, 421, 433, 439, 457, 463, 487, 499, 523, 541, 547, 571, 577, 601, 607, 613, 619, 631, 643, 661, 673, 691, 709, 727, 733, 739, 751, 757, 769, 787, 811, 823, 829, 853, 859, 877, 883, 907, 919, 937, 967, 991, 997 (has inner twists)
4	2	29	29

k	d	N	reducible $p < 1000$
4	3	17	17
4	3	19	19
4	3	35	5, 7
4	3	39	3, 13
4	3	41	41

Table 6: Newforms on $\Gamma_0(389)$ of Weight 2

k	d	N	reducible $p < 10000$
2	1	389	none (degree 1 polynomials are all irreducible)
2	2	389	5, 11, 59, 97, 157, 173, 223, 389, 653, 739, 859, 947, 1033, 1283, 1549, 1667, 2207, 2417, 2909, 3121, 4337, 5431, 5647, 5689, 5879, 6151, 6323, 6373, 6607, 6763, 7583, 7589, 8363, 9013, 9371, 9767
2	3	389	7, 13, 389, 503, 1303, 1429, 1877, 5443
2	6	389	19, 389
2	20	389	389

References

[BM03] Baba, Srinath and Murty, Ram, *Irreducibility of Hecke Polynomials*, Math. Research Letters, 10(2003), no.5-6, pp.709-715.

[FJ02] D. W. Farmer and K. James, *The irreducibility of some level 1 Hecke polynomials*, Math. Comp. **71** (2002), no. 239, 1263–1270.

[FH91] Fulton, William and Harris, Joe, *Representation Theory, A First Course*, Springer, 1991.

[JO98] James, Kevin and Ono, Ken, *A note on the Irreducibility of Hecke Polynomials*, Journal of Number Theory 73, 1998, pp.527-532.

[R80] Ribet, Kenneth A., *Twists of modular forms and endomorphisms of abelian varieties.*, Math. Ann. 253 (1980), no. 1, 43–62.

[R85] Ribet, Kenneth A., *On l-adic representations attached to modular forms. II.*, Glasgow Math. J. 27 (1985), 185–194.

[SAGE] Stein, William, *Sage Mathematics Software (Version 2.8.12)*, The SAGE Group, 2007, `http://www.sagemath.org`.

34 Three Lectures about Explicit Methods in Number Theory Using Sage

Three Lectures about Explicit Methods in Number Theory Using Sage

William Stein

October 2008

Abstract

This article is about using the mathematical software Sage to do computations with number fields and modular forms. It was written for the October 2008 Bordeaux meeting on explicit methods in number theory (`http://www.math.u-bordeaux.fr/gtem2008/`). It assumes no prior knowledge about Sage, but assumes a graduate level background in algebraic number theory.

Contents

Introduction

Sage (see http://sagemath.org) is a comprehensive mathematical software system for computations in many areas of pure and applied mathematics. We program Sage using the mainstream programming language Python (see http://python.org), or its compiled variant Cython. It is also very easy to efficiently use code written in C/C++ from Sage.

The author of this article started the Sage project in 2005. Sage is free and open source, meaning you can change any part of Sage and redistribute the result without having to pay any license fees, and Sage can also leverage the power of commercial mathematical software such as Magma and Mathematica, if you happen to have access to those closed source commercial systems.

This paper assumes no prior knowledge of either Python or Sage. Our goal is to help number theorists do computations involving number fields and modular forms using Sage.

As you read this article, please try every example in Sage, and make sure things works as I claim, and do all of the exercises. Moreover, you should experiment by typing in similar examples and checking that the output you get agrees with what you expect.

To use Sage, install it on your computer, and use either the command line or start the Sage notebook by typing `notebook()` at the command line.

We show Sage sessions as follows:

```
sage: factor(123456)
2^6 * 3 * 643
```

This means that if you type `factor(123456)` as input to Sage, then you'll get `2^6 * 3 * 643` as output. If you're using the Sage command line, you type `factor(123456)` and press enter; if you're using the Sage notebook via your web browser, you type `factor(123456)` into an input cell and press shift-enter; in the output cell you'll see `2^6 * 3 * 643`.

After trying the `factor` command in the previous paragraph (do this now!), you should try factoring some other numbers.

Exercise 0.1. What happens if you factor a negative number? a rational number?

You can also draw both 2d and 3d pictures using Sage. For example, the following input plots the number of prime divisors of each positive integer up to 500.

```
sage: line([(n, len(factor(n))) for n in [1..500]])
```

And, this example draws a similar 3d plot:

```
sage: v = [[len(factor(n*m)) for n in [1..15]] for m in [1..15]]
sage: list_plot3d(v, interpolation_type='nn')
```

The main difference between Sage and Pari is that Sage is vastly larger than Pari with a much wider range of functionality, and has many more datatypes and much more structured objects. Sage in fact includes Pari, and a typical Sage install takes nearly a gigabyte of disk space, whereas a typical Pari install is much more nimble, using only a few megabytes. There are many number-theoretic algorithms that are included in Sage, which have never been implemented in Pari, and Sage has 2d and 3d graphics which can be helpful for visualizing number theoretic ideas, and a graphical user interface. Both Pari and Sage are free and open source, which means anybody can read or change anything in either program, and the software is free.

The biggest difference between Sage and Magma is that Magma is closed source, not free, and difficult for users to extend. This means that most of Magma cannot be changed except by the core Magma developers, since Magma itself is well over two million lines of compiled C code, combined with about a half million lines of interpreted Magma code (that anybody can read and modify). In designing Sage, we carried over some of the excellent design ideas from Magma, such as the parent, element, category hierarchy.

Any mathematician who is serious about doing extensive computational work in algebraic number theory and arithmetic geometry is strongly urged to become familiar with all three systems, since they all have their pros and cons. Pari is sleek and small, Magma has much unique functionality for computations in arithmetic geometry, and Sage has a wide range of functionality in most areas of mathematics, a large developer community, and much unique new code.

1 Number Fields

In Sage, we can create the number field $\mathbb{Q}(\sqrt[3]{2})$ as follows.

```
sage: K.<alpha> = NumberField(x^3 - 2)
```

The above creates *two* Sage objects, K and α. Here K "is" (isomorphic to) the number field $\mathbb{Q}(\sqrt[3]{2})$, as we confirm below:

```
sage: K
Number Field in alpha with defining polynomial x^3 - 2
```

and α is a root of $x^3 - 2$, so α is an abstract choice of $\sqrt[3]{2}$ (no specific embedding of the number field K into $\mathbb{C}$ is chosen by default in Sage-3.1.2):

```
sage: alpha^3
2
sage: (alpha+1)^3
3*alpha^2 + 3*alpha + 3
```

Note that we did *not* define x above before using it. You could "break" the above example by redefining x to be something funny:

```
sage: x = 1
sage: K.<alpha> = NumberField(x^3 - 2)
Traceback (most recent call last):
...
TypeError: polynomial (=-1) must be a polynomial.
```

The *Traceback* above indicates that there was an error. Potentially lots of detailed information about the error (a "traceback") may be given after the word `Traceback` and before the last line, which contains the actual error messages.

Important: *whenever you use Sage and get a big error, look at the last line for the actual error, and only look at the rest if you are feeling adventurous.* In the notebook, the part indicated by ... above is not displayed; to see it, click just to the left of the word *Traceback* and the traceback will appear.

If you redefine x as above, but need to define a number field using the indeterminate x, you have several options. You can reset x to its default value at the start of Sage, you can redefine x to be a symbolic variable, or you can define x to be a polynomial indeterminant (a polygen):

```
sage: reset('x')
sage: x
x
sage: x = 1
sage: x = var('x')
sage: x
x
sage: x = 1
sage: x = polygen(QQ, 'x')
sage: x
x
sage: x = 1
sage: R.<x> = PolynomialRing(QQ)
sage: x
x
```

One you have created a number field K, type `K.[tab key]` to see a list of functions. Type, e.g., `K.Minkowski_embedding?[tab key]` to see help on the `Minkowski_embedding` command. To see source code, type `K.Minkowski_embedding??[tab key]`.

```
sage: K.<alpha> = NumberField(x^3 - 2)
sage: K.[tab key]
```

1.1 Symbolic Expressions

Another natural way for us to create certain number fields is to create a symbolic expression and adjoin it to the rational numbers. Unlike Pari and Magma (and like Mathematica and Maple), Sage also supports manipulation of symbolic expressions and solving equations, without defining abstract structures such as a number fields. For example, we can define a variable $a = \sqrt{2}$ as an abstract symbolic object by simply typing `a = sqrt(2)`. When we type `parent(a)` below, Sage tells us the mathematical object that it views a as being an element of; in this case, it's the ring of all symbolic expressions.

```
sage: a = sqrt(2)
sage: parent(a)
Symbolic Ring
```

In particular, typing `sqrt(2)` does *not* numerically extract an approximation to $\sqrt{2}$, like it would in Pari or Magma. We illustrate this below by calling Pari (via the gp interpreter) and Magma directly from within Sage. After we evaluate the following two input lines, copies of GP/Pari and Magma are running, and there is a persistent connection between Sage and those sessions.

```
sage: gp('sqrt(2)')
1.414213562373095048801688724
sage: magma('Sqrt(2)')                    # optional
1.41421356237309504880168872421
```

You probably noticed a pause when evaluated the second line as Magma started up. Also, note the `# optional` comment, which indicates that the line won't work if you don't have Magma installed.

Incidentally, if you want to numerically evaluate $\sqrt{2}$ in Sage, just give the optional `prec` argument to the `sqrt` function, which takes the required number of *bits* (binary digits) of precision.

```
sage: sqrt(2, prec=100)
1.4142135623730950488016887242
```

It's important to note in computations like this that there is not an *a priori* guarantee that `prec` bits of the *answer* are all correct. Instead, what happens is that Sage creates the number 2 as a floating point number with 100 bits of accuracy, then asks Paul Zimmerman's MPFR C library to compute the square root of that approximate number.

We return now to our symbolic expression $a = \sqrt{2}$. If you ask to square $a + 1$ you simply get the formal square. To expand out this formal square, we use the expand command.

```
sage: a = sqrt(2)
sage: (a+1)^2
(sqrt(2) + 1)^2
sage: expand((a+1)^2)
2*sqrt(2) + 3
```

Given any symbolic expression for which Sage can computes its minimal polynomial, you can construct the number field obtained by adjoining that expression to $\mathbb{Q}$. The notation is quite simple – just type `QQ[a]` where `a` is the symbolic expression.

```
sage: a = sqrt(2)
sage: K.<b> = QQ[a]
sage: K
Number Field in sqrt2 with defining polynomial x^2 - 2
sage: b
sqrt2
sage: (b+1)^2
2*sqrt2 + 3
sage: QQ[a/3 + 5]
Number Field in a with defining polynomial x^2 - 10*x + 223/9
```

You can't create the number field $\mathbb{Q}(a)$ in Sage by typing `QQ(a)`, which has a *very different* meaning in Sage. It means "try to create a rational number from a." Thus `QQ(a)` in Sage is the analogue of `QQ!a` in Magma (Pari has no notion of rings such as `QQ`).

```
sage: a = sqrt(2)
sage: QQ(a)
Traceback (most recent call last):
...
TypeError: unable to convert sqrt(2) to a rational
```

In general, if X is a ring, or vector space or other "parent structure" in Sage, and a is an element, type `X(a)` to make an element of X from a. For example, if X is the finite field of order 7, and $a = 2/5$ is a rational number, then `X(a)` is the finite field element 6 (as a quick exercise, check that this is mathematically the correct interpretation).

```
sage: X = GF(7); a = 2/5
sage: X(a)
6
```

As a slightly less trivial illustration of symbolic manipulation, consider the cubic equation

$$x^3 + \sqrt{2}x + 5 = 0. \tag{1.1}$$

In Sage, we can create this equation, and find an exact symbolic solution.

```
sage: x = var('x')
sage: eqn =  x^3 + sqrt(2)*x + 5 == 0
sage: a = solve(eqn, x)[0].rhs()
```

The first line above makes sure that the symbolic variable x is defined, the second creates the equation `eqn`, and the third line solves `eqn` for x, extracts the first solution (there are three), and takes the right hand side of that solution and assigns it to the variable `a`.

To see the solution nicely typeset, use the `show` command:

```
sage: show(a)
{{\left(...
```

$$\left(\frac{\sqrt{8\sqrt{2}+675}}{6\sqrt{3}}-\frac{5}{2}\right)^{\frac{1}{3}}\left(\frac{-\sqrt{3}i}{2}-\frac{1}{2}\right)-\frac{\sqrt{2}\left(\frac{\sqrt{3}i}{2}-\frac{1}{2}\right)}{3\left(\frac{\sqrt{8\sqrt{2}+675}}{6\sqrt{3}}-\frac{5}{2}\right)^{\frac{1}{3}}}$$

You can also see the latex needed to paste a into a paper by typing `latex(a)`. The `latex` command works on most Sage objects.

```
sage: latex(a)
{{\left( \frac{\sqrt{ {8 \sqrt{ 2 }} ...
```

Next, we construct the number field obtained by adjoining the solution `a` to $\mathbb{Q}$. Notice that the minimal polynomial of the root is $x^6 + 10x^3 - 2x^2 + 25$.

```
sage: K.<b> = QQ[a]
sage: K
Number Field in a with defining
polynomial x^6 + 10*x^3 - 2*x^2 + 25
sage: a.minpoly()
x^6 + 10*x^3 - 2*x^2 + 25
sage: b.minpoly()
x^6 + 10*x^3 - 2*x^2 + 25
```

We can now compute interesting invariants of the number field K:

```
sage: K.class_number()
5
sage: K.galois_group().order()
72
```

1.2 Galois Groups

We can compute the Galois group of the Galois closure as an abstract "Pari group" using the `galois_group` function, which by default calls Pari (`http://pari.math.u-bordeaux.fr/`). You do not have to worry about installing Pari, since *Pari is part of Sage.* In fact, despite appearances much of the difficult algebraic number theory in Sage is actually done by the Pari C library (be sure to also cite Pari in papers that use Sage).

```
sage: K.<alpha> = NumberField(x^3 - 2)
sage: G = K.galois_group()
sage: G
Galois group PARI group [6, -1, 2, "S3"] of degree 3 of the
Number Field in alpha with defining polynomial x^3 - 2
```

We can find out more about G, too:

```
sage: G.order()
6
```

We compute two more Galois groups of degree 5 extensions, and see that one has Galois group S_5, so is not solvable by radicals:

```
sage: NumberField(x^5 - 2, 'a').galois_group()
Galois group PARI group [20, -1, 3, "F(5) = 5:4"] of
degree 5 of the Number Field in a with defining
polynomial x^5 - 2
sage: NumberField(x^5 - x + 2, 'a').galois_group()
Galois group PARI group [120, -1, 5, "S5"] of degree 5 of
the Number Field in a with defining polynomial x^5 - x + 2
```

Recent versions of Magma have an algorithm for computing Galois groups that in theory applies when the input polynomial has any degree. There are no open source implementation of this algorithm (as far as I know). If you have Magma, you can use this algorithm from Sage by calling the `galois_group` function and giving the `algorithm='magma'` option.

```
sage: K.<a> = NumberField(x^3 - 2)
sage: K.galois_group(algorithm='magma')    # optional
verbose...
Galois group Transitive group number 2 of degree 3 of
the Number Field in a with defining polynomial x^3 - 2
```

We emphasize that the above example should not work if you don't have Magma.

It is also possible to work explicitly with the group of automorphisms of a field (though the link in Sage between abstract groups and automorphisms of fields is currently poor[1]). For example, here we first define $\mathbb{Q}(\sqrt[3]{2})$, then compute its Galois closure, which we represent as $\mathbb{Q}(b)$, where $b^6 + 40b^3 + 1372 = 0$. Then we compute the automorphism group of the field L, and explicitly list its elements.

```
sage: K.<a> = NumberField(x^3 - 2)
sage: L.<b> = K.galois_closure()
sage: L
Number Field in b with defining polynomial x^6 + 40*x^3 + 1372
sage: G = Hom(L, L)
sage: G
Automorphism group of Number Field in b ...
sage: G.list()
[
Ring endomorphism of Number Field in b ...
  Defn: b |--> b,
Ring endomorphism of Number Field in b ...
```

```
  Defn: b |--> 1/36*b^4 + 1/18*b,
...
Ring endomorphism of Number Field in b ...
  Defn: b |--> -2/63*b^4 - 31/63*b
]
```

You can explicitly apply any of the automorphisms above to any elements of L.

```
sage: phi = G.list()[1]
sage: phi
Ring endomorphism of Number Field in b ...
  Defn: b |--> 1/36*b^4 + 1/18*b
sage: phi(b^2 + 2/3*b)
-1/36*b^5 + 1/54*b^4 - 19/18*b^2 + 1/27*b
```

You can also enumerate all complex embeddings of a number field:

```
sage: K.complex_embeddings()
[
Ring morphism:
  From: Number Field in a with defining polynomial x^3 - 2
  To:   Complex Double Field
  Defn: a |--> -0.629960524947 - 1.09112363597*I,
Ring morphism:
  From: Number Field in a with defining polynomial x^3 - 2
  To:   Complex Double Field
  Defn: a |--> -0.629960524947 + 1.09112363597*I,
Ring morphism:
  From: Number Field in a with defining polynomial x^3 - 2
  To:   Complex Double Field
  Defn: a |--> 1.25992104989
]
```

1.3 Class Numbers and Class Groups

The class group C_K of a number field K is the group of fractional ideals of the maximal order R of K modulo the subgroup of principal fractional ideals. One of the main theorems of algebraic number theory asserts that C_K is a finite group. For example, the quadratic number field $\mathbb{Q}(\sqrt{-23})$ has class number 3, as we see using the Sage class_number command.

```
sage: L.<a> = NumberField(x^2 + 23)
sage: L.class_number()
3
```

There are only 9 quadratic imaginary field $\mathbb{Q}(\sqrt{D})$ that have class number 1:

$$D = -3, -4, -7, -8, -11, -19, -43, -67, -163.$$

To find this list using Sage, we first experiment with making lists in Sage. For example, typing [1..10] makes the list of integers between 1 and 10.

```
sage: [1..10]
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

We can also make the list of odd integers between 1 and 11, by typing [1,3,..,11], i.e., by giving the second term in the arithmetic progression.

```
sage: [1,3,..,11]
[1, 3, 5, 7, 9, 11]
```

Applying this idea, we make the list of negative numbers from -1 down to -10.

```
sage: [-1,-2,..,-10]
[-1, -2, -3, -4, -5, -6, -7, -8, -9, -10]
```

The first two lines below makes a list v of every D from -1 down to -200 such that D is a fundamental discriminant (the discriminant of a quadratic imaginary field). Note that you will not see the ... in the output below; this ... notation just means that part of the output is omitted below.

```
sage: w = [-1,-2,..,-200]
sage: v = [D for D in w if is_fundamental_discriminant(D)]
sage: v
[-3, -4, -7, -8, -11, -15, -19, -20, ..., -195, -199]
```

Finally, we make the list of D in our list v such that the quadratic number field $\mathbb{Q}(\sqrt{D})$ has class number 1. Notice that QuadraticField(D) is a shorthand for NumberField(x^2 - D).

```
sage: [D for D in v if QuadraticField(D,'a').class_number()==1]
[-3, -4, -7, -8, -11, -19, -43, -67, -163]
```

Of course, we have *not* proved that this is the list of all negative D so that $\mathbb{Q}(\sqrt{D})$ has class number 1.

A frustrating open problem is to prove that there are infinitely many number fields with class number 1. It is quite easy to be convinced that this is probably true by computing a bunch of class numbers of real quadratic fields. For example, over 58 percent of the real quadratic number fields with discriminant $D < 1000$ have class number 1!

```
sage: w = [1..1000]
sage: v = [D for D in w if is_fundamental_discriminant(D)]
sage: len(v)
302
sage: len([D for D in v if QuadraticField(D,'a').class_number() == 1])
176
sage: 176.0/302
0.582781456953642
```

For more intuition about what is going on, read about the Cohen-Lenstra heuristics.

Sage can also compute class numbers of extensions of higher degree, within reason. Here we use the shorthand `CyclotomicField(n)` to create the number field $\mathbb{Q}(\zeta_n)$.

```
sage: CyclotomicField(7)
Cyclotomic Field of order 7 and degree 6
sage: for n in [2..15]: print n, CyclotomicField(n).class_number()
2 1
3 1
...
15 1
```

In the code above, the notation `for n in [2..15]: ...` means "do ... for n equal to each of the integers $2, 3, 4, \ldots, 15$."

Exercise 1.1. Compute what is omitted (replaced by ...) in the output of the previous example.

Computations of class numbers and class groups in Sage is done by the Pari C library, and *unlike in Pari*, by default Sage tells Pari *not to assume* any conjectures. This can make some commands vastly slower than they might be directly in Pari, which *does assume unproved conjectures* by default. Fortunately, it is easy to tell Sage to be more permissive and allow Pari to assume conjectures, either just for this one call or henceforth for all number field functions. For example, with `proof=False` it takes only a few seconds to verify, modulo the conjectures assumed by Pari, that the class number of $\mathbb{Q}(\zeta_{23})$ is 3.

```
sage: CyclotomicField(23).class_number(proof=False)
3
```

Exercise 1.2. What is the smallest n such that $\mathbb{Q}(\zeta_n)$ has class number bigger than 1?

In addition to computing class numbers, Sage can also compute the group structure and generators for class groups. For example, the quadratic field $\mathbb{Q}(\sqrt{-30})$ has class group $C = (\mathbb{Z}/2\mathbb{Z})^{\oplus 2}$, with generators the ideal classes containing $(5, \sqrt{-30})$ and $(3, \sqrt{-30})$.

```
sage: K.<a> = QuadraticField(-30)
sage: C = K.class_group()
sage: C
Class group of order 4 with structure C2 x C2 of Number Field
in a with defining polynomial x^2 + 30
sage: category(C)
Category of groups
sage: C.gens()
[Fractional ideal class (5, a), Fractional ideal class (3, a)]
```

In Sage, the notation C.i means "the ith generator of the object C," where the generators are indexed by numbers $0, 1, 2, \ldots$. Below, when we write C.0 * C.1, this means "the product of the 0th and 1st generators of the class group C."

```
sage: K.<a> = QuadraticField(-30)
sage: C = K.class_group()
sage: C.0
Fractional ideal class (5, a)
sage: C.0.ideal()
Fractional ideal (5, a)
sage: I = C.0 * C.1
sage: I
Fractional ideal class (2, a)
```

Next we find that the class of the fractional ideal $(2, \sqrt{-30} + 4/3)$ is equal to the ideal class I.

```
sage: A = K.ideal([2, a+4/3])
sage: J = C(A)
sage: J
Fractional ideal class (2/3, 1/3*a)
sage: J == I
True
```

Unfortunately, there is currently no Sage function that writes a fractional ideal class in terms of the generators for the class group.

1.4 Orders in Number Fields

An *order* in a number field K is a subring of K whose rank over $\mathbb{Z}$ equals the degree of K. For example, if $K = \mathbb{Q}(\sqrt{-1})$, then $\mathbb{Z}[7i]$ is an order in K. A good first exercise is to prove that every element of an order is an algebraic integer.

```
sage: K.<I> = NumberField(x^2 + 1)
sage: R = K.order(7*I)
sage: R
Order in Number Field in I with defining polynomial x^2 + 1
sage: R.basis()
[1, 7*I]
```

Using the discriminant command, we compute the discriminant of this order:

```
sage: factor(R.discriminant())
-1 * 2^2 * 7^2
```

You can give any list of elements of the number field, and it will generate the smallest ring R that contains them.

```
sage: K.<a> = NumberField(x^4 + 2)
sage: K.order([12*a^2, 4*a + 12]).basis()
[1, 4*a, 4*a^2, 16*a^3]
```

If R isn't of rank equal to the degree of the number field (i.e., R isn't an order), then you'll get an error message.

```
sage: K.order([a^2])
Traceback (most recent call last):
...
ValueError: the rank of the span of gens is wrong
```

We can also compute the maximal order, using the `maxima_order` command, which behind the scenes finds an integral basis using Pari's `nfbasis` command. For example, $\mathbb{Q}(\sqrt[4]{2})$ has maximal order $\mathbb{Z}[\sqrt[4]{2}]$, and if α is a root of $x^3 + x^2 - 2x + 8$, then $\mathbb{Q}(\alpha)$ has maximal order with $\mathbb{Z}$-basis

$$1, \frac{1}{2}a^2 + \frac{1}{2}a, a^2.$$

```
sage: K.<a> = NumberField(x^4 + 2)
sage: K.maximal_order().basis()
[1, a, a^2, a^3]
sage: L.<a> = NumberField(x^3 + x^2 - 2*x+8)
sage: L.maximal_order().basis()
[1, 1/2*a^2 + 1/2*a, a^2]
sage: L.maximal_order().basis()[1].minpoly()
x^3 - 2*x^2 + 3*x - 10
```

There is still much important functionality for computing with non-maximal orders that is missing in Sage. For example, there is no support at all in Sage for computing with modules over orders or with ideals in non-maximal orders.

```
sage: K.<a> = NumberField(x^3 + 2)
sage: R = K.order(3*a)
sage: R.ideal(5)
Traceback (most recent call last):
...
NotImplementedError: ideals of non-maximal orders not
yet supported.
```

1.5 Relative Extensions

A *relative number field* L is a number field of the form $K(\alpha)$, where K is a number field, and an *absolute number field* is a number field presented in the form $\mathbb{Q}(\alpha)$. By the primitive element theorem, any relative number field $K(\alpha)$ can be written as $\mathbb{Q}(\beta)$ for some $\beta \in L$. However, in practice it is often convenient to view L as $K(\alpha)$. In Section 1.1 we constructed the number field $\mathbb{Q}(\sqrt{2})(\alpha)$, where α is a root of $x^3 + \sqrt{2}x + 5$, but *not* as a relative field—we obtained just the number field defined by a root of $x^6 + 10x^3 - 2x^2 + 25$.

To construct this number field as a relative number field, first we let K be $\mathbb{Q}(\sqrt{2})$.

```
sage: K.<sqrt2> = QuadraticField(2)
```

Next we create the univariate polynomial ring $R = K[X]$. In Sage, we do this by typing `R.<X> = K[]`. Here `R.<X>` means "create the object R with generator X" and `K[]` means a "polynomial ring over K", where the generator is named based on the afformentioned X (to create a polynomial ring in two variables X, Y simply replace `R.<X>` by `R.<X,Y>`).

```
sage: R.<X> = K[]
sage: R
Univariate Polynomial Ring in X over Number Field in sqrt2
with defining polynomial x^2 - 2
```

Now we can make a polynomial over the number field $K = \mathbb{Q}(\sqrt{2})$, and construct the extension of K obtained by adjoining a root of that polynomial to K.

```
sage: L.<a> = K.extension(X^3 + sqrt2*X + 5)
sage: L
Number Field in a with defining polynomial X^3 + sqrt2*X + 5...
```

Finally, L is the number field $\mathbb{Q}(\sqrt{2})(\alpha)$, where α is a root of $X^3 + \sqrt{2}\alpha + 5$. We can do now do arithmetic in this number field, and of course include $\sqrt{2}$ in expressions.

```
sage: a^3
(-sqrt2)*a - 5
sage: a^3 + sqrt2*a
-5
```

The relative number field L also has numerous functions, many of which are by default relative. For example the `degree` function on L returns the relative degree of L over K; for the degree of L over $\mathbb{Q}$ use the `absolute_degree` function.

```
sage: L.degree()
3
sage: L.absolute_degree()
6
```

Given any relative number field you can also an absolute number field that is isomorphic to it. Below we create $M = \mathbb{Q}(b)$, which is isomorphic to L, but is an absolute field over $\mathbb{Q}$.

```
sage: M.<b> = L.absolute_field()
sage: M
Number Field in b with defining
polynomial x^6 + 10*x^3 - 2*x^2 + 25
```

The `structure` function returns isomorphisms in both directions between M and L.

```
sage: M.structure()
(Isomorphism from Number Field in b ...,
 Isomorphism from Number Field in a ...)
```

In Sage one can create arbitrary towers of relative number fields (unlike in Pari, where a relative extension must be a single extension of an absolute field).

```
sage: R.<X> = L[]
sage: Z.<b> = L.extension(X^3 - a)
sage: Z
Number Field in b with defining polynomial
X^3 + (-1)*a over its base field
sage: Z.absolute_degree()
18
```

Exercise 1.3. Construct the relative number field $L = K(\sqrt[3]{\sqrt{2}+\sqrt{3}})$, where $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

One shortcoming with relative extensions in Sage is that behind the scenes all arithmetic is done in terms of a single absolute defining polynomial, and in some cases this can be very slow (much slower than Magma). Perhaps this could be fixed by using Singular's multivariate polynomials modulo an appropriate ideal, since Singular polynomial arithmetic is extremely flast. Also, Sage has very little direct support for constructive class field theory, which is a major motivation for explicit computation with relative orders; it would be good to expose more of Pari's functionality in this regard.

2 A Birds Eye View

We now take a whirlwind tour of some of the number theoretical functionality of Sage. There is much that we won't cover here, but this should help give you a flavor for some of the number theoretic capabilities of Sage, much of which is unique to Sage.

2.1 Integer Factorization

Bill Hart's quadratic sieve is included with Sage. The quadratic sieve is the best algorithm for factoring numbers of the form pq up to around 100 digits. It involves searching for relations, solving a linear algebra problem modulo 2, then factoring n using a relation $x^2 \equiv y^2 \mod n$.

```
sage: qsieve(next_prime(2^90)*next_prime(2^91), time=True)   # not tested
([1237940039285380274899124357, 2475880078570760549798248507],
 '14.94user 0.53system 0:15.72elapsed 98%CPU (0avgtext+0avgdata 0maxresident)k')
```

Using *qsieve* is twice as fast as Sage's *general* factor command in this example. Note that Sage's general factor command does nothing but call Pari's factor C library function.

```
sage: time factor(next_prime(2^90)*next_prime(2^91))     # not tested
CPU times: user 28.71 s, sys: 0.28 s, total: 28.98 s
Wall time: 29.38 s
1237940039285380274899124357 * 2475880078570760549798248507
```

Obviously, Sage's `factor` command should not just call Pari, but nobody has gotten around to rewriting it yet.

Paul Zimmerman's GMP-ECM is included in Sage. The elliptic curve factorization (ECM) algorithm is the best algorithm for factoring numbers of the form $n = pm$, where p is not "too big". ECM is an algorithm due to Hendrik Lenstra, which works by "pretending" that n is prime, chosing a random elliptic curve over $\mathbb{Z}/n\mathbb{Z}$, and doing arithmetic on that curve—if something goes wrong when doing arithmetic, we factor n.

In the following example, GMP-ECM is over 10 times faster than Sage's generic factor function. Again, this emphasizes that Sage's generic `factor` command would benefit from a rewrite that uses GMP-ECM and qsieve.

```
sage: time ecm.factor(next_prime(2^40) * next_prime(2^300))    # not tested
CPU times: user 0.85 s, sys: 0.01 s, total: 0.86 s
Wall time: 1.73 s
[1099511627791,
 2037035976334486086268445688409378161051468393665936250636140449354381299763336706183397533
sage: time factor(next_prime(2^40) * next_prime(2^300))        # not tested
CPU times: user 23.82 s, sys: 0.04 s, total: 23.86 s
Wall time: 24.35 s
1099511627791 * 2037035976334486086268445688409378161051468393665936250636140449354381299763
```

2.2 Elliptic Curves

Cremona's databases of elliptic curves is part of Sage. The curves up to conductor 10,000 come standard with Sage, and an optional 75MB download gives all his tables up to conductor 130,000. Type `sage -i database_cremona_ellcurve-20071019` to automatically download and install this extended table.

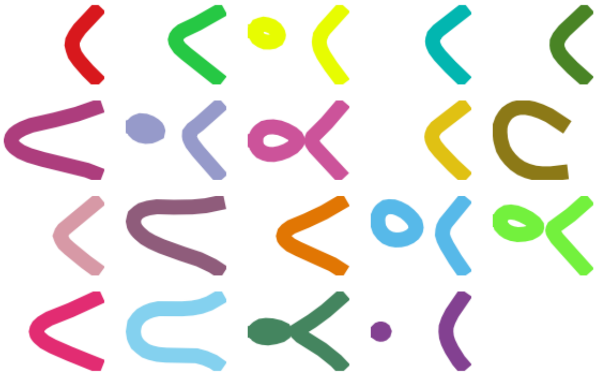
To use the database, just create a curve by giving

```
sage: EllipticCurve('5077a1')
Elliptic Curve defined by y^2 + y = x^3 - 7*x + 6 over Rational Field
sage: C = CremonaDatabase()
sage: C.number_of_curves()
847550
sage: C[37]
{'a': {'a1': [[0, 0, 1, -1, 0], 1, 1],
       'b1': [[0, 1, 1, -23, -50], 0, 3], ...
sage: C.isogeny_class('37b')
[Elliptic Curve defined by y^2 + y = x^3 + x^2 - 23*x - 50
over Rational Field, ...]
```

There is also a Stein-Watkins database that contains hundreds of millions of elliptic curves. It's over a 2GB download though!

Bryan Birch's recently had a birthday conference, and I used Sage to draw the cover of his birthday card by enumerating all optimal elliptic curves of conductor up to 37, then plotting them with thick randomly colored lines. As you can see below, plotting an elliptic curve is as simple as calling the plot method on it. Also, the `graphics_array` command allows us to easily combine numerous plots into a single graphics object.

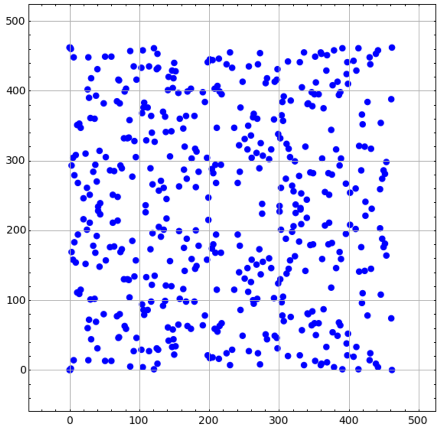
```
sage: v = cremona_optimal_curves([11..37])
sage: w = [E.plot(thickness=10,
   rgbcolor=(random(),random(),random())) for E in v]
sage: graphics_array(w, 4, 5).show(axes=False)
```



We can use Sage's interact feature to draw a plot of an elliptic curve modulo p, with a slider that one drags to change the prime p. The interact feature of Sage is very helpful for interactively changing parameters and viewing the results. Type `interact?` for more help and examples and visit the webpage `http://wiki.sagemath.org/interact`.

In the code below we first define the elliptic curve E using the Cremona label `37a`. Then we define an interactive function f, which is made interactive using the `@interact` Python *decorator*. Because the default for p is `primes(2,500)`, the Sage notebook constructs a slider that varies over the primes up to 500. When you drag the slider and let go, a plot is drawn of the affine $\mathbb{F}_p$ points on the curve $E_{\mathbb{F}_p}$. Of course, one *should* never plot curves over finite fields, which makes this even more fun.

```
E = EllipticCurve('37a')
@interact
def f(p=primes(2,500)):
    show(plot(E.change_ring(GF(p)),pointsize=30),
    axes=False, frame=True, gridlines="automatic",
    aspect_ratio=1, gridlinesstyle={'rgbcolor':(0.7,0.7,0.7)})
```



Sage includes `sea.gp`, which is a fast implementation of the SEA (Schoff-Elkies-Atkin) algorithm for counting the number of points on an elliptic curve over $\mathbb{F}_p$.

We create the finite field $k = \mathbb{F}_p$, where p is the next prime after 10^{20}. The `next_prime` command uses Pari's `nextprime` function, but proves primality of the result (unlike Pari which gives only the next probable prime after a number). Sage also has a `next_probable_prime` function.

```
sage: k = GF(next_prime(10^20))
```

compute its cardinality, which behind the scenes uses SEA.

```
sage: E = EllipticCurve(k.random_element())
sage: E.cardinality()                   # less than a second
100000000005466254167
```

To see how Sage chooses when to use SEA versus other methods, type `E.cardinality??` and read the source code. As of this writing, it simply uses SEA whenever $p > 10^{18}$.

Sage has the world's best code for computing p-adic regulators of elliptic curves, thanks to work of David Harvey and Robert Bradshaw. The p-adic regulator of an elliptic curve E at a good ordinary prime p is the determinant of the global p-adic height pairing matrix on the Mordell-Weil group $E(\mathbb{Q})$. (This has nothing to do with local or archimedean heights.) This is the analogue of the regulator in the Mazur-Tate-Teitelbaum p-adic analogue of the Birch and Swinnerton-Dyer conjecture.

In particular, Sage implements Harvey's improvement on an algorithm of Mazur-Stein-Tate, which builds on Kiran Kedlaya's Monsky-Washnitzer approach to computing p-adic cohomology groups.

We create the elliptic curve with Cremona label 389a, which is the curve of smallest conductor and rank 2. We then compute both the 5-adic and 997-adic regulators of this curve.

```
sage: E = EllipticCurve('389a')
sage: E.padic_regulator(5, 10)
5^2 + 2*5^3 + 2*5^4 + 4*5^5 + 3*5^6 + 4*5^7 + 3*5^8 + 5^9 + O(5^11)
sage: E.padic_regulator(997, 10)
740*997^2 + 916*997^3 + 472*997^4 + 325*997^5 + 697*997^6
          + 642*997^7 + 68*997^8 + 860*997^9 + 884*997^10 + O(997^11)
```

Before the new algorithm mentioned above, even computing a 7-adic regulator to 3 digits of precision was a nontrivial computational challenge. Now in Sage computing the 100003-adic regulator is routine:

```
sage: E.padic_regulator(100003,5)  # a couple of seconds
42582*100003^2 + 35250*100003^3 + 12790*100003^4 + 64078*100003^5 + O(100003^6)
```

p-adic L-functions play a central role in the arithmetic study of elliptic curves. They are p-adic analogues of complex analytic L-function, and their leading coefficient (at 0) is the analogue of $L^{(r)}(E,1)/\Omega_E$ in the p-adic analogue of the Birch and Swinnerton-Dyer conjecture. They also appear in theorems of Kato, Schneider, and others that prove partial results toward p-adic BSD using Iwasawa theory.

The implementation in Sage is mainly due to work of myself, Christian Wuthrich, and Robert Pollack. We use Sage to compute the 5-adic L-series of the elliptic curve 389a of rank 2.

```
sage: E = EllipticCurve('389a')
sage: L = E.padic_lseries(5)
sage: L
```

```
5-adic L-series of Elliptic Curve defined
by y^2 + y = x^3 + x^2 - 2*x over Rational Field
sage: L.series(3)
O(5^5) + O(5^2)*T + (4 + 4*5 + O(5^2))*T^2 +
(2 + 4*5 + O(5^2))*T^3 + (3 + O(5^2))*T^4 + O(T^5)
```

Sage implements code to compute numerous explicit bounds on Shafarevich-Tate Groups of elliptic curves. This functionality is *only* available in Sage, and uses results Kolyvagin, Kato, Perrin-Riou, etc., and unpublished papers of Wuthrich and me.

```
sage: E = EllipticCurve('11a1')
sage: E.sha().bound()           # so only 2,3,5 could divide sha
[2, 3, 5]
sage: E = EllipticCurve('37a1')  # so only 2 could divide sha
sage: E.sha().bound()
([2], 1)
sage: E = EllipticCurve('389a1')
sage: E.sha().bound()
(0, 0)
```

The $(0, 0)$ in the last output above indicates that the Euler systems results of Kolyvagin and Kato give no information about finiteness of the Shafarevich-Tate group of the curve E. In fact, it is an open problem to prove this finiteness, since E has rank 2, and finiteness is only known for elliptic curves for which $L(E, 1) \neq 0$ or $L'(E, 1) \neq 0$.

Partial results of Kato, Schneider and others on the p-adic analogue of the BSD conjecture yield algorithms for bounding the p-part of the Shafarevich-Tate group. These algorithms require as input explicit computation of p-adic L-functions, p-adic regulators, etc., as explained in Stein-Wuthrich. For example, below we use Sage to prove that 5 and 7 do not divide the Shafarevich-Tate group of our rank 2 curve 389a.

```
sage: E = EllipticCurve('389a1')
sage: sha = E.sha()
sage: sha.p_primary_bound(5)  # iwasawa theory ==> 5 doesn't divide sha
0
sage: sha.p_primary_bound(7)  # iwasawa theory ==> 7 doesn't divide sha
0
```

This is consistent with the Birch and Swinnerton-Dyer conjecture, which predicts that the Shafarevich-Tate group is trivial. Below we compute this predicted order, which is the floating point number 1.000000 to some precision. That the result is a floating point number helps emphasize that it is an open problem to show that the *conjectural order* of the Shafarevich-Tate group is even a rational number in general!

```
sage: E.sha().an()
1.00000000000000
```

2.3 Mordell-Weil Groups and Integral Points

Sage includes both Cremona's mwrank library and Simon's 2-descent GP scripts for computing Mordell-Weil groups of elliptic curves.

```
sage: E = EllipticCurve([1,2,5,7,17])
sage: E.conductor()       # not in the Tables
154907
sage: E.gens()            # a few seconds
[(1 : 3 : 1), (67/4 : 507/8 : 1)]
```

Sage can also compute the torsion subgroup, isogeny class, determine images of Galois representations, determine reduction types, and includes a full implementation of Tate's algorithm over number fields.

Sage has the world's fastest implementation of computation of all integral points on an elliptic curve over $\mathbb{Q}$, due to work of Cremona, Michael Mardaus, and Tobias Nagel. This is also the only free open source implementation available.

```
sage: E = EllipticCurve([1,2,5,7,17])
sage: E.integral_points(both_signs=True)
[(1 : -9 : 1), (1 : 3 : 1)]
```

A very impressive example is the lowest conductor elliptic curve of rank 3, which has 36 integral points.

```
sage: E = elliptic_curves.rank(3)[0]
sage: E.integral_points(both_signs=True)   # less than 3 seconds
[(-3 : -1 : 1), (-3 : 0 : 1), (-2 : -4 : 1), (-2 : 3 : 1),
 ...(816 : -23310 : 1), (816 : 23309 : 1)]
```

The algorithm to compute all integral points involves first computing the Mordell-Weil group, then bounding the integral points, and listing all integral points satisfying those bounds. See Cohen's new GTM 239 for complete details.

The complexity grows exponentially in the rank of the curve. We can do the above calculation, but with the first known curve of rank 4, and it finishes in about a minute (and outputs 64 points).

```
sage: E = elliptic_curves.rank(4)[0]
sage: E.integral_points(both_signs=True)   # about a minute
[(-10 : 3 : 1), (-10 : 7 : 1), ...
 (19405 : -2712802 : 1), (19405 : 2693397 : 1)]
```

2.4 Elliptic Curve L-functions

We next compute with the complex L-function

$$L(E,s) = \prod_{p\nmid\Delta=389} \frac{1}{1-a_p p^{-s}+pp^{-2s}} \cdot \prod_{p\mid\Delta=389} \frac{1}{1-a_p p^{-s}}$$

of E. Though the above Euler product only defines an analytic function on the right half plane where $\mathrm{Re}(s) > 3/2$, a deep theorem of Wiles et al. (the **Modularity Theorem**) implies that it has an analytic continuation to the whole complex plane and functional equation. We can evaluate the function L anywhere on the complex plane using Sage (via code of Tim Dokchitser).

```
sage: E = EllipticCurve('389a1')
sage: L = E.lseries()
sage: L
Complex L-series of the Elliptic Curve defined by
       y^2 + y = x^3 + x^2 - 2*x over Rational Field
sage: L(1)
-1.04124792770327e-19
sage: L(1+I)
-0.638409938588039 + 0.715495239204667*I
sage: L(100)
1.00000000000000
```

We can also compute the Taylor series of L about *any* point, thanks to Tim Dokchitser's code.

```
sage: E = EllipticCurve('389a1')
sage: L = E.lseries()
sage: Ld = L.dokchitser()
sage: Ld.taylor_series(1,4)
-1.28158145691931e-23 + (7.26268290635587e-24)*z + 0.759316500288427*z^2
                      - 0.430302337583362*z^3 + O(z^4)
```

The Generalized Riemann Hypothesis asserts that all nontrivial zeros of $L(E, s)$ are of the form $1 + iy$. Mike Rubinstein has written a C++ program that is part of Sage that can for any n compute the first n values of y such that $1 + iy$ is a zero of $L(E, s)$. It also verifies the Riemann Hypothesis for these zeros (I think). Rubinstein's program can also do similar computations for a wide class of L-functions, though not all of this functionality is as easy to use from Sage as for elliptic curves. Below we compute the first 10 zeros of $L(E, s)$, where E is still the rank 2 curve 389a.

```
sage: L.zeros(10)
[0.000000000, 0.000000000, 2.87609907, 4.41689608, 5.79340263,
 6.98596665, 7.47490750, 8.63320525, 9.63307880, 10.3514333]
```

2.5 The Matrix of Frobenius on Hyperelliptic Curves

Sage has a highly optimized implementation of the Harvey-Kedlaya algorithm for computing the matrix of Frobenius associated to a curve over a finite field. This is an implementation by David Harvey, which is GPL'd and depends only on NTL and `zn_poly` (a C library in Sage for fast arithmetic $(\mathbb{Z}/n\mathbb{Z})[x]$).

We import the `hypellfrob` function and call it on a polynomial over $\mathbb{Z}$.

```
sage: from sage.schemes.hyperelliptic_curves.hypellfrob import hypellfrob
sage: R.<x> = PolynomialRing(ZZ)
sage: f = x^5 + 2*x^2 + x + 1; p = 101
sage: M = hypellfrob(p, 1, f); M
[ 0 + O(101)  0 + O(101) 93 + O(101) 62 + O(101)]
[ 0 + O(101)  0 + O(101) 55 + O(101) 19 + O(101)]
[ 0 + O(101)  0 + O(101) 65 + O(101) 42 + O(101)]
[ 0 + O(101)  0 + O(101) 89 + O(101) 29 + O(101)]
```

We do the same calculation but in $\mathbb{Z}/101^4\mathbb{Z}$, which gives enough precision to recognize the exact characteristic polynomial in $\mathbb{Z}[x]$ of Frobenius as an element of the endomorphism ring. This computation is still very fast, taking only a fraction of a second.

```
sage: M = hypellfrob(p, 4, f)   # about 0.25 seconds
sage: M[0,0]
91844754 + O(101^4)
```

The characteristic polynomial of Frobenius is $x^4 + 7x^3 + 167x^2 + 707x + 10201$, which determines the ζ function of the curve $y^2 = f(x)$.

```
sage: M.charpoly()
(1 + O(101^4))*x^4 + (7 + O(101^3))*x^3 + (167 + O(101^3))*x^2
   + (707 + O(101^3))*x + (10201 + O(101^4))
```

2.6 Modular Symbols

Modular symbols play a key role in algorithms for computing with modular forms, special values of L-functions, elliptic curves, and modular abelian varieties. Sage has the most general implementation of modular symbols available, thanks to work of myself, Jordi Quer (of Barcelona) and Craig Citro (a student of Hida). Moreover, computation with modular symbols is by far my *most favorite* part of computational mathematics. There is still a lot of tuning and optimization work to be done for modular symbols in Sage, in order for it to be across the board the fastest implementation in the world, since my Magma implementation is still better in some important cases.

We create the space M of weight 4 modular symbols for a certain congruence subgroup $\Gamma_H(13)$ of level 13. Then we compute a basis for this space, expressed in terms of *Manin symbols*. Finally, we compute the Hecke operator T_2 acting on M, find its characteristic polynomial and factor it. We also compute the dimension of the cuspidal subspace.

```
sage: M = ModularSymbols(GammaH(13,[3]), weight=4)
sage: M
Modular Symbols space of dimension 14 for Congruence Subgroup
Gamma_H(13) with H generated by [3] of weight 4 with sign 0
and over Rational Field
sage: M.basis()
```

```
([X^2,(0,1)], [X^2,(0,7)], [X^2,(2,5)], [X^2,(2,8)], [X^2,(2,9)],
 [X^2,(2,10)], [X^2,(2,11)], [X^2,(2,12)], [X^2,(4,0)], [X^2,(4,3)],
 [X^2,(4,6)], [X^2,(4,8)], [X^2,(4,12)], [X^2,(7,1)])
sage: factor(charpoly(M.T(2)))
(x - 7) * (x + 7) * (x - 9)^2 * (x + 5)^2
        * (x^2 - x - 4)^2 * (x^2 + 9)^2
sage: dimension(M.cuspidal_subspace())
10
```

Sage includes John Cremona's specialized and *insanely fast* implementation of modular symbols for weight 2 and trivial character. We illustrate below computing the space of modular symbols of level 20014, which has dimension 5005, along with a Hecke operator on this space. The whole computation below takes only a few seconds; a similar computation takes a few minutes using Sage's generic modular symbols code. Moreover, Cremona has done computations at levels over 200,000 using his library, so the code is known to scale well to large problems. The new code in Sage for modular symbols is much more general, but doesn't scale nearly so well (yet).

```
sage: M = CremonaModularSymbols(20014)       # few seconds
sage: M
Cremona Modular Symbols space of dimension 5005 for
Gamma_0(20014) of weight 2 with sign 0
sage: t = M.hecke_matrix(3)              # few seconds
```

2.7 Enumerating Totally Real Number Fields

As part of his project to enumerate Shimura curves, John Voight has contributed code to Sage for enumerating totally real number fields. The algorithm isn't extremely complicated, but it involves some "inner loops" that have to be coded to run very quickly. Using *Cython*, Voight was able to implement exactly the variant of Newton iteration that he needed for his problem.

The function `enumerate_totallyreal_fields_prim(n, B, ...)` enumerates without using a database (!) primitive (no proper subfield) totally real fields of degree $n > 1$ with discriminant $d \leq B$.

We compute the totally real quadratic fields of discriminant ≤ 50. The calculation below, which is almost instant, is done in real time and is not a table lookup.

```
sage: enumerate_totallyreal_fields_prim(2,50)
[[5, x^2 - x - 1], [8, x^2 - 2], [12, x^2 - 3], [13, x^2 - x - 3],
 [17, x^2 - x - 4], [21, x^2 - x - 5], [24, x^2 - 6], [28, x^2 - 7],
 [29, x^2 - x - 7], [33, x^2 - x - 8], [37, x^2 - x - 9],
 [40, x^2 - 10], [41, x^2 - x - 10], [44, x^2 - 11]]
```

We compute all totally real quintic fields of discriminant $\leq 10^5$. Again, this is done in real time – it's not a table lookup!

```
sage: enumerate_totallyreal_fields_prim(5,10^5)
[[14641, x^5 - x^4 - 4*x^3 + 3*x^2 + 3*x - 1],
 [24217, x^5 - 5*x^3 - x^2 + 3*x + 1],
 [36497, x^5 - 2*x^4 - 3*x^3 + 5*x^2 + x - 1],
 [38569, x^5 - 5*x^3 + 4*x - 1],
 [65657, x^5 - x^4 - 5*x^3 + 2*x^2 + 5*x + 1],
 [70601, x^5 - x^4 - 5*x^3 + 2*x^2 + 3*x - 1],
 [81509, x^5 - x^4 - 5*x^3 + 3*x^2 + 5*x - 2],
 [81589, x^5 - 6*x^3 + 8*x - 1],
 [89417, x^5 - 6*x^3 - x^2 + 8*x + 3]]
```

2.8 Bernoulli Numbers

From the mathematica website:

> "**Today We Broke the Bernoulli Record: From the Analytical Engine to Mathematica**
> April 29, 2008
> Oleksandr Pavlyk, Kernel Technology
> A week ago, I took our latest development version of Mathematica, and I typed `BernoulliB[10^7]`.
> And then I waited.
> Yesterday—5 days, 23 hours, 51 minutes, and 37 seconds later—I got the result!"

Tom Boothby did that same computation in Sage, which uses Pari's `bernfrac` command that uses evaluation of ζ and factorial to high precision, and it took 2 days, 12 hours.

Then David Harvey came up with an entirely new algorithm that parallelizes well. He gives these timings for computing B_{10^7} on his machine (it takes 59 minutes, 57 seconds on my 16-core 1.8ghz Opteron box):

```
PARI: 75 h, Mathematica: 142 h
bernmm (1 core) = 11.1 h, bernmm (10 cores) = 1.3 h
```

> "Running on 10 cores for 5.5 days, I [David Harvey] computed [the Bernoulli number] B_k for $k = 10^8$, which I believe is a new record. Essentially it's the multimodular algorithm I suggested earlier on this thread, but I figured out some tricks to optimise the crap out of the computation of $B_k \text{mod} p$."

So now Sage is the fastest in the world for large Bernoulli numbers. The timings below are on a 16-core 1.8Ghz Opteron box.

```
sage: w = bernoulli(100000, num_threads=16)     # 1.87 seconds
sage: w = bernoulli(100000, algorithm='pari')   # 28 seconds
```

2.9 Polynomial Arithmetic

Sage uses Bill Hart and David Harvey's GPL'd Flint C library for arithmetic in $\mathbb{Z}[x]$. Its main claim to fame is that it is the world's fastest for polynomial multiplication, e.g., in the benchmark below it is 3 times faster than NTL and twice as fast as Magma. Behind the scenes it contains some carefully tuned discrete Fourier transform code (which I know nearly nothing about).

```
sage: Rflint = PolynomialRing(ZZ, 'x')
sage: f = Rflint([ZZ.random_element(2^64) for _ in [1..32]])
sage: g = Rflint([ZZ.random_element(2^64) for _ in [1..32]])
sage: timeit('f*g')                # random output
625 loops, best of 3: 105 microseconds per loop
sage: Rntl = PolynomialRing(ZZ, 'x', implementation='NTL')
sage: f = Rntl([ZZ.random_element(2^64) for _ in [1..32]])
sage: g = Rntl([ZZ.random_element(2^64) for _ in [1..32]])
sage: timeit('f*g')                # random output
625 loops, best of 3: 310 microseconds per loop
sage: ff = magma(f); gg = magma(g)
sage: s = 'time v := [%s * %s for _ in [1..10^5]];'%(ff.name(), gg.name())
sage: magma.eval(s)      # random output
'Time: 17.120'
sage: (17.120/10^5)*10^(6)    # convert to microseconds
171.200000000000
```

Multivariate polynomial arithmetic in many cases uses Singular in library mode (Martin Albrecht), which is quite fast. For example, below we do the Fateman benchmark over the finite field of order 32003.

```
sage: P.<x,y,z> = GF(32003)[]
sage: p = (x+y+z+1)^20
sage: q = p+1
sage: timeit('p*q')   # random output
5 loops, best of 3: 384 ms per loop
sage: pp = magma(p); qq = magma(q)
sage: s = 'time w := %s*%s;'%(pp.name(),qq.name())
sage: magma.eval(s)
'Time: 1.480'
```

Notice that the multiplication takes about four times as long in Magma.

3 Modular Forms

This section is about computing with modular forms, modular symbols, and modular abelian varieties. Most of the Sage functionality we describe below is new code written for Sage by myself, Craig Citro, Robert Bradshaw, and Jordi Quer in consultation with John Cremona. It has much overlap in functionality with the modular forms code in Magma, which I developed during 1998–2004.

3.1 Modular Forms and Hecke Operators

A *congruence subgroup* is a subgroup of the group $\mathrm{SL}_2(\mathbb{Z})$ of determinant ±1 integer matrices that contains

$$\Gamma(N) = \mathrm{Ker}(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$$

for some positive integer N. Since $\Gamma(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$, all congruence subgroups have finite index. The converse is *not* true, though in many other settings it is true (see [paper of Serre]).

The inverse image $\Gamma_0(N)$ of the subgroup of upper triangular matrices in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is a congruence subgroup, as is the inverse image $\Gamma_1(N)$ of the subgroup of matrices of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$. Also, for any subgroup $H \subset (\mathbb{Z}/N\mathbb{Z})^*$, the inverse image $\Gamma_H(N)$ of the subgroup of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ of all elements of the form $\left(\begin{smallmatrix} a & * \\ 0 & d \end{smallmatrix}\right)$ with $d \in H$ is a congruence subgroup.

We can create each of the above congruence subgroups in Sage, using the `Gamma0`, `Gamma1`, and `GammaH` commands.

```
sage: Gamma0(8)
Congruence Subgroup Gamma0(8)
sage: Gamma1(13)
Congruence Subgroup Gamma1(13)
sage: GammaH(11,[2])
Congruence Subgroup Gamma_H(11) with H generated by [2]
```

The second argument to the `GammaH` command is a list of generators of the subgroup H of $(\mathbb{Z}/N\mathbb{Z})^*$.
Sage can compute a list of generators for these subgroups. The algorithm Sage uses is a straightforward generic procedure that uses coset representatives for the congruence subgroup (which are easy to enumerate) to obtain a list of generators [[ref my modular forms book]].

```
sage: Gamma0(2).gens()
([1 1]
 [0 1],
 [-1  0]
 [ 0 -1],
 [ 1 -1]
 [ 0  1],
 [ 1 -1]
 [ 2 -1],
 [-1  1]
 [-2  1])
sage: len(Gamma1(13).gens())
284
```

As you can see above, the list of generators Sage computes is unfortunately large. Improving this would be an excellent Sage development project, which would involve much beautiful mathematics.

A *modular form* on a congruence subgroup Γ of integer weight k is a holomorphic function $f(z)$ on the upper half plane

$$\mathfrak{h}^* = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\} \cup \mathbb{Q} \cup \{i\infty\}$$

such that for every matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, we have

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z). \qquad (3.1)$$

A *cusp form* is a modular form that vanishes at all of the *cusps* $\mathbb{Q} \cup \{i\infty\}$.

If Γ contains $\Gamma_1(N)$ for some N, then $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$, so (3.1) implies that $f(z) = f(z+1)$. This, coupled with the holomorphicity condition, implies that $f(z)$ has a Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} a_n e^{2\pi i n z},$$

with $a_n \in \mathbb{C}$. We let $q = e^{2\pi i z}$, and call $f = \sum_{n=0}^{\infty} a_n q^n$ the *q-expansion* of f.

Henceforth we assume that Γ is either $\Gamma_1(N)$, $\Gamma_0(N)$, or $\Gamma_H(N)$ for some H and N. The complex vector space $M_k(\Gamma)$ of all modular forms of weight k on Γ is a finite dimensional vector space.

We create the space $M_k(\Gamma)$ in Sage by typing `ModularForms(G, k)` where G is the congruence subgroup and k is the weight.

```
sage: ModularForms(Gamma0(25), 4)
Modular Forms space of dimension 11 for ...
sage: S = CuspForms(Gamma0(25),4, prec=15); S
Cuspidal subspace of dimension 5 of Modular Forms space ...
sage: S.basis()
[
q + q^9 - 8*q^11 - 8*q^14 + O(q^15),
q^2 - q^7 - q^8 - 7*q^12 + 7*q^13 + O(q^15),
q^3 + q^7 - 2*q^8 - 6*q^12 - 5*q^13 + O(q^15),
q^4 - q^6 - 3*q^9 + 5*q^11 - 2*q^14 + O(q^15),
q^5 - 4*q^10 + O(q^15)
]
```

Sage computes the dimensions of all these spaces using simple arithmetic formulas instead of actually computing bases for the spaces in question. In fact, Sage has the most general collection of modular forms dimension formulas of any software; type `help(sage.modular.dims)` to see a list of arithmetic functions that are used to implement these dimension formulas.

```
sage: ModularForms(Gamma1(949284), 456).dimension()
11156973844800
sage: a = [dimension_cusp_forms(Gamma0(N),2) for N in [1..25]]; a
[0, 0, ..., 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 2, 2, 1, 0]
```

```
sage: sloane_find(a)
Searching Sloane’s online database...
[[1617,
  ’Genus of modular group GAMMA_0 (n). Or, genus of
   modular curve X_0(n).’,...
```

Sage doesn’t have simple formulas for dimensions of spaces of modular forms of weight 1, since such formulas perhaps do not exist.

The space $M_k(\Gamma_1(N))$ is equipped with an action of $(\mathbb{Z}/N\mathbb{Z})^*$ by *diamond bracket operators* $\langle d\rangle$, and this induces a decomposition

$$M_k(\Gamma_1(N)) = \bigoplus_{\varepsilon:(\mathbb{Z}/N\mathbb{Z})^*\to\mathbb{C}^*} M_k(N,\varepsilon),$$

where the sum is over all complex characters of the finite abelian group $(\mathbb{Z}/N\mathbb{Z})^*$. These characters are called *Dirichlet characters*, which are central in number theory.

Theorem 3.1. *The space $M_k(\Gamma_1(N))$ has a basis of elements whose q-expansions $f(q)$ are all elements of $\mathbb{Z}[[q]]$.*

The factors $M_k(N,\varepsilon)$ then have bases whose q-expansions are elements of $R[[q]]$, where $R = \mathbb{Z}[\varepsilon]$ is the ring generated over $\mathbb{Z}$ by the image of ε. We illustrate this with $N = k = 5$ below, where `DirichletGroup` will be described later.

```
sage: CuspForms(DirichletGroup(5).0, 5).basis()
[q + (-zeta4 - 1)*q^2 + (6*zeta4 - 6)*q^3 - ... + O(q^6)]
```

Use the command `DirichletGroup(N,R)` to create the group of all Dirichlet characters of modulus N taking values in the ring R. If R is omited, it defaults to a cyclotomic field.

```
sage: G = DirichletGroup(8); G
Group of Dirichlet characters of modulus 8 over Cyclotomic
Field of order 2 and degree 1
sage: v = G.list(); v
[[1, 1], [-1, 1], [1, -1], [-1, -1]]
sage: eps = G.0; eps
[-1, 1]
sage: [eps(3), eps(5)]
[-1, 1]
```

Sage both represents Dirichlet characters by giving a “matrix”, i.e., the list of images of canonical generators of $(\mathbb{Z}/N\mathbb{Z})^*$, and as vectors modulo and integer n. For years, I was torn between these two representations, until J. Quer and I realized that the best approach is to use both and make it easy to convert between them.

```
sage: parent(eps.element())
Vector space of dimension 2 over Ring of integers modulo 2
```

Given a Dirichlet character, Sage also lets you compute the associated Jacobi and Gauss sums, generalized Bernoulli numbers, the conductor, Galois orbit, etc.

Recall that Dirichlet characters give a decomposition

$$M_k(\Gamma_1(N)) = \bigoplus_{\varepsilon:(\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*} M_k(N, \varepsilon).$$

Given a Dirichlet character ε we type `ModularForms(eps, weight)` to create the space of modular forms with that character and a given integer weight. For example, we create the space of forms of weight 5 with the character modulo 8 above that is -1 on 3 and 1 on 5 as follows.

```
sage: ModularForms(eps,5)
Modular Forms space of dimension 6, character [-1, 1] and
weight 5 over Rational Field
sage: sum([ModularForms(eps,5).dimension() for eps in v])
11
sage: ModularForms(Gamma1(8),5)
Modular Forms space of dimension 11 ...
```

Exercise 3.2. Compute the dimensions of all spaces $M_2(37, \varepsilon)$ for all Dirichlet characters ε.

The space $M_k(\Gamma)$ is equipped with an action of a commuting ring $\mathbb{T}$ of Hecke operators T_n for $n \geq 1$. A standard computational problem in the theory of modular forms is to compute an explicit basis of q-expansion for $M_k(\Gamma)$ along with matrices for the action of any Hecke operator T_n, and to compute the subspace $S_k(\Gamma)$ of cusp forms.

```
sage: M = ModularForms(Gamma0(11),4)
sage: M.basis()
[
q + 3*q^3 - 6*q^4 - 7*q^5 + O(q^6),
q^2 - 4*q^3 + 2*q^4 + 8*q^5 + O(q^6),
1 + O(q^6),
q + 9*q^2 + 28*q^3 + 73*q^4 + 126*q^5 + O(q^6)
]
sage: M.hecke_matrix(2)
[0 2 0 0]
[1 2 0 0]
[0 0 9 0]
[0 0 0 9]
```

We can also compute Hecke operators on the cuspidal subspace.

```
sage: S = M.cuspidal_subspace()
sage: S.hecke_matrix(2)
[0 2]
[1 2]
sage: S.hecke_matrix(3)
[ 3 -8]
[-4 -5]
```

Unfortunately, Sage doesn't yet implement computation of the Hecke operators on $M_k(\Gamma_1(N))$.

```
sage: M = ModularForms(Gamma1(5),2)
sage: M
Modular Forms space of dimension 3 for Congruence Subgroup
Gamma1(5) of weight 2 over Rational Field
sage: M.hecke_matrix(2)
Traceback (most recent call last):
...
NotImplementedError
```

However, we can compute Hecke operators on *modular symbols* for $\Gamma_1(N)$, which is a $\mathbb{T}$-module that is isomorphic to $M_k(\Gamma_1(N))$ (see Section 3.2).

```
sage: ModularSymbols(Gamma1(5),2,sign=1).hecke_matrix(2)
[ 2  1  1]
[ 1  2 -1]
[ 0  0 -1]
```

3.2 Modular Symbols

Modular symbols are a beautiful piece of mathematics that was developed since the 1960s by Birch, Manin, Shokorov, Mazur, Merel, Cremona, and others. Not only are modular symbols a powerful computational tool as we will see, they have also been used to prove rationality results for special values of L-series, to construct p-adic L-series, and they play a key role in Merel's proof of the uniform boundedness theorem for torsion points on elliptic curves over number fields.

We view modular symbols as a remarkably flexible computational tool that provides a single uniform algorithm for computing $M_k(N,\varepsilon)$ for any N,ε and $k \geq 2$. There are ways to use computation of those spaces to obtain explicit basis for spaces of weight 1 and half-integral weight, so in a sense modular symbols yield everything. There are also generalizations of modular symbols to higher rank groups, though Sage currently has no code for modular symbols on higher rank groups.

A *modular symbol* of weight k, and level N, with character ε is a sum of terms $X^iY^{k-2-i}\{\alpha,\beta\}$, where $0 \leq i \leq k-2$ and $\alpha,\beta \in \mathbb{P}^1(\mathbb{Q}) = \mathbb{Q}\cup\{\infty\}$. Modular symbols satisfy the relations

$$X^iY^{k-2-i}\{\alpha,\beta\} + X^iY^{k-2-i}\{\beta,\gamma\} + X^iY^{k-2-i}\{\gamma,\alpha\} = 0,$$

$$X^iY^{k-2-i}\{\alpha,\beta\} = -X^iY^{k-2-i}\{\beta,\alpha\},$$

and for every $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$(dX - bY)^i(-cX + aY)^{k-2-i}\{\gamma(\alpha),\gamma(\beta)\} = \varepsilon(d)X^iY^{k-2-i}\{\alpha,\beta\}.$$

The modular symbols space $\mathcal{M}_k(N,\varepsilon)$ is the torsion free $\mathbb{Q}[\varepsilon]$-module generated by all sums of modular symbols, modulo the relations listed above. Here $\mathbb{Q}[\varepsilon]$ is the ring generated by the values of the character ε, so it is of the form $\mathbb{Q}[\zeta_m]$ for some integer m.

The amazing theorem that makes modular symbols useful is that there is an explicit description of an action of a Hecke algebra $\mathbb{T}$ on $\mathcal{M}_k(N,\varepsilon)$, and there is an isomorphism

$$\mathcal{M}_k(N,\varepsilon;\mathbb{C}) \xrightarrow{\approx} M_k(N,\varepsilon) \oplus S_k(N,\varepsilon).$$

This means that if modular symbols are computable (they are!), then they can be used to compute a lot about the $\mathbb{T}$-module $M_k(N,\varepsilon)$.

Though $\mathcal{M}_k(N,\varepsilon)$ as described above is not explicitly generated by finitely many elements, it is finitely generated. Manin, Shokoruv, and Merel give an explicit description of finitely many generators (Manin symbols) for this space, along with all explicit relations that these generators satisfy (see my book). In particular, if we let

$$(i,c,d) = [X^iY^{2-k-i},(c,d)] = (dX - bY)^i(-cX + aY)^{k-2-i}\{\gamma(0),\gamma(\infty)\},$$

where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then the Manin symbols (i,c,d) with $0 \leq i \leq k-2$ and $(c,d) \in \mathbb{P}^1(N)$ generate $\mathcal{M}_k(N,\varepsilon)$.

We compute a basis for the space of weight 4 modular symbols for $\Gamma_0(11)$, then coerce in $(2,0,1)$ and $(1,1,3)$.

```
sage: M = ModularSymbols(11,4)
sage: M.basis()
([X^2,(0,1)], [X^2,(1,6)], [X^2,(1,7)], [X^2,(1,8)],
 [X^2,(1,9)], [X^2,(1,10)])
sage: M( (2,0,1) )
[X^2,(0,1)]
sage: M( (1,1,3) )
2/7*[X^2,(1,6)] + 1/14*[X^2,(1,7)] - 4/7*[X^2,(1,8)]
                  + 3/14*[X^2,(1,10)]
```

We compute a modular symbols representation for the Manin symbol $(2,1,6)$, and verify this by converting back.

```
sage: a = M.1; a
[X^2,(1,6)]
sage: a.modular_symbol_rep()
36*X^2*{5/6,1} - 60*X*Y*{5/6,1} + 25*Y^2*{5/6,1}
sage: 36*M([2,5/6,1]) - 60*M([1,5/6,1]) + 25*M([0,5/6,1])
[X^2,(1,6)]
```

3.3 Method of Graphs

The Mestre Method of Graphs is an intriguing algorithm for computing the action of Hecke operators on yet another module X that is isomorphic to $M_2(\Gamma_0(N))$. The implementation in Sage unfortunately only works when N is prime; in contrast, my implementation in Magma works when $N = pM$ and $S_2(\Gamma_0(M)) = 0$.

The matrices of Hecke operators on X are *vastly* sparser than on any basis of $M_2(\Gamma_0(N))$ that you are likely to use.

```
sage: X = SupersingularModule(389); X
Module of supersingular points on X_0(1)/F_389 over Integer Ring
sage: t2 = X.T(2).matrix(); t2[0]
(1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
sage: factor(charpoly(t2))
(x - 3) * (x + 2) * (x^2 - 2) * (x^3 - 4*x - 2) * ...
sage: t2 = ModularSymbols(389,sign=1).hecke_matrix(2); t2[0]
(3, 0, -1, 0, 0, -1, 1, 0, 0, 0, -1, 1, 0, 1, -1, 0, 1, 1,
 0, 1, -1, 1, -1, 1, 0, 0, 0, 0, 0, 0, 1, -1, -1)
sage: factor(charpoly(t2))
(x - 3) * (x + 2) * (x^2 - 2) * (x^3 - 4*x - 2) * ...
```

The method of graphs is also used in computer science to construct *expander graphs* with good properties. And it is important in my algorithm for computing Tamagawa numbers of purely toric modular abelian varieties. This algorithm is not implemented in Sage yet, since it is only interesting in the case of non-prime level, as it turns out.

3.4 Level One Modular Forms

The modular form

$$\Delta = q \prod (1 - q^n)^{24} = \sum \tau(n) q^n$$

is perhaps the world's most famous modular form. We compute some terms from the definition.

```
sage: R.<q> = QQ[[]]
sage: q * prod( 1-q^n+O(q^6) for n in (1..5) )^24
q - 24*q^2 + 252*q^3 - 1472*q^4 + 4830*q^5 - 6048*q^6 + O(q^7)
```

There are much better ways to compute Δ, which amount to just a few polynomial multiplactions over $\mathbb{Z}$.

```
sage: D = delta_qexp(10^5)      # less than 10 seconds
sage: D[:10]
q - 24*q^2 + 252*q^3 - 1472*q^4 + ...
sage: [p for p in primes(10^5) if D[p] % p == 0]
```

```
[2, 3, 5, 7, 2411]
sage: D[2411]
4542041100095889012
sage: f = eisenstein_series_qexp(12,6) - D[:6]; f
691/65520 + 2073*q^2 + 176896*q^3 + 4197825*q^4 + 48823296*q^5 + O(q^6)
sage: f % 691
O(q^6)
```

The Victor Miller basis for $M_k(\mathrm{SL}_2(\mathbb{Z}))$ is the reduced row echelon basis. It's a lemma that it has all integer coefficients, and a rather nice diagonal shape.

```
sage: victor_miller_basis(24, 6)
[
1 + 52416000*q^3 + 39007332000*q^4 + 6609020221440*q^5 + O(q^6),
q + 195660*q^3 + 12080128*q^4 + 44656110*q^5 + O(q^6),
q^2 - 48*q^3 + 1080*q^4 - 15040*q^5 + O(q^6)
]
sage: dimension_modular_forms(1,200)
17
sage: time B = victor_miller_basis(200, 18)
CPU time: 4.43 s,  Wall time: 5.07 s
sage: B
[
1 + 79288314420681734048660707200000*q^17 + O(q^18),
q + 2687602718106772837928968846869*q^17 + O(q^18),
...
q^16 + 96*q^17 + O(q^18)
]
```

Note: Craig Citro has made the above computation an order of magnitude faster in code he hasn't quite got into Sage yet. "I'll clean those up and submit them soon, since I need them for something I'm working on ... I'm currently in the process of making spaces of modular forms of level one subclass the existing code, and actually take advantage of all our fast E_k and Δ computation code, as well as cleaning things up a bit."

3.5 Half Integral Weight Forms

ALGORITHM: Basmaji (page 55 of his Essen thesis, "Ein Algorithmus zur Berechnung von Hecke-Operatoren und Anwendungen auf modulare Kurven", `http://wstein.org/scans/papers/basmaji/`).

Let $S = S_{k+1}(\varepsilon)$ be the space of cusp forms of even integer weight $k+1$ and character $\varepsilon = \chi\psi^{(k+1)/2}$, where ψ is the nontrivial mod-4 Dirichlet character. Let U be the subspace of $S \times S$ of elements (a, b) such that $\Theta_2 a = \Theta_3 b$. Then U is isomorphic to $S_{k/2}(\chi)$ via the map $(a, b) \mapsto a/\Theta_3$.

This algorithm is implemented in Sage. I'm sure it could be implemented in a way that is much faster than the current implementation...

```
sage: half_integral_weight_modform_basis(DirichletGroup(16,QQ).1, 3, 10)
[]
sage: half_integral_weight_modform_basis(DirichletGroup(16,QQ).1, 5, 10)
[q - 2*q^3 - 2*q^5 + 4*q^7 - q^9 + O(q^10)]
sage: half_integral_weight_modform_basis(DirichletGroup(16*7).0^2,3,30)
[q - 2*q^2 - q^9 + 2*q^14 + 6*q^18 - 2*q^21 - 4*q^22 - q^25 + O(q^30),
 q^2 - q^14 - 3*q^18 + 2*q^22 + O(q^30),
 q^4 - q^8 - q^16 + q^28 + O(q^30), q^7 - 2*q^15 + O(q^30)]
```

3.6 Generators for Rings of Modular Forms

For any congruence subgroup Γ, the direct sum

$$M(\Gamma) = \bigoplus_{k\geq 0} M_k(\Gamma)$$

is a ring, since the product of modular forms $f \in M_k(\Gamma)$ and $g \in M_{k'}(\Gamma)$ is an element $fg \in M_{k+k'}(\Gamma)$. Sage can compute likely generators for rings of modular forms, but currently doesn't prove any of these results.

We verify the statement proved in Serre's "A Course in Arithmetic" that E_4 and E_6 generate the space of level one modular forms.

```
sage: from sage.modular.modform.find_generators import modform_generators
sage: modform_generators(1)
[(4, 1 + 240*q + 2160*q^2 + 6720*q^3 + O(q^4)),
 (6, 1 - 504*q - 16632*q^2 - 122976*q^3 + O(q^4))]
```

Have you ever wondered which forms generate the ring $M(\Gamma_0(2))$? it turns out a form of weight 2 and two forms of weight 4 together generate.

```
sage: modform_generators(2)
[(2, 1 + 24*q + 24*q^2 + ... + 288*q^11 + O(q^12)),
 (4, 1 + 240*q^2 + .. + 30240*q^10 + O(q^12)),
 (4, q + 8*q^2 + .. + 1332*q^11 + O(q^12))]
```

Here's generators for $M(\Gamma_0(3))$. Notice that elements of weight 6 are now required, in addition to weights 2 and 4.

```
sage: modform_generators(3)
[(2, 1 + 12*q + 36*q^2 + .. + 168*q^13 + O(q^14)),
 (4, 1 + 240*q^3 + 2160*q^6 + 6720*q^9 + 17520*q^12 + O(q^14)),
 (4, q + 9*q^2 + 27*q^3 + 73*q^4 + .. + O(q^14)),
 (6, q - 6*q^2 + 9*q^3 + 4*q^4 + .. + O(q^14)),
 (6, 1 - 504*q^3 - 16632*q^6 .. + O(q^14)),
 (6, q + 33*q^2 + 243*q^3 + .. + O(q^14))]
```

3.7 L-series

Thanks to wrapping work of Jennifer Balakrishnan of M.I.T., we can compute explicitly with the L-series of the modular form Δ. Like for elliptic curves, behind these scenes this uses Dokchitsers L-functions calculation Pari program.

```
sage: L = delta_lseries(); L
L-series associated to the modular form Delta
sage: L(1)
0.0374412812685155
```

In some cases we can also compute with L-series attached to a cusp form.

```
 sage: f = CuspForms(2,8).0
 sage: L = f.cuspform_lseries()
 sage: L(1)
 0.0884317737041015
 sage: L(0.5)
 0.0296568512531983
```

Unfortunately, computing with the L-series of a general newform is not yet implemented.

```
sage: S = CuspForms(23,2); S
Cuspidal subspace of dimension 2 of Modular Forms space of
dimension 3 for Congruence Subgroup Gamma0(23) of weight
2 over Rational Field
sage: f = S.newforms('a')[0]; f
q + a0*q^2 + (-2*a0 - 1)*q^3 + (-a0 - 1)*q^4 + 2*a0*q^5 + O(q^6)
```

Computing with $L(f, s)$ totally not implemented yet, though should be easy via Dokchitser.

3.8 Modular Abelian Varieties

The quotient of the extended upper half plane $\mathfrak{h}^*$ by the congruence subgroup $\Gamma_1(N)$ is the modular curve $X_1(N)$. Its Jacobian $J_1(N)$ is an abelian variety that is canonically defined over $\mathbb{Q}$. Likewise, one defines a modular abelian variety $J_0(N)$ associated to $\Gamma_0(N)$.

Definition 3.3. A *modular abelian variety* is an abelian variety over $\mathbb{Q}$ that is a quotient of $J_1(N)$ for some N.

The biggest recent theorem in number theory is the proof of Serre's conjecture by Khare and Wintenberger. According to an argument of Ribet and Serre, this implies the following modularity theorem, which generalizes the modularity theorem that Taylor-Wiles proved in the course of proving Fermat's Last Theorem.

Theorem 3.4 (Modularity Theorem)**.** *Let A be a simple abelian variety defined over $\mathbb{Q}$. Then* $\mathrm{End}(A) \otimes \mathbb{Q}$ *is a number field of degree* $\dim(A)$ *if and only if A is modular.*

One of my longterm research goals is to develop a systematic theory for computing with modular abelian varieties. A good start is the observation using the Abel-Jacobi theorem that every modular abelian variety (up to isomorphism) can be specified by giving a lattice in a space of modular symbols.

We define some modular abelian varieties of level 39, and compute some basic invariants.

```
sage: D = J0(39).decomposition(); D
[
Simple abelian subvariety 39a(1,39) of dimension 1 of J0(39),
Simple abelian subvariety 39b(1,39) of dimension 2 of J0(39)
]
sage: D[1].lattice()
Free module of degree 6 and rank 4 over Integer Ring
Echelon basis matrix:
[ 1  0  0  1 -1  0]
[ 0  1  1  0 -1  0]
[ 0  0  2  0 -1  0]
[ 0  0  0  0  0  1]
sage: G = D[1].rational_torsion_subgroup(); G
Torsion subgroup of Simple abelian subvariety 39b(1,39)
of dimension 2 of J0(39)
sage: G.order()
28
sage: G.gens()
[[(1/14, 2/7, 0, 1/14, -3/14, 1/7)], [(0, 1, 0, 0, -1/2, 0)],
 [(0, 0, 1, 0, -1/2, 0)]]
sage: B, phi = D[1]/G
sage: B
Abelian variety factor of dimension 2 of J0(39)
sage: phi.kernel()
(Finite subgroup with invariants [2, 14] ...
```

There is an algorithm in Sage for computing the exact endomorphism ring of any modular abelian variety.

```
sage: A = J0(91)[2]; A
Simple abelian subvariety 91c(1,91) of dimension 2 of J0(91)
sage: R = End(A); R
Endomorphism ring of Simple abelian subvariety 91c(1,91)
of dimension 2 of J0(91)
sage: for x in R.gens(): print x.matrix(),'\n'
[1 0 0 0]
```

```
[0 1 0 0]
[0 0 1 0]
[0 0 0 1]

[ 0  4 -2  0]
[-1  5 -2  1]
[-1  2  0  2]
[-1  1  0  3]
```

It is also possible to test isomorphism of two modular abelian varieties. But much exciting theoretical and computational work remains to be done.

35 Book – Elementary Number Theory: Primes, Congruences, and Secrets

This book is published by Springer-Verlag:

`http://www.springer.com/math/numbers/book/978-0-387-85524-0`

You can show your support by buying a copy. See also the Amazon.com page:

```
http://www.amazon.com/
Elementary-Number-Theory-Computational-Undergraduate/dp/
0387855246/ref=sr_1_1?ie=UTF8&s=books&qid=1227154537&sr=8-1
```

Elementary Number Theory: Primes, Congruences, and Secrets

William Stein

July 22, 2009

To my wife Clarita Lefthand

Contents

Preface

This is a book about prime numbers, congruences, secret messages, and elliptic curves that you can read cover to cover. It grew out of undergraduate courses that the author taught at Harvard, UC San Diego, and the University of Washington.

The systematic study of number theory was initiated around 300B.C. when Euclid proved that there are infinitely many prime numbers, and also cleverly deduced the fundamental theorem of arithmetic, which asserts that every positive integer factors uniquely as a product of primes. Over a thousand years later (around 972A.D.) Arab mathematicians formulated the *congruent number problem* that asks for a way to decide whether or not a given positive integer n is the area of a right triangle, all three of whose sides are rational numbers. Then another thousand years later (in 1976), Diffie and Hellman introduced the first ever public-key cryptosystem, which enabled two people to communicate secretely over a public communications channel with no predetermined secret; this invention and the ones that followed it revolutionized the world of digital communication. In the 1980s and 1990s, elliptic curves revolutionized number theory, providing striking new insights into the congruent number problem, primality testing, public-key cryptography, attacks on public-key systems, and playing a central role in Andrew Wiles' resolution of Fermat's Last Theorem.

Today, pure and applied number theory is an exciting mix of simultaneously broad and deep theory, which is constantly informed and motivated by algorithms and explicit computation. Active research is underway that promises to resolve the congruent number problem, deepen our understanding into the structure of prime numbers, and both challenge and improve our ability to communicate securely. The goal of this book is to bring the reader closer to this world.

The reader is strongly encouraged to do every exercise in this book, checking their answers in the back (where many, but not all, solutions are given). Also, throughout the text there, are examples of calculations done using the powerful free open source mathematical software system Sage (`http://www.sagemath.org`), and the reader should try every such example and experiment with similar examples.

Background. The reader should know how to read and write mathematical proofs and must have know the basics of groups, rings, and fields. Thus, the prerequisites for this book are more than the prerequisites for most elementary number theory books, while still being aimed at undergraduates.

Notation and Conventions. We let $\mathbf{N} = \{1, 2, 3, \ldots\}$ denote the natural numbers, and use the standard notation $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ for the rings of integer, rational, real, and complex numbers, respectively. In this book, we will use the words proposition, theorem, lemma, and corollary as follows. Usually a proposition is a less important or less fundamental assertion, a theorem is a deeper culmination of ideas, a lemma is something that we will use later in this book to prove a proposition or theorem, and a corollary is an easy consequence of a proposition, theorem, or lemma. More difficult exercises are marked with a (*).

Acknowledgements. I would like to thank Brian Conrad, Carl Pomerance, and Ken Ribet for many clarifying comments and suggestions. Baurzhan Bektemirov, Lawrence Cabusora, and Keith Conrad read drafts of this book and made many comments, and Carl Witty commented extensively on the first two chapters. Frank Calegari used the course when teaching Math 124 at Harvard, and he and his students provided much feedback. Noam Elkies made comments and suggested Exercise 4.6. Seth Kleinerman wrote a version of Section 5.4 as a class project. Hendrik Lenstra made helpful remarks about how to present his factorization algorithm. Michael Abshoff, Sabmit Dasgupta, David Joyner, Arthur Patterson, George Stephanides, Kevin Stern, Eve Thompson, Ting-You Wang, and Heidi Williams all suggested corrections. I also benefited from conversations with Henry Cohn and David Savitt. I used Sage ([Sag08]), emacs, and LaTeX in the preparation of this book.

1
Prime Numbers

Every positive integer can be written uniquely as a product of prime numbers, e.g., $100 = 2^2 \cdot 5^2$. This is surprisingly difficult to prove, as we will see below. Even more astounding is that actually *finding* a way to write certain 1,000-digit numbers as a product of primes seems out of the reach of present technology, an observation that is used by millions of people every day when they buy things online.

Since prime numbers are the building blocks of integers, it is natural to wonder how the primes are distributed among the integers.

> "There are two facts about the distribution of prime numbers. The first is that, [they are] the most arbitrary and ornery objects studied by mathematicians: they grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision."
>
> — Don Zagier [Zag75]

The Riemann Hypothesis, which is the most famous unsolved problem in number theory, postulates a very precise answer to the question of how the prime numbers are distributed.

This chapter lays the foundations for our study of the theory of numbers by weaving together the themes of prime numbers, integer factorization, and the distribution of primes. In Section 1.1, we rigorously prove that the every positive integer is a product of primes, and give examples of specific integers for which finding such a decomposition would win one a large cash bounty. In Section 1.2, we discuss theorems about the set of prime numbers, starting with Euclid's proof that this set is infinite, and discuss the largest known prime. Finally we discuss the distribution of primes via the prime number theorem and the Riemann Hypothesis.

1.1 Prime Factorization

1.1.1 Primes

The set of *natural numbers* is

$$\mathbf{N} = \{1, 2, 3, 4, \dots\},$$

and the set of *integers* is

$$\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}.$$

Definition 1.1.1 (Divides)**.** If $a, b \in \mathbf{Z}$ we say that *a divides b*, written $a \mid b$, if $ac = b$ for some $c \in \mathbf{Z}$. In this case, we say a is a *divisor* of b. We say that *a does not divide b*, written $a \nmid b$, if there is no $c \in \mathbf{Z}$ such that $ac = b$.

For example, we have $2 \mid 6$ and $-3 \mid 15$. Also, all integers divide 0, and 0 divides only 0. However, 3 does not divide 7 in $\mathbf{Z}$.

Remark 1.1.2. The notation $b \,\vdots\, a$ for "b is divisible by a" is common in Russian literature on number theory.

Definition 1.1.3 (Prime and Composite)**.** An integer $n > 1$ is *prime* if the only positive divisors of n are 1 and n. We call n *composite* if n is not prime.

The number 1 is neither prime nor composite. The first few primes of $\mathbf{N}$ are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, \dots,$$

and the first few composites are

$$4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 32, 33, 34, \dots.$$

Remark 1.1.4. J. H. Conway argues in [Con97, viii] that -1 should be considered a prime, and in the 1914 table [Leh14], Lehmer considers 1 to be a prime. In this book, we consider neither -1 nor 1 to be prime.

SAGE Example 1.1.5. We use Sage to compute all prime numbers between a and $b - 1$.

```
sage: prime_range(10,50)
[11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47]
```

We can also compute the composites in an interval.

```
sage: [n for n in range(10,30) if not is_prime(n)]
[10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28]
```

Every natural number is built, in a unique way, out of prime numbers:

Theorem 1.1.6 (Fundamental Theorem of Arithmetic)**.** *Every natural number can be written as a product of primes uniquely up to order.*

Note that primes are the products with only one factor and 1 is the empty product.

Remark 1.1.7. Theorem 1.1.6, which we will prove in Section 1.1.4, is trickier to prove than you might first think. For example, unique factorization fails in the *ring*

$$\mathbf{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbf{Z}\} \subset \mathbf{C},$$

where 6 factors in two different ways:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

1.1.2 The Greatest Common Divisor

We will use the notion of the greatest common divisor of two integers to prove that if p is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. Proving this is the key step in our proof of Theorem 1.1.6.

Definition 1.1.8 (Greatest Common Divisor)**.** Let

$$\gcd(a, b) = \max\left\{d \in \mathbf{Z} : d \mid a \text{ and } d \mid b\right\},$$

unless both a and b are 0 in which case $\gcd(0, 0) = 0$.

For example, $\gcd(1, 2) = 1$, $\gcd(6, 27) = 3$, and for any a, $\gcd(0, a) = \gcd(a, 0) = a$.

If $a \neq 0$, the greatest common divisor exists because if $d \mid a$ then $d \leq |a|$, and there are only $|a|$ positive integers $\leq |a|$. Similarly, the gcd exists when $b \neq 0$.

Lemma 1.1.9. *For any integers a and b, we have*

$$\gcd(a, b) = \gcd(b, a) = \gcd(\pm a, \pm b) = \gcd(a, b - a) = \gcd(a, b + a).$$

Proof. We only prove that $\gcd(a, b) = \gcd(a, b - a)$, since the other cases are proved in a similar way. Suppose $d \mid a$ and $d \mid b$, so there exist integers c_1 and c_2 such that $dc_1 = a$ and $dc_2 = b$. Then $b - a = dc_2 - dc_1 = d(c_2 - c_1)$, so $d \mid b-a$. Thus $\gcd(a,b) \leq \gcd(a, b-a)$, since the set over which we are taking the max for $\gcd(a,b)$ is a subset of the set for $\gcd(a,b-a)$. The same argument with a replaced by $-a$ and b replaced by $b-a$, shows that $\gcd(a,b-a) = \gcd(-a,b-a) \leq \gcd(-a,b) = \gcd(a,b)$, which proves that $\gcd(a,b) = \gcd(a,b-a)$. □

Lemma 1.1.10. *Suppose* $a, b, n \in \mathbf{Z}$. *Then* $\gcd(a,b) = \gcd(a, b-an)$.

Proof. By repeated application of Lemma 1.1.9, we have

$$\gcd(a,b) = \gcd(a,b-a) = \gcd(a,b-2a) = \cdots = \gcd(a,b-an).$$

□

Assume for the moment that we have already proved Theorem 1.1.6. A naive way to compute $\gcd(a,b)$ is to factor a and b as a product of primes using Theorem 1.1.6; then the prime factorization of $\gcd(a,b)$ can be read off from that of a and b. For example, if $a = 2261$ and $b = 1275$, then $a = 7 \cdot \mathbf{17} \cdot 19$ and $b = 3 \cdot 5^2 \cdot \mathbf{17}$, so $\gcd(a,b) = 17$. It turns out that the greatest common divisor of two integers, even huge numbers (millions of digits), is surprisingly easy to compute using Algorithm 1.1.13 below, which computes $\gcd(a,b)$ without factoring a or b.

To motivate Algorithm 1.1.13, we compute $\gcd(2261, 1275)$ in a different way. First, we recall a helpful fact.

Proposition 1.1.11. *Suppose that* a *and* b *are integers with* $b \neq 0$. *Then there exists unique integers* q *and* r *such that* $0 \leq r < |b|$ *and* $a = bq + r$.

Proof. For simplicity, assume that both a and b are positive (we leave the general case to the reader). Let Q be the set of all nonnegative integers n such that $a - bn$ is nonnegative. Then Q is nonempty because $0 \in Q$ and Q is bounded because $a - bn < 0$ for all $n > a/b$. Let q be the largest element of Q. Then $r = a - bq < b$, otherwise $q+1$ would also be in Q. Thus q and r satisfy the existence conclusion.

To prove uniqueness, suppose that q' and r' also satisfy the conclusion. Then $q' \in Q$ since $r' = a - bq' \geq 0$, so $q' \leq q$, and we can write $q' = q - m$ for some $m \geq 0$. If $q' \neq q$, then $m \geq 1$ so

$$r' = a - bq' = a - b(q-m) = a - bq + bm = r + bm \geq b$$

since $r \geq 0$, a contradiction. Thus $q = q'$ and $r' = a - bq' = a - bq = r$, as claimed. □

For us, an *algorithm* is a finite sequence of instructions that can be followed to perform a specific task, such as a sequence of instructions in a computer program, which must terminate on any valid input. The word "algorithm" is sometimes used more loosely (and sometimes more precisely) than defined here, but this definition will suffice for us.

Algorithm 1.1.12 (Division Algorithm)**.** Suppose a and b are integers with $b \neq 0$. This algorithm computes integers q and r such that $0 \leq r < |b|$ and $a = bq + r$.

We will not describe the actual steps of Algorithm 1.1.12, since it is just the familiar long division algorithm. Note that it might not be exactly the same as the standard long division algorithm you learned in school, because we make the remainder positive even when dividing a negative number by a positive number.

We use the division algorithm repeatedly to compute $\gcd(2261, 1275)$. Dividing 2261 by 1275 we find that

$$2261 = 1 \cdot 1275 + 986,$$

so $q = 1$ and $r = 986$. Notice that if a natural number d divides both 2261 and 1275, then d divides their difference 986 and d still divides 1275. On the other hand, if d divides both 1275 and 986, then it has to divide their sum 2261 as well! We have made progress:

$$\gcd(2261, 1275) = \gcd(1275, 986).$$

This equality also follows by applying Lemma 1.1.9. Repeating, we have

$$1275 = 1 \cdot 986 + 289,$$

so $\gcd(1275, 986) = \gcd(986, 289)$. Keep going:

$$\begin{aligned} 986 &= 3 \cdot 289 + 119 \\ 289 &= 2 \cdot 119 + 51 \\ 119 &= 2 \cdot 51 + 17. \end{aligned}$$

Thus $\gcd(2261, 1275) = \cdots = \gcd(51, 17)$, which is 17 because $17 \mid 51$. Thus

$$\gcd(2261, 1275) = 17.$$

Aside from some tedious arithmetic, that computation was systematic, and it was not necessary to factor any integers (which is something we do not know how to do quickly if the numbers involved have hundreds of digits).

Algorithm 1.1.13 (Greatest Common Division)**.** Given integers a, b, this algorithm computes $\gcd(a, b)$.

1. [Assume $a > b > 0$] We have $\gcd(a, b) = \gcd(|a|, |b|) = \gcd(|b|, |a|)$, so we may replace a and b by their absolute values and hence assume $a, b \geq 0$. If $a = b$, output a and terminate. Swapping if necessary, we assume $a > b$. If $b = 0$, we output a.
2. [Quotient and Remainder] Using Algorithm 1.1.12, write $a = bq + r$, with $0 \leq r < b$ and $q \in \mathbf{Z}$.

3. [Finished?] If $r = 0$, then $b \mid a$, so we output b and terminate.
4. [Shift and Repeat] Set $a \leftarrow b$ and $b \leftarrow r$, then go to Step 2.

Proof. Lemmas 1.1.9–1.1.10 imply that $\gcd(a, b) = \gcd(b, r)$ so the gcd does not change in Step 4. Since the remainders form a decreasing sequence of nonnegative integers, the algorithm terminates. □

Example 1.1.14. Set $a = 15$ and $b = 6$.

$$\begin{aligned} 15 &= 6 \cdot 2 + 3 \qquad \gcd(15, 6) = \gcd(6, 3) \\ 6 &= 3 \cdot 2 + 0 \qquad \gcd(6, 3) = \gcd(3, 0) = 3 \end{aligned}$$

Note that we can just as easily do an example that is ten times as big, an observation that will be important in the proof of Theorem 1.1.19 below.

Example 1.1.15. Set $a = 150$ and $b = 60$.

$$\begin{aligned} 150 &= 60 \cdot 2 + 30 \qquad \gcd(150, 60) = \gcd(60, 30) \\ 60 &= 30 \cdot 2 + 0 \qquad \gcd(60, 30) = \gcd(30, 0) = 30 \end{aligned}$$

SAGE Example 1.1.16. Sage uses the `gcd` command to compute the greatest common divisor of two integers. For example,

```
sage: gcd(97,100)
1
sage: gcd(97 * 10^15, 19^20 * 97^2)
97
```

Lemma 1.1.17. *For any integers a, b, n, we have*

$$\gcd(an, bn) = \gcd(a, b) \cdot |n|.$$

Proof. The idea is to follow Example 1.1.15; we step through Euclid's algorithm for $\gcd(an, bn)$ and note that at every step the equation is the equation from Euclid's algorithm for $\gcd(a, b)$ but multiplied through by n. For simplicity, assume that both a and b are positive. We will prove the lemma by induction on $a + b$. The statement is true in the base case when $a + b = 2$, since then $a = b = 1$. Now assume a, b are arbitrary with $a \geq b$. Let q and r be such that $a = bq + r$ and $0 \leq r < b$. Then by Lemmas 1.1.9–1.1.10, we have $\gcd(a, b) = \gcd(b, r)$. Multiplying $a = bq + r$ by n we see that $an = bnq + rn$, so $\gcd(an, bn) = \gcd(bn, rn)$. Then

$$b + r = b + (a - bq) = a - b(q - 1) \leq a < a + b,$$

so by induction $\gcd(bn, rn) = \gcd(b, r) \cdot |n|$. Since $\gcd(a, b) = \gcd(b, r)$, this proves the lemma. □

Lemma 1.1.18. *Suppose $a, b, n \in \mathbf{Z}$ are such that $n \mid a$ and $n \mid b$. Then $n \mid \gcd(a, b)$.*

Proof. Since $n \mid a$ and $n \mid b$, there are integers c_1 and c_2, such that $a = nc_1$ and $b = nc_2$. By Lemma 1.1.17, $\gcd(a, b) = \gcd(nc_1, nc_2) = n \gcd(c_1, c_2)$, so n divides $\gcd(a, b)$. □

With Algorithm 1.1.13, we can prove that if a prime divides the product of two numbers, then it has got to divide one of them. This result is the key to proving that prime factorization is unique.

Theorem 1.1.19 (Euclid)**.** *Let p be a prime and $a, b \in \mathbf{N}$. If $p \mid ab$ then $p \mid a$ or $p \mid b$.*

You might think this theorem is "intuitively obvious," but that might be because the fundamental theorem of arithmetic (Theorem 1.1.6) is deeply ingrained in your intuition. Yet Theorem 1.1.19 will be needed in our proof of the fundamental theorem of arithmetic.

Proof of Theorem 1.1.19. If $p \mid a$ we are done. If $p \nmid a$ then $\gcd(p, a) = 1$, since only 1 and p divide p. By Lemma 1.1.17, $\gcd(pb, ab) = b$. Since $p \mid pb$ and, by hypothesis, $p \mid ab$, it follows (using Lemma 1.1.17) that

$$p \mid \gcd(pb, ab) = b \gcd(p, a) = b \cdot 1 = b.$$

□

1.1.3 Numbers Factor as Products of Primes

In this section, we prove that every natural number factors as a product of primes. Then we discuss the difficulty of finding such a decomposition in practice. We will wait until Section 1.1.4 to prove that factorization is unique.

As a first example, let $n = 1275$. The sum of the digits of n is divisible by 3, so n is divisible by 3 (see Proposition 2.1.9), and we have $n = 3 \cdot 425$. The number 425 is divisible by 5, since its last digit is 5, and we have $1275 = 3 \cdot 5 \cdot 85$. Again, dividing 85 by 5, we have $1275 = 3 \cdot 5^2 \cdot 17$, which is the prime factorization of 1275. Generalizing this process proves the following proposition.

Proposition 1.1.20. *Every natural number is a product of primes.*

Proof. Let n be a natural number. If $n = 1$, then n is the empty product of primes. If n is prime, we are done. If n is composite, then $n = ab$ with $a, b < n$. By induction, a and b are products of primes, so n is also a product of primes. □

Two questions immediately arise: (1) is this factorization unique, and (2) how quickly can we find such a factorization? Addressing (1), what if we had done something differently when breaking apart 1275 as a product of primes? Could the primes that show up be different? Let's try: we have

$1275 = 5 \cdot 255$. Now $255 = 5 \cdot 51$ and $51 = 17 \cdot 3$, and again the factorization is the same, as asserted by Theorem 1.1.6. We will prove the uniqueness of the prime factorization of any integer in Section 1.1.4.

SAGE Example 1.1.21. The `factor` command in Sage factors an integer as a product of primes with multiplicities. For example,

```
sage: factor(1275)
3 * 5^2 * 17
sage: factor(2007)
3^2 * 223
sage: factor(31415926535898)
2 * 3 * 53 * 73 * 2531 * 534697
```

Regarding (2), there are algorithms for integer factorization. It is a major open problem to decide how fast integer factorization algorithms can be. We say that an algorithm to factor n is *polynomial time* if there is a polynomial $f(x)$ such that for any n the number of steps needed by the algorithm to factor n is less than $f(\log_{10}(n))$. Note that $\log_{10}(n)$ is an approximation for the number of digits of the input n to the algorithm.

Open Problem 1.1.22. *Is there an algorithm that can factor any integer n in polynomial time?*

Peter Shor [Sho97] devised a polynomial time algorithm for factoring integers on quantum computers. We will not discuss his algorithm further, except to note that in 2001 IBM researchers built a quantum computer that used Shor's algorithm to factor 15 (see [LMG$^+$01, IBM01]). Building much larger quantum computers appears to be extremely difficult.

You can earn money by factoring certain large integers. Many cryptosystems would be easily broken if factoring certain large integers was easy. Since nobody has proven that factoring integers is difficult, one way to increase confidence that factoring is difficult is to offer cash prizes for factoring certain integers. For example, until recently there was a \$10,000 bounty on factoring the following 174-digit integer (see [RSA]):

$$\begin{array}{l}
188198812920607963838697239461650439807163563379417382700 7\\
633564229888597152346654853190606065047430453173880113033 9\\
671619969232120573403187955065699622130516875930765025705 9
\end{array}$$

This number is known as RSA-576 since it has 576 digits when written in binary (see Section 2.3.2 for more on binary numbers). It was factored at the German Federal Agency for Information Technology Security in December 2003 (see [Wei03]):

$$\begin{array}{l}
398075086424064937397125500550386491199064362342526708406\\
\quad 385189575946388957261768583317\\
\times\\
472772146107435302536223071973048224632914695302097116459\\
\quad 852171130520711256363590397527
\end{array}$$

The previous RSA challenge was the 155-digit number

10941738641570527421809707322040357612003732945449205990913
84213147634998428893478471799725789126733249762575289978183
3797076537244027146743531593354333897.

It was factored on 22 August 1999 by a group of sixteen researchers in four months on a cluster of 292 computers (see [ACD+99]). They found that RSA-155 is the product of the following two 78-digit primes:

$$\begin{aligned} p &= 102639592829741105772054196573991675900716567808038066\\ &\quad 803341933521790711307779\\ q &= 106603488380168454820927220360012878679207958575989291\\ &\quad 522270608237193062808643. \end{aligned}$$

The next RSA challenge is RSA-640:

3107418240490043721350750035888567930037346022842727545720161948
8232064405180815045563468296717232867824379162728380334154710731
0850191954852900733772482278352574238645401469173660247765
2346609,

and its factorization was worth $20,000 until November 2005 when it was factored by F. Bahr, M. Boehm, J. Franke, and T. Kleinjun. This factorization took five months. Here is one of the prime factors (you can find the other):

16347336458092538484431338838650908598417836700330923121811108
52389333100104508151212118167511579.

(This team also factored a 663-bit RSA challenge integer.)

The smallest currently open challenge is RSA-704, worth $30,000:

74037563479561712828046796097429573142593188889231289084936232
63897276503402826627689199641962511784399589433050212758537011
89680982867331732731089309005525051168770632990723963807867100
86096962537934650563796359

SAGE Example 1.1.23. Using Sage, we see that the above number has 212 decimal digits and is definitely composite:

```
sage: n = 740375634795617128280467960974295731425931888\
...923128908493623263897276503402826627689199641962511 7\
...843995894330502127585370118968098286733173273108930 9\
...005525051168770632990723963807867100860969625379346 5\
...0563796359
sage: len(n.str(2))
```

```
704
sage: len(n.str(10))
212
sage: n.is_prime()              # this is instant
False
```

These RSA numbers were factored using an algorithm called the number field sieve (see [LL93]), which is the best-known general purpose factorization algorithm. A description of how the number field sieve works is beyond the scope of this book. However, the number field sieve makes extensive use of the elliptic curve factorization method, which we will describe in Section 6.3.

1.1.4 The Fundamental Theorem of Arithmetic

We are ready to prove Theorem 1.1.6 using the following idea. Suppose we have two factorizations of n. Using Theorem 1.1.19, we cancel common primes from each factorization, one prime at a time. At the end, we discover that the factorizations must consist of exactly the same primes. The technical details are given below.

Proof. If $n = 1$, then the only factorization is the empty product of primes, so suppose $n > 1$.

By Proposition 1.1.20, there exist primes $p_1, \ldots, p_d$ such that

$$n = p_1 p_2 \cdots p_d.$$

Suppose that

$$n = q_1 q_2 \cdots q_m$$

is another expression of n as a product of primes. Since

$$p_1 \mid n = q_1(q_2 \cdots q_m),$$

Euclid's theorem implies that $p_1 = q_1$ or $p_1 \mid q_2 \cdots q_m$. By induction, we see that $p_1 = q_i$ for some i.

Now cancel p_1 and q_i, and repeat the above argument. Eventually, we find that, up to order, the two factorizations are the same. □

1.2 The Sequence of Prime Numbers

This section is concerned with three questions:

1. Are there infinitely many primes?

2. Given $a, b \in \mathbf{Z}$, are there infinitely many primes of the form $ax + b$?

3. How are the primes spaced along the number line?

We first show that there are infinitely many primes, then state Dirichlet's theorem that if $\gcd(a, b) = 1$, then $ax + b$ is a prime for infinitely many values of x. Finally, we discuss the Prime Number Theorem which asserts that there are asymptotically $x/\log(x)$ primes less than x, and we make a connection between this asymptotic formula and the Riemann Hypothesis.

1.2.1 There Are Infinitely Many Primes

Each number on the left in the following table is prime. We will see soon that this pattern does not continue indefinitely, but something similar works.

$$\begin{aligned} 3 &= 2 + 1 \\ 7 &= 2 \cdot 3 + 1 \\ 31 &= 2 \cdot 3 \cdot 5 + 1 \\ 211 &= 2 \cdot 3 \cdot 5 \cdot 7 + 1 \\ 2311 &= 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 \end{aligned}$$

Theorem 1.2.1 (Euclid)**.** *There are infinitely many primes.*

Proof. Suppose that $p_1, p_2, \ldots, p_n$ are n distinct primes. We construct a prime p_{n+1} not equal to any of $p_1, \ldots, p_n$, as follows. If

$$N = p_1 p_2 p_3 \cdots p_n + 1, \tag{1.2.1}$$

then by Proposition 1.1.20 there is a factorization

$$N = q_1 q_2 \cdots q_m$$

with each q_i prime and $m \geq 1$. If $q_1 = p_i$ for some i, then $p_i \mid N$. Because of (1.2.1), we also have $p_i \mid N - 1$, so $p_i \mid 1 = N - (N - 1)$, which is a contradiction. Thus the prime $p_{n+1} = q_1$ is not in the list $p_1, \ldots, p_n$, and we have constructed our new prime. □

For example,

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509.$$

Multiplying together the first six primes and adding 1 doesn't produce a prime, but it produces an integer that is merely divisible by a new prime.

Joke 1.2.2 (Hendrik Lenstra)**.** *There are infinitely many composite numbers. Proof.* To obtain a new composite number, multiply together the first n composite numbers and don't add 1.

1.2.2 Enumerating Primes

In this section we describe a sieving process that allows us to enumerate all primes up to n. The sieve works by first writing down all numbers up to n, noting that 2 is prime, and crossing off all multiples of 2. Next, note that the first number not crossed off is 3, which is prime, and cross off all multiples of 3, etc. Repeating this process, we obtain a list of the primes up to n. Formally, the algorithm is as follows:

Algorithm 1.2.3 (Prime Sieve)**.** Given a positive integer n, this algorithm computes a list of the primes up to n.

1. [Initialize] Let $X = [3, 5, \ldots]$ be the list of all odd integers between 3 and n. Let $P = [2]$ be the list of primes found so far.
2. [Finished?] Let p be the first element of X. If $p \geq \sqrt{n}$, append each element of X to P and terminate. Otherwise append p to P.
3. [Cross Off] Set X equal to the sublist of elements in X that are not divisible by p. Go to Step 2.

For example, to list the primes ≤ 40 using the sieve, we proceed as follows. First $P = [2]$ and

$$X = [3, 5, 7, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39].$$

We append 3 to P and cross off all multiples of 3 to obtain the new list

$$X = [5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37].$$

Next we append 5 to P, obtaining $P = [2, 3, 5]$, and cross off the multiples of 5, to obtain $X = [7, 11, 13, 17, 19, 23, 29, 31, 37]$. Because $7^2 \geq 40$, we append X to P and find that the primes less than 40 are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37.$$

Proof of Algorithm 1.2.3. The part of the algorithm that is not clear is that when the first element a of X satisfies $a \geq \sqrt{n}$, then each element of X is prime. To see this, suppose m is in X, so $\sqrt{n} \leq m \leq n$ and that m is divisible by no prime that is $\leq \sqrt{n}$. Write $m = \prod p_i^{e_i}$ with the p_i distinct primes ordered so that $p_1 < p_2 < \ldots$. If $p_i > \sqrt{n}$ for each i and there is more than one p_i, then $m > n$, a contradiction. Thus some p_i is less than $\sqrt{n}$, which also contradicts our assumptions on m. □

1.2.3 The Largest Known Prime

Though Theorem 1.2.1 implies that there are infinitely many primes, it still makes sense to ask the question "What is the largest *known* prime?"

A *Mersenne prime* is a prime of the form $2^q - 1$. According to [Cal] the largest known prime as of March 2007 is the 44th known Mersenne prime

$$p = 2^{32582657} - 1,$$

which has 9,808,358 decimal digits[1]. This would take over 2000 pages to print, assuming a page contains 60 lines with 80 characters per line. The Electronic Frontier Foundation has offered a $100,000 prize to the first person who finds a 10,000,000 digit prime.

Euclid's theorem implies that there definitely are infinitely many primes bigger than p. Deciding whether or not a number is prime is interesting, as a theoretical problem, and as a problem with applications to cryptography, as we will see in Section 2.4 and Chapter 3.

SAGE Example 1.2.4. We can compute the decimal expansion of p in Sage, although watch out as this is a serious computation that may take around a minute on your computer. Also, do not print out p or s below, because both would take a very long time to scroll by.

```
sage: p = 2^32582657 - 1
sage: p.ndigits()
9808358
```

Next we convert p to a decimal string and look at some of the digits.

```
sage: s = p.str(10) # this takes a long time
sage: len(s)        # s is a very long string  (long time)
9808358
sage: s[:20]        # the first 20 digits of p (long time)
'12457502601536945540'
sage: s[-20:]       # the last 20 digits       (long time)
'11752880154053967871'
```

1.2.4 Primes of the Form $ax + b$

Next we turn to primes of the form $ax+b$, where a and b are fixed integers with $a > 1$ and x varies over the natural numbers $\mathbf{N}$. We assume that $\gcd(a, b) = 1$, because otherwise there is no hope that $ax + b$ is prime infinitely often. For example, $2x + 2 = 2(x + 1)$ is only prime if $x = 0$, and is not prime for any $x \in \mathbf{N}$.

Proposition 1.2.5. *There are infinitely many primes of the form* $4x - 1$.

Why might this be true? We list numbers of the form $4x-1$ and underline those that are prime.

$$\underline{3},\ \underline{7},\ \underline{11},\ 15,\ \underline{19},\ \underline{23},\ 27,\ \underline{31},\ 35,\ 39,\ \underline{43},\ \underline{47},\ \ldots$$

[1] The 45th known Mersenne prime may have been found on August 23, 2008 as this book goes to press.

Not only is it plausible that underlined numbers will continue to appear indefinitely, it is something we can easily prove.

Proof. Suppose $p_1, p_2, \ldots, p_n$ are distinct primes of the form $4x-1$. Consider the number

$$N = 4p_1p_2\cdots p_n - 1.$$

Then $p_i \nmid N$ for any i. Moreover, not every prime $p \mid N$ is of the form $4x+1$; if they all were, then N would be of the form $4x+1$. Since N is odd, each prime divisor p_i is odd so there is a $p \mid N$ that is of the form $4x-1$. Since $p \neq p_i$ for any i, we have found a new prime of the form $4x-1$. We can repeat this process indefinitely, so the set of primes of the form $4x-1$ cannot be finite. □

Note that this proof does not work if $4x-1$ is replaced by $4x+1$, since a product of primes of the form $4x-1$ can be of the form $4x+1$.

Example 1.2.6. Set $p_1 = 3$, $p_2 = 7$. Then

$$N = 4 \cdot 3 \cdot 7 - 1 = \underline{83}$$

is a prime of the form $4x-1$. Next

$$N = 4 \cdot 3 \cdot 7 \cdot 83 - 1 = \underline{6971},$$

which is again a prime of the form $4x-1$. Again,

$$N = 4 \cdot 3 \cdot 7 \cdot 83 \cdot 6971 - 1 = 48601811 = 61 \cdot \underline{796751}.$$

This time 61 is a prime, but it is of the form $4x+1 = 4 \cdot 15 + 1$. However, 796751 is prime and $796751 = 4 \cdot 199188 - 1$. We are unstoppable.

$$N = 4 \cdot 3 \cdot 7 \cdot 83 \cdot 6971 \cdot 796751 - 1 = \underline{5591} \cdot 6926049421.$$

This time the small prime, 5591, is of the form $4x-1$ and the large one is of the form $4x+1$.

Theorem 1.2.7 (Dirichlet)**.** *Let a and b be integers with* $\gcd(a,b) = 1$. *Then there are infinitely many primes of the form $ax+b$.*

Proofs of this theorem typically use tools from advanced number theory, and are beyond the scope of this book (see e.g., [FT93, §VIII.4]).

1.2.5 How Many Primes are There?

We saw in Section 1.2.1 that there are infinitely many primes. In order to get a sense of just how many primes there are, we consider a few warm-up questions. Then we consider some numerical evidence and state the prime number theorem, which gives an asymptotic answer to our question, and connect this theorem with a form of the famous Riemann Hypothesis. Our discussion of counting primes in this section is very cursory; for more details, read Crandall and Pomerance's excellent book [CP01, §1.1.5].

The following vague discussion is meant to motivate a precise way to measure the number (or percentage) of primes. What percentage of natural numbers are even? Answer: Half of them. What percentage of natural numbers are of the form $4x-1$? Answer: One fourth of them. What percentage of natural numbers are perfect squares? Answer: Zero percent of all natural numbers, in the sense that the limit of the proportion of perfect squares to all natural numbers converges to 0. More precisely,

$$\lim_{x\to\infty} \frac{\#\{n \in \mathbf{N} : n \leq x \text{ and } n \text{ is a perfect square}\}}{x} = 0,$$

since the numerator is roughly $\sqrt{x}$ and $\lim_{x\to\infty} \frac{\sqrt{x}}{x} = 0$. Likewise, it is an easy consequence of Theorem 1.2.10 that zero percent of all natural numbers are prime (see Exercise 1.4).

We are thus led to ask another question: How many positive integers $\leq x$ are perfect squares? Answer: Roughly $\sqrt{x}$. In the context of primes, we ask,

Question 1.2.8. How many natural numbers $\leq x$ are prime?

Let

$$\pi(x) = \#\{p \in \mathbf{N} : p \leq x \text{ is a prime}\}.$$

For example,

$$\pi(6) = \#\{2, 3, 5\} = 3.$$

Some values of $\pi(x)$ are given in Table 1.1, and Figures 1.1 and 1.2 contain graphs of $\pi(x)$. These graphs look like straight lines, which maybe bend down slightly.

SAGE Example 1.2.9. To compute $\pi(x)$ in Sage use the `prime_pi(x)` command:

```
sage: prime_pi(6)
3
sage: prime_pi(100)
25
sage: prime_pi(3000000)
216816
```

We can also draw a plot of $\pi(x)$ using the `plot` command:

```
sage: plot(prime_pi, 1,1000, rgbcolor=(0,0,1))
```

Gauss was an inveterate computer: he wrote in an 1849 letter that there are $216,745$ primes less than $3,000,000$ (this is wrong but close; the correct count is $216,816$).

TABLE 1.1. Values of $\pi(x)$

x	100	200	300	400	500	600	700	800	900	1000
$\pi(x)$	25	46	62	78	95	109	125	139	154	168

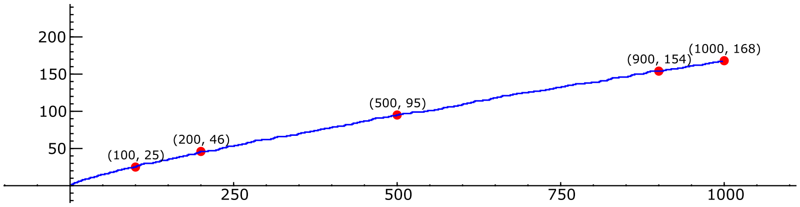


FIGURE 1.1. Graph of $\pi(x)$ for $x < 1000$

Gauss conjectured the following asymptotic formula for $\pi(x)$, which was later proved independently by Hadamard and Vallée Poussin in 1896 (but will not be proved in this book).

Theorem 1.2.10 (Prime Number Theorem)**.** *The function $\pi(x)$ is asymptotic to $x/\log(x)$, in the sense that*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} = 1.$$

We do nothing more here than motivate this deep theorem with a few further observations. The theorem implies that

$$\lim_{x\to\infty} \frac{\pi(x)}{x} = \lim_{x\to\infty} \frac{1}{\log(x)} = 0,$$

so for any a,

$$\lim_{x\to\infty} \frac{\pi(x)}{x/(\log(x)-a)} = \lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} - \frac{a\pi(x)}{x} = 1.$$

Thus $x/(\log(x)-a)$ is also asymptotic to $\pi(x)$ for any a. See [CP01, §1.1.5] for a discussion of why $a = 1$ is the best choice. Table 1.2 compares $\pi(x)$ and $x/(\log(x)-1)$ for several $x < 10000$.

The record for counting primes is

$$\pi(10^{23}) = 1925320391606803968923.$$

Note that such computations are very difficult to get exactly right, so the above might be slightly wrong.

For the reader familiar with complex analysis, we mention a connection between $\pi(x)$ and the Riemann Hypothesis. The Riemann zeta function $\zeta(s)$ is a complex analytic function on $\mathbf{C} \setminus \{1\}$ that extends the function

TABLE 1.2. Comparison of $\pi(x)$ and $x/(\log(x) - 1)$

x	$\pi(x)$	$x/(\log(x) - 1)$ (approx)
1000	168	169.2690290604408165186256278
2000	303	302.9888734545463878029800994
3000	430	428.1819317975237043747385740
4000	550	548.3922097278253264133400985
5000	669	665.1418784486502172369455815
6000	783	779.2698885854778626863677374
7000	900	891.3035657223339974352567759
8000	1007	1001.602962794770080754784281
9000	1117	1110.428422963188172310675011
10000	1229	1217.976301461550279200775705

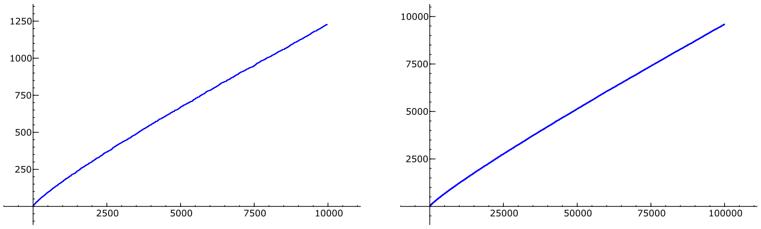


FIGURE 1.2. Graphs of $\pi(x)$ for $x < 10000$ and $x < 100000$

defined on a right half plane by $\sum_{n=1}^{\infty} n^{-s}$. The Riemann Hypothesis is the conjecture that the zeros in $\mathbf{C}$ of $\zeta(s)$ with positive real part lie on the line $\mathrm{Re}(s) = 1/2$. This conjecture is one of the Clay Math Institute million dollar millennium prize problems [Cla].

According to [CP01, §1.4.1], the Riemann Hypothesis is equivalent to the conjecture that

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\log(t)} dt$$

is a "good" approximation to $\pi(x)$, in the following precise sense.

Conjecture 1.2.11 (Equivalent to the Riemann Hypothesis)**.**
For all $x \geq 2.01$,

$$|\pi(x) - \mathrm{Li}(x)| \leq \sqrt{x}\log(x).$$

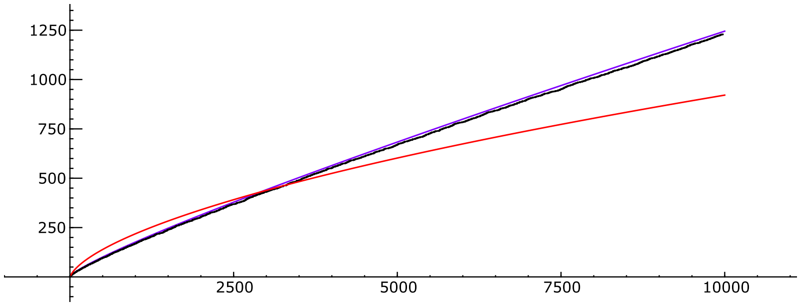
If $x = 2$, then $\pi(2) = 1$ and $\mathrm{Li}(2) = 0$, but $\sqrt{2}\log(2) = 0.9802\ldots$, so the inequality is not true for $x \geq 2$, but 2.01 is big enough. We will do nothing more to explain this conjecture, and settle for one numerical example.

Example 1.2.12. Let $x = 4 \cdot 10^{22}$. Then

$$\begin{aligned}
\pi(x) &= 783964159847056303858, \\
\mathrm{Li}(x) &= 783964159852157952242.7155276025801473\ldots, \\
|\pi(x) - \mathrm{Li}(x)| &= 5101648384.71552760258014\ldots, \\
\sqrt{x}\log(x) &= 10408633281397.77913344605\ldots, \\
x/(\log(x) - 1) &= 783650443647303761503.5237113087392967\ldots.
\end{aligned}$$

SAGE Example 1.2.13. We use Sage to graph $\pi(x)$, $\mathrm{Li}(x)$, and $\sqrt{x}\log(x)$.

```
sage: P = plot(Li, 2,10000, rgbcolor='purple')
sage: Q = plot(prime_pi, 2,10000, rgbcolor='black')
sage: R = plot(sqrt(x)*log(x),2,10000,rgbcolor='red')
sage: show(P+Q+R,xmin=0, figsize=[8,3])
```



The topmost line is $\mathrm{Li}(x)$, the next line is $\pi(x)$, and the bottom line is $\sqrt{x}\log(x)$.

For more on the prime number theorem and the Riemann hypothesis see [Zag75] and [MS08].

1.3 Exercises

1.1 Compute the greatest common divisor $\gcd(455, 1235)$ by hand.

1.2 Use the prime enumeration sieve to make a list of all primes up to 100.

1.3 Prove that there are infinitely many primes of the form $6x - 1$.

1.4 Use Theorem 1.2.10 to deduce that $\displaystyle\lim_{x\to\infty} \frac{\pi(x)}{x} = 0$.

1.5 Let $\psi(x)$ be the number of primes of the form $4k-1$ that are $\leq x$. Use a computer to make a conjectural guess about $\lim_{x\to\infty} \psi(x)/\pi(x)$.

1.6 So far 44 Mersenne primes $2^p - 1$ have been discovered. Give a guess, backed up by an argument, about when the next Mersenne prime might be discovered (you will have to do some online research).

1.7 (a) Let $y = 10000$. Compute $\pi(y) = \#\{\text{primes } p \leq y\}$.

(b) The prime number theorem implies $\pi(x)$ is asymptotic to $\frac{x}{\log(x)}$. How close is $\pi(y)$ to $y/\log(y)$, where y is as in (a)?

1.8 Let a, b, c, n be integers. Prove that

(a) if $a \mid n$ and $b \mid n$ with $\gcd(a, b) = 1$, then $ab \mid n$.

(b) if $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.

1.9 Let a, b, c, d, and m be integers. Prove that

(a) if $a \mid b$ and $b \mid c$ then $a \mid c$.

(b) if $a \mid b$ and $c \mid d$ then $ac \mid bd$.

(c) if $m \neq 0$, then $a \mid b$ if and only if $ma \mid mb$.

(d) if $d \mid a$ and $a \neq 0$, then $|d| \leq |a|$.

1.10 In each of the following, apply the division algorithm to find q and r such that $a = bq + r$ and $0 \leq r < |b|$:

$$a = 300, b = 17,\ \ a = 729, b = 31,\ \ a = 300, b = -17,\ \ a = 389, b = 4.$$

1.11 (a) (Do this part by hand.) Compute the greatest common divisor of 323 and 437 using the algorithm described in class that involves quotients and remainders (i.e., do not just factor a and b).

(b) Compute by any means the greatest common divisor of

$$31415926535897932384626433 8$$

and

$$2718281828459045235360287 47.$$

1.12 (a) Suppose a, b and n are positive integers. Prove that if $a^n \mid b^n$, then $a \mid b$.

(b) Suppose p is a prime and a and k are positive integers. Prove that if $p \mid a^k$, then $p^k \mid a^k$.

1.13 (a) Prove that if a positive integer n is a perfect square, then n cannot be written in the form $4k + 3$ for k an integer. (Hint: Compute the remainder upon division by 4 of each of $(4m)^2$, $(4m + 1)^2$, $(4m + 2)^2$, and $(4m + 3)^2$.)

(b) Prove that no integer in the sequence

$$11, 111, 1111, 11111, 111111, \ldots$$

is a perfect square. (Hint: $111 \cdots 111 = 111 \cdots 108 + 3 = 4k + 3$.)

1.14 Prove that a positive integer n is prime if and only if n is not divisible by any prime p with $1 < p \leq \sqrt{n}$.

2
The Ring of Integers Modulo n

A startling fact about numbers is that it takes less than a second to decide with near certainty whether or not any given 1,000 digit number n is a prime, *without actually factoring* n. The algorithm for this involves doing some arithmetic with n that works differently depending on whether n is prime or composite. In particular, we do arithmetic with the set (in fact, "ring") of integers $\{0, 1, \ldots, n-1\}$ using an innovative rule for addition and multiplication, where the sum and product of two elements of that set is again in that set.

Another surprising fact is that one can almost instantly compute the last 1,000 digits of a massive multi-billion digit number like $n = 1234^{1234567890}$ without explicitly writing down all the digits of n. Again, this calculation involves arithmetic with the ring $\{0, 1, \ldots, n-1\}$.

This chapter is about the ring $\mathbf{Z}/n\mathbf{Z}$ of integers modulo n, the beautiful structure this ring has, and how to apply it to the above mentioned problems, among others. It is foundational for the rest of this book. In Section 2.1, we discuss when linear equations modulo n have a solution, then introduce the Euler φ function and prove Euler's Theorem and Wilson's theorem. In Section 2.2, we prove the Chinese Remainer Theorem, which addresses simultaneous solubility of several linear equations modulo coprime moduli. With these theoretical foundations in place, in Section 2.3, we introduce algorithms for doing powerful computations modulo n, including computing large powers quickly, and solving linear equations. We finish in Section 2.4 with a discussion of recognizing prime numbers using arithmetic modulo n.

2.1 Congruences Modulo n

Definition 2.1.1 (Group)**.** A *group* is a set G equipped with a binary operation $G \times G \to G$ (denoted by multiplication below) and an identity element $1 \in G$ such that:

1. For all $a, b, c \in G$, we have $(ab)c = a(bc)$.
2. For each $a \in G$, we have $1a = a1 = a$, and there exists $b \in G$ such that $ab = 1$.

Definition 2.1.2 (Abelian Group)**.** An *abelian group* is a group G such that $ab = ba$ for every $a, b \in G$.

Definition 2.1.3 (Ring)**.** A *ring* R is a set equipped with binary operations $+$ and $\times$ and elements $0, 1 \in R$ such that R is an abelian group under $+$, and for all $a, b, c \in R$ we have

- $1a = a1 = a$
- $(ab)c = a(bc)$
- $a(b + c) = ab + ac$.

If, in addition, $ab = ba$ for all $a, b \in R$, then we call R a *commutative ring.*

In this section, we define the ring $\mathbf{Z}/n\mathbf{Z}$ of integers modulo n, introduce the Euler φ-function, and relate it to the multiplicative order of certain elements of $\mathbf{Z}/n\mathbf{Z}$.

If $a, b \in \mathbf{Z}$ and $n \in \mathbf{N}$, we say that a is *congruent to b modulo n* if $n \mid a - b$, and write $a \equiv b \pmod{n}$. Let $n\mathbf{Z} = (n)$ be the subset of $\mathbf{Z}$ consisting of all multiples of n (this is called the "ideal of $\mathbf{Z}$ generated by n").

Definition 2.1.4 (Integers Modulo n)**.** The ring $\mathbf{Z}/n\mathbf{Z}$ of *integers modulo n* is the set of equivalence classes of integers modulo n. It is equipped with its natural ring structure:

$$(a + n\mathbf{Z}) + (b + n\mathbf{Z}) = (a + b) + n\mathbf{Z}$$

$$(a + n\mathbf{Z}) \cdot (b + n\mathbf{Z}) = (a \cdot b) + n\mathbf{Z}.$$

Example 2.1.5. For example,

$$\mathbf{Z}/3\mathbf{Z} = \{\{\ldots, -3, 0, 3, \ldots\}, \{\ldots, -2, 1, 4, \ldots\}, \{\ldots, -1, 2, 5, \ldots\}\}$$

SAGE Example 2.1.6. In Sage, we list the elements of $\mathbf{Z}/n\mathbf{Z}$ as follows:

```
sage: R = Integers(3)
sage: list(R)
[0, 1, 2]
```

We use the notation $\mathbf{Z}/n\mathbf{Z}$ because $\mathbf{Z}/n\mathbf{Z}$ is the quotient of the ring $\mathbf{Z}$ by the "ideal" $n\mathbf{Z}$ of multiples of n. Because $\mathbf{Z}/n\mathbf{Z}$ is the quotient of a ring by an ideal, the ring structure on $\mathbf{Z}$ induces a ring structure on $\mathbf{Z}/n\mathbf{Z}$. We often let a or $a \pmod{n}$ denote the equivalence class $a + n\mathbf{Z}$ of a.

Definition 2.1.7 (Field)**.** A *field* K is a ring such that for every nonzero element $a \in K$ there is an element $b \in K$ such that $ab = 1$.

For example, if p is a prime, then $\mathbf{Z}/p\mathbf{Z}$ is a field (see Exercise 2.12).

Definition 2.1.8 (Reduction Map and Lift)**.** We call the natural reduction map $\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$, which sends a to $a + n\mathbf{Z}$, *reduction modulo n*. We also say that a is a *lift* of $a + n\mathbf{Z}$. Thus, e.g., 7 is a lift of 1 mod 3, since $7 + 3\mathbf{Z} = 1 + 3\mathbf{Z}$.

We can use that arithmetic in $\mathbf{Z}/n\mathbf{Z}$ is well defined is to derive tests for divisibility by n (see Exercise 2.8).

Proposition 2.1.9. *A number $n \in \mathbf{Z}$ is divisible by* 3 *if and only if the sum of the digits of n is divisible by* 3*.*

Proof. Write

$$n = a + 10b + 100c + \cdots ,$$

where the digits of n are a, b, c, etc. Since $10 \equiv 1 \pmod{3}$,

$$n = a + 10b + 100c + \cdots \equiv a + b + c + \cdots \pmod{3},$$

from which the proposition follows. □

2.1.1 Linear Equations Modulo n

In this section, we are concerned with how to decide whether or not a linear equation of the form $ax \equiv b \pmod{n}$ has a solution modulo n. Algorithms for *computing* solutions to $ax \equiv b \pmod{n}$ are the topic of Section 2.3.

First, we prove a proposition that gives a criterion under which one can cancel a quantity from both sides of a congruence.

Proposition 2.1.10 (Cancellation)**.** *If* $\gcd(c, n) = 1$ *and*

$$ac \equiv bc \pmod{n},$$

then $a \equiv b \pmod{n}$*.*

Proof. By definition

$$n \mid ac - bc = (a - b)c.$$

Since $\gcd(n, c) = 1$, it follows from Theorem 1.1.6 that $n \mid a - b$, so

$$a \equiv b \pmod{n},$$

as claimed. □

When a has a multiplicative inverse a' in $\mathbf{Z}/n\mathbf{Z}$ (i.e., $aa' \equiv 1 \pmod{n}$) then the equation $ax \equiv b \pmod{n}$ has a unique solution $x \equiv a'b \pmod{n}$. Thus, it is of interest to determine the units in $\mathbf{Z}/n\mathbf{Z}$, i.e., the elements which have a multiplicative inverse.

We will use complete sets of residues to prove that the units in $\mathbf{Z}/n\mathbf{Z}$ are exactly the $a \in \mathbf{Z}/n\mathbf{Z}$ such that $\gcd(\tilde{a}, n) = 1$ for any lift $\tilde{a}$ of a to $\mathbf{Z}$ (it doesn't matter which lift).

Definition 2.1.11 (Complete Set of Residues)**.** We call a subset $R \subset \mathbf{Z}$ of size n whose reductions modulo n are pairwise distinct a *complete set of residues* modulo n. In other words, a complete set of residues is a choice of representative for each equivalence class in $\mathbf{Z}/n\mathbf{Z}$.

For example,

$$R = \{0, 1, 2, \ldots, n-1\}$$

is a complete set of residues modulo n. When $n = 5$, $R = \{0, 1, -1, 2, -2\}$ is a complete set of residues.

Lemma 2.1.12. *If R is a complete set of residues modulo n and $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$, then $aR = \{ax : x \in R\}$ is also a complete set of residues modulo n.*

Proof. If $ax \equiv ax' \pmod{n}$ with $x, x' \in R$, then Proposition 2.1.10 implies that $x \equiv x' \pmod{n}$. Because R is a complete set of residues, this implies that $x = x'$. Thus the elements of aR have distinct reductions modulo n. It follows, since $\#aR = n$, that aR is a complete set of residues modulo n. □

Proposition 2.1.13 (Units)**.** *If $\gcd(a, n) = 1$, then the equation $ax \equiv b \pmod{n}$ has a solution, and that solution is unique modulo n.*

Proof. Let R be a complete set of residues modulo n, so there is a unique element of R that is congruent to b modulo n. By Lemma 2.1.12, aR is also a complete set of residues modulo n, so there is a unique element $ax \in aR$ that is congruent to b modulo n, and we have $ax \equiv b \pmod{n}$. □

Algebraically, this proposition asserts that if $\gcd(a, n) = 1$, then the map $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$ given by left multiplication by a is a bijection.

Example 2.1.14. Consider the equation $2x \equiv 3 \pmod{7}$, and the complete set $R = \{0, 1, 2, 3, 4, 5, 6\}$ of coset representatives. We have

$$2R = \{0, 2, 4, 6, 8 \equiv 1, 10 \equiv 3, 12 \equiv 5\},$$

so $2 \cdot 5 \equiv 3 \pmod{7}$.

When $\gcd(a, n) \neq 1$, then the equation $ax \equiv b \pmod{n}$ may or may not have a solution. For example, $2x \equiv 1 \pmod{4}$ has no solution, but $2x \equiv 2 \pmod{4}$ does, and in fact it has more than one mod 4 ($x = 1$ and $x = 3$). Generalizing Proposition 2.1.13, we obtain the following more general criterion for solvability.

Proposition 2.1.15 (Solvability)**.** *The equation* $ax \equiv b \pmod{n}$ *has a solution if and only if* $\gcd(a, n)$ *divides* b*.*

Proof. Let $g = \gcd(a, n)$. If there is a solution x to the equation $ax \equiv b \pmod{n}$, then $n \mid (ax - b)$. Since $g \mid n$ and $g \mid a$, it follows that $g \mid b$.

Conversely, suppose that $g \mid b$. Then $n \mid (ax - b)$ if and only if

$$\frac{n}{g} \mid \left(\frac{a}{g}x - \frac{b}{g}\right).$$

Thus $ax \equiv b \pmod{n}$ has a solution if and only if $\frac{a}{g}x \equiv \frac{b}{g} \pmod{\frac{n}{g}}$ has a solution. Since $\gcd(a/g, n/g) = 1$, Proposition 2.1.13 implies this latter equation does have a solution. □

In Chapter 4, we will study quadratic reciprocity, which gives a nice criterion for whether or not a quadratic equation modulo n has a solution.

2.1.2 Euler's Theorem

Let $(\mathbf{Z}/n\mathbf{Z})^*$ denote the set of elements $[x] \in \mathbf{Z}/n\mathbf{Z}$ such that $\gcd(x, n) = 1$.

The set $(\mathbf{Z}/n\mathbf{Z})^*$ is a group, called the *group of units of the ring* $\mathbf{Z}/n\mathbf{Z}$; it will be of great interest to us. Each element of this group has an order, and Lagrange's theorem from group theory implies that each element of $(\mathbf{Z}/n\mathbf{Z})^*$ has an order that divides the order of $(\mathbf{Z}/n\mathbf{Z})^*$. In elementary number theory, this fact goes by the monicker "Fermat's Little Theorem" when n is prime and "Euler's Theorem" in general, and we reprove it from basic principles in this section.

Definition 2.1.16 (Order of an Element)**.** Let $n \in \mathbf{N}$ and $x \in \mathbf{Z}$ and suppose that $\gcd(x, n) = 1$. The *order* of x modulo n is the smallest $m \in \mathbf{N}$ such that

$$x^m \equiv 1 \pmod{n}.$$

To show that the definition makes sense, we verify that such an m exists. Consider $x, x^2, x^3, \ldots$ modulo n. There are only finitely many residue classes modulo n, so we must eventually find two integers i, j with $i < j$ such that

$$x^j \equiv x^i \pmod{n}.$$

Since $\gcd(x, n) = 1$, Proposition 2.1.10 implies that we can cancel x's and conclude that

$$x^{j-i} \equiv 1 \pmod{n}.$$

SAGE Example 2.1.17. Use `x.multiplicative_order()` to compute the order of an element of $\mathbf{Z}/n\mathbf{Z}$ in Sage.

```
sage: R = Integers(10)
sage: a = R(3)                 # create an element of Z/10Z
sage: a.multiplicative_order()
4
```

Notice that the powers of a are periodic with period 4, i.e., there are four powers and they repeat:

```
sage: [a^i for i in range(15)]
[1, 3, 9, 7, 1, 3, 9, 7, 1, 3, 9, 7, 1, 3, 9]
```

The command `range(n)` we use above returns the list of integers between 0 and $n-1$, inclusive.

Definition 2.1.18 (Euler's φ-function)**.** For $n \in \mathbf{N}$, let

$$\varphi(n) = \#\{a \in \mathbf{N} : a \leq n \text{ and } \gcd(a,n) = 1\}.$$

For example,

$$\begin{aligned}\varphi(1) &= \#\{1\} = 1,\\ \varphi(2) &= \#\{1\} = 1,\\ \varphi(5) &= \#\{1,2,3,4\} = 4,\\ \varphi(12) &= \#\{1,5,7,11\} = 4.\end{aligned}$$

Also, if p is any prime number then

$$\varphi(p) = \#\{1,2,\ldots,p-1\} = p-1.$$

In Section 2.2.1, we prove that if $\gcd(m,r) = 1$, then $\varphi(mr) = \varphi(m)\varphi(r)$. This will yield an easy way to compute $\varphi(n)$ in terms of the prime factorization of n.

SAGE Example 2.1.19. Use the `euler_phi(n)` command to compute $\varphi(n)$ in Sage:

```
sage: euler_phi(2007)
1332
```

Theorem 2.1.20 (Euler's Theorem)**.** *If* $\gcd(x,n) = 1$*, then*

$$x^{\varphi(n)} \equiv 1 \pmod{n}.$$

Proof. As mentioned above, Euler's Theorem has the following group-theoretic interpretation. The set of units in $\mathbf{Z}/n\mathbf{Z}$ is a group

$$(\mathbf{Z}/n\mathbf{Z})^* = \{a \in \mathbf{Z}/n\mathbf{Z} : \gcd(a,n) = 1\}$$

that has order $\varphi(n)$. The theorem then asserts that the order of an element of $(\mathbf{Z}/n\mathbf{Z})^*$ divides the order $\varphi(n)$ of $(\mathbf{Z}/n\mathbf{Z})^*$. This is a special case of the more general fact (Lagrange's Theorem) that if G is a finite group and $g \in G$, then the order of g divides the cardinality of G.

We now give an elementary proof of the theorem. Let

$$P = \{a : 1 \leq a \leq n \text{ and } \gcd(a,n) = 1\}.$$

In the same way that we proved Lemma 2.1.12, we see that the reductions modulo n of the elements of xP are the same as the reductions of the elements of P. Thus

$$\prod_{a \in P} (xa) \equiv \prod_{a \in P} a \pmod{n},$$

since the products are over the same numbers modulo n. Now cancel the a's on both sides to get

$$x^{\#P} \equiv 1 \pmod{n},$$

as claimed. □

SAGE Example 2.1.21. We illustrate Euler's Theorem using Sage. The `Mod(x,n)` command returns the equivalence class of x in $\mathbf{Z}/n\mathbf{Z}$.

```
sage: n = 20
sage: k = euler_phi(n); k
8
sage: [Mod(x,n)^k for x in range(n) if gcd(x,n) == 1]
[1, 1, 1, 1, 1, 1, 1, 1]
```

2.1.3 Wilson's Theorem

The following characterization of prime numbers, from the 1770s, is called "Wilson's Theorem," though it was first proved by Lagrange.

Proposition 2.1.22 (Wilson's Theorem). *An integer $p > 1$ is prime if and only if $(p-1)! \equiv -1 \pmod{p}$.*

For example, if $p = 3$, then $(p-1)! = 2 \equiv -1 \pmod{3}$. If $p = 17$, then

$$(p-1)! = 20922789888000 \equiv -1 \pmod{17}.$$

But if $p = 15$, then

$$(p-1)! = 87178291200 \equiv 0 \pmod{15},$$

so 15 is composite. Thus Wilson's theorem could be viewed as a primality test, though, from a computational point of view, it is probably one of the world's *least efficient* primality tests since computing $(n-1)!$ takes so many steps.

Proof. The statement is clear when $p = 2$, so henceforth we assume that $p > 2$. We first assume that p is prime and prove that $(p-1)! \equiv -1 \pmod{p}$. If $a \in \{1, 2, \ldots, p-1\}$, then the equation

$$ax \equiv 1 \pmod{p}$$

has a unique solution $a' \in \{1, 2, \ldots, p-1\}$. If $a = a'$, then $a^2 \equiv 1 \pmod{p}$, so $p \mid a^2 - 1 = (a-1)(a+1)$, so $p \mid (a-1)$ or $p \mid (a+1)$, so $a \in \{1, p-1\}$. We can thus pair off the elements of $\{2, 3, \ldots, p-2\}$, each with their inverse. Thus

$$2 \cdot 3 \cdot \cdots \cdot (p-2) \equiv 1 \pmod{p}.$$

Multiplying both sides by $p-1$ proves that $(p-1)! \equiv -1 \pmod{p}$.

Next, we assume that $(p-1)! \equiv -1 \pmod{p}$ and prove that p must be prime. Suppose not, so that $p \geq 4$ is a composite number. Let ℓ be a prime divisor of p. Then $\ell < p$, so $\ell \mid (p-1)!$. Also, by assumption,

$$\ell \mid p \mid ((p-1)! + 1).$$

This is a contradiction, because a prime can not divide a number a and also divide $a+1$, since it would then have to divide $(a+1) - a = 1$. □

Example 2.1.23. We illustrate the key step in the above proof in the case $p = 17$. We have

$$2 \cdot 3 \cdots 15 = (2 \cdot 9) \cdot (3 \cdot 6) \cdot (4 \cdot 13) \cdot (5 \cdot 7) \cdot (8 \cdot 15) \cdot (10 \cdot 12) \cdot (14 \cdot 11) \equiv 1 \pmod{17},$$

where we have paired up the numbers a, b for which $ab \equiv 1 \pmod{17}$.

SAGE Example 2.1.24. We use Sage to create a table of triples; the first column contains n, the second column contains $(n-1)!$ modulo n, and the third contains -1 modulo n. Notice that the first columns contains a prime precisely when the second and third columns are equal. (The ... notation indicates a multi-line command in Sage; you should not type the dots in explicitly.)

```
sage: for n in range(1,10):
...     print n, factorial(n-1) % n, -1 % n
1 0 0
2 1 1
3 2 2
4 2 3
5 4 4
6 0 5
7 6 6
8 0 7
9 0 8
```

2.2 The Chinese Remainder Theorem

In this section, we prove the Chinese Remainder Theorem, which gives conditions under which a system of linear equations is guaranteed to have a solution. In the 4th century a Chinese mathematician asked the following:

Question 2.2.1. There is a quantity whose number is unknown. Repeatedly divided by 3, the remainder is 2; by 5 the remainder is 3; and by 7 the remainder is 2. What is the quantity?

In modern notation, Question 2.2.1 asks us to find a positive integer solution to the following system of three equations:

$$\begin{aligned} x &\equiv 2 \pmod{3}\\ x &\equiv 3 \pmod{5}\\ x &\equiv 2 \pmod{7} \end{aligned}$$

The Chinese Remainder Theorem asserts that a solution exists, and the proof gives a method to find one. (See Section 2.3 for the necessary algorithms.)

Theorem 2.2.2 (Chinese Remainder Theorem)**.** *Let* $a, b \in \mathbf{Z}$ *and* $n, m \in \mathbf{N}$ *such that* $\gcd(n, m) = 1$*. Then there exists* $x \in \mathbf{Z}$ *such that*

$$\begin{aligned} x &\equiv a \pmod{m},\\ x &\equiv b \pmod{n}. \end{aligned}$$

Moreover x *is unique modulo* mn*.*

Proof. If we can solve for t in the equation

$$a + tm \equiv b \pmod{n},$$

then $x = a + tm$ will satisfy both congruences. To see that we can solve, subtract a from both sides and use Proposition 2.1.13 together with our assumption that $\gcd(n, m) = 1$ to see that there is a solution.

For uniqueness, suppose that x and y solve both congruences. Then $z = x - y$ satisfies $z \equiv 0 \pmod{m}$ and $z \equiv 0 \pmod{n}$, so $m \mid z$ and $n \mid z$. Since $\gcd(n, m) = 1$, it follows that $nm \mid z$, so $x \equiv y \pmod{nm}$. □

Algorithm 2.2.3 (Chinese Remainder Theorem)**.** Given coprime integers m and n and integers a and b, this algorithm find an integer x such that $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$.

1. [Extended GCD] Use Algorithm 2.3.7 below to find integers c, d such that $cm + dn = 1$.
2. [Answer] Output $x = a + (b - a)cm$ and terminate.

Proof. Since $c \in \mathbf{Z}$, we have $x \equiv a \pmod{m}$, and using that $cm + dn = 1$, we have $a + (b - a)cm \equiv a + (b - a) \equiv b \pmod{n}$. □

Now we can answer Question 2.2.1. First, we use Theorem 2.2.2 to find a solution to the pair of equations

$$\begin{aligned} x &\equiv 2 \pmod{3},\\ x &\equiv 3 \pmod{5}. \end{aligned}$$

Set $a = 2$, $b = 3$, $m = 3$, $n = 5$. Step 1 is to find a solution to $t \cdot 3 \equiv 3 - 2 \pmod{5}$. A solution is $t = 2$. Then $x = a + tm = 2 + 2 \cdot 3 = 8$. Since any x' with $x' \equiv x \pmod{15}$ is also a solution to those two equations, we can solve all three equations by finding a solution to the pair of equations

$$\begin{aligned} x &\equiv 8 \pmod{15}\\ x &\equiv 2 \pmod{7}. \end{aligned}$$

Again, we find a solution to $t \cdot 15 \equiv 2 - 8 \pmod{7}$. A solution is $t = 1$, so

$$x = a + tm = 8 + 15 = 23.$$

Note that there are other solutions. Any $x' \equiv x \pmod{3 \cdot 5 \cdot 7}$ is also a solution; e.g., $23 + 3 \cdot 5 \cdot 7 = 128$.

SAGE Example 2.2.4. The `CRT(a,b,m,n)` command in Sage computes an integer x such that $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$. For example,

```
sage: CRT(2,3, 3, 5)
-7
```

The `CRT_list` command computes a number that reduces to several numbers modulo coprime moduli. We use it to answer Question 2.2.1:

```
sage: CRT_list([2,3,2], [3,5,7])
23
```

2.2.1 Multiplicative Functions

Recall from Definition 2.1.18 that the *Euler φ-function* is

$$\varphi(n) = \#\{a : 1 \leq a \leq n \text{ and } \gcd(a, n) = 1\}.$$

Lemma 2.2.5. *Suppose that $m, n \in \mathbf{N}$ and* $\gcd(m, n) = 1$. *Then the map*

$$\psi : (\mathbf{Z}/mn\mathbf{Z})^* \to (\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*. \tag{2.2.1}$$

defined by

$$\psi(c) = (c \textit{ mod } m,\ c \textit{ mod } n)$$

is a bijection.

Proof. We first show that ψ is injective. If $\psi(c) = \psi(c')$, then $m \mid c-c'$ and $n \mid c-c'$, so $nm \mid c-c'$ because $\gcd(n,m) = 1$. Thus $c = c'$ as elements of $(\mathbf{Z}/mn\mathbf{Z})^*$.

Next we show that ψ is surjective, i.e., that every element of $(\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*$ is of the form $\psi(c)$ for some c. Given a and b with $\gcd(a,m) = 1$ and $\gcd(b,n) = 1$, Theorem 2.2.2 implies that there exists c with $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$. We may assume that $1 \leq c \leq nm$, and since $\gcd(a,m) = 1$ and $\gcd(b,n) = 1$, we must have $\gcd(c,nm) = 1$. Thus $\psi(c) = (a,b)$. □

Definition 2.2.6 (Multiplicative Function)**.** A function $f : \mathbf{N} \to \mathbf{C}$ is *multiplicative* if, whenever $m, n \in \mathbf{N}$ and $\gcd(m,n) = 1$, we have

$$f(mn) = f(m) \cdot f(n).$$

Proposition 2.2.7 (Multiplicativity of φ)**.** *The function φ is multiplicative.*

Proof. The map ψ of Lemma 2.2.5 is a bijection, so the set on the left in (2.2.1) has the same size as the product set on the right in (2.2.1). Thus

$$\varphi(mn) = \varphi(m) \cdot \varphi(n).$$

□

The proposition is helpful in computing $\varphi(n)$, at least if we assume we can compute the factorization of n (see Section 3.4.1 for a connection between factoring n and computing $\varphi(n)$). For example,

$$\varphi(12) = \varphi(2^2) \cdot \varphi(3) = 2 \cdot 2 = 4.$$

Also, for $n \geq 1$, we have

$$\varphi(p^n) = p^n - \frac{p^n}{p} = p^n - p^{n-1} = p^{n-1}(p-1), \tag{2.2.2}$$

since $\varphi(p^n)$ is the number of numbers less than p^n minus the number of those that are divisible by p. Thus, e.g.,

$$\varphi(389 \cdot 11^2) = 388 \cdot (11^2 - 11) = 388 \cdot 110 = 42680.$$

2.3 Quickly Computing Inverses and Huge Powers

This section is about how to solve the equation $ax \equiv 1 \pmod{n}$ when we know it has a solution, and how to efficiently compute $a^m \pmod{n}$. We also discuss a simple probabilistic primality test that relies on our ability to compute $a^m \pmod{n}$ quickly. All three of these algorithms are of fundamental importance to the cryptography algorithms of Chapter 3.

2.3.1 How to Solve $ax \equiv 1 \pmod{n}$

Suppose $a, n \in \mathbf{N}$ with $\gcd(a, n) = 1$. Then by Proposition 2.1.13 the equation $ax \equiv 1 \pmod{n}$ has a unique solution. How can we find it?

Proposition 2.3.1 (Extended Euclidean Representation)**.** *Suppose* $a, b \in \mathbf{Z}$ *and let* $g = \gcd(a, b)$*. Then there exists* $x, y \in \mathbf{Z}$ *such that*

$$ax + by = g.$$

Remark 2.3.2. If $e = cg$ is a multiple of g, then $cax + cby = cg = e$, so $e = (cx)a + (cy)b$ can also be written in terms of a and b.

Proof of Proposition 2.3.1. Let $g = \gcd(a, b)$. Then $\gcd(a/g, b/g) = 1$, so by Proposition 2.1.15, the equation

$$\frac{a}{g} \cdot x \equiv 1 \left(\mathrm{mod}\ \frac{b}{g}\right) \tag{2.3.1}$$

has a solution $x \in \mathbf{Z}$. Multiplying (2.3.1) through by g yields $ax \equiv g \pmod{b}$, so there exists y such that $b \cdot (-y) = ax - g$. Then $ax + by = g$, as required. □

Given a, b and $g = \gcd(a, b)$, our proof of Proposition 2.3.1 gives a way to explicitly find x, y such that $ax + by = g$, assuming one knows an algorithm to solve linear equations modulo n. Since we do not know such an algorithm, we now discuss a way to explicitly find x and y. This algorithm will in fact enable us to solve linear equations modulo n. To solve $ax \equiv 1 \pmod{n}$ when $\gcd(a, n) = 1$, use the Algorithm 2.3.7 to find x and y such that $ax + ny = 1$. Then $ax \equiv 1 \pmod{n}$.

Example 2.3.3. Suppose $a = 5$ and $b = 7$. The steps of Algorithm 1.1.13 to compute $\gcd(5, 7)$ are as follows. Here we underline certain numbers, because it clarifies the subsequent back substitution we will use to find x and y.

$$\begin{aligned}
&\underline{7} = 1 \cdot \underline{5} + \underline{2} && \text{so } \underline{2} = \underline{7} - \underline{5}\\
&\underline{5} = 2 \cdot \underline{2} + \underline{1} && \text{so } \underline{1} = \underline{5} - 2 \cdot \underline{2} = \underline{5} - 2(\underline{7} - \underline{5}) = 3 \cdot \underline{5} - 2 \cdot \underline{7}
\end{aligned}$$

On the right, we have back-substituted in order to write each partial remainder as a linear combination of a and b. In the last step, we obtain $\gcd(a, b)$ as a linear combination of a and b, as desired.

Example 2.3.4. That example was not too complicated, so we try another one. Let $a = 130$ and $b = 61$. We have

$$\begin{aligned}
\underline{130} &= 2 \cdot \underline{61} + \underline{8} & \underline{8} &= \underline{130} - 2 \cdot \underline{61}\\
\underline{61} &= 7 \cdot \underline{8} + \underline{5} & \underline{5} &= -7 \cdot \underline{130} + 15 \cdot \underline{61}\\
\underline{8} &= 1 \cdot \underline{5} + \underline{3} & \underline{3} &= 8 \cdot \underline{130} - 17 \cdot \underline{61}\\
\underline{5} &= 1 \cdot \underline{3} + \underline{2} & \underline{2} &= -15 \cdot \underline{130} + 32 \cdot \underline{61}\\
\underline{3} &= 1 \cdot \underline{2} + \underline{1} & \underline{1} &= 23 \cdot \underline{130} - 49 \cdot \underline{61}
\end{aligned}$$

Thus $x = 23$ and $y = -49$ is a solution to $130x + 61y = 1$.

Example 2.3.5. This example is just like Example 2.3.4 above, except we make the notation on the right more compact.

$$\begin{aligned}
\underline{130} &= 2 \cdot \underline{61} + \underline{8} & \qquad \underline{8} &= (1, -2)\\
\underline{61} &= 7 \cdot \underline{8} + \underline{5} & \underline{5} &= (-7, 15) = (0, 1) - 7(1, -2)\\
\underline{8} &= 1 \cdot \underline{5} + \underline{3} & \underline{3} &= (8, -17) = (1, -2) - (-7, 15)\\
\underline{5} &= 1 \cdot \underline{3} + \underline{2} & \underline{2} &= (-15, 32) = (-7, 15) - (8, -17)\\
\underline{3} &= 1 \cdot \underline{2} + \underline{1} & \underline{1} &= (23, -49) = (8, -17) - (-15, 32)
\end{aligned}$$

Notice at each step that the vector on the right is just the vector from two steps ago minus a multiple of the vector from one step ago, where the multiple is the cofficient of what we divide by.

SAGE Example 2.3.6. The `xgcd(a,b)` command computes the greatest common divisor g of a and b along with x, y such that $ax + by = g$.

```
sage: xgcd(5,7)
(1, -4, 3)
sage: xgcd(130,61)
(1, 23, -49)
```

Algorithm 2.3.7 (Extended Euclidean Algorithm)**.** Suppose a and b are integers and let $g = \gcd(a, b)$. This algorithm finds g, x and y such that $ax + by = g$. We describe only the steps when $a > b \geq 0$, since one can easily reduce to this case.

1. [Initialize] Set $x = 1$, $y = 0$, $r = 0$, $s = 1$.
2. [Finished?] If $b = 0$, set $g = a$ and terminate.
3. [Quotient and Remainder] Use Algorithm 1.1.12 to write $a = qb + c$ with $0 \leq c < b$.
4. [Shift] Set $(a, b, r, s, x, y) = (b, c, x - qr, y - qs, r, s)$ and go to Step 2. (This shift step is nicely illustrated in Example 2.3.5.)

Proof. This algorithm is the same as Algorithm 1.1.13, except that we keep track of extra variables x, y, r, s, so it terminates and when it terminates $d = \gcd(a, b)$. We omit the rest of the inductive proof that the algorithm is correct, and instead refer the reader to [Knu97, §1.2.1]. □

Algorithm 2.3.8 (Inverse Modulo n)**.** Suppose a and n are integers and $\gcd(a, n) = 1$. This algorithm finds an x such that $ax \equiv 1 \pmod{n}$.

1. [Compute Extended GCD] Use Algorithm 2.3.7 to compute integers x, y such that $ax + ny = \gcd(a, n) = 1$.
2. [Finished] Output x.

Proof. Reduce $ax+ny = 1$ modulo n to see that x satisfies $ax \equiv 1 \pmod{n}$. □

Example 2.3.9. Solve $17x \equiv 1 \pmod{61}$. First, we use Algorithm 2.3.7 to find x, y such that $17x + 61y = 1$:

$$\begin{array}{rlcrl}
\underline{61} &= 3 \cdot \underline{17} + \underline{10} & \qquad & \underline{10} &= \underline{61} - 3 \cdot \underline{17} \\
\underline{17} &= 1 \cdot \underline{10} + \underline{7} & & \underline{7} &= -\underline{61} + 4 \cdot \underline{17} \\
\underline{10} &= 1 \cdot \underline{7} + \underline{3} & & \underline{3} &= 2 \cdot \underline{61} - 7 \cdot \underline{17} \\
\underline{3} &= 2 \cdot \underline{3} + \underline{1} & & \underline{1} &= -5 \cdot \underline{61} + 18 \cdot \underline{17}
\end{array}$$

Thus $17 \cdot 18 + 61 \cdot (-5) = 1$ so $x = 18$ is a solution to $17x \equiv 1 \pmod{61}$.

SAGE Example 2.3.10. Sage implements the above algorithm for quickly computing inverses modulo n. For example,

```
sage: a = Mod(17, 61)
sage: a^(-1)
18
```

2.3.2 How to Compute $a^m \pmod{n}$

Let a and n be integers, and m a nonnegative integer. In this section, we describe an efficient algorithm to compute $a^m \pmod{n}$. For the cryptography applications in Chapter 3, m will have hundreds of digits.

The naive approach to computing $a^m \pmod{n}$ is to simply compute $a^m = a \cdot a \cdots a \pmod{n}$ by repeatedly multiplying by a and reducing modulo m. Note that after each arithmetic operation is completed, we reduce the result modulo n so that the sizes of the numbers involved do not get too large. Nonetheless, this algorithm is horribly inefficient because it takes $m - 1$ multiplications, which is huge if m has hundreds of digits.

A much more efficient algorithm for computing $a^m \pmod{n}$ involves writing m in binary, then expressing a^m as a product of expressions a^{2^i}, for various i. These latter expressions can be computed by repeatedly squaring a^{2^i}. This more clever algorithm is not “simpler,” but it is vastly more efficient since the number of operations needed grows with the number of binary digits of m, whereas with the naive algorithm in the previous paragraph, the number of operations is $m - 1$.

Algorithm 2.3.11 (Write a number in binary)**.** Let m be a nonnegative integer. This algorithm writes m in binary, so it finds $\varepsilon_i \in \{0, 1\}$ such that $m = \sum_{i=0}^{r} \varepsilon_i 2^i$ with each $\varepsilon_i \in \{0, 1\}$.

1. [Initialize] Set $i = 0$.
2. [Finished?] If $m = 0$, terminate.
3. [Digit] If m is odd, set $\varepsilon_i = 1$, otherwise $\varepsilon_i = 0$. Increment i.

4. [Divide by 2] Set $m = \left\lfloor \frac{m}{2} \right\rfloor$, the greatest integer $\leq m/2$. Goto Step 2.

SAGE Example 2.3.12. To write a number in binary using Sage, use the `str` command:

```
sage: 100.str(2)
'1100100'
```

Notice the above is the correct binary expansion:

```
sage: 0*2^0 + 0*2^1 + 1*2^2 + 0*2^3 + 0*2^4 + 1*2^5 + 1*2^6
100
```

Algorithm 2.3.13 (Compute Power)**.** Let a and n be integers and m a nonnegative integer. This algorithm computes a^m modulo n.

1. [Write in Binary] Write m in binary using Algorithm 2.3.11, so $a^m = \prod_{\varepsilon_i=1} a^{2^i} \pmod{n}$.
2. [Compute Powers] Compute a, a^2, $a^{2^2} = (a^2)^2$, $a^{2^3} = (a^{2^2})^2$, etc., up to a^{2^r}, where $r+1$ is the number of binary digits of m.
3. [Multiply Powers] Multiply together the a^{2^i} such that $\varepsilon_i = 1$, always working modulo n.

Example 2.3.14. We can compute the last 2 digits of 7^{91}, by finding 7^{91} (mod 100). First, because $\gcd(7, 100) = 1$, we have by Theorem 2.1.20 that $7^{\varphi(100)} \equiv 1 \pmod{100}$. Because φ is multiplicative,

$$\varphi(100) = \varphi(2^2 \cdot 5^2) = (2^2 - 2) \cdot (5^2 - 5) = 40.$$

Thus $7^{40} \equiv 1 \pmod{100}$, hence

$$7^{91} \equiv 7^{40+40+11} \equiv 7^{11} \pmod{100}.$$

We now compute $7^{11} \pmod{100}$ using the above algorithm. First, write 11 in binary by repeatedly dividing by 2.

$$\begin{aligned} 11 &= 5 \cdot 2 + 1 \\ 5 &= 2 \cdot 2 + 1 \\ 2 &= 1 \cdot 2 + 0 \\ 1 &= 0 \cdot 2 + 1 \end{aligned}$$

So in binary, $(11)_2 = 1011$, which we check:

$$11 = 1 \cdot 8 + 1 \cdot 2 + 1.$$

Next, compute a, a^2, a^4, a^8 and output $a^8 \cdot a^2 \cdot a$. We have

$$\begin{aligned} a &= 7\\ a^2 &\equiv 49\\ a^4 &\equiv 49^2 \equiv 1\\ a^8 &\equiv 1^2 \equiv 1 \end{aligned}$$

Note: it is easiest to square 49 by working modulo 4 and 25 and using the Chinese Remainder Theorem. Finally,

$$7^{91} \equiv 7^{11} \equiv a^8 \cdot a^2 \cdot a \equiv 1 \cdot 49 \cdot 7 \equiv 43 \pmod{100}.$$

SAGE Example 2.3.15. Sage implements the above algorithm for computing powers efficiently. For example,

```
sage: Mod(7,100)^91
43
```

We can also, of course, directly compute 7^{91} in Sage, though we would not want to do this by hand:

```
sage: 7^91
80153343160247310515380886994816022539378033762994852
007501964604841680190743
```

2.4 Primality Testing

Theorem 2.4.1 (Pseudoprimality)**.** *An integer $p > 1$ is prime if and only if for* every $a \not\equiv 0 \pmod{p}$,

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof. If p is prime, then the statement follows from Proposition 2.1.22. If p is composite, then there is a divisor a of p with $2 \leq a < p$. If $a^{p-1} \equiv 1 \pmod{p}$, then $p \mid a^{p-1} - 1$. Since $a \mid p$, we have $a \mid a^{p-1} - 1$, hence there exists an integer k such that $ak = a^{p-1} - 1$. Subtracting, we see that $a^{p-1} - ak = 1$, so $a(a^{p-2} - k) = 1$. This implies that $a \mid 1$, which is a contradiction since $a \geq 2$. □

Suppose $n \in \mathbf{N}$. Using Theorem 2.4.1 and Algorithm 2.3.13, we can either quickly prove that n is not prime, or convince ourselves that n is likely prime (but not quickly prove that n is prime). For example, if $2^{n-1} \not\equiv 1 \pmod{n}$, then we have proved that n is not prime. On the other hand, if $a^{n-1} \equiv 1 \pmod{n}$ for a few a, it "seems likely" that n is prime, and we loosely refer to such a number that seems prime for several bases as a *pseudoprime.*

There are composite numbers n (called *Carmichael numbers*) with the amazing property that $a^{n-1} \equiv 1 \pmod{n}$ for *all* a with $\gcd(a, n) = 1$. The first Carmichael number is 561, and it is a theorem that there are infinitely many such numbers ([AGP94]).

Example 2.4.2. Is $p = 323$ prime? We compute $2^{322} \pmod{323}$. Making a table as above, we have

i	m	ε_i	$2^{2^i} \bmod 323$
0	322	0	2
1	161	1	4
2	80	0	16
3	40	0	256
4	20	0	290
5	10	0	120
6	5	1	188
7	2	0	137
8	1	1	35

Thus

$$2^{322} \equiv 4 \cdot 188 \cdot 35 \equiv 157 \pmod{323},$$

so 323 is not prime, though this computation gives no information about how 323 factors as a product of primes. In fact, one finds that $323 = 17 \cdot 19$.

SAGE Example 2.4.3. It's possible to easily prove that a large number is composite, but the proof does not easily yield a factorization. For example if

$$n = 95468093486093450983409583409850934850938459083,$$

then $2^{n-1} \not\equiv 1 \pmod{n}$, so n is composite.

```
sage: n = 95468093486093450983409583409850934850938459083
sage: Mod(2,n)^(n-1)
34173444139265553870830266378598407069248687241
```

Note that factoring n actually takes much longer than the above computation (which was essentially instant).

```
sage: factor(n)                # takes up to a few seconds.
1610302526747 * 59285812386415488446397191791023889
```

Another practical primality test is the Miller-Rabin test, which has the property that each time it is run on a number n it either correctly asserts that the number is definitely not prime, or that it is probably prime, and the probability of correctness goes up with each successive call. If Miller-Rabin is called m times on n and in each case claims that n is probably prime, then one can in a precise sense bound the probability that n is composite in terms of m.

We state the Miller-Rabin algorithm precisely, but do not prove anything about the probability that it will succeed.

Algorithm 2.4.4 (Miller-Rabin Primality Test)**.** Given an integer $n \geq 5$ this algorithm outputs either true or false. If it outputs true, then n is "probably prime," and if it outputs false, then n is definitely composite.

1. [Split Off Power of 2] Compute the unique integers m and k such that m is odd and $n - 1 = 2^k \cdot m$.
2. [Random Base] Choose a random integer a with $1 < a < n$.
3. [Odd Power] Set $b = a^m \pmod{n}$. If $b \equiv \pm 1 \pmod{n}$ output true and terminate.
4. [Even Powers] If $b^{2^r} \equiv -1 \pmod{n}$ for any r with $1 \leq r \leq k-1$, output true and terminate. Otherwise output false.

If Miller-Rabin outputs true for n, we can call it again with n and if it again outputs true then the probability that we have incorrectly determined that n is prime (when n is actually composite) decreases.

Proof. We will prove that the algorithm is correct, but will prove nothing about how likely the algorithm is to assert that a composite is prime. We must prove that if the algorithm pronounces an integer n composite, then n really is composite. Thus suppose n is prime, yet the algorithm pronounces n composite. Then $a^m \not\equiv \pm 1 \pmod{n}$, and for all r with $1 \leq r \leq k-1$ we have $a^{2^r m} \not\equiv -1 \pmod{n}$. Since n is prime and $2^{k-1}m = (n-1)/2$, Proposition 4.2.1 implies that $a^{2^{k-1}m} \equiv \pm 1 \pmod{n}$, so by our hypothesis $a^{2^{k-1}m} \equiv 1 \pmod{n}$. But then $(a^{2^{k-2}m})^2 \equiv 1 \pmod{n}$, so by Proposition 2.5.3 (which is proved right after it is stated, and whose proof does not depend on this argument), we have $a^{2^{k-2}m} \equiv \pm 1 \pmod{n}$. Again, by our hypothesis, this implies $a^{2^{k-2}} \equiv 1 \pmod{n}$. Repeating this argument inductively, we see that $a^m \equiv \pm 1 \pmod{n}$, which contradicts our hypothesis on a. □

Until recently it was an open problem to give an algorithm (with proof) that decides whether or not any integer is prime in time bounded by a polynomial in the number of digits of the integer. Agrawal, Kayal, and Saxena recently found the first polynomial-time primality test (see [AKS02]). We will not discuss their algorithm further, because for our applications to cryptography Miller-Rabin or pseudoprimality tests will be sufficient. See [Sho05, Ch. 21] for a book that gives a detailed exposition of this algorithm.

SAGE Example 2.4.5. The `is_prime` command uses a combination of techniques to determines (provably correctly!) whether or not an integer is prime.

```
sage: n = 95468093486093450983409583409850934850938459083
```

```
sage: is_prime(n)
False
```

We use the is_prime function to make a table of the first few Mersenne primes (see Section 1.2.3).

```
sage: for p in primes(100):
...   if is_prime(2^p - 1):
...       print p, 2^p - 1
2 3
3 7
5 31
7 127
13 8191
17 131071
19 524287
31 2147483647
61 2305843009213693951
89 618970019642690137449562111
```

There is a specialized test for primality of Mersenne numbers called the Lucas-Lehmer test. This remarkably simple algorithm determines provably correctly whether or not a number $2^p - 1$ is prime. We implement it in a few lines of code and use the Lucas-Lehmer test to check for primality of two Mersenne numbers:

```
sage: def is_prime_lucas_lehmer(p):
...    s = Mod(4, 2^p - 1)
...    for i in range(3, p+1):
...        s = s^2 - 2
...    return s == 0
sage: # Check primality of 2^9941 - 1
sage: is_prime_lucas_lehmer(9941)
True
sage: # Check primality of 2^next_prime(1000)-1
sage: is_prime_lucas_lehmer(next_prime(1000))
False
```

For more on Mersenne primes, see the Great Internet Mersenne Prime Search (GIMPS) project at http://www.mersenne.org/.

2.5 The Structure of $(\mathbf{Z}/p\mathbf{Z})^*$

This section is about the structure of the group $(\mathbf{Z}/p\mathbf{Z})^*$ of units modulo a prime number p. The main result is that this group is always cyclic. We will use this result later in Chapter 4 in our proof of quadratic reciprocity.

Definition 2.5.1 (Primitive root)**.** A *primitive root* modulo an integer n is an element of $(\mathbf{Z}/n\mathbf{Z})^*$ of order $\varphi(n)$.

We will prove that there is a primitive root modulo every prime p. Since the unit group $(\mathbf{Z}/p\mathbf{Z})^*$ has order $p-1$, this implies that $(\mathbf{Z}/p\mathbf{Z})^*$ is a cyclic group, a fact that will be extremely useful, since it completely determines the structure of $(\mathbf{Z}/p\mathbf{Z})^*$ as a group.

If n is an odd prime power, then there is a primitive root modulo n (see Exercise 2.28), but there is no primitive root modulo the prime power 2^3, and hence none mod 2^n for $n \geq 3$ (see Exercise 2.27).

Section 2.5.1 is the key input to our proof that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic; here we show that for every divisor d of $p-1$ there are exactly d elements of $(\mathbf{Z}/p\mathbf{Z})^*$ whose order divides d. We then use this result in Section 2.5.2 to produce an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order q^r when q^r is a prime power that exactly divides $p-1$ (i.e., q^r divides $p-1$, but q^{r+1} does not divide $p-1$), and multiply together these elements to obtain an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order $p-1$.

SAGE Example 2.5.2. Use the `primitive_root` command to compute the smallest positive integer that is a primitive root modulo n. For example, below we compute primitive roots modulo p for each prime $p < 20$.

```
sage: for p in primes(20):
...     print p, primitive_root(p)
2 1
3 2
5 2
7 3
11 2
13 2
17 3
19 2
```

2.5.1 Polynomials over $\mathbf{Z}/p\mathbf{Z}$

The polynomials $x^2 - 1$ has four roots in $\mathbf{Z}/8\mathbf{Z}$, namely 1, 3, 5, and 7. In contrast, the following proposition shows that a polynomial of degree d over a field, such as $\mathbf{Z}/p\mathbf{Z}$, can have at most d roots.

Proposition 2.5.3 (Root Bound)**.** *Let $f \in k[x]$ be a nonzero polynomial over a field k. Then there are at most* $\deg(f)$ *elements $\alpha \in k$ such that $f(\alpha) = 0$.*

Proof. We prove the proposition by induction on $\deg(f)$. The cases in which $\deg(f) \leq 1$ are clear. Write $f = a_n x^n + \cdots a_1 x + a_0$. If $f(\alpha) = 0$, then

$$\begin{aligned} f(x) &= f(x) - f(\alpha) \\ &= a_n(x^n - \alpha^n) + \cdots + a_1(x - \alpha) + a_0(1 - 1) \\ &= (x - \alpha)(a_n(x^{n-1} + \cdots + \alpha^{n-1}) + \cdots + a_2(x + \alpha) + a_1) \\ &= (x - \alpha)g(x), \end{aligned}$$

for some polynomial $g(x) \in k[x]$. Next, suppose that $f(\beta) = 0$ with $\beta \neq \alpha$. Then $(\beta - \alpha)g(\beta) = 0$, so, since $\beta - \alpha \neq 0$ and k is a field, we have $g(\beta) = 0$. By our inductive hypothesis, g has at most $n - 1$ roots, so there are at most $n - 1$ possibilities for β. It follows that f has at most n roots. □

SAGE Example 2.5.4. We use Sage to find the roots of a polynomials over $\mathbf{Z}/13\mathbf{Z}$.

```
sage: R.<x> = PolynomialRing(Integers(13))
sage: f = x^15 + 1
sage: f.roots()
[(12, 1), (10, 1), (4, 1)]
sage: f(12)
0
```

The output of the roots command above lists each root along with its multiplicity (which is 1 in each case above).

Proposition 2.5.5. *Let p be a prime number and let d be a divisor of $p - 1$. Then $f = x^d - 1 \in (\mathbf{Z}/p\mathbf{Z})[x]$ has exactly d roots in $\mathbf{Z}/p\mathbf{Z}$.*

Proof. Let $e = (p - 1)/d$. We have

$$\begin{aligned} x^{p-1} - 1 &= (x^d)^e - 1 \\ &= (x^d - 1)((x^d)^{e-1} + (x^d)^{e-2} + \cdots + 1) \\ &= (x^d - 1)g(x), \end{aligned}$$

where $g \in (\mathbf{Z}/p\mathbf{Z})[x]$ and $\deg(g) = de - d = p - 1 - d$. Theorem 2.1.20 implies that $x^{p-1} - 1$ has exactly $p - 1$ roots in $\mathbf{Z}/p\mathbf{Z}$, since every nonzero element of $\mathbf{Z}/p\mathbf{Z}$ is a root! By Proposition 2.5.3, g has *at most* $p - 1 - d$ roots and $x^d - 1$ has at most d roots. Since a root of $(x^d - 1)g(x)$ is a root of either $x^d - 1$ or $g(x)$ and $x^{p-1} - 1$ has $p - 1$ roots, g must have exactly $p - 1 - d$ roots and $x^d - 1$ must have exactly d roots, as claimed. □

SAGE Example 2.5.6. We use Sage to illustrate the proposition.

```
sage: R.<x> = PolynomialRing(Integers(13))
sage: f = x^6 + 1
sage: f.roots()
[(11, 1), (8, 1), (7, 1), (6, 1), (5, 1), (2, 1)]
```

We pause to reemphasize that the analog of Proposition 2.5.5 is false when p is replaced by a composite integer n, since a root mod n of a product of two polynomials need not be a root of either factor. For example, $f = x^2 - 1 = (x-1)(x+1) \in \mathbf{Z}/15\mathbf{Z}[x]$ has the four roots 1, 4, 11, and 14.

2.5.2 Existence of Primitive Roots

Recall from Section 2.1.2 that the *order* of an element x in a finite group is the smallest $m \geq 1$ such that $x^m = 1$. In this section, we prove that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic by using the results of Section 2.5.1 to produce an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order d for each prime power divisor d of $p-1$, and then we multiply these together to obtain an element of order $p-1$.

We will use the following lemma to assemble elements of each order dividing $p-1$ to produce an element of order $p-1$.

Lemma 2.5.7. *Suppose $a, b \in (\mathbf{Z}/n\mathbf{Z})^*$ have orders r and s, respectively, and that* $\gcd(r,s) = 1$. *Then ab has order rs.*

Proof. This is a general fact about commuting elements of any group; our proof only uses that $ab = ba$ and nothing special about $(\mathbf{Z}/n\mathbf{Z})^*$. Since

$$(ab)^{rs} = a^{rs}b^{rs} = 1,$$

the order of ab is a divisor of rs. Write this divisor as $r_1 s_1$ where $r_1 \mid r$ and $s_1 \mid s$. Raise both sides of the equation

$$a^{r_1 s_1} b^{r_1 s_1} = (ab)^{r_1 s_1} = 1$$

to the power $r_2 = r/r_1$ to obtain

$$a^{r_1 r_2 s_1} b^{r_1 r_2 s_1} = 1.$$

Since $a^{r_1 r_2 s_1} = (a^{r_1 r_2})^{s_1} = 1$, we have

$$b^{r_1 r_2 s_1} = 1,$$

so $s \mid r_1 r_2 s_1$. Since $\gcd(s, r_1 r_2) = \gcd(s, r) = 1$, it follows that $s = s_1$. Similarly $r = r_1$, so the order of ab is rs. □

Theorem 2.5.8 (Primitive Roots)**.** *There is a primitive root modulo any prime p. In particular, the group $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic.*

Proof. The theorem is true if $p = 2$, since 1 is a primitive root, so we may assume $p > 2$. Write $p - 1$ as a product of distinct prime powers $q_i^{n_i}$:

$$p - 1 = q_1^{n_1} q_2^{n_2} \cdots q_r^{n_r}.$$

By Proposition 2.5.5, the polynomial $x^{q_i^{n_i}} - 1$ has exactly $q_i^{n_i}$ roots, and the polynomial $x^{q_i^{n_i - 1}} - 1$ has exactly $q_i^{n_i - 1}$ roots. There are $q_i^{n_i} - q_i^{n_i - 1} =$

$q_i^{n_i-1}(q_i-1)$ elements $a \in \mathbf{Z}/p\mathbf{Z}$ such that $a^{q_i^{n_i}} = 1$ but $a^{q_i^{n_i-1}} \neq 1$; each of these elements has order $q_i^{n_i}$. Thus for each $i = 1, \ldots, r$, we can choose an a_i of order $q_i^{n_i}$. Then, using Lemma 2.5.7 repeatedly, we see that

$$a = a_1 a_2 \cdots a_r$$

has order $q_1^{n_1} \cdots q_r^{n_r} = p - 1$, so a is a primitive root modulo p. □

Example 2.5.9. We illustrate the proof of Theorem 2.5.8 when $p = 13$. We have

$$p - 1 = 12 = 2^2 \cdot 3.$$

The polynomial $x^4 - 1$ has roots $\{1, 5, 8, 12\}$ and $x^2 - 1$ has roots $\{1, 12\}$, so we may take $a_1 = 5$. The polynomial $x^3 - 1$ has roots $\{1, 3, 9\}$, and we set $a_2 = 3$. Then $a = 5 \cdot 3 = 15 \equiv 2$ is a primitive root. To verify this, note that the successive powers of $2 \pmod{13}$ are

$$2,\ 4,\ 8,\ 3,\ 6,\ 12,\ 11,\ 9,\ 5,\ 10,\ 7,\ 1.$$

Example 2.5.10. Theorem 2.5.8 is false if, for example, p is replaced by a power of 2 bigger than 4. For example, the four elements of $(\mathbf{Z}/8\mathbf{Z})^*$ each have order dividing 2, but $\varphi(8) = 4$.

Theorem 2.5.11 (Primitive Roots mod p^n)**.** *Let p^n be a power of an odd prime. Then there is a primitive root modulo p^n.*

The proof is left as Exercise 2.28.

Proposition 2.5.12 (Number of primitive roots)**.** *If there is a primitive root modulo n, then there are exactly $\varphi(\varphi(n))$ primitive roots modulo n.*

Proof. The primitive roots modulo n are the generators of $(\mathbf{Z}/n\mathbf{Z})^*$, which by assumption is cyclic of order $\varphi(n)$. Thus they are in bijection with the generators of any cyclic group of order $\varphi(n)$. In particular, the number of primitive roots modulo n is the same as the number of elements of $\mathbf{Z}/\varphi(n)\mathbf{Z}$ with additive order $\varphi(n)$. An element of $\mathbf{Z}/\varphi(n)\mathbf{Z}$ has additive order $\varphi(n)$ if and only if it is coprime to $\varphi(n)$. There are $\varphi(\varphi(n))$ such elements, as claimed. □

Example 2.5.13. For example, there are $\varphi(\varphi(17)) = \varphi(16) = 2^4 - 2^3 = 8$ primitive roots mod 17, namely $3, 5, 6, 7, 10, 11, 12, 14$. The $\varphi(\varphi(9)) = \varphi(6) = 2$ primitive roots modulo 9 are 2 and 5. There are no primitive roots modulo 8, even though $\varphi(\varphi(8)) = \varphi(4) = 2 > 0$.

2.5.3 Artin's Conjecture

Conjecture 2.5.14 (Emil Artin)**.** *Suppose $a \in \mathbf{Z}$ is not -1 or a perfect square. Then there are infinitely many primes p such that a is a primitive root modulo p.*

There is no single integer a such that Artin's conjecture is known to be true. For any given a, Pieter [Mor93] proved that there are infinitely many p such that the order of a is divisible by the largest prime factor of $p-1$. Hooley [Hoo67] proved that something called the Generalized Riemann Hypothesis implies Conjecture 2.5.14.

Remark 2.5.15. Artin conjectured more precisely that if $N(x,a)$ is the number of primes $p \leq x$ such that a is a primitive root modulo p, then $N(x,a)$ is asymptotic to $C(a)\pi(x)$, where $C(a)$ is a positive constant that depends only on a and $\pi(x)$ is the number of primes up to x.

2.5.4 Computing Primitive Roots

Theorem 2.5.8 does not suggest an efficient algorithm for finding primitive roots. To actually find a primitive root mod p in practice, we try $a = 2$, then $a = 3$, etc., until we find an a that has order $p-1$. Computing the order of an element of $(\mathbf{Z}/p\mathbf{Z})^*$ requires factoring $p-1$, which we do not know how to do quickly in general, so finding a primitive root modulo p for large p seems to be a difficult problem.

Algorithm 2.5.16 (Primitive Root)**.** Given a prime p, this algorithm computes the smallest positive integer a that generates $(\mathbf{Z}/p\mathbf{Z})^*$.

1. [$p = 2$?] If $p = 2$ output 1 and terminate. Otherwise set $a = 2$.
2. [Prime Divisors] Compute the prime divisors $p_1, \ldots, p_r$ of $p-1$.
3. [Generator?] If for every p_i, we have $a^{(p-1)/p_i} \not\equiv 1 \pmod{p}$, then a is a generator of $(\mathbf{Z}/p\mathbf{Z})^*$, so output a and terminate.
4. [Try next] Set $a = a + 1$ and go to Step 3.

Proof. Let $a \in (\mathbf{Z}/p\mathbf{Z})^*$. The order of a is a divisor d of the order $p-1$ of the group $(\mathbf{Z}/p\mathbf{Z})^*$. Write $d = (p-1)/n$, for some divisor n of $p-1$. If a is not a generator of $(\mathbf{Z}/p\mathbf{Z})^*$, then since $n \mid (p-1)$, there is a prime divisor p_i of $p-1$ such that $p_i \mid n$. Then

$$a^{(p-1)/p_i} = (a^{(p-1)/n})^{n/p_i} \equiv 1 \pmod{p}.$$

Conversely, if a is a generator, then $a^{(p-1)/p_i} \not\equiv 1 \pmod{p}$ for any p_i. Thus the algorithm terminates with Step 3 if and only if the a under consideration is a primitive root. By Theorem 2.5.8, there is at least one primitive root, so the algorithm terminates. □

2.6 Exercises

2.1 Prove that for any positive integer n, the set $(\mathbf{Z}/n\mathbf{Z})^*$ under multiplication modulo n is a group.

2.2 Compute the following gcd's using Algorithm 1.1.13:

$$\gcd(15,35) \quad \gcd(247,299) \quad \gcd(51,897) \quad \gcd(136,304)$$

2.3 Use Algorithm 2.3.7 to find x, $y \in \mathbf{Z}$ such that $2261x + 1275y = 17$.

2.4 Prove that if a and b are integers and p is a prime, then $(a+b)^p \equiv a^p + b^p \pmod{p}$. You may assume that the binomial coefficient

$$\frac{p!}{r!(p-r)!}$$

is an integer.

2.5 (a) Prove that if x, y is a solution to $ax + by = d$, then for all $c \in \mathbf{Z}$,

$$x' = x + c \cdot \frac{b}{d}, \qquad y' = y - c \cdot \frac{a}{d} \tag{2.6.1}$$

is also a solution to $ax + by = d$.

(b) Find two distinct solutions to $2261x + 1275y = 17$.

(c) Prove that all solutions are of the form (2.6.1) for some c.

2.6 Let $f(x) = x^2 + ax + b \in \mathbf{Z}[x]$ be a quadratic polynomial with integer coefficients, for example, $f(x) = x^2 + x + 6$. Formulate a conjecture about when the set

$$\{f(n) : n \in \mathbf{Z} \text{ and } f(n) \text{ is prime}\}$$

is infinite. Give numerical evidence that supports your conjecture.

2.7 Find four complete sets of residues modulo 7, where the ith set satisfies the ith condition: (1) nonnegative, (2) odd, (3) even, (4) prime.

2.8 Find rules in the spirit of Proposition 2.1.9 for divisibility of an integer by 5, 9, and 11, and prove each of these rules using arithmetic modulo a suitable n.

2.9 (*) *(The following problem is from the 1998 Putnam Competition.)* Define a sequence of decimal integers a_n as follows: $a_1 = 0$, $a_2 = 1$, and a_{n+2} is obtained by writing the digits of a_{n+1} immediately followed by those of a_n. For example, $a_3 = 10$, $a_4 = 101$, and $a_5 = 10110$. Determine the n such that a_n is a multiple of 11, as follows:

(a) Find the smallest integer $n > 1$ such that a_n is divisible by 11.

(b) Prove that a_n is divisible by 11 if and only if $n \equiv 1 \pmod{6}$.

2.10 Find an integer x such that $37x \equiv 1 \pmod{101}$.

2.11 What is the order of 2 modulo 17?

2.12 Let p be a prime. Prove that $\mathbf{Z}/p\mathbf{Z}$ is a field.

2.13 Find an $x \in \mathbf{Z}$ such that $x \equiv -4 \pmod{17}$ and $x \equiv 3 \pmod{23}$.

2.14 Prove that if $n > 4$ is composite then

$$(n-1)! \equiv 0 \pmod{n}.$$

2.15 For what values of n is $\varphi(n)$ odd?

2.16 (a) Prove that φ is multiplicative as follows. Suppose m, n are positive integers and $\gcd(m, n) = 1$. Show that the natural map $\psi : \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is an injective homomorphism of rings, hence bijective by counting, then look at unit groups.

(b) Prove conversely that if $\gcd(m, n) > 1$, then the natural map $\psi : \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is not an isomorphism.

2.17 Seven competitive math students try to share a huge hoard of stolen math books equally between themselves. Unfortunately, six books are left over, and in the fight over them, one math student is expelled. The remaining six math students, still unable to share the math books equally since two are left over, again fight, and another is expelled. When the remaining five share the books, one book is left over, and it is only after yet another math student is expelled that an equal sharing is possible. What is the minimum number of books that allows this to happen?

2.18 Show that if p is a positive integer such that both p and $p^2 + 2$ are prime, then $p = 3$.

2.19 Let $\varphi : \mathbf{N} \to \mathbf{N}$ be the Euler φ function.

(a) Find all natural numbers n such that $\varphi(n) = 1$.

(b) Do there exist natural numbers m and n such that $\varphi(mn) \neq \varphi(m) \cdot \varphi(n)$?

2.20 Find a formula for $\varphi(n)$ directly in terms of the prime factorization of n.

2.21 (a) Prove that if $\varphi : G \to H$ is a group homomorphism, then $\ker(\varphi)$ is a subgroup of G.

(b) Prove that $\ker(\varphi)$ is *normal*, i.e., if $a \in G$ and $b \in \ker(\varphi)$, then $a^{-1}ba \in \ker(\varphi)$.

2.22 Is the set $\mathbf{Z}/5\mathbf{Z} = \{0, 1, 2, 3, 4\}$ with binary operation multiplication modulo 5 a group?

2.23 Find all *four* solutions to the equation

$$x^2 - 1 \equiv 0 \pmod{35}.$$

2.24 Prove that for any positive integer n the fraction $(12n+1)/(30n+2)$ is in reduced form.

2.25 Suppose a and b are positive integers.

(a) Prove that $\gcd(2^a - 1,\ 2^b - 1) = 2^{\gcd(a,b)} - 1$.

(b) Does it matter if 2 is replaced by an arbitrary prime p?

(c) What if 2 is replaced by an arbitrary positive integer n?

2.26 For every positive integer b, show that there exists a positive integer n such that the polynomial $x^2 - 1 \in (\mathbf{Z}/n\mathbf{Z})[x]$ has at least b roots.

2.27 (a) Prove that there is no primitive root modulo 2^n for any $n \geq 3$.

(b) (*) Prove that $(\mathbf{Z}/2^n\mathbf{Z})^*$ is generated by -1 and 5.

2.28 Let p be an odd prime.

(a) (*) Prove that there is a primitive root modulo p^2. (Hint: Use that if a, b have orders n, m, with $\gcd(n, m) = 1$, then ab has order nm.)

(b) Prove that for any n, there is a primitive root modulo p^n.

(c) Explicitly find a primitive root modulo 125.

2.29 (*) In terms of the prime factorization of n, characterize the integers n such that there is a primitive root modulo n.

2.30 Compute the last two digits of 3^{45}.

2.31 Find the integer a such that $0 \leq a < 113$ and

$$102^{70} + 1 \equiv a^{37} \pmod{113}.$$

2.32 Find the proportion of primes $p < 1000$ such that 2 is a primitive root modulo p.

2.33 Find a prime p such that the smallest primitive root modulo p is 37.

3
Public-key Cryptography

In the 1970s, techniques from number theory changed the world forever by providing, for the first time ever, a way for two people to communicate secret messages under the assumption that *all* of their communication is intercepted and read by an adversary. This idea has stood the test of time. In fact, whenever you buy something online, you use such a system, which typically involves working in the ring of integers modulo n. This chapter tells the story of several such systems.

3.1 Playing with Fire

I recently watched a TV show called La Femme Nikita about a woman named Nikita who is forced to be an agent for a shady anti-terrorist organization called Section One. Nikita has strong feelings for fellow agent Michael, and she most trusts Walter, Section One's ex-biker gadgets and explosives expert. Often Nikita's worst enemies are her superiors and coworkers at Section One. A synopsis for a Season Three episode is as follows:

> PLAYING WITH FIRE
>
> On a mission to secure detonation chips from a terrorist organization's heavily armed base camp, Nikita is captured as a hostage by the enemy. Or so it is made to look. Michael and Nikita have actually created the scenario in order to secretly rendezvous with each other. The ruse works, but when Birkoff

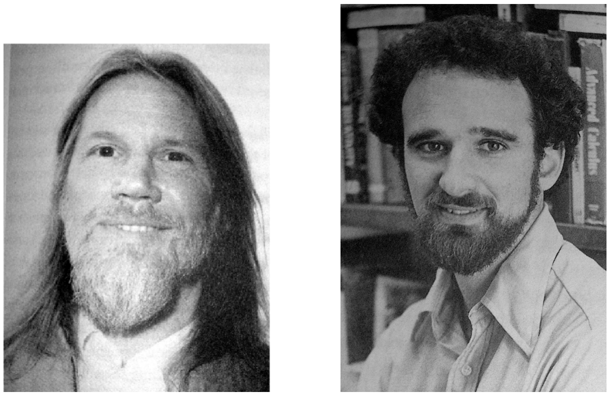


FIGURE 3.1. Diffie and Hellman (photos from [Sin99])

> [Section One's master hacker] accidentally discovers encrypted messages between Michael and Nikita sent with Walter's help, Birkoff is forced to tell Madeline. Suspecting that Michael and Nikita may be planning a coup d'état, Operations and Madeline use a second team of operatives to track Michael and Nikita's next secret rendezvous... killing them if necessary.

What sort of encryption might Walter have helped them to use? I let my imagination run free, and this is what I came up with. After being captured at the base camp, Nikita is given a phone by her captors in hopes that she'll use it and they'll be able to figure out what she is really up to. Everyone is eagerly listening in on her calls.

Remark 3.1.1. In this book, we will assume a method is available for producing random integers. Methods for generating random integers are involved and interesting, but we will not discuss them in this book. For an in-depth treatment of random numbers, see [Knu98, Ch. 3].

Nikita remembers a conversation with Walter about a public-key cryptosystem called the "Diffie-Hellman key exchange." She remembers that it allows two people to agree on a secret key in the presence of eavesdroppers. Moreover, Walter mentioned that though Diffie-Hellman was the first ever public-key exchange system, it is still in common use today (for example, in OpenSSH protocol version 2, see `http://www.openssh.com/`).

Nikita pulls out her handheld computer and phone, calls up Michael, and they do the following, which is *wrong* (try to figure out what is wrong as you read it).

1. Together they choose a big prime number p and a number g with $1 < g < p$.

2. Nikita *secretly* chooses an integer n.

3. Michael *secretly* chooses an integer m.
4. Nikita tells Michael $ng \pmod{p}$.
5. Michael tells $mg \pmod{p}$ to Nikita.
6. The “secret key” is $s = nmg \pmod{p}$, which both Nikita and Michael can easily compute.

Here’s a very simple example with small numbers that illustrates what Michael and Nikita do. (They really used much larger numbers.)

1. $p = 97$, $g = 5$
2. $n = 31$
3. $m = 95$
4. $ng \equiv 58 \pmod{97}$
5. $mg \equiv 87 \pmod{97}$
6. $s = nmg = 78 \pmod{97}$

Nikita and Michael are foiled because everyone easily figures out s:

1. Everyone knows p, g, $ng \pmod{p}$, and $mg \pmod{p}$.
2. Using Algorithm 2.3.7, anyone can easily find $a, b \in \mathbf{Z}$ such that $ag + bp = 1$, which exists because $\gcd(g, p) = 1$.
3. Then, $ang \equiv n \pmod{p}$, so everyone knows Nikita’s secret key n, and hence can easily compute the shared secret s.

To taunt her, Nikita’s captors give her a paragraph from a review of Diffie and Hellman’s 1976 paper “New Directions in Cryptography” [DH76]:

> “The authors discuss some recent results in communications theory [...] The first [method] has the feature that an unauthorized ‘eavesdropper’ will find it computationally infeasible to decipher the message [...] They propose a couple of techniques for implementing the system, but the reviewer was unconvinced.”

3.2 The Diffie-Hellman Key Exchange

As night darkens Nikita’s cell, she reflects on what has happened. Upon realizing that she mis-remembered how the system works, she phones Michael and they do the following:

1. Together Michael and Nikita choose a 200-digit integer p that is likely to be prime (see Section 2.4), and choose a number g with $1 < g < p$.
2. Nikita *secretly* chooses an integer n.
3. Michael *secretly* chooses an integer m.
4. Nikita computes $g^n \pmod{p}$ on her handheld computer and tells Michael the resulting number over the phone.
5. Michael tells Nikita $g^m \pmod{p}$.
6. The shared secret key is then
$$s \equiv (g^n)^m \equiv (g^m)^n \equiv g^{nm} \pmod{p},$$
which both Nikita and Michael can compute.

Here is a simplified example that illustrates what they did, that involves only relatively simple arithmetic.

1. $p = 97$, $g = 5$
2. $n = 31$
3. $m = 95$
4. $g^n \equiv 7 \pmod{p}$
5. $g^m \equiv 39 \pmod{p}$
6. $s \equiv (g^n)^m \equiv 14 \pmod{p}$

3.2.1 The Discrete Log Problem

Nikita communicates with Michael by encrypting everything using their agreed upon secret key (for example, using a standard symmetric cipher such as AES, Arcfour, Cast128, 3DES, or Blowfish). In order to understand the conversation, the eavesdropper needs s, but it takes a long time to compute s given only p, g, g^n, and g^m. One way would be to compute n from knowledge of g and g^n; this is possible, but appears to be "computationally infeasible," in the sense that it would take too long to be practical.

Let a, b, and n be real numbers with $a, b > 0$ and $n \geq 0$. Recall that the "log to the base b" function is characterized by
$$\log_b(a) = n \text{ if and only if } a = b^n.$$
We use the $\log_b$ function in algebra to solve the following problem: Given a base b and a power a of b, find an exponent n such that
$$a = b^n.$$

That is, given $a = b^n$ and b, find n.

SAGE Example 3.2.1. The number $a = 19683$ is the nth power of $b = 3$ for some n. We quickly find that

$$n = \log_3(19683) = \log(19683)/\log(3) = 9.$$

```
sage: log(19683.0)
9.88751059801299
sage: log(3.0)
1.09861228866811
sage: log(19683.0) / log(3.0)
9.00000000000000
```

Sage can quickly compute a numerical approximation for $\log(x)$, for any x, by computing a partial sum of an appropriate rapidly-converging infinite series (at least for x in a certain range).

The discrete log problem is the analog of computing $\log_b(a)$ but where both b and a are elements of a finite group.

Problem 3.2.2 (Discrete Log Problem)**.** Let G be a finite group, for example, $G = (\mathbf{Z}/p\mathbf{Z})^*$. Given $b \in G$ and a power a of b, find a positive integer n such that $b^n = a$.

As far as we know, finding discrete logarithms in $(\mathbf{Z}/p\mathbf{Z})^*$ when p is large is "very difficult" in practice. Over the years, many people have been very motivated to try. For example, if Nikita's captors could efficiently solve Problem 3.2.2, then they could read the messages she exchanges with Michael. Unfortunately, we have no formal proof that computing discrete logarithms on a classical computer is difficult. Also, Peter Shor [Sho97] showed that if one could build a sufficiently complicated quantum computer, it could solve the discrete logarithm problem in time bounded by a polynomial function of the number of digits of $\#G$.

It is easy to give an inefficient algorithm that solves the discrete log problem. Simply try b^1, b^2, b^3, etc., until we find an exponent n such that $b^n = a$. For example, suppose $a = 18$, $b = 5$, and $p = 23$. Working modulo 23, we have

$$b^1 = 5,\ b^2 = 2,\ b^3 = 10,\ \ldots,\ b^{12} = 18,$$

so $n = 12$. When p is large, computing the discrete log this way soon becomes impractical, because increasing the number of digits of the modulus makes the computation take vastly longer.

SAGE Example 3.2.3. Perhaps part of the reason that computing discrete logarithms is difficult, is that the logarithm in the real numbers is continuous, but the (minimum) logarithm of a number mod n bounces around at random. We illustrate this exotic behavior in Figure 3.2.

This draws the continuous plot.

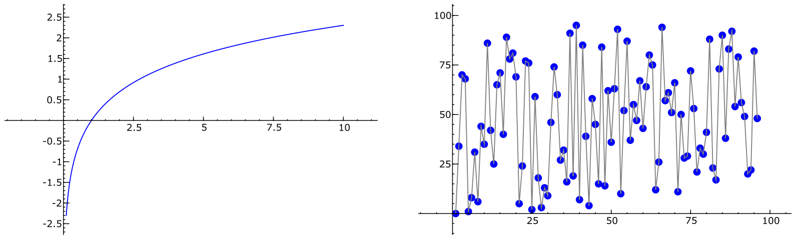


FIGURE 3.2. Graphs of the continuous log and of the discrete log modulo 53. Which picture looks easier to predict?

```
sage: plot(log, 0.1,10, rgbcolor=(0,0,1))
```

This draws the discrete plot.

```
sage: p = 53
sage: R = Integers(p)
sage: a = R.multiplicative_generator()
sage: v = sorted([(a^n, n) for n in range(p-1)])
sage: G = plot(point(v,pointsize=50,rgbcolor=(0,0,1)))
sage: H = plot(line(v,rgbcolor=(0.5,0.5,0.5)))
sage: G + H
```

3.2.2 Realistic Diffie-Hellman Example

In this section, we present an example that uses bigger numbers. First, we prove a proposition that we can use to choose a prime p in such a way that it is easy to find a $g \in (\mathbf{Z}/p\mathbf{Z})^*$ with order $p-1$. We have already seen in Section 2.5 that for every prime p there exists an element g of order $p-1$, and we gave Algorithm 2.5.16 for finding a primitive root for any prime. The significance of Proposition 3.2.4 below is that it suggests an algorithm for finding a primitive root that is easier to use in practice when p is large, because it does not require factoring $p-1$. Of course, one could also just use a random g for Diffie-Hellman; it is not essential that g generates $(\mathbf{Z}/p\mathbf{Z})^*$.

Proposition 3.2.4. *Suppose p is a prime such that $(p-1)/2$ is also prime. Then each element of $(\mathbf{Z}/p\mathbf{Z})^*$ has order one of 1, 2, $(p-1)/2$, or $p-1$.*

Proof. Since p is prime, the group $(\mathbf{Z}/p\mathbf{Z})^*$ is of order $p-1$. By assumption, the prime factorization of $p-1$ is $2 \cdot ((p-1)/2)$. Let $a \in (\mathbf{Z}/p\mathbf{Z})^*$. Then by Theorem 2.1.20, $a^{p-1} = 1$, so the order of a is a divisor of $p-1$, which proves the proposition. □

Given a prime p with $(p-1)/2$ prime, find an element of order $p-1$ as follows. If 2 has order $p-1$, we are done. If not, 2 has order $(p-1)/2$ since 2 does not have order either 1 or 2. Then -2 has order $p-1$.

Let $p = 93450983094850938450983409611$. Then p is prime, but $(p-1)/2$ is not. So we keep adding 2 to p and testing pseudoprimality using algorithms from Section 2.4 until we find that the next pseudoprime after p is

$$q = 93450983094850938450983409623.$$

It turns out that q pseudoprime and $(q-1)/2$ is also pseudoprime. We find that 2 has order $(q-1)/2$, so $g = -2$ has order $q-1$ modulo q, and is hence a generator of $(\mathbf{Z}/q\mathbf{Z})^*$, at least assuming that q is really prime.

The secret random numbers generated by Nikita and Michael are

$$n = 18319922375531859171613379181$$

and

$$m = 82335836243866695680141440300.$$

Nikita sends

$$g^n = 45416776270485369791375944998 \in (\mathbf{Z}/p\mathbf{Z})^*$$

to Michael, and Michael sends

$$g^m = 15048074151770884271824225393 \in (\mathbf{Z}/p\mathbf{Z})^*$$

to Nikita. They agree on the secret key

$$g^{nm} = 85771409470770521212346739540 \in (\mathbf{Z}/p\mathbf{Z})^*.$$

SAGE Example 3.2.5. We illustrate the above computations using Sage.

```
sage: q = 93450983094850938450983409623
sage: q.is_prime()
True
sage: is_prime((q-1)//2)
True
sage: g = Mod(-2, q)
sage: g.multiplicative_order()
93450983094850938450983409622
sage: n = 18319922375531859171613379181
sage: m = 82335836243866695680141440300
sage: g^n
45416776270485369791375944998
sage: g^m
15048074151770884271824225393
sage: (g^n)^m
85771409470770521212346739540
sage: (g^m)^n
85771409470770521212346739540
```

3.2.3 The Man in the Middle Attack

Since their first system was broken, instead of talking on the phone, Michael and Nikita can now only communicate via text messages. One of her captors, The Man, is watching each of the transmissions; moreover, he can intercept messages and send false messages. When Nikita sends a message to Michael announcing $g^n \pmod{p}$, The Man intercepts this message, and sends his own number $g^t \pmod{p}$ to Michael. Eventually, Michael and The Man agree on the secret key $g^{tm} \pmod{p}$, and Nikita and The Man agree on the key $g^{tn} \pmod{p}$. When Nikita sends a message to Michael she unwittingly uses the secret key $g^{tn} \pmod{p}$; The Man then intercepts it, decrypts it, changes it, and re-encrypts it using the key $g^{tm} \pmod{p}$, and sends it on to Michael. This is bad because now The Man can read every message sent between Michael and Nikita, and moreover, he can change them in transmission in subtle ways.

One way to get around this attack is to use a digital signature scheme based on the RSA cryptosystem. We will not discuss digital signatures further in this book, but will discuss RSA in the next section.

3.3 The RSA Cryptosystem

The Diffie-Hellman key exchange has drawbacks. As discussed in Section 3.2.3, it is susceptible to the man in the middle attack. This section is about the RSA public-key cryptosystem of Rivest, Shamir, and Adleman [RSA78], which is an alternative to Diffie-Hellman that is more flexible in some ways.

We first describe the RSA cryptosystem, then discuss several ways to attack it. It is important to be aware of such weaknesses, in order to avoid foolish mistakes when implementing RSA. We barely scratched the surface here of the many possible attacks on specific implementations of RSA or other cryptosystems.

3.3.1 How RSA works

The fundamental idea behind RSA is to try to construct a trap-door or one-way function on a set X. This is an invertible function

$$E : X \to X$$

such that it is easy for Nikita to compute E^{-1}, but extremely difficult for anybody else to do so.

Here is how Nikita makes a one-way function E on the set of integers modulo n.

1. Using a method hinted at in Section 2.4, Nikita picks two large primes p and q, and lets $n = pq$.

2. It is then easy for Nikita to compute
$$\varphi(n) = \varphi(p) \cdot \varphi(q) = (p-1) \cdot (q-1).$$
3. Nikita next chooses a random integer e with
$$1 < e < \varphi(n) \text{ and } \gcd(e, \varphi(n)) = 1.$$
4. Nikita uses the algorithm from Section 2.3.2 to find a solution $x = d$ to the equation
$$ex \equiv 1 \pmod{\varphi(n)}.$$
5. Finally, Nikita defines a function $E : \mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$ by
$$E(x) = x^e \in \mathbf{Z}/n\mathbf{Z}.$$

Note that anybody can compute E fairly quickly using the repeated-squaring algorithm from Section 2.3.2. Nikita's *public key* is the pair of integers (n, e), which is just enough information for people to easily compute E. Nikita knows a number d such that $ed \equiv 1 \pmod{\varphi(n)}$, so, as we will see, she can quickly compute E^{-1}.

To send Nikita a message, proceed as follows. Encode your message, in some way, as a sequence of numbers modulo n (see Section 3.3.2)
$$m_1, \ldots, m_r \in \mathbf{Z}/n\mathbf{Z},$$
then send
$$E(m_1), \ldots, E(m_r)$$
to Nikita. (Recall that $E(m) = m^e$ for $m \in \mathbf{Z}/n\mathbf{Z}$.)

When Nikita receives $E(m_i)$, she finds each m_i by using that $E^{-1}(m) = m^d$, a fact that follows from Proposition 3.3.1

Proposition 3.3.1 (Decryption Key)**.** *Let n be an integer that is a product of distinct primes and let $d, e \in \mathbf{N}$ be such that $p-1 \mid de-1$ for each prime $p \mid n$. Then $a^{de} \equiv a \pmod{n}$ for all $a \in \mathbf{Z}$.*

Proof. Since $n \mid a^{de} - a$, if and only if $p \mid a^{de} - a$ for each prime divisor p of n, it suffices to prove that $a^{de} \equiv a \pmod{p}$ for each prime divisor p of n. If $\gcd(a, p) \neq 1$, then $a \equiv 0 \pmod{p}$, so $a^{de} \equiv a \pmod{p}$. If $\gcd(a, p) = 1$, then Theorem 2.1.20 asserts that $a^{p-1} \equiv 1 \pmod{p}$. Since $p - 1 \mid de - 1$, we have $a^{de-1} \equiv 1 \pmod{p}$ as well. Multiplying both sides by a shows that $a^{de} \equiv a \pmod{p}$. □

Thus to decrypt $E(m_i)$ Nikita computes
$$E(m_i)^d = (m_i^e)^d = m_i.$$

SAGE Example 3.3.2. We implement the RSA cryptosystem using Sage. The rsa function creates a key with (at most) the given number of bits, i.e., if bits equals 20, it creates a key $n = pq$ such that n is approximately 2^{20}. Typical real-life cryptosystems would choose keys that are 512, 1024, or 2048 bits long. Try generating large keys yourself using Sage; how long does it take?

```
sage: def rsa(bits):
...     # only prove correctness up to 1024 bits
...     proof = (bits <= 1024)
...     p = next_prime(ZZ.random_element(2**(bits//2 +1)),
...                 proof=proof)
...     q = next_prime(ZZ.random_element(2**(bits//2 +1)),
...                 proof=proof)
...     n = p * q
...     phi_n = (p-1) * (q-1)
...     while True:
...         e = ZZ.random_element(1,phi_n)
...         if gcd(e,phi_n) == 1: break
...     d = lift(Mod(e,phi_n)^(-1))
...     return e, d, n
...
sage: def encrypt(m,e,n):
...     return lift(Mod(m,n)^e)
...
sage: def decrypt(c,d,n):
...     return lift(Mod(c,n)^d)
...
sage: e,d,n = rsa(20)
sage: c = encrypt(123, e, n)
sage: decrypt(c, d, n)
123
```

3.3.2 Encoding a Phrase in a Number

In order to use the RSA cryptosystem to encrypt messages, it is necessary to encode them as a sequence of numbers of size less than $n = pq$. We now describe a simple way to do this. Note that in any actual deployed implementation, it is crucial that you add extra random characters ("salt") at the beginning of each block of the message, so that the same plain text encodes differently each time. This helps thwart chosen plain text attacks.

Suppose s is a sequence of capital letters and spaces, and that s does not begin with a space. We encode s as a number in base 27 as follows: a single space corresponds to 0, the letter A to 1, B to 2, …, Z to 26. Thus "RUN

NIKITA” is a number written in base 27.

$$\begin{aligned}\text{RUN NIKITA} \quad \leftrightarrow \quad & 27^9 \cdot 18 + 27^8 \cdot 21 + 27^7 \cdot 14 + 27^6 \cdot 0 + 27^5 \cdot 14 \\ & + 27^4 \cdot 9 + 27^3 \cdot 11 + 27^2 \cdot 9 + 27 \cdot 20 + 1 \\ & = 143338425831991 \text{ (in decimal).}\end{aligned}$$

To recover the letters from the decimal number, repeatedly divide by 27 and read off the letter corresponding to each remainder.

$$\begin{array}{rcrcrc}
143338425831991 & = & 5308830586370 \cdot 27 & + & 1 & \text{“A”} \\
5308830586370 & = & 196623355050 \cdot 27 & + & 20 & \text{“T”} \\
196623355050 & = & 7282346483 \cdot 27 & + & 9 & \text{“I”} \\
7282346483 & = & 269716536 \cdot 27 & + & 11 & \text{“K”} \\
269716536 & = & 9989501 \cdot 27 & + & 9 & \text{“I”} \\
9989501 & = & 369981 \cdot 27 & + & 14 & \text{“N”} \\
369981 & = & 13703 \cdot 27 & + & 0 & \text{“ ”} \\
13703 & = & 507 \cdot 27 & + & 14 & \text{“N”} \\
507 & = & 18 \cdot 27 & + & 21 & \text{“U”} \\
18 & = & 0 \cdot 27 & + & 18 & \text{“R”}
\end{array}$$

If $27^k \leq n$, then any sequence of k letters can be encoded as above using a positive integer $\leq n$. Thus if we can encrypt integers of size at most n, then we must break our message up into blocks of size at most $\log_{27}(n)$.

SAGE Example 3.3.3. We use Sage to implement conversion between a string and a number, though in a bit more generally than in the toy illustration above (which used only base 27). The input string s on a computer is stored in a format called ASCII, so each “letter” corresponds to an integer between 0 and 255, inclusive. This number is obtained from the letter using the `ord` command.

```
sage: def encode(s):
...     s = str(s)      # make input a string
...     return sum(ord(s[i])*256^i for i in range(len(s)))
sage: def decode(n):
...     n = Integer(n)  # make input an integer
...     v = []
...     while n != 0:
...         v.append(chr(n % 256))
...         n //= 256    # this replaces n by floor(n/256).
...     return ''.join(v)
sage: m = encode('Run Nikita!'); m
40354769014714649421968722
sage: decode(m)
'Run Nikita!'
```

3.3.3 Some Complete Examples

To make the arithmetic easier to follow, we use small prime numbers p and q and encrypt the single letter "X" using the RSA cryptosystem. First, we compute the parameters of an RSA cryptosystem.

1. Choose p and q: Let $p = 17$, $q = 19$, so $n = pq = 323$.
2. Compute $\varphi(n)$:
$$\begin{aligned}\varphi(n) &= \varphi(p \cdot q) = \varphi(p) \cdot \varphi(q) = (p-1)(q-1) \\ &= pq - p - q + 1 = 323 - 17 - 19 + 1 = 288.\end{aligned}$$
3. Randomly choose an $e < 288$: We choose $e = 95$.
4. Solve
$$95x \equiv 1 \pmod{288}.$$
Using the GCD algorithm, we find that $d = 191$ solves the equation.

We have thus computed the parameters of an RSA public key cryptosystem. The public key is $(323, 95)$, so the encryption function is

$$E(x) = x^{95},$$

and the decryption function is $D(x) = x^{191}$.

Next, we encrypt the letter "X". It is encoded as the number 24, since X is the 24th letter of the alphabet. We have

$$E(24) = 24^{95} = 294 \in \mathbf{Z}/323\mathbf{Z}.$$

To decrypt, we compute E^{-1}:

$$E^{-1}(294) = 294^{191} = 24 \in \mathbf{Z}/323\mathbf{Z}.$$

This next example illustrates RSA but with bigger numbers. Let

$$p = 738873402423833494183027176953, \; q = 3787776806865662882378273.$$

Then,

$$n = p \cdot q = 2798687536910915970127263606347911460948554197853542169$$

and

$$\begin{aligned}\varphi(n) &= (p-1)(q-1) \\ &= 2798687536910915970127262867470721260308194351943986944.\end{aligned}$$

Using a pseudo-random number generator on a computer, the author randomly chose the integer

$$e = 148395919486620417934853601028471665544213902491572069 9.$$

Then,

$$d = 2113367928496305469541348387088632973457802358781610803$$

Since $\log_{27}(n) \approx 38.04$, we can encode then encrypt single blocks of up to 38 letters. Let's encrypt the string `RUN NIKITA`, which encodes as $m = 143338425831991$. We have

$$\begin{aligned} E(m) &= m^e \\ &= 1504554432996568133393088878600948101773726800878873990. \end{aligned}$$

Remark 3.3.4. In practice, one usually choses e to be small, since that does not seem to reduce the security of RSA, and makes the key size smaller. For example, in the OpenSSL documentation (see `http://www.openssl.org/`) about their implementation of RSA, it states that "The exponent is an odd number, typically 3, 17 or 65537."

3.4 Attacking RSA

Suppose Nikita's public key is (n, e) and her decryption key is d, so $ed \equiv 1 \pmod{\varphi(n)}$. If somehow we compute the factorization $n = pq$, then we can compute $\varphi(n) = (p-1)(q-1)$ and hence compute d. Thus, if we can factor n then we can break the corresponding RSA public-key cryptosystem.

3.4.1 Factoring n Given $\varphi(n)$

Suppose $n = pq$. Given $\varphi(n)$, it is very easy to compute p and q. We have

$$\varphi(n) = (p-1)(q-1) = pq - (p+q) + 1,$$

so we know both $pq = n$ and $p + q = n + 1 - \varphi(n)$. Thus, we know the polynomial

$$x^2 - (p+q)x + pq = (x-p)(x-q)$$

whose roots are p and q. These roots can be found using the quadratic formula.

Example 3.4.1. The number $n = pq = 31615577110997599711$ is a product of two primes, and $\varphi(n) = 31615577098574867424$. We have

$$\begin{aligned} f &= x^2 - (n + 1 - \varphi(n))x + n \\ &= x^2 - 12422732288x + 31615577110997599711 \\ &= (x - 3572144239)(x - 8850588049), \end{aligned}$$

where the factorization step is easily accomplished using the quadratic formula:

$$\begin{aligned}\frac{-b+\sqrt{b^2-4ac}}{2a}\\ &= \frac{12422732288+\sqrt{12422732288^2-4\cdot 31615577110997599711}}{2}\\ &= 8850588049.\end{aligned}$$

We conclude that $n = 3572144239 \cdot 8850588049$.

SAGE Example 3.4.2. The following Sage function factors $n = pq$ given n and $\varphi(n)$.

```
sage: def crack_rsa(n, phi_n):
...     R.<x> = PolynomialRing(QQ)
...     f = x^2 - (n+1 -phi_n)*x + n
...     return [b for b, _ in f.roots()]
sage: crack_rsa(31615577110997599711, 31615577098574867424)
[8850588049, 3572144239]
```

3.4.2 When p and q are Close

Suppose that p and q are "close" to each other. Then it is easy to factor n using a factorization method of Fermat called the *Fermat Factorization Method*.

Suppose $n = pq$ with $p > q$. Then,

$$n = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2.$$

Since p and q are "close,"

$$s = \frac{p-q}{2}$$

is small,

$$t = \frac{p+q}{2}$$

is only slightly larger than $\sqrt{n}$, and $t^2 - n = s^2$ is a perfect square. So, we just try

$$t = \lceil\sqrt{n}\rceil, \quad t = \lceil\sqrt{n}\rceil + 1, \quad t = \lceil\sqrt{n}\rceil + 2, \ldots$$

until $t^2 - n$ is a perfect square s^2. (Here $\lceil x\rceil$ denotes the least integer $n \geq x$.) Then

$$p = t + s, \qquad q = t - s.$$

Example 3.4.3. Suppose $n = 23360947609$. Then

$$\sqrt{n} = 152842.88\ldots.$$

If $t = 152843$, then $\sqrt{t^2 - n} = 187.18\ldots$.
If $t = 152844$, then $\sqrt{t^2 - n} = 583.71\ldots$.
If $t = 152845$, then $\sqrt{t^2 - n} = 804 \in \mathbf{Z}$.
Thus $s = 804$. We find that $p = t + s = 153649$ and $q = t - s = 152041$.

SAGE Example 3.4.4. We implement the above algorithm for factoring an RSA modulus $n = pq$, when one of p and q is close to $\sqrt{n}$.

```
sage: def crack_when_pq_close(n):
...     t = Integer(ceil(sqrt(n)))
...     while True:
...        k = t^2 - n
...        if k > 0:
...           s = Integer(int(round(sqrt(t^2 - n))))
...           if s^2 + n == t^2:
...              return t+s, t-s
...
...        t += 1
...
sage: crack_when_pq_close(23360947609)
(153649, 152041)
```

For example, you might think that choosing a random prime, and the next prime after would be a good idea, but instead it creates an easy-to-crack cryptosystem.

```
sage: p = next_prime(2^128); p
340282366920938463463374607431768211507
sage: q = next_prime(p)
sage: crack_when_pq_close(p*q)
(340282366920938463463374607431768211537,
      340282366920938463463374607431768211507)
```

3.4.3 Factoring n Given d

In this section, we show that finding the decryption key d for an RSA cryptosystem is, in practice, at least as difficult as factoring n. We give a probabilistic algorithm that given a decryption key determines the factorization of n.

Consider an RSA cryptosystem with modulus n and encryption key e. Suppose we somehow finding an integer d such that

$$a^{ed} \equiv a \pmod{n}$$

for all a. Then $m = ed - 1$ satisfies $a^m \equiv 1 \pmod{n}$ for all a that are coprime to n. As we saw in Section 3.4.1, knowing $\varphi(n)$ leads directly to a factorization of n. Unfortunately, knowing d does not seem to lead easily to a factorization of n. However, there is a probabilistic procedure that, given an m such that $a^m \equiv 1 \pmod{n}$, will find a factorization of n with "high probability" (we will not analyze the probability here).

Algorithm 3.4.5 (Probabilistic Algorithm to Factor n)**.** Let $n = pq$ be the product of two distinct odd primes, and suppose m is an integer such that $a^m \equiv 1 \pmod{n}$ for all a coprime to n. This probabilistic algorithm factors n with "high probability." In the steps below, a always denotes an integer coprime to $n = pq$.

1. [Divide out powers of 2] If m is even and $a^{m/2} \equiv 1 \pmod{n}$ for several randomly chosen a, set $m = m/2$, and go to Step 1, otherwise let a be such that $a^{m/2} \not\equiv 1 \pmod{n}$.
2. [Compute GCD] Choose a random a and compute $g = \gcd(a^{m/2} - 1, n)$.
3. [Terminate?] If g is a proper divisor of n, output g and terminate. Otherwise go to Step 2.

Before giving the proof, we introduce some more terminology from algebra.

Definition 3.4.6 (Group Homomorphism)**.** Let G and H be groups. A map $\varphi : G \to H$ is a *group homomorphism* if for all $a, b \in G$ we have $\varphi(ab) = \varphi(a)\varphi(b)$. A group homomorphism is called *surjective* if for every $c \in H$ there is $a \in G$ such that $\varphi(a) = c$. The *kernel* of a group homomorphism $\varphi : G \to H$ is the set $\ker(\varphi)$ of elements $a \in G$ such that $\varphi(a) = 1$. A group homomorphism is *injective* if $\ker(\varphi) = \{1\}$.

Definition 3.4.7 (Subgroup)**.** If G is a group and H is a subset of G, then H is a *subgroup* if H is a group under the group operation on G.

For example, if $\varphi : G \to H$ is a group homomorphism, then $\ker(\varphi)$ is a subgroup of G (see Exercise 2.21).

We now return to discussing Algorithm 3.4.5. In Step 1, note that m is even since $(-1)^m \equiv 1 \pmod{n}$, so it makes sense to consider $m/2$. It is not practical to determine whether or not $a^{m/2} \equiv 1 \pmod{n}$ for all a, because it would require doing a computation for too many a. Instead, we try a few random a; if $a^{m/2} \equiv 1 \pmod{n}$ for the a we check, we divide m by 2. Also note that if there exists even a single a such that $a^{m/2} \not\equiv 1 \pmod{n}$, then half the a have this property, since then $a \mapsto a^{m/2}$ is a surjective homomorphism $(\mathbf{Z}/n\mathbf{Z})^* \to \{\pm 1\}$ and the kernel has index 2.

Proposition 2.5.3 implies that if $x^2 \equiv 1 \pmod{p}$ then $x = \pm 1 \pmod{p}$. In Step 2, since $(a^{m/2})^2 \equiv 1 \pmod{n}$, we also have $(a^{m/2})^2 \equiv 1 \pmod{p}$ and $(a^{m/2})^2 \equiv 1 \pmod{q}$, so $a^{m/2} \equiv \pm 1 \pmod{p}$ and $a^{m/2} \equiv \pm 1 \pmod{q}$. Since $a^{m/2} \not\equiv 1 \pmod{n}$, there are three possibilities for these signs, so with positive probability one of the following two possibilities occurs:

1. $a^{m/2} \equiv +1 \pmod{p}$ and $a^{m/2} \equiv -1 \pmod{q}$

2. $a^{m/2} \equiv -1 \pmod{p} \qquad \text{and} \qquad a^{m/2} \equiv +1 \pmod{q}.$

The only other possibility is that both signs are -1. In the first case,

$$p \mid a^{m/2} - 1 \qquad \text{but} \qquad q \nmid a^{m/2} - 1,$$

so $\gcd(a^{m/2} - 1, pq) = p$, and we have factored n. Similarly, in the second case, $\gcd(a^{m/2} - 1, pq) = q$, and we again factor n.

Example 3.4.8. Somehow we discover that the RSA cryptosystem with

$$n = 32295194023343 \qquad \text{and} \qquad e = 29468811804857$$

has decryption key $d = 11127763319273$. We use this information and Algorithm 3.4.5 to factor n. If

$$m = ed - 1 = 327921963064646896263108960,$$

then $\varphi(pq) \mid m$, so $a^m \equiv 1 \pmod{n}$ for all a coprime to n. For each $a \leq 20$ we find that $a^{m/2} \equiv 1 \pmod{n}$, so we replace m with

$$\frac{m}{2} = 163960981532323448131554480.$$

Again, we find with this new m that for each $a \leq 20$, $a^{m/2} \equiv 1 \pmod{n}$, so we replace m by 81980490766161724065777240. Yet again, for each $a \leq 20$, $a^{m/2} \equiv 1 \pmod{n}$, so we replace m by 40990245383080862032888620. This is enough, since $2^{m/2} \equiv 4015382800099 \pmod{n}$. Then,

$$\gcd(2^{m/2} - 1, n) = \gcd(4015382800098, 32295194023343) = 737531,$$

and we have found a factor of n. Dividing, we find that

$$n = 737531 \cdot 43788253.$$

SAGE Example 3.4.9. We implement Algorithm 3.4.5 in Sage.

```
sage: def crack_given_decrypt(n, m):
...     n = Integer(n); m = Integer(m);  # some type checking
...     # Step 1: divide out powers of 2
...     while True:
...         if is_odd(m): break
...         divide_out = True
...         for i in range(5):
...            a = randrange(1,n)
...            if gcd(a,n) == 1:
...               if Mod(a,n)^(m//2) != 1:
...                   divide_out = False
...                   break
...         if divide_out:
...             m = m//2
...         else:
...             break
...     # Step 2: Compute GCD
...     while True:
...         a = randrange(1,n)
...         g = gcd(lift(Mod(a, n)^(m//2)) - 1, n)
...         if g != 1 and g != n:
...             return g
...
```

We show how to verify Example 3.4.8 using Sage.

```
sage: n=32295194023343; e=29468811804857; d=11127763319273
sage: crack_given_decrypt(n, e*d - 1)
737531
sage: factor(n)
737531 * 43788253
```

We try a much larger example.

```
sage: e = 22601762315966221465875845336488389513
sage: d = 31940292321834506197902778067109010093
sage: n = 268494924039590992469444675130990465673
sage: p = crack_given_decrypt(n, e*d - 1)
sage: p   # random output (could be other prime divisor)
13432418150982799907
sage: n % p
0
```

3.4.4 Further Remarks

If one were to implement an actual RSA cryptosystem, there are many additional tricks and ideas to keep in mind. For example, one can add some extra random letters to each block of text, so that a given string will encrypt differently each time it is encrypted. This makes it more difficult for an attacker who knows the encrypted and plaintext versions of one message to gain information about subsequent encrypted messages. In any particular implementation, there might be attacks that would be devastating in practice, but which would not require factorization of the RSA modulus.

RSA is in common use, for example, it is used in OpenSSH protocol version 1 (see `http://www.openssh.com/`).

We will consider the ElGamal cryptosystem in Sections 6.4.2. It has a similar flavor to RSA, but is more flexible in some ways.

Probably the best general purpose attack on RSA is the number field sieve, which is a general algorithm for factoring integers of the form pq. A description of the sieve is beyond the scope of this book. The elliptic curve method is another related general algorithm that we will discuss in detail in Section 6.3.

SAGE Example 3.4.10. Here is a simple example of using a variant of the number field sieve (called the quadratic sieve) in Sage to factor an RSA key with about 192 bits:

```
sage: set_random_seed(0)
sage: p = next_prime(randrange(2^96))
sage: q = next_prime(randrange(2^97))
sage: n = p * q
sage: qsieve(n)
([6340271405786663791648052309,
  46102313108592180286398757159], '')
```

3.5 Exercises

3.1 This problem concerns encoding phrases using numbers using the encoding of Section 3.3.2. What is the longest that an arbitrary sequence of letters (no spaces) can be if it must fit in a number that is less than 10^{20}?

3.2 Suppose Michael creates an RSA cryptosystem with a very large modulus n for which the factorization of n cannot be found in a reasonable amount of time. Suppose that Nikita sends messages to Michael by representing each alphabetic character as an integer between 0 and 26 (`A` corresponds to 1, `B` to 2, etc., and a space ␣ to 0), then encrypts each number *separately* using Michael's RSA cryptosystem. Is this method secure? Explain your answer.

3.3 For any $n \in \mathbf{N}$, let $\sigma(n)$ be the sum of the divisors of n; for example, $\sigma(6) = 1 + 2 + 3 + 6 = 12$ and $\sigma(10) = 1 + 2 + 5 + 10 = 18$. Suppose that $n = pqr$ with p, q, and r distinct primes. Devise an "efficient" algorithm that given n, $\varphi(n)$ and $\sigma(n)$, computes the factorization of n. For example, if $n = 105$, then $p = 3$, $q = 5$, and $r = 7$, so the input to the algorithm would be

$$n = 105, \qquad \varphi(n) = 48, \qquad \text{and} \quad \sigma(n) = 192,$$

and the output would be 3, 5, and 7.

3.4 You and Nikita wish to agree on a secret key using the Diffie-Hellman key exchange. Nikita announces that $p = 3793$ and $g = 7$. Nikita secretly chooses a number $n < p$ and tells you that $g^n \equiv 454 \pmod{p}$. You choose the random number $m = 1208$. What is the secret key?

3.5 You see Michael and Nikita agree on a secret key using the Diffie-Hellman key exchange. Michael and Nikita choose $p = 97$ and $g = 5$. Nikita chooses a random number n and tells Michael that $g^n \equiv 3 \pmod{97}$, and Michael chooses a random number m and tells Nikita that $g^m \equiv 7 \pmod{97}$. Brute force crack their code: What is the secret key that Nikita and Michael agree upon? What is n? What is m?

3.6 In this problem, you will "crack" an RSA cryptosystem. What is the secret decoding number d for the RSA cryptosystem with public key $(n, e) = (5352381469067, 4240501142039)$?

3.7 Nikita creates an RSA cryptosystem with public key

$$(n, e) = (1433811615146881, 329222149569169).$$

In the following two problems, show the steps you take to factor n. (Don't simply factor n directly using a computer.)

(a) Somehow you discover that $d = 116439879930113$. Show how to use the probabilistic algorithm of Section 3.4.3 to factor n.

(b) In part (a) you found that the factors p and q of n are very close. Show how to use the Fermat Factorization Method of Section 3.4.2 to factor n.

4
Quadratic Reciprocity

A linear equation

$$ax \equiv b \pmod{n}$$

has a solution if and only if $\gcd(a, n)$ divides b (see Proposition 2.1.15). This chapter is about some amazing mathematics motivated by the search for a criterion for whether or not a given quadratic equation

$$ax^2 + bx + c \equiv 0 \pmod{n}$$

has a solution. In many cases, the Chinese Remainder Theorem and the quadratic formula reduce this to the key question of whether a given integer a is a perfect square modulo a prime p.

The Quadratic Reciprocity Law of Gauss provides a precise answer to the following question: For which primes p is the image of a in $(\mathbf{Z}/p\mathbf{Z})^*$ a perfect square? A deep fact, which we will completely prove in this chapter, is that the answer depends only on the reduction of p modulo $4a$. Thus *to decide if a is a square modulo p, one only needs to consider the residue of p modulo $4a$*, which is extremely surprising. It turns out that this "reciprocity law" goes to the heart of modern number theory and touches on advanced topics such as class field theory and the Langlands program.

There are over a hundred proofs of the Quadratic Reciprocity Law (see [Lem] for a long list). In this chapter, we give two proofs. The first, which we give in Section 4.3, is completely elementary and involves keeping track of integer points in intervals. It is satisfying because one can understand every detail without much abstraction, but it might be unsatisfying if you find it difficult to conceptualize what is going on. In contrast, our second proof, which we give in Section 4.4, is more abstract and uses a conceptual development of properties of Gauss sums. You should read Sections 4.1 and 4.2, then at least one of Section 4.3 or Section 4.4, depending on your taste and how much abstract algebra you know.

In Section 4.5, we return to the computational question of actually finding square roots and solving quadratic equations in practice.

4.1 Statement of the Quadratic Reciprocity Law

In this section, we state the Quadratic Reciprocity Law.

Definition 4.1.1 (Quadratic Residue)**.** Fix a prime p. An integer a not divisible by p is a *quadratic residue* modulo p if a is a square modulo p; otherwise, a is a *quadratic nonresidue.*

For example, the squares modulo 5 are

$$1^2 = 1, \quad 2^2 = 4, \quad 3^2 = 4, \quad 4^2 = 1, \qquad \pmod 5$$

so 1 and 4 are both quadratic residues and 2 and 3 are quadratic nonresidues.

The quadratic reciprocity theorem is the deepest theorem that we will prove in this book. It connects the question of whether or not a is a quadratic residue modulo p to the question of whether p is a quadratic residue modulo each of the prime divisors of a. To express it precisely, we introduce some new notation.

Definition 4.1.2 (Legendre Symbol)**.** Let p be an odd prime and let a be an integer. Set

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } \gcd(a,p) \neq 1, \\ +1 & \text{if } a \text{ is a quadratic residue, and} \\ -1 & \text{if } a \text{ is a quadratic nonresidue.} \end{cases}$$

We call this symbol the *Legendre Symbol.*

For example, we have

$$\left(\frac{1}{5}\right) = 1, \quad \left(\frac{2}{5}\right) = -1, \quad \left(\frac{3}{5}\right) = -1, \quad \left(\frac{4}{5}\right) = 1, \quad \left(\frac{5}{5}\right) = 1.$$

This notation is well entrenched in the literature even though it is also the notation for “a divided by p;” be careful not to confuse the two.

SAGE Example 4.1.3. Use the `legendre_symbol` command to compute the Legendre symbol in Sage.

```
sage: legendre_symbol(2,3)
-1
sage: legendre_symbol(1,3)
1
sage: legendre_symbol(3,5)
-1
sage: legendre_symbol(Mod(3,5), 5)
-1
```

Since $\left(\frac{a}{p}\right)$ only depends on $a \pmod{p}$, it makes sense to define $\left(\frac{a}{p}\right)$ for $a \in \mathbf{Z}/p\mathbf{Z}$ to be $\left(\frac{\tilde{a}}{p}\right)$ for any lift $\tilde{a}$ of a to $\mathbf{Z}$.

Recall (see Definition 3.4.6) that a group homomorphism $\varphi : G \to H$ is a map such that for every $a, b \in G$ we have $\varphi(ab) = \varphi(a)\varphi(b)$. Moreover, we say that φ is surjective if for every $c \in H$ there is an $a \in G$ with $\varphi(a) = c$. The next lemma explains how the quadratic residue symbol defines a surjective group homomorphism.

Lemma 4.1.4. *The map $\psi : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$ given by $\psi(a) = \left(\frac{a}{p}\right)$ is a surjective group homomorphism.*

Proof. By Theorem 2.5.8, primitive roots exist, so there is $g \in (\mathbf{Z}/p\mathbf{Z})^*$ such that the elements of $(\mathbf{Z}/p\mathbf{Z})^*$ are

$$g, g^2, \ldots, g^{(p-1)/2}, g^{(p+1)/2}, \ldots, g^{p-1} = 1.$$

Since $p - 1$ is even, the squares of elements of $(\mathbf{Z}/p\mathbf{Z})^*$ are

$$g^2, g^4, \ldots, g^{(p-1)/2 \cdot 2} = 1, g^{p+1} = g^2, \ldots, g^{2(p-1)}.$$

Note that the powers of g starting with $g^{p+1} = g^2$ all appeared earlier on the list. Thus, the perfect squares in $(\mathbf{Z}/p\mathbf{Z})^*$ are exactly the powers g^n with $n = 2, 4, \ldots, p - 1$, even, and the nonsquares the powers g^n with $n = 1, 3, \ldots, p - 2$, odd. It follows that ψ is a homomorphism since an odd plus an odd is even, the sum of two evens is even, and odd plus an even is odd. Moreover, since g is not a square, $\psi(g) = -1$, so ψ is surjective.

□

Remark 4.1.5. We rephrase the above proof in the language of group theory. The group $G = (\mathbf{Z}/p\mathbf{Z})^*$ of order $p - 1$ is a cyclic group. Since p is odd, $p - 1$ is even, so the subgroup H of squares of elements of G has index 2 in G. (See Exercise 4.2 for why H is a subgroup.) Since $\left(\frac{a}{p}\right) = 1$ if and only if $a \in H$, we see that ψ is the composition $G \to G/H \cong \{\pm 1\}$, where we identify the nontrivial element of G/H with -1.

Remark 4.1.6. We can alternatively prove that ψ is surjective without using that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic, as follows. If $a \in (\mathbf{Z}/p\mathbf{Z})^*$ is a square, say $a \equiv b^2$

TABLE 4.1. When is 5 a square modulo p?

p	$\left(\frac{5}{p}\right)$	p mod 5
7	-1	2
11	1	1
13	-1	3
17	-1	2
19	1	4
23	-1	3
29	1	4
31	1	1
37	-1	2
41	1	1
43	-1	3
47	-1	2

(mod p), then $a^{(p-1)/2} = b^{p-1} \equiv 1 \pmod{p}$, so a is a root of $f = x^{(p-1)/2} - 1$. By Proposition 2.5.3, the polynomial f has at most $(p-1)/2$ roots. Thus, there must be an $a \in (\mathbf{Z}/p\mathbf{Z})^*$ that is not a root of f, and for that a, we have $\psi(a) = \left(\frac{a}{p}\right) = -1$, and trivially $\psi(1) = 1$, so the map ψ is surjective. Note that this argument does not prove that ψ is a homomorphism.

The symbol $\left(\frac{a}{p}\right)$ only depends on the residue class of a modulo p, so making a table of values $\left(\frac{a}{5}\right)$ for many values of a would be easy. Would it be easy to make a table of $\left(\frac{5}{p}\right)$ for many p? Perhaps, since there *appears* to be a simple pattern in Table 4.1. It seems that $\left(\frac{5}{p}\right)$ depends only on the congruence class of p modulo 5. More precisely, $\left(\frac{5}{p}\right) = 1$ if and only if $p \equiv 1, 4 \pmod{5}$, i.e., $\left(\frac{5}{p}\right) = 1$ if and only if p is a square modulo 5.

Based on similar observations, in the 18th century various mathematicians found a conjectural explanation for the mystery suggested by Table 4.1. Finally, on April 8, 1796, at the age of 19, Gauss proved the following theorem.

Theorem 4.1.7 (Gauss's Quadratic Reciprocity Law)**.** *Suppose p and q are distinct odd primes. Then*

$$\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}\left(\frac{q}{p}\right).$$

Also

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} \qquad \text{and} \qquad \left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod{8} \\ -1 & \text{if } p \equiv \pm 3 \pmod{8}. \end{cases}$$

We will give two proofs of Gauss's formula relating $\left(\frac{p}{q}\right)$ to $\left(\frac{q}{p}\right)$. The first elementary proof is in Section 4.3, and the second more algebraic proof is in Section 4.4.

In our example, Gauss's theorem implies that

$$\left(\frac{5}{p}\right) = (-1)^{2\cdot\frac{p-1}{2}}\left(\frac{p}{5}\right) = \left(\frac{p}{5}\right) = \begin{cases} +1 & \text{if } p \equiv 1,4 \pmod 5\\ -1 & \text{if } p \equiv 2,3 \pmod 5.\end{cases}$$

As an application, the following example illustrates how to answer questions like "is a a square modulo b" using Theorem 4.1.7.

Example 4.1.8. Is 69 a square modulo the prime 389? We have

$$\left(\frac{69}{389}\right) = \left(\frac{3\cdot 23}{389}\right) = \left(\frac{3}{389}\right)\cdot\left(\frac{23}{389}\right) = (-1)\cdot(-1) = 1.$$

Here

$$\left(\frac{3}{389}\right) = \left(\frac{389}{3}\right) = \left(\frac{2}{3}\right) = -1,$$

and

$$\begin{aligned}\left(\frac{23}{389}\right) &= \left(\frac{389}{23}\right) = \left(\frac{21}{23}\right) = \left(\frac{-2}{23}\right)\\ &= \left(\frac{-1}{23}\right)\left(\frac{2}{23}\right) = (-1)^{\frac{23-1}{2}}\cdot 1 = -1.\end{aligned}$$

Thus 69 is a square modulo 389.

SAGE Example 4.1.9. We could also do this computation in Sage as follows:

```
sage: legendre_symbol(69,389)
1
```

Though we know that 69 is a square modulo 389, we don't know an explicit x such that $x^2 \equiv 69 \pmod{389}$! This is reminiscent of how we proved using Theorem 2.1.20 that certain numbers are composite without knowing a factorization.

Remark 4.1.10. The Jacobi symbol is an extension of the Legendre symbol to composite moduli. For more details, see Exercise 4.9.

4.2 Euler's Criterion

Let p be an odd prime and a an integer not divisible by p. Euler used the existence of primitive roots to show that $\left(\frac{a}{p}\right)$ is congruent to $a^{(p-1)/2}$ modulo p. We will use this fact repeatedly below in both proofs of Theorem 4.1.7.

Proposition 4.2.1 (Euler's Criterion)**.** *We have* $\left(\frac{a}{p}\right) = 1$ *if and only if*

$$a^{(p-1)/2} \equiv 1 \pmod p.$$

Proof. The map $\varphi : (\mathbf{Z}/p\mathbf{Z})^* \to (\mathbf{Z}/p\mathbf{Z})^*$ given by $\varphi(a) = a^{(p-1)/2}$ is a group homomorphism, since powering is a group homomorphism of any abelian group (see Exercise 4.2). Let $\psi : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$ be the homomorphism $\psi(a) = \left(\frac{a}{p}\right)$ of Lemma 4.1.4. If $a \in \ker(\psi)$, then $a = b^2$ for some $b \in (\mathbf{Z}/p\mathbf{Z})^*$, so

$$\varphi(a) = a^{(p-1)/2} = (b^2)^{(p-1)/2} = b^{p-1} = 1.$$

Thus $\ker(\psi) \subset \ker(\varphi)$. By Lemma 4.1.4, $\ker(\psi)$ has index 2 in $(\mathbf{Z}/p\mathbf{Z})^*$, i.e., $\#(\mathbf{Z}/p\mathbf{Z})^* = 2 \cdot \#\ker(\psi)$. Since the kernel of a homomorphism is a group, and the order of a subgroup divides the order of the group, we have either $\ker(\varphi) = \ker(\psi)$ or $\varphi = 1$. If $\varphi = 1$, the polynomial $x^{(p-1)/2} - 1$ has $p - 1$ roots in the field $\mathbf{Z}/p\mathbf{Z}$, which contradicts Proposition 2.5.3. Thus $\ker(\varphi) = \ker(\psi)$, which proves the proposition. □

SAGE Example 4.2.2. From a computational point of view, Corollary 4.2.3 provides a convenient way to compute $\left(\frac{a}{p}\right)$, which we illustrate in Sage:

```
sage: def kr(a, p):
...    if Mod(a,p)^((p-1)//2) == 1:
...       return 1
...    else:
...       return -1
sage: for a in range(1,5):
...    print a, kr(a,5)
1 1
2 -1
3 -1
4 1
```

Corollary 4.2.3. *The equation* $x^2 \equiv a \pmod{p}$ *has no solution if and only if* $a^{(p-1)/2} \equiv -1 \pmod{p}$. *Thus* $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$.

Proof. This follows from Proposition 4.2.1 and the fact that the polynomial $x^2 - 1$ has no roots besides $+1$ and -1 (which follows from Proposition 2.5.5). □

As additional computational motivation for the value of Corollary 4.2.3, note that to evaluate $\left(\frac{a}{p}\right)$ using Theorem 4.1.7 would not be practical if a and p are both very large, because it would require factoring a. However, Corollary 4.2.3 provides a method for evaluating $\left(\frac{a}{p}\right)$ without factoring a.

Example 4.2.4. Suppose $p = 11$. By squaring each element of $(\mathbf{Z}/11\mathbf{Z})^*$, we see that the squares modulo 11 are $\{1, 3, 4, 5, 9\}$. We compute $a^{(p-1)/2} = a^5$

for each $a \in (\mathbf{Z}/11\mathbf{Z})^*$ and get

$$\begin{aligned}&1^5 = 1,\ 2^5 = -1,\ 3^5 = 1,\ 4^5 = 1,\ 5^5 = 1,\\ &6^5 = -1,\ 7^5 = -1,\ 8^5 = -1,\ 9^5 = 1,\ 10^5 = -1.\end{aligned}$$

Thus the a with $a^5 = 1$ are $\{1, 3, 4, 5, 9\}$, just as Proposition 4.2.1 predicts.

Example 4.2.5. We determine whether or not 3 is a square modulo the prime $p = 726377359$.

```
sage: p = 726377359
sage: Mod(3, p)^((p-1)//2)
726377358
```

so

$$3^{(p-1)/2} \equiv -1 \pmod{726377359}.$$

Thus 3 is not a square modulo p. This computation wasn't difficult, but it would have been tedious by hand. Since 3 is small, the Quadratic Reciprocity Law provides a way to answer this question, which could easily be carried out by hand:

$$\begin{aligned}\left(\frac{3}{726377359}\right) &= (-1)^{(3-1)/2\cdot(726377359-1)/2}\left(\frac{726377359}{3}\right)\\ &= (-1)\cdot\left(\frac{1}{3}\right) = -1.\end{aligned}$$

4.3 First Proof of Quadratic Reciprocity

Our first proof of quadratic reciprocity is elementary. The proof involves keeping track of integer points in intervals. Proving Gauss's lemma is the first step; this lemma computes $\left(\frac{a}{p}\right)$ in terms of the number of integers of a certain type that lie in a certain interval. We next prove Lemma 4.3.3, which controls how the parity of the number of integer points in an interval changes when an endpoint of the interval is changed. We then prove that $\left(\frac{a}{p}\right)$ depends only on p modulo $4a$ by applying Gauss's Lemma and keeping careful track of intervals as they are rescaled and their endpoints are changed. Finally, in Section 4.3.2, we use some basic algebra to deduce the Quadratic Reciprocity Law using the tools we've just developed. Our proof follows the one given in [Dav99] closely.

Lemma 4.3.1 (Gauss's Lemma)**.** *Let p be an odd prime and let a be an integer $\not\equiv 0 \pmod{p}$. Form the numbers*

$$a,\ 2a,\ 3a,\ \ldots,\ \frac{p-1}{2}a$$

and reduce them modulo p to lie in the interval $(-\frac{p}{2},\ \frac{p}{2})$, i.e., for each of the above products $k \cdot a$ find a number in the interval $(-\frac{p}{2},\ \frac{p}{2})$ that is congruent to $k \cdot a$ modulo p. Let ν be the number of negative numbers in the resulting set. Then

$$\left(\frac{a}{p}\right) = (-1)^{\nu}.$$

Proof. In defining ν, we expressed each number in

$$S = \left\{a, 2a, \ldots, \frac{p-1}{2}a\right\}$$

as congruent to a number in the set

$$\left\{1, -1, 2, -2, \ldots, \frac{p-1}{2}, -\frac{p-1}{2}\right\}.$$

No number $1, 2, \ldots, \frac{p-1}{2}$ appears more than once, with either choice of sign, because if it did then either two elements of S are congruent modulo p or 0 is the sum of two elements of S, and both events are impossible (the former case cannot occur because of cancellation modulo p, and in the latter case we would have $ka + ja \equiv 0 \pmod{p}$ for $1 \leq k, j \leq (p-1)/2$, so $k + j \equiv 0 \pmod{p}$, a contradiction). The resulting set must be of the form

$$T = \left\{\varepsilon_1 \cdot 1, \varepsilon_2 \cdot 2, \ldots, \varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right\},$$

where each ε_i is either $+1$ or -1. Multiplying together the elements of S and of T, we see that

$$(1a) \cdot (2a) \cdot (3a) \cdots \left(\frac{p-1}{2}a\right) \equiv \\ (\varepsilon_1 \cdot 1) \cdot (\varepsilon_2 \cdot 2) \cdots \left(\varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right) \pmod{p},$$

so

$$a^{(p-1)/2} \equiv \varepsilon_1 \cdot \varepsilon_2 \cdots \varepsilon_{(p-1)/2} \pmod{p}.$$

The lemma then follows from Proposition 4.2.1, since $\left(\frac{a}{p}\right) = a^{(p-1)/2}$. □

SAGE Example 4.3.2. We illustrate Gauss's Lemma using Sage. The `gauss` function below prints out a list of the normalized numbers appearing in the statement of Gauss's Lemma, and returns $(-1)^{\nu}$. In each case below, $(-1)^{\nu} = \left(\frac{a}{p}\right)$.

```
sage: def gauss(a, p):
...    # make the list of numbers reduced modulo p
...     v = [(n*a)%p for n in range(1, (p-1)//2 + 1)]
...     # normalize them to be in the range -p/2 to p/2
...     v = [(x if (x < p/2) else x - p) for x in v]
...     # sort and print the resulting numbers
...     v.sort()
...     print v
...     # count the number that are negative
...     num_neg = len([x for x in v if x < 0])
...     return (-1)^num_neg
sage: gauss(2, 13)
[-5, -3, -1, 2, 4, 6]
-1
sage: legendre_symbol(2,13)
-1
sage: gauss(4, 13)
[-6, -5, -2, -1, 3, 4]
1
sage: legendre_symbol(4,13)
1
sage: gauss(2,31)
[-15, -13, -11, -9, -7, -5, -3, -1, 2, 4, 6, 8, 10, 12, 14]
1
sage: legendre_symbol(2,31)
1
```

4.3.1 Euler's Proposition

For rational numbers $a, b \in \mathbf{Q}$, let

$$(a, b) \cap \mathbf{Z} = \{x \in \mathbf{Z} : a \leq x \leq b\}$$

be the set of integers between a and b. The following lemma will help us to keep track of how many integers lie in certain intervals.

Lemma 4.3.3. *Let $a, b \in \mathbf{Q}$. Then for any integer n,*

$$\#\left((a, b) \cap \mathbf{Z}\right) \equiv \#\left((a, b + 2n) \cap \mathbf{Z}\right) \pmod{2}$$

and

$$\#\left((a, b) \cap \mathbf{Z}\right) \equiv \#\left((a - 2n, b) \cap \mathbf{Z}\right) \pmod{2},$$

provided that each interval involved in the congruence is nonempty.

Note that if one of the intervals is empty, then the statement may be false; for example, if $(a, b) = (-1/2, 1/2)$ and $n = -1$, then $\#((a, b) \cap \mathbf{Z}) = 1$ but $\#(a, b - 2) \cap \mathbf{Z} = 0$.

Proof. Let $\lceil x \rceil$ denotes the least integer $\geq x$. Since $n > 0$,

$$(a, b + 2n) = (a, b) \cup [b, b + 2n),$$

where the union is disjoint. There are $2n$ integers

$$\lceil b \rceil, \lceil b \rceil + 1, \ldots, \lceil b \rceil + 2n - 1$$

in the interval $[b, b + 2n)$, so the first congruence of the lemma is true in this case. We also have

$$(a, b - 2n) = (a, b) \text{ minus } [b - 2n, b)$$

and $[b-2n, b)$ contains exactly $2n$ integers, so the lemma is also true when n is negative. The statement about $\#\left((a - 2n, b) \cap \mathbf{Z}\right)$ is proved in a similar manner. □

Once we have proved the following proposition, it will be easy to deduce the Quadratic Reciprocity Law.

Proposition 4.3.4 (Euler). *Let p be an odd prime and let a be a positive integer with $p \nmid a$. If q is a prime with $q \equiv \pm p \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.*

Proof. We will apply Lemma 4.3.1 to compute $\left(\frac{a}{p}\right)$. Let

$$S = \left\{ a, 2a, 3a, \ldots, \frac{p-1}{2} a \right\}$$

and

$$I = \left(\frac{1}{2}p, p\right) \cup \left(\frac{3}{2}p, 2p\right) \cup \cdots \cup \left(\left(b - \frac{1}{2}\right)p, bp\right),$$

where $b = \frac{1}{2}a$ or $\frac{1}{2}(a - 1)$, whichever is an integer.

We check that every element of S that is equivalent modulo p to something in the interval $(-\frac{p}{2}, 0)$ lies in I. First suppose that $b = \frac{1}{2}a$. Then

$$bp = \frac{1}{2}ap = \frac{p}{2}a > \frac{p-1}{2}a,$$

so each element of S that is equivalent modulo p to an element of $(-\frac{p}{2}, 0)$ lies in I. Next suppose that $b = \frac{1}{2}(a - 1)$. Then

$$bp + \frac{p}{2} = \frac{a-1}{2}p + \frac{p}{2} = \frac{p - 1 + a}{2} > \frac{p-1}{2}a,$$

so $((b - \frac{1}{2})p, bp)$ is the last interval that could contain an element of S that reduces to $(-\frac{p}{2}, 0)$. Note that the integer endpoints of I are not in S, since those endpoints are divisible by p, but no element of S is divisible by p. Thus, by Lemma 4.3.1,

$$\left(\frac{a}{p}\right) = (-1)^{\#(S\cap I)}.$$

To compute $\#(S \cap I)$, first rescale by a to see that

$$\#(S\cap I) = \#\left(\frac{1}{a}S \cap \frac{1}{a}I\right) = \#\left(\mathbf{Z} \cap \frac{1}{a}I\right),$$

where

$$\frac{1}{a}I = \left(\left(\frac{p}{2a}, \frac{p}{a}\right) \cup \left(\frac{3p}{2a}, \frac{2p}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)p}{2a}, \frac{bp}{a}\right)\right),$$

$\frac{1}{a}S = \{1, 2, 3, 4, \ldots, (p-1)/2\}$, and the second equality is because $\frac{1}{a}I \subset (0, (p-1)/2 + 1/2]$, since

$$\frac{pb}{a} \leq \frac{p\frac{a}{2}}{a} = \frac{p}{2} = \frac{p-1}{2} + \frac{1}{2}.$$

Write $p = 4ac + r$, and let

$$J = \left(\left(\frac{r}{2a}, \frac{r}{a}\right) \cup \left(\frac{3r}{2a}, \frac{2r}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)r}{2a}, \frac{br}{a}\right)\right).$$

The only difference between $\frac{1}{a}I$ and J is that the endpoints of intervals are changed by addition of an even integer, since

$$\frac{r}{2a} - \frac{p}{2a} = \frac{p}{2a} - 2c - \frac{p}{2a} = -2c.$$

By Lemma 4.3.3,

$$\nu = \#\left(\mathbf{Z} \cap \frac{1}{a}I\right) \equiv \#(\mathbf{Z} \cap J) \pmod{2}.$$

Thus $\left(\frac{a}{p}\right) = (-1)^\nu$ depends only on r and a, i.e., only on p modulo $4a$. Thus if $q \equiv p \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.

If $q \equiv -p \pmod{4a}$, then the only change in the above computation is that r is replaced by $4a - r$. This changes J into

$$K = \left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right) \cup \left(6 - \frac{3r}{2a}, 8 - \frac{2r}{a}\right) \cup \cdots \\ \cup \left(4b - 2 - \frac{(2b-1)r}{2a}, 4b - \frac{br}{a}\right).$$

Thus K is the same as $-J$, except even integers have been added to the endpoints. By Lemma 4.3.3,

$$\#(K \cap \mathbf{Z}) \equiv \# \left(\frac{1}{a} I \cap \mathbf{Z} \right) \pmod{2},$$

so $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$ again, which completes the proof. □

The following more careful analysis in the special case when $a = 2$ helps illustrate the proof of the above lemma, and the result is frequently useful in computations. For an alternative proof of the proposition, see Exercise 4.6.

Proposition 4.3.5 (Legendre Lymbol of 2)**.** *Let p be an odd prime. Then*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod{8} \\ -1 & \text{if } p \equiv \pm 3 \pmod{8}. \end{cases}$$

Proof. When $a = 2$, the set $S = \{a, 2a, \ldots, 2 \cdot \frac{p-1}{2}\}$ is

$$\{2, 4, 6, \ldots, p - 1\}.$$

We must count the parity of the number of elements of S that lie in the interval $I = (\frac{p}{2}, p)$. Writing $p = 8c + r$, we have

$$\begin{aligned} \#\,(I \cap S) &= \# \left(\frac{1}{2} I \cap \mathbf{Z} \right) = \# \left(\left(\frac{p}{4}, \frac{p}{2} \right) \cap \mathbf{Z} \right) \\ &= \# \left(\left(2c + \frac{r}{4}, 4c + \frac{r}{2} \right) \cap \mathbf{Z} \right) \equiv \# \left(\left(\frac{r}{4}, \frac{r}{2} \right) \cap \mathbf{Z} \right) \pmod{2}, \end{aligned}$$

where the last equality comes from Lemma 4.3.3. The possibilities for r are $1, 3, 5, 7$. When $r = 1$, the cardinality is 0; when $r = 3, 5$ it is 1; and when $r = 7$ it is 2. □

4.3.2 Proof of Quadratic Reciprocity

It is now straightforward to deduce the Quadratic Reciprocity Law.

First Proof of Theorem 4.1.7. First suppose that $p \equiv q \pmod{4}$. By swapping p and q if necessary, we may assume that $p > q$, and write $p - q = 4a$. Since $p = 4a + q$,

$$\left(\frac{p}{q}\right) = \left(\frac{4a + q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{4}{q}\right) \left(\frac{a}{q}\right) = \left(\frac{a}{q}\right),$$

and

$$\left(\frac{q}{p}\right) = \left(\frac{p - 4a}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right) \cdot \left(\frac{a}{p}\right).$$

Proposition 4.3.4 implies that $\left(\frac{a}{q}\right) = \left(\frac{a}{p}\right)$, since $p \equiv q \pmod{4a}$. Thus

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}},$$

where the last equality is because $\frac{p-1}{2}$ is even if and only if $\frac{q-1}{2}$ is even.

Next suppose that $p \not\equiv q \pmod 4$, so $p \equiv -q \pmod 4$. Write $p + q = 4a$. We have

$$\left(\frac{p}{q}\right) = \left(\frac{4a-q}{q}\right) = \left(\frac{a}{q}\right), \quad \text{and} \quad \left(\frac{q}{p}\right) = \left(\frac{4a-p}{p}\right) = \left(\frac{a}{p}\right).$$

Since $p \equiv -q \pmod{4a}$, Proposition 4.3.4 implies that $\left(\frac{a}{q}\right) = \left(\frac{a}{p}\right)$. Since $(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}} = 1$, the proof is complete. □

4.4 A Proof of Quadratic Reciprocity Using Gauss Sums

In this section, we present a beautiful proof of Theorem 4.1.7 using algebraic identities satisfied by sums of "roots of unity." The objects we introduce in the proof are of independent interest, and provide a powerful tool to prove higher-degree analogs of quadratic reciprocity. (For more on higher reciprocity, see [IR90]. See also Section 6 of [IR90], on which the proof below is modeled.)

Definition 4.4.1 (Root of Unity)**.** An nth *root of unity* is a complex number ζ such that $\zeta^n = 1$. A root of unity ζ is a *primitive* nth root of unity if n is the smallest positive integer such that $\zeta^n = 1$.

For example, -1 is a primitive second root of unity, and $\zeta = \frac{\sqrt{-3}-1}{2}$ is a primitive cube root of unity. More generally, for any $n \in \mathbf{N}$ the complex number

$$\zeta_n = \cos(2\pi/n) + i\sin(2\pi/n)$$

is a primitive nth root of unity (this follows from the identity $e^{i\theta} = \cos(\theta) + i\sin(\theta)$). For the rest of this section, we fix an odd prime p and the primitive pth root $\zeta = \zeta_p$ of unity.

SAGE Example 4.4.2. In Sage, use the `CyclotomicField` command to create an exact pth root of ζ unity. Expressions in ζ are always re-expressed as polynomials in ζ of degree at most $p - 1$.

```
sage: K.<zeta> = CyclotomicField(5)
sage: zeta^5
1
```

```
sage: 1/zeta
-zeta^3 - zeta^2 - zeta - 1
```

Definition 4.4.3 (Gauss Sum)**.** Fix an odd prime p. The *Gauss sum* associated to an integer a is

$$g_a = \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) \zeta^{an},$$

where $\zeta = \zeta_p = \cos(2\pi/p) + i\sin(2\pi/p) = e^{2\pi i/p}$.

Note that p is implicit in the definition of g_a. If we were to change p, then the Gauss sum g_a associated to a would be different. The definition of g_a also depends on our choice of ζ; we've chosen $\zeta = \zeta_p$, but could have chosen a different ζ and then g_a could be different.

SAGE Example 4.4.4*.* We define a gauss_sum function and compute the Gauss sum g_2 for $p = 5$:

```
sage: def gauss_sum(a,p):
...    K.<zeta> = CyclotomicField(p)
...    return sum(legendre_symbol(n,p) * zeta^(a*n)
...               for n in range(1,p))
sage: g2 = gauss_sum(2,5); g2
2*zeta^3 + 2*zeta^2 + 1
sage: g2.complex_embedding()
-2.2360679775 + 3.33066907388e-16*I
sage: g2^2
5
```

Here, g_2 is initially output as a polynomial in ζ_5, so there is no loss of precision. The complex_embedding command shows some embedding of g_2 into the complex numbers, which is only correct to about the first 15 digits. Note that $g_2^2 = 5$, so $g_2 = -\sqrt{5}$.

We compute a graphical representation of the Gauss sum g_2 as follows (see Figure 4.1):

```
zeta = CDF(exp(2*pi*I/5))
v = [legendre_symbol(n,5) * zeta^(2*n) for n in range(1,5)]
S = sum([point(tuple(z), pointsize=100) for z in v])
show(S + point(tuple(sum(v)), pointsize=100, rgbcolor='red'))
```

Figure 4.1 illustrates the Gauss sum g_2 for $p = 5$. The Gauss sum is obtained by adding the points on the unit circle, with signs as indicated, to obtain the real number $-\sqrt{5}$. This suggests the following proposition, whose proof will require some work.

Proposition 4.4.5 (Gauss Sum)**.** *For any a not divisible by p,*

$$g_a^2 = (-1)^{(p-1)/2} p.$$

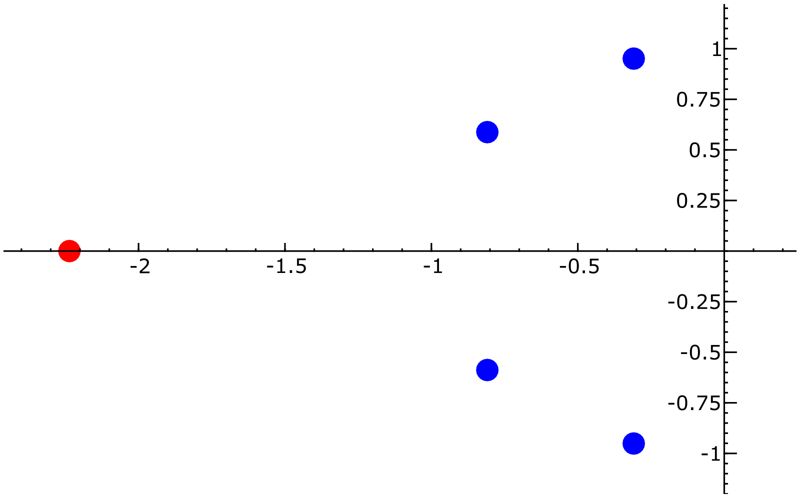


FIGURE 4.1. The red dot is the Gauss sum g_2 for $p = 5$

SAGE Example 4.4.6. We illustrate using Sage that the proposition is correct for $p = 7$ and $p = 13$:

```
sage: [gauss_sum(a, 7)^2 for a in range(1,7)]
[-7, -7, -7, -7, -7, -7]
sage: [gauss_sum(a, 13)^2 for a in range(1,13)]
[13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13]
```

In order to prove the proposition, we introduce a few lemmas.

Lemma 4.4.7. *For any integer a,*

$$\sum_{n=0}^{p-1} \zeta^{an} = \begin{cases} p & \text{if } a \equiv 0 \pmod{p}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If $a \equiv 0 \pmod{p}$, then $\zeta^a = 1$, so the sum equals the number of summands, which is p. If $a \not\equiv 0 \pmod{p}$, then we use the identity

$$x^p - 1 = (x-1)(x^{p-1} + \cdots + x + 1)$$

with $x = \zeta^a$. We have $\zeta^a \neq 1$, so $\zeta^a - 1 \neq 0$ and

$$\sum_{n=0}^{p-1} \zeta^{an} = \frac{\zeta^{ap} - 1}{\zeta^a - 1} = \frac{1-1}{\zeta^a - 1} = 0.$$

□

Lemma 4.4.8. *If x and y are arbitrary integers, then*

$$\sum_{n=0}^{p-1} \zeta^{(x-y)n} = \begin{cases} p & \text{if } x \equiv y \pmod{p}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. This follows from Lemma 4.4.7 by setting $a = x - y$. □

Lemma 4.4.9. *We have $g_0 = 0$.*

Proof. By definition

$$g_0 = \sum_{n=0}^{p-1} \left(\frac{n}{p}\right). \tag{4.4.1}$$

By Lemma 4.1.4, the map

$$\left(\frac{\cdot}{p}\right) : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$$

is a surjective homomorphism of groups. Thus, half the elements of $(\mathbf{Z}/p\mathbf{Z})^*$ map to $+1$ and half map to -1 (the subgroup that maps to $+1$ has index 2). Since $\left(\frac{0}{p}\right) = 0$, the sum (4.4.1) is 0. □

Lemma 4.4.10. *For any integer a,*

$$g_a = \left(\frac{a}{p}\right) g_1.$$

Proof. When $a \equiv 0 \pmod{p}$, the lemma follows from Lemma 4.4.9, so suppose that $a \not\equiv 0 \pmod{p}$. Then,

$$\left(\frac{a}{p}\right) g_a = \left(\frac{a}{p}\right) \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^{an} = \sum_{n=0}^{p-1} \left(\frac{an}{p}\right) \zeta^{an} = \sum_{m=0}^{p-1} \left(\frac{m}{p}\right) \zeta^{m} = g_1.$$

Here, we use that multiplication by a is an automorphism of $\mathbf{Z}/p\mathbf{Z}$. Finally, multiply both sides by $\left(\frac{a}{p}\right)$ and use that $\left(\frac{a}{p}\right)^2 = 1$. □

We have enough lemmas to prove Proposition 4.4.5.

Proof of Proposition 4.4.5. We evaluate the sum $\sum_{a=0}^{p-1} g_a g_{-a}$ in two different ways. By Lemma 4.4.10, since $a \not\equiv 0 \pmod{p}$ we have

$$g_a g_{-a} = \left(\frac{a}{p}\right) g_1 \left(\frac{-a}{p}\right) g_1 = \left(\frac{-1}{p}\right) \left(\frac{a}{p}\right)^2 g_1^2 = (-1)^{(p-1)/2} g_1^2,$$

where the last step follows from Proposition 4.2.1 and that $\left(\frac{a}{p}\right) \in \{\pm 1\}$. Thus

$$\sum_{a=0}^{p-1} g_a g_{-a} = (p-1)(-1)^{(p-1)/2} g_1^2. \tag{4.4.2}$$

On the other hand, by definition

$$\begin{aligned}
g_a g_{-a} &= \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^{an} \cdot \sum_{m=0}^{p-1} \left(\frac{m}{p}\right) \zeta^{-am} \\
&= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an} \zeta^{-am} \\
&= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an-am}.
\end{aligned}$$

Let $\delta(n, m) = 1$ if $n \equiv m \pmod{p}$ and 0 otherwise. By Lemma 4.4.8,

$$\begin{aligned}
\sum_{a=0}^{p-1} g_a g_{-a} &= \sum_{a=0}^{p-1} \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an-am} \\
&= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \sum_{a=0}^{p-1} \zeta^{an-am} \\
&= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) p\delta(n, m) \\
&= \sum_{n=0}^{p-1} \left(\frac{n}{p}\right)^2 p \\
&= p(p-1).
\end{aligned}$$

Equate (4.4.2) and the above equality, then cancel $(p-1)$ to see that

$$g_1^2 = (-1)^{(p-1)/2} p.$$

Since $a \not\equiv 0 \pmod{p}$, we have $\left(\frac{a}{p}\right)^2 = 1$, so by Lemma 4.4.10,

$$g_a^2 = \left(\frac{a}{p}\right)^2 g_1^2 = g_1^2,$$

and the proposition is proved. □

4.4.1 Proof of Quadratic Reciprocity

We are now ready to prove Theorem 4.1.7 using Gauss sums.

Proof. Let q be an odd prime with $q \neq p$. Set $p^* = (-1)^{(p-1)/2} p$ and recall that Proposition 4.4.5 asserts that $p^* = g^2$, where $g = g_1 = \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^n$.

Proposition 4.2.1 implies that

$$(p^*)^{(q-1)/2} \equiv \left(\frac{p^*}{q}\right) \pmod{q}.$$

We have $g^{q-1} = (g^2)^{(q-1)/2} = (p^*)^{(q-1)/2}$, so multiplying both sides of the displayed equation by g yields a congruence

$$g^q \equiv g\left(\frac{p^*}{q}\right) \pmod{q}. \tag{4.4.3}$$

But wait, what does this congruence mean, given that g^q is not an integer? It means that the difference $g^q - g\left(\frac{p^*}{q}\right)$ is a multiple of q in the ring $\mathbf{Z}[\zeta]$ of all polynomials in ζ with coefficients in $\mathbf{Z}$.

The ring $\mathbf{Z}[\zeta]/(q)$ has characteristic q, so if $x, y \in \mathbf{Z}[\zeta]$, then $(x+y)^q \equiv x^q + y^q \pmod{q}$. Applying this to (4.4.3), we see that

$$g^q = \left(\sum_{n=0}^{p-1}\left(\frac{n}{p}\right)\zeta^n\right)^q \equiv \sum_{n=0}^{p-1}\left(\frac{n}{p}\right)^q \zeta^{nq} \equiv \sum_{n=0}^{p-1}\left(\frac{n}{p}\right)\zeta^{nq} \equiv g_q \pmod{q}.$$

By Lemma 4.4.10,

$$g^q \equiv g_q \equiv \left(\frac{q}{p}\right) g \pmod{q}.$$

Combining this with (4.4.3) yields

$$\left(\frac{q}{p}\right) g \equiv \left(\frac{p^*}{q}\right) g \pmod{q}.$$

Since $g^2 = p^*$ and $p \neq q$, we can cancel g from both sides to find that $\left(\frac{q}{p}\right) \equiv \left(\frac{p^*}{q}\right) \pmod{q}$. Since both residue symbols are ± 1 and q is odd, it follows that $\left(\frac{q}{p}\right) = \left(\frac{p^*}{q}\right)$. Finally, we note using Corollary 4.2.3 that

$$\left(\frac{p^*}{q}\right) = \left(\frac{(-1)^{(p-1)/2}p}{q}\right) = \left(\frac{-1}{q}\right)^{(p-1)/2}\left(\frac{p}{q}\right) = (-1)^{\frac{q-1}{2}\cdot\frac{p-1}{2}} \cdot \left(\frac{p}{q}\right).$$

□

4.5 Finding Square Roots

We return in this section to the question of computing square roots. If K is a field in which $2 \neq 0$, and $a, b, c \in K$, with $a \neq 0$, then the two solutions to the quadratic equation $ax^2 + bx + c = 0$ are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Now assume $K = \mathbf{Z}/p\mathbf{Z}$, with p an odd prime. Using Theorem 4.1.7, we can decide whether or not $b^2 - 4ac$ is a perfect square in $\mathbf{Z}/p\mathbf{Z}$, and hence whether or not $ax^2 + bx + c = 0$ has a solution in $\mathbf{Z}/p\mathbf{Z}$. However, Theorem 4.1.7 says nothing about how to actually find a solution when there is one. Also note that for this problem we do *not* need the full Quadratic Reciprocity Law; in practice, deciding whether an element of $\mathbf{Z}/p\mathbf{Z}$ is a perfect square with Proposition 4.2.1 is quite fast, in view of Section 2.3.

Suppose $a \in \mathbf{Z}/p\mathbf{Z}$ is a nonzero quadratic residue. If $p \equiv 3 \pmod 4$, then $b = a^{\frac{p+1}{4}}$ is a square root of a because

$$b^2 = a^{\frac{p+1}{2}} = a^{\frac{p-1}{2}+1} = a^{\frac{p-1}{2}} \cdot a = \left(\frac{a}{p}\right) \cdot a = a.$$

We can compute b in time polynomial in the number of digits of p using the powering algorithm of Section 2.3.

Suppose next that $p \equiv 1 \pmod 4$. Unfortunately, we do not know a deterministic algorithm that takes a and p as input, outputs a square root of a modulo p when one exists, and is polynomial-time in $\log(p)$.

Remark 4.5.1. There is an algorithm due to Schoof [Sch85] that computes the square root of a in time $O((\sqrt{(}|a|)^{1/2+\varepsilon} \cdot \log(p))^9)$. This beautiful algorithm (which makes use of elliptic curves) is not polynomial time in the sense described above, since for large a it takes exponentially longer than for small a.

We next describe a probabilistic algorithm to compute a square root of a modulo p, which is very quick in practice. Recall the notion of ring from Definition 2.1.3. We will also need the notion of ring homomorphism and isomorphism.

Definition 4.5.2 (Homomorphism of Rings)**.** Let R and S be rings. A *homomorphism of rings* $\varphi : R \to S$ is a map such that for all $a, b \in R$, we have

- $\varphi(ab) = \varphi(a)\varphi(b)$,
- $\varphi(a + b) = \varphi(a) + \varphi(b)$, and
- $\varphi(1) = 1$.

An *isomorphism* $\varphi : R \to S$ of rings is a ring homomorphism that is bijective.

Consider the ring

$$R = (\mathbf{Z}/p\mathbf{Z})[x]/(x^2 - a)$$

defined as follows. We have

$$R = \{u + v\alpha : u, v \in \mathbf{Z}/p\mathbf{Z}\}$$

with multiplication defined by

$$(u+v\alpha)(z+w\alpha) = (uz+awv) + (uw+vz)\alpha.$$

Here α corresponds to the class of x in R.

SAGE Example 4.5.3. We define and work with the ring R above in Sage as follows (for $p = 13$):

```
sage: S.<x> = PolynomialRing(GF(13))
sage: R.<alpha> = S.quotient(x^2 - 3)
sage: (2+3*alpha)*(1+2*alpha)
7*alpha + 7
```

Let b and c be the square roots of a in $\mathbf{Z}/p\mathbf{Z}$ (though we cannot easily compute b and c yet, we can consider them in order to deduce an algorithm to find them). We have ring homomorphisms $f : R \to \mathbf{Z}/p\mathbf{Z}$ and $g : R \to \mathbf{Z}/p\mathbf{Z}$ given by $f(u+v\alpha) = u+vb$ and $g(u+v\alpha) = u+vc$. Together, these define a ring isomorphism

$$\varphi : R \longrightarrow \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}$$

given by $\varphi(u+v\alpha) = (u+vb, u+vc)$. Choose in some way a random element z of $(\mathbf{Z}/p\mathbf{Z})^*$, and define $u, v \in \mathbf{Z}/p\mathbf{Z}$ by

$$u + v\alpha = (1+z\alpha)^{\frac{p-1}{2}},$$

where we compute $(1+z\alpha)^{\frac{p-1}{2}}$ quickly using an analog of the binary powering algorithm of Section 2.3.2. If $v = 0$, we try again with another random z. If $v \neq 0$, we can quickly find the desired square roots b and c as follows. The quantity $u+vb$ is a $(p-1)/2$ power in $\mathbf{Z}/p\mathbf{Z}$, so it equals either 0, 1, or -1, so $b = -u/v$, $(1-u)/v$, or $(-1-u)/v$, respectively. Since we know u and v, we can try each of $-u/v$, $(1-u)/v$, and $(-1-u)/v$ and see which is a square root of a.

Example 4.5.4. Continuing Example 4.1.8, we find a square root of 69 modulo 389. We apply the algorithm described above in the case $p \equiv 1 \pmod 4$. We first choose the random $z = 24$ and find that $(1+24\alpha)^{194} = -1$. The coefficient of α in the power is 0, and we try again with $z = 51$. This time, we have $(1+51\alpha)^{194} = 239\alpha = u + v\alpha$. The inverse of 239 in $\mathbf{Z}/389\mathbf{Z}$ is 153, so we consider the following three possibilities for a square root of 69:

$$-\frac{u}{v} = 0 \qquad \frac{1-u}{v} = 153 \qquad -\frac{1-u}{v} = -153.$$

Thus, 153 and -153 are the square roots of 69 in $\mathbf{Z}/389\mathbf{Z}$.

SAGE Example 4.5.5. We implement the above algorithm in Sage and illustrate it with some examples.

```
sage: def find_sqrt(a, p):
...     assert (p-1)%4 == 0
...     assert legendre_symbol(a,p) == 1
...     S.<x> = PolynomialRing(GF(p))
...     R.<alpha> = S.quotient(x^2 - a)
...     while True:
...         z = GF(p).random_element()
...         w = (1 + z*alpha)^((p-1)//2)
...         (u, v) = (w[0], w[1])
...         if v != 0: break
...     if (-u/v)^2 == a: return -u/v
...     if ((1-u)/v)^2 == a: return (1-u)/v
...     if ((-1-u)/v)^2 == a: return (-1-u)/v
...
sage: b = find_sqrt(3,13)
sage: b                         # random: either 9 or 3
9
sage: b^2
3
sage: b = find_sqrt(3,13)
sage: b                         # see, it's random
4
sage: find_sqrt(5,389)          # random: either 303 or 86
303
sage: find_sqrt(5,389)          # see, it's random
86
```

4.6 Exercises

4.1 Calculate the following by hand: $\left(\frac{3}{97}\right)$, $\left(\frac{3}{389}\right)$, $\left(\frac{22}{11}\right)$, and $\left(\frac{5!}{7}\right)$.

4.2 Let G be an abelian group, and let n be a positive integer.

(a) Prove that the map $\varphi : G \to G$ given by $\varphi(x) = x^n$ is a group homomorphism.

(b) Prove that the subset H of G of squares of elements of G is a subgroup.

4.3 Use Theorem 4.1.7 to show that for $p \geq 5$ prime,

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 11 \pmod{12}, \\ -1 & \text{if } p \equiv 5, 7 \pmod{12}. \end{cases}$$

4.4 (*) Use that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic to give a direct proof that $\left(\frac{-3}{p}\right) = 1$ when $p \equiv 1 \pmod 3$. (Hint: There is an element $c \in (\mathbf{Z}/p\mathbf{Z})^*$ of order 3. Show that $(2c+1)^2 = -3$.)

4.5 (*) If $p \equiv 1 \pmod 5$, show directly that $\left(\frac{5}{p}\right) = 1$ by the method of Exercise 4.4. (Hint: Let $c \in (\mathbf{Z}/p\mathbf{Z})^*$ be an element of order 5. Show that $(c+c^4)^2 + (c+c^4) - 1 = 0$, etc.)

4.6 (*) Let p be an odd prime. In this exercise, you will prove that $\left(\frac{2}{p}\right) = 1$ if and only if $p \equiv \pm 1 \pmod 8$.

(a) Prove that
$$x = \frac{1-t^2}{1+t^2}, \qquad y = \frac{2t}{1+t^2}$$
is a parameterization of the set of solutions to $x^2 + y^2 \equiv 1 \pmod p$, in the sense that the solutions $(x, y) \in \mathbf{Z}/p\mathbf{Z}$ are in bijection with the $t \in \mathbf{Z}/p\mathbf{Z} \cup \{\infty\}$ such that $1+t^2 \not\equiv 0 \pmod p$. Here, $t = \infty$ corresponds to the point $(-1, 0)$. (Hint: if (x_1, y_1) is a solution, consider the line $y = t(x+1)$ through (x_1, y_1) and $(-1, 0)$, and solve for x_1, y_1 in terms of t.)

(b) Prove that the number of solutions to $x^2 + y^2 \equiv 1 \pmod p$ is $p+1$ if $p \equiv 3 \pmod 4$ and $p-1$ if $p \equiv 1 \pmod 4$.

(c) Consider the set S of pairs $(a, b) \in (\mathbf{Z}/p\mathbf{Z})^* \times (\mathbf{Z}/p\mathbf{Z})^*$ such that $a + b = 1$ and $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = 1$. Prove that $\#S = (p+1-4)/4$ if $p \equiv 3 \pmod 4$ and $\#S = (p-1-4)/4$ if $p \equiv 1 \pmod 4$. Conclude that $\#S$ is odd if and only if $p \equiv \pm 1 \pmod 8$.

(d) The map $\sigma(a, b) = (b, a)$ that swaps coordinates is a bijection of the set S. It has exactly one fixed point if and only if there is an $a \in \mathbf{Z}/p\mathbf{Z}$ such that $2a = 1$ and $\left(\frac{a}{p}\right) = 1$. Also, prove that $2a = 1$ has a solution $a \in \mathbf{Z}/p\mathbf{Z}$ with $\left(\frac{a}{p}\right) = 1$ if and only if $\left(\frac{2}{p}\right) = 1$.

(e) Finish by showing that σ has exactly one fixed point if and only if $\#S$ is odd, i.e., if and only if $p \equiv \pm 1 \pmod 8$.

Remark: The method of proof of this exercise can be generalized to give a proof of the full Quadratic Reciprocity Law.

4.7 How many natural numbers $x < 2^{13}$ satisfy the equation
$$x^2 \equiv 5 \pmod{2^{13} - 1}?$$
You may assume that $2^{13} - 1$ is prime.

4.8 Find the natural number $x < 97$ such that $x \equiv 4^{48} \pmod{97}$. Note that 97 is prime.

4.9 In this problem, we will formulate an analog of quadratic reciprocity for a symbol like $\left(\frac{a}{q}\right)$, but without the restriction that q be a prime. Suppose n is an odd positive integer, which we factor as $\prod_{i=1}^{k} p_i^{e_i}$. We define the Jacobi symbol $\left(\frac{a}{n}\right)$ as follows:

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{e_i}.$$

(a) Give an example to show that $\left(\frac{a}{n}\right) = 1$ need not imply that a is a perfect square modulo n.

(b) (*) Let n be odd and a and b be integers. Prove that the following holds:

i. $\left(\frac{a}{n}\right)\left(\frac{b}{n}\right) = \left(\frac{ab}{n}\right)$. (Thus $a \mapsto \left(\frac{a}{n}\right)$ induces a homomorphism from $(\mathbf{Z}/n\mathbf{Z})^*$ to $\{\pm 1\}$.)

ii. $\left(\frac{-1}{n}\right) \equiv n \pmod{4}$.

iii. $\left(\frac{2}{n}\right) = 1$ if $n \equiv \pm 1 \pmod{8}$ and -1 otherwise.

iv. Assume a is positive and odd. Then $\left(\frac{a}{n}\right) = (-1)^{\frac{a-1}{2}\cdot\frac{n-1}{2}}\left(\frac{n}{a}\right)$

4.10 (*) Prove that for any $n \in \mathbf{Z}$, the integer $n^2 + n + 1$ does not have any divisors of the form $6k - 1$.

5
Continued Fractions

The golden ratio $\frac{1+\sqrt{5}}{2}$ is equal to the infinite fraction

$$1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cdots,}}}$$

and the fraction

$$\frac{103993}{33102} = 3.14159265301190260407\ldots$$

is an excellent approximation to π. Both of these observations are explained by continued fractions.

Continued fractions are theoretically beautiful and provide tools that yield powerful algorithms for solving problems in number theory. For example, continued fractions provide a fast way to write a prime—even a hundred digit prime—as a sum of two squares, when possible.

Continued fractions are thus a beautiful algorithmic and conceptual tool in number theory that has many applications. For example, they provide a surprisingly efficient way to recognize a rational number given just the first few digits of its decimal expansion, and they give a sense in which e is “less complicated” than π (see Example 5.3.4 and Section 5.4).

In Section 5.2, we study continued fractions of finite length and lay the foundations for our later investigations. In Section 5.3, we give the continued fraction procedure, which associates to a real number x a continued fraction that converges to x. In Section 5.5, we characterize (eventually) periodic continued fractions as the continued fractions of nonrational roots of quadratic polynomials, then discuss an unsolved mystery concerning continued fractions of roots of irreducible polynomials of degree greater than 2. We conclude the chapter with applications of continued fractions to recognizing approximations to rational numbers (Section 5.6) and writing integers as sums of two squares (Section 5.7).

The reader is encouraged to read more about continued fractions in [HW79, Ch. X], [Khi63], [Bur89, §13.3], and [NZM91, Ch. 7].

5.1 The Definition

A *continued fraction* is an expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots\ .}}}$$

In this book, we will assume that the a_i are real numbers and $a_i > 0$ for $i \geq 1$, and the expression may or may not go on indefinitely. More general notions of continued fractions have been extensively studied, but they are beyond the scope of this book. We will be most interested in the case when the a_i are all integers.

We denote the continued fraction displayed above by

$$[a_0, a_1, a_2, \ldots].$$

For example,

$$[1, 2] = 1 + \frac{1}{2} = \frac{3}{2},$$

$$\begin{aligned} [3, 7, 15, 1, 292] &= 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292}}}} \\ &= \frac{103993}{33102} = 3.14159265301190260407\ldots, \end{aligned}$$

and

$$\begin{aligned}[2,1,2,1,1,4,1,1,6] &= 2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{6}}}}}}}}\\ &= \frac{1264}{465}\\ &= 2.71827956989247311827956989\ldots\end{aligned}$$

The second two examples were chosen to foreshadow that continued fractions can be used to obtain good rational approximations to irrational numbers. Note that the first approximates π, and the second e.

5.2 Finite Continued Fractions

This section is about continued fractions of the form $[a_0, a_1, \ldots, a_m]$ for some $m \geq 0$. We give an inductive definition of numbers p_n and q_n such that for all $n \leq m$

$$[a_0, a_1, \ldots, a_n] = \frac{p_n}{q_n}. \tag{5.2.1}$$

We then give related formulas for the determinants of the 2×2 matrices $\left(\begin{smallmatrix} p_n & p_{n-1} \\ q_n & q_{n-1}\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} p_n & p_{n-2} \\ q_n & q_{n-2}\end{smallmatrix}\right)$, which we will repeatedly use to deduce properties of the sequence of partial convergents $[a_0, \ldots, a_k]$. We will use Algorithm 1.1.13 to prove that every rational number is represented by a continued fraction, as in (5.2.1).

Definition 5.2.1 (Finite Continued Fraction)**.** A *finite continued fraction* is an expression

$$a_0+\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{\cdots+\frac{1}{a_n}}}}$$

where each a_m is a real number and $a_m > 0$ for all $m \geq 1$.

Definition 5.2.2 (Simple Continued Fraction)**.** A *simple continued fraction* is a finite or infinite continued fraction in which the a_i are all integers.

To get a feeling for continued fractions, observe that

$$\begin{aligned}
[a_0] &= a_0, \\
[a_0, a_1] &= a_0 + \frac{1}{a_1} = \frac{a_0a_1 + 1}{a_1}, \\
[a_0, a_1, a_2] &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}} = \frac{a_0a_1a_2 + a_0 + a_2}{a_1a_2 + 1}.
\end{aligned}$$

Also,

$$\begin{aligned}
[a_0, a_1, \ldots, a_{n-1}, a_n] &= \left[a_0, a_1, \ldots, a_{n-2}, a_{n-1} + \frac{1}{a_n}\right] \\
&= a_0 + \frac{1}{[a_1, \ldots, a_n]} \\
&= [a_0, [a_1, \ldots, a_n]].
\end{aligned}$$

SAGE Example 5.2.3. The `continued_fraction` command computes continued fractions:

```
sage: continued_fraction(17/23)
[0, 1, 2, 1, 5]
sage: continued_fraction(e)
[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1,
 12, 1, 1, 11]
```

Use the optional second argument `bits = n` to determine the precision (in bits) of the input number that is used to compute the continued fraction.

```
sage: continued_fraction(e, bits=20)
[2, 1, 2, 1, 1, 4, 1, 1, 6]
sage: continued_fraction(e, bits=30)
[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1]
```

You can obtain the value of a continued fraction and even do arithmetic with continued fractions:

```
sage: a = continued_fraction(17/23); a
[0, 1, 2, 1, 5]
sage: a.value()
17/23
sage: b = continued_fraction(6/23); b
[0, 3, 1, 5]
sage: a + b
[1]
```

5.2.1 Partial Convergents

Fix a finite continued fraction $[a_0, \dots, a_m]$. We do not assume at this point that the a_i are integers.

Definition 5.2.4 (Partial convergents)**.** For $0 \leq n \leq m$, the nth *convergent* of the continued fraction $[a_0, \dots, a_m]$ is $[a_0, \dots, a_n]$. These convergents for $n < m$ are also called *partial convergents.*

For each n with $-2 \leq n \leq m$, define real numbers p_n and q_n as follows:

$$\begin{array}{llllll} p_{-2} = 0, & p_{-1} = 1, & p_0 = a_0, & \cdots & p_n = a_n p_{n-1} + p_{n-2} & \cdots, \\ q_{-2} = 1, & q_{-1} = 0, & q_0 = 1, & \cdots & q_n = a_n q_{n-1} + q_{n-2} & \cdots. \end{array}$$

Proposition 5.2.5 (Partial Convergents)**.** *For $n \geq 0$ with $n \leq m$ we have*

$$[a_0, \dots, a_n] = \frac{p_n}{q_n}.$$

Proof. We use induction. The assertion is obvious when $n = 0, 1$. Suppose the proposition is true for all continued fractions of length $n - 1$. Then

$$\begin{aligned} [a_0, \dots, a_n] &= [a_0, \dots, a_{n-2}, a_{n-1} + \frac{1}{a_n}] \\ &= \frac{\left(a_{n-1} + \frac{1}{a_n}\right) p_{n-2} + p_{n-3}}{\left(a_{n-1} + \frac{1}{a_n}\right) q_{n-2} + q_{n-3}} \\ &= \frac{(a_{n-1}a_n + 1)p_{n-2} + a_n p_{n-3}}{(a_{n-1}a_n + 1)q_{n-2} + a_n q_{n-3}} \\ &= \frac{a_n(a_{n-1}p_{n-2} + p_{n-3}) + p_{n-2}}{a_n(a_{n-1}q_{n-2} + q_{n-3}) + q_{n-2}} \\ &= \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}} \\ &= \frac{p_n}{q_n}. \end{aligned}$$

□

SAGE Example 5.2.6. If `c` is a continued fraction, use `c.convergents()` to compute a list of the partial convergents of `c`.

```
sage: c = continued_fraction(pi,bits=33); c
[3, 7, 15, 1, 292, 2]
sage: c.convergents()
[3, 22/7, 333/106, 355/113, 103993/33102, 208341/66317]
```

As we will see, the convergents of a continued fraction are the best rational approximations to the value of the continued fraction. In the example above, the listed convergents are the best rational approximations of π with given denominator size.

Proposition 5.2.7. *For $n \geq 0$ with $n \leq m$ we have*

$$p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1} \tag{5.2.2}$$

and

$$p_n q_{n-2} - q_n p_{n-2} = (-1)^n a_n. \tag{5.2.3}$$

Equivalently,

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}}$$

and

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}}.$$

Proof. The case for $n = 0$ is obvious from the definitions. Now suppose $n > 0$ and the statement is true for $n - 1$. Then

$$\begin{aligned} p_n q_{n-1} - q_n p_{n-1} &= (a_n p_{n-1} + p_{n-2}) q_{n-1} - (a_n q_{n-1} + q_{n-2}) p_{n-1} \\ &= p_{n-2} q_{n-1} - q_{n-2} p_{n-1} \\ &= -(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) \\ &= -(-1)^{n-2} = (-1)^{n-1}. \end{aligned}$$

This completes the proof of (5.2.2). For (5.2.3), we have

$$\begin{aligned} p_n q_{n-2} - p_{n-2} q_n &= (a_n p_{n-1} + p_{n-2}) q_{n-2} - p_{n-2}(a_n q_{n-1} + q_{n-2}) \\ &= a_n (p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) \\ &= (-1)^n a_n. \end{aligned}$$

□

Remark 5.2.8. Expressed in terms of matrices, the proposition asserts that the determinant of $\left(\begin{smallmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{smallmatrix}\right)$ is $(-1)^{n-1}$, and of $\left(\begin{smallmatrix} p_n & p_{n-2} \\ q_n & q_{n-2} \end{smallmatrix}\right)$ is $(-1)^n a_n$.

SAGE Example 5.2.9. We use Sage to verify Proposition 5.2.7 for the first few terms of the continued fraction of π.

```
sage: c = continued_fraction(pi); c
[3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 3]
sage: for n in range(-1, len(c)):
...    print c.pn(n)*c.qn(n-1) - c.qn(n)*c.pn(n-1),
1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1
sage: for n in range(len(c)):
...    print c.pn(n)*c.qn(n-2) - c.qn(n)*c.pn(n-2),
3 -7 15 -1 292 -1 1 -1 2 -1 3 -1 14 -3
```

Corollary 5.2.10 (Convergents in lowest terms)**.** *If $[a_0, a_1, \ldots, a_m]$ is a simple continued fraction, so each a_i is an integer, then the p_n and q_n are integers and the fraction p_n/q_n is in lowest terms.*

Proof. It is clear that the p_n and q_n are integers, from the formula that defines them. If d is a positive divisor of both p_n and q_n, then $d \mid (-1)^{n-1}$, so $d = 1$. □

SAGE Example 5.2.11. We illustrate Corollary 5.2.10 using Sage.

```
sage: c = continued_fraction([1,2,3,4,5])
sage: c.convergents()
[1, 3/2, 10/7, 43/30, 225/157]
sage: [c.pn(n) for n in range(len(c))]
[1, 3, 10, 43, 225]
sage: [c.qn(n) for n in range(len(c))]
[1, 2, 7, 30, 157]
```

5.2.2 The Sequence of Partial Convergents

Let $[a_0, \ldots, a_m]$ be a continued fraction and for $n \leq m$ let

$$c_n = [a_0, \ldots, a_n] = \frac{p_n}{q_n}$$

denote the nth convergent. Recall that by definition of continued fraction, $a_n > 0$ for $n > 0$, which gives the partial convergents of a continued fraction additional structure. For example, the partial convergents of $[2, 1, 2, 1, 1, 4, 1, 1, 6]$ are

$$2, 3, 8/3, 11/4, 19/7, 87/32, 106/39, 193/71, 1264/465.$$

To make the size of these numbers clearer, we approximate them using decimals. We also underline every other number, to illustrate some extra structure.

$$\underline{2}, 3, \underline{2.66667}, 2.75000, \underline{2.71429}, 2.71875, \underline{2.71795}, 2.71831, \underline{2.71828}$$

The underlined numbers are smaller than all of the nonunderlined numbers, and the sequence of underlined numbers is strictly increasing, whereas the nonunderlined numbers strictly decrease.

SAGE Example 5.2.12. Figure 5.1 illustrates the above pattern on another continued fraction using Sage.

```
sage: c = continued_fraction([1,1,1,1,1,1,1,1])
sage: v = [(i, c.pn(i)/c.qn(i)) for i in range(len(c))]
sage: P = point(v, rgbcolor=(0,0,1), pointsize=40)
sage: L = line(v, rgbcolor=(0.5,0.5,0.5))
sage: L2 = line([(0,c.value()),(len(c)-1,c.value())], \
...      thickness=0.5, rgbcolor=(0.7,0,0))
sage: (L+L2+P).show(xmin=0,ymin=1)
```

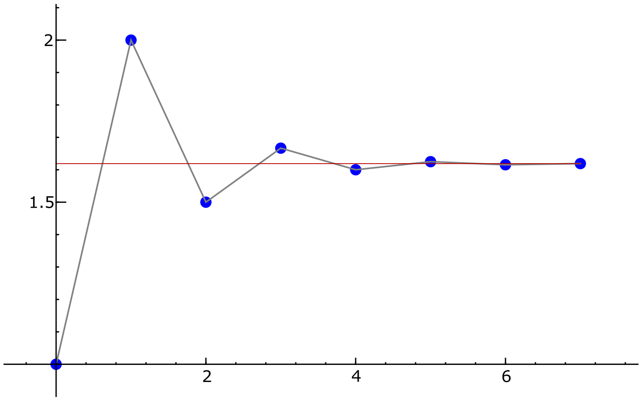


FIGURE 5.1. Graph of a Continued Fraction

We next prove that this extra structure is a general phenomenon.

Proposition 5.2.13 (How Convergents Converge)**.** *The even indexed convergents c_{2n} increase strictly with n, and the odd indexed convergents c_{2n+1} decrease strictly with n. Also, the odd indexed convergents c_{2n+1} are greater than all of the even indexed convergents c_{2m}.*

Proof. The a_n are positive for $n \geq 1$, so the q_n are positive. By Proposition 5.2.7, for $n \geq 2$,

$$c_n - c_{n-2} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}},$$

which proves the first claim.

Suppose for the sake of contradiction that there exist integers r and m such that $c_{2m+1} < c_{2r}$. Proposition 5.2.7 implies that for $n \geq 1$,

$$c_n - c_{n-1} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}}$$

has sign $(-1)^{n-1}$, so for all $s \geq 0$ we have $c_{2s+1} > c_{2s}$. Thus it is impossible that $r = m$. If $r < m$, then by what we proved in the first paragraph, $c_{2m+1} < c_{2r} < c_{2m}$, a contradiction (with $s = m$). If $r > m$, then $c_{2r+1} < c_{2m+1} < c_{2r}$, which is also a contradiction (with $s = r$). □

5.2.3 Every Rational Number is Represented

Proposition 5.2.14 (Rational Continued Fractions)**.** *Every nonzero rational number can be represented by a simple continued fraction.*

Proof. Without loss of generality, we may assume that the rational number is a/b, with $b \geq 1$ and $\gcd(a, b) = 1$. Algorithm 1.1.13 gives:

$$\begin{aligned} a &= b \cdot a_0 + r_1, & 0 < r_1 < b \\ b &= r_1 \cdot a_1 + r_2, & 0 < r_2 < r_1 \\ &\cdots \\ r_{n-2} &= r_{n-1} \cdot a_{n-1} + r_n, & 0 < r_n < r_{n-1} \\ r_{n-1} &= r_n \cdot a_n + 0. \end{aligned}$$

Note that $a_i > 0$ for $i > 0$ (also $r_n = 1$, since $\gcd(a, b) = 1$). Rewrite the equations as follows:

$$\begin{aligned} a/b &= a_0 + r_1/b = a_0 + 1/(b/r_1), \\ b/r_1 &= a_1 + r_2/r_1 = a_1 + 1/(r_1/r_2), \\ r_1/r_2 &= a_2 + r_3/r_2 = a_2 + 1/(r_2/r_3), \\ &\cdots \\ r_{n-1}/r_n &= a_n. \end{aligned}$$

It follows that

$$\frac{a}{b} = [a_0, a_1, \ldots, a_n].$$

□

The proof of Proposition 5.2.14 leads to an algorithm for computing the continued fraction of a rational number.

A nonzero rational number can be represented in exactly two ways; for example, $2 = [1, 1] = [2]$ (see Exercise 5.2).

5.3 Infinite Continued Fractions

This section begins with the continued fraction procedure, which associates a sequence $a_0, a_1, \ldots$ of integers to a real number x. After giving several examples, we prove that $x = \lim_{n\to\infty}[a_0, a_1, \ldots, a_n]$ by proving that the odd and even partial convergents become arbitrarily close to each other. We also show that if $a_0, a_1, \ldots$ is any infinite sequence of positive integers, then the sequence of $c_n = [a_0, a_1, \ldots, a_n]$ converges. More generally, if a_n is an arbitrary sequence of positive reals such that $\sum_{n=0}^{\infty} a_n$ diverges then (c_n) converges.

5.3.1 The Continued Fraction Procedure

Let $x \in \mathbf{R}$ and write

$$x = a_0 + t_0$$

with $a_0 \in \mathbf{Z}$ and $0 \leq t_0 < 1$. We call the number a_0 the *floor* of x, and we also sometimes write $a_0 = \lfloor x \rfloor$. If $t_0 \neq 0$, write

$$\frac{1}{t_0} = a_1 + t_1$$

with $a_1 \in \mathbf{N}$ and $0 \leq t_1 < 1$. Thus $t_0 = \frac{1}{a_1+t_1} = [0, a_1 + t_1]$, which is a continued fraction expansion of t_0, which need not be simple. Continue in this manner so long as $t_n \neq 0$ writing

$$\frac{1}{t_n} = a_{n+1} + t_{n+1}$$

with $a_{n+1} \in \mathbf{N}$ and $0 \leq t_{n+1} < 1$. We call this procedure, which associates to a real number x the sequence of integers $a_0, a_1, a_2, \ldots$, the *continued fraction process.*

Example 5.3.1. Let $x = \frac{8}{3}$. Then $x = 2 + \frac{2}{3}$, so $a_0 = 2$ and $t_0 = \frac{2}{3}$. Then $\frac{1}{t_0} = \frac{3}{2} = 1 + \frac{1}{2}$, so $a_1 = 1$ and $t_1 = \frac{1}{2}$. Then $\frac{1}{t_1} = 2$, so $a_2 = 2$, $t_2 = 0$, and the sequence terminates. Notice that

$$\frac{8}{3} = [2, 1, 2],$$

so the continued fraction procedure produces the continued fraction of $\frac{8}{3}$.

Example 5.3.2. Let $x = \frac{1+\sqrt{5}}{2}$. Then

$$x = 1 + \frac{-1+\sqrt{5}}{2},$$

so $a_0 = 1$ and $t_0 = \frac{-1+\sqrt{5}}{2}$. We have

$$\frac{1}{t_0} = \frac{2}{-1+\sqrt{5}} = \frac{-2-2\sqrt{5}}{-4} = \frac{1+\sqrt{5}}{2},$$

so $a_1 = 1$ and $t_1 = \frac{-1+\sqrt{5}}{2}$. Likewise, $a_n = 1$ for all n. As we will see below, the following exciting equality makes sense.

$$\frac{1+\sqrt{5}}{2} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}}$$

SAGE Example 5.3.3. The equality of Example 5.3.2 is consistent with the following Sage calculation:

```
sage: def cf(bits):
...     x = (1 + sqrt(RealField(bits)(5))) / 2
...     return continued_fraction(x)
sage: cf(10)
[1, 1, 1, 1, 1, 1, 1, 3]
sage: cf(30)
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 2]
sage: cf(50)
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

Example 5.3.4. Suppose $x = e = 2.71828182\ldots$. Using the continued fraction procedure, we find that

$$a_0, a_1, a_2, \ldots = 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots$$

For example, $a_0 = 2$ is the floor of 2. Subtracting 2 and inverting, we obtain $1/0.718\ldots = 1.3922\ldots$, so $a_1 = 1$. Subtracting 1 and inverting yields $1/0.3922\ldots = 2.5496\ldots$, so $a_2 = 2$. We will prove in Section 5.4 that the continued fraction of e obeys a simple pattern.

The 5th partial convergent of the continued fraction of e is

$$[a_0, a_1, a_2, a_3, a_4, a_5] = \frac{87}{32} = 2.71875,$$

which is a good rational approximation to e, in the sense that

$$\left|\frac{87}{32} - e\right| = 0.000468\ldots.$$

Note that $0.000468\ldots < 1/32^2 = 0.000976\ldots$, which illustrates the bound in Corollary 5.3.11.

Let's do the same thing with $\pi = 3.14159265358979\ldots$. Applying the continued fraction procedure, we find that the continued fraction of π is

$$a_0, a_1, a_2, \ldots = 3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, \ldots$$

The first few partial convergents are

$$3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \cdots$$

These are good rational approximations to π; for example,

$$\frac{103993}{33102} = 3.14159265301\ldots.$$

Notice that the continued fraction of e exhibits a nice pattern (see Section 5.4 for a proof), whereas the continued fraction of π exhibits no pattern that is obvious to the author. The continued fraction of π has been extensively studied, and over 20 million terms have been computed. The data suggests that every integer appears infinitely often as a partial convergent. For much more about the continued fraction of π, or of any other sequence in this book, type the first few terms of the sequence into [Slo].

5.3.2 Convergence of Infinite Continued Fractions

Lemma 5.3.5. *For every n such that a_n is defined, we have*

$$x = [a_0, a_1, \ldots, a_n + t_n],$$

and if $t_n \neq 0$, then $x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}]$.

Proof. We use induction. The statements are both true when $n = 0$. If the second statement is true for $n - 1$, then

$$\begin{aligned} x &= \left[a_0, a_1, \ldots, a_{n-1}, \frac{1}{t_{n-1}}\right] \\ &= [a_0, a_1, \ldots, a_{n-1}, a_n + t_n] \\ &= \left[a_0, a_1, \ldots, a_{n-1}, a_n, \frac{1}{t_n}\right]. \end{aligned}$$

Similarly, the first statement is true for n if it is true for $n - 1$. □

Theorem 5.3.6 (Continued Fraction Limit)**.** *Let $a_0, a_1, \ldots$ be a sequence of integers such that $a_n > 0$ for all $n \geq 1$, and for each $n \geq 0$, set $c_n = [a_0, a_1, \ldots a_n]$. Then $\lim\limits_{n\to\infty} c_n$ exists.*

Proof. For any $m \geq n$, the number c_n is a partial convergent of $[a_0, \ldots, a_m]$. By Proposition 5.2.13, the even convergents c_{2n} form a strictly *increasing* sequence and the odd convergents c_{2n+1} form a strictly *decreasing* sequence. Moreover, the even convergents are all $\leq c_1$ and the odd convergents are all $\geq c_0$. Hence $\alpha_0 = \lim_{n\to\infty} c_{2n}$ and $\alpha_1 = \lim_{n\to\infty} c_{2n+1}$ both exist, and $\alpha_0 \leq \alpha_1$. Finally, by Proposition 5.2.7

$$|c_{2n} - c_{2n-1}| = \frac{1}{q_{2n} \cdot q_{2n-1}} \leq \frac{1}{2n(2n-1)} \to 0,$$

so $\alpha_0 = \alpha_1$. □

We define

$$[a_0, a_1, \ldots] = \lim_{n\to\infty} c_n.$$

Example 5.3.7. We illustrate the theorem with $x = \pi$. As in the proof of Theorem 5.3.6, let c_n be the nth partial convergent to π. The c_n with n odd converge down to π

$$c_1 = 3.1428571\ldots,\ c_3 = 3.1415929\ldots,\ c_5 = 3.1415926\ldots$$

whereas the c_n with n even converge up to π

$$c_2 = 3.1415094\ldots,\ c_4 = 3.1415926\ldots,\ c_6 = 3.1415926\ldots.$$

Theorem 5.3.8. *Let $a_0, a_1, a_2, \ldots$ be a sequence of real numbers such that $a_n > 0$ for all $n \geq 1$, and for each $n \geq 0$, set $c_n = [a_0, a_1, \ldots a_n]$. Then $\lim\limits_{n\to\infty} c_n$ exists if and only if the sum $\sum_{n=0}^{\infty} a_n$ diverges.*

Proof. We only prove that if $\sum a_n$ diverges, then $\lim_{n\to\infty} c_n$ exists. A proof of the converse can be found in [Wal48, Ch. 2, Thm. 6.1].

Let q_n be the sequence of "denominators" of the partial convergents, as defined in Section 5.2.1, so $q_{-2} = 1$, $q_{-1} = 0$, and for $n \geq 0$, we have

$$q_n = a_n q_{n-1} + q_{n-2}.$$

As we saw in the proof of Theorem 5.3.6, the limit $\lim_{n\to\infty} c_n$ exists provided that the sequence $\{q_n q_{n-1}\}$ diverges to positive infinity.

For n even,

$$\begin{aligned} q_n &= a_n q_{n-1} + q_{n-2} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + q_{n-4} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + a_{n-4} q_{n-5} + q_{n-6} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + \cdots + a_2 q_1 + q_0 \end{aligned}$$

and for n odd,

$$q_n = a_n q_{n-1} + a_{n-2} q_{n-3} + \cdots + a_1 q_0 + q_{-1}.$$

Since $a_n > 0$ for $n > 0$, the sequence $\{q_n\}$ is increasing, so $q_i \geq 1$ for all $i \geq 0$. Applying this fact to the above expressions for q_n, we see that for n even

$$q_n \geq a_n + a_{n-2} + \cdots + a_2,$$

and for n odd

$$q_n \geq a_n + a_{n-2} + \cdots + a_1.$$

If $\sum a_n$ diverges, then at least one of $\sum a_{2n}$ or $\sum a_{2n+1}$ must diverge. The above inequalities then imply that at least one of the sequences $\{q_{2n}\}$ or $\{q_{2n+1}\}$ diverge to infinity. Since $\{q_n\}$ is an increasing sequence, it follows that $\{q_n q_{n-1}\}$ diverges to infinity. □

Example 5.3.9. Let $a_n = \frac{1}{n\log(n)}$ for $n \geq 2$ and $a_0 = a_1 = 0$. By the integral test, $\sum a_n$ diverges, so by Theorem 5.3.8, the continued fraction $[a_0, a_1, a_2, \ldots]$ converges. This convergence is very slow, since, e.g.

$$[a_0, a_1, \ldots, a_{9999}] = 0.5750039671012225425930\ldots$$

yet

$$[a_0, a_1, \ldots, a_{10000}] = 0.7169153932917378550424\ldots.$$

Theorem 5.3.10. *Let $x \in \mathbf{R}$ be a real number. Then x is the value of the (possibly infinite) simple continued fraction $[a_0, a_1, a_2, \ldots]$ produced by the continued fraction procedure.*

Proof. If the sequence is finite, then some $t_n = 0$ and the result follows by Lemma 5.3.5. Suppose the sequence is infinite. By Lemma 5.3.5,

$$x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}].$$

By Proposition 5.2.5 (which we apply in a case when the partial quotients of the continued fraction are not integers), we have

$$x = \frac{\dfrac{1}{t_n} \cdot p_n + p_{n-1}}{\dfrac{1}{t_n} \cdot q_n + q_{n-1}}.$$

Thus, if $c_n = [a_0, a_1, \ldots, a_n]$, then

$$\begin{aligned} x - c_n &= x - \frac{p_n}{q_n} \\ &= \frac{\frac{1}{t_n} p_n q_n + p_{n-1} q_n - \frac{1}{t_n} p_n q_n - p_n q_{n-1}}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)}. \\ &= \frac{p_{n-1} q_n - p_n q_{n-1}}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)} \\ &= \frac{(-1)^n}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)}. \end{aligned}$$

Thus

$$\begin{aligned} |x - c_n| &= \frac{1}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)} \\ &< \frac{1}{q_n (a_{n+1} q_n + q_{n-1})} \\ &= \frac{1}{q_n \cdot q_{n+1}} \leq \frac{1}{n(n+1)} \to 0. \end{aligned}$$

In the inequality, we use that a_{n+1} is the integer part of $\frac{1}{t_n}$, and is hence $\leq \frac{1}{t_n} < 1$, since $t_n < 1$. □

This corollary follows from the proof of Theorem 5.3.10.

Corollary 5.3.11 (Convergence of continued fraction)**.** *Let $a_0, a_1, \ldots$ define a simple continued fraction, and let $x = [a_0, a_1, \ldots] \in \mathbf{R}$ be its value. Then for all m,*

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}}.$$

Proposition 5.3.12. *If x is a rational number, then the sequence $a_0, a_1, \ldots$ produced by the continued fraction procedure terminates.*

Proof. Let $[b_0, b_1, \ldots, b_m]$ be the continued fraction representation of x that we obtain using Algorithm 1.1.13, so the b_i are the partial quotients at each step. If $m = 0$, then x is an integer, so we may assume $m > 0$. Then

$$x = b_0 + 1/[b_1, \ldots, b_m].$$

If $[b_1, \ldots, b_m] = 1$, then $m = 1$ and $b_1 = 1$, which will not happen using Algorithm 1.1.13, since it would give $[b_0 + 1]$ for the continued fraction of the integer $b_0 + 1$. Thus $[b_1, \ldots, b_m] > 1$, so in the continued fraction algorithm we choose $a_0 = b_0$ and $t_0 = 1/[b_1, \ldots, b_m]$. Repeating this argument enough times proves the claim. □

5.4 The Continued Fraction of e

The continued fraction expansion of e begins $[2, 1, 2, 1, 1, 4, 1, 1, 6, \ldots]$. The obvious pattern in fact does continue, as Euler proved in 1737 (see [Eul85]), and we will prove in this section. As an application, Euler gave a proof that e is irrational by noting that its continued fraction is infinite.

The proof we give below draws heavily on the proof in [Coh], which describes a slight variant of a proof of Hermite (see [Old70]). The continued fraction representation of e is also treated in the German book [Per57], but the proof requires substantial background from elsewhere in that text.

5.4.1 Preliminaries

First, we write the continued fraction of e in a slightly different form. Instead of $[2, 1, 2, 1, 1, 4, \ldots]$, we can start the sequence of coefficients

$$[1, 0, 1, 1, 2, 1, 1, 4, \ldots]$$

to make the pattern the same throughout. (Everywhere else in this chapter we assume that the partial quotients a_n for $n \geq 1$ are positive, but temporarily relax that condition here and allow $a_1 = 0$.) The numerators and denominators of the convergents given by this new sequence satisfy a simple recurrence. Using r_i as a stand-in for p_i or q_i, we have

$$\begin{aligned} r_{3n} &= r_{3n-1} + r_{3n-2} \\ r_{3n-1} &= r_{3n-2} + r_{3n-3} \\ r_{3n-2} &= 2(n-1)r_{3n-3} + r_{3n-4}. \end{aligned}$$

Our first goal is to collapse these three recurrences into one recurrence that only makes mention of r_{3n}, r_{3n-3}, and r_{3n-6}. We have

$$\begin{aligned} r_{3n} &= r_{3n-1} + r_{3n-2} \\ &= (r_{3n-2} + r_{3n-3}) + (2(n-1)r_{3n-3} + r_{3n-4}) \\ &= (4n-3)r_{3n-3} + 2r_{3n-4}. \end{aligned}$$

This same method of simplification also shows us that

$$r_{3n-3} = 2r_{3n-7} + (4n-7)r_{3n-6}.$$

To get rid of $2r_{3n-4}$ in the first equation, we make the substitutions

$$\begin{aligned} 2r_{3n-4} &= 2(r_{3n-5} + r_{3n-6}) \\ &= 2((2(n-2)r_{3n-6} + r_{3n-7}) + r_{3n-6}) \\ &= (4n-6)r_{3n-6} + 2r_{3n-7}. \end{aligned}$$

Substituting for $2r_{3n-4}$ and then $2r_{3n-7}$, we finally have the needed collapsed recurrence,

$$r_{3n} = 2(2n-1)r_{3n-3} + r_{3n-6}.$$

5.4.2 Two Integral Sequences

We define the sequences $x_n = p_{3n}$, $y_n = q_{3n}$. Since the $3n$-convergents will converge to the same real number that the n convergents do, x_n/y_n also converges to the limit of the continued fraction. Each sequence $\{x_n\}$, $\{y_n\}$ will obey the recurrence relation derived in the previous section (where z_n is a stand-in for x_n or y_n):

$$z_n = 2(2n-1)z_{n-1} + z_{n-2}, \text{ for all } n \geq 2. \qquad (5.4.1)$$

The two sequences can be found in Table 5.1. (The initial conditions $x_0 = 1$, $x_1 = 3$, $y_0 = y_1 = 1$ are taken straight from the first few convergents of the original continued fraction.) Notice that since we are skipping several convergents at each step, the ratio x_n/y_n converges to e very quickly.

TABLE 5.1. Convergents

n	0	1	2	3	4	$\cdots$
x_n	1	3	19	193	2721	$\cdots$
y_n	1	1	7	71	1001	$\cdots$
x_n/y_n	1	3	2.714...	2.71830...	2.7182817...	$\cdots$

5.4.3 A Related Sequence of Integrals

Now, we define a sequence of real numbers $T_0, T_1, T_2, \ldots$ by the following integrals:

$$T_n = \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt.$$

Below, we compute the first two terms of this sequence explicitly. (When we compute T_1, we are doing the integration by parts $u = t(t-1)$, $dv = e^t dt$. Since the integral runs from 0 to 1, the boundary condition is 0 when evaluated at each of the endpoints. This vanishing will be helpful when we do the integral in the general case.)

$$\begin{aligned} T_0 &= \int_0^1 e^t dt = e - 1, \\ T_1 &= \int_0^1 t(t-1)e^t dt \\ &= -\int_0^1 ((t-1)+t)e^t dt \\ &= -(t-1)e^t \Big|_0^1 - te^t \Big|_0^1 + 2\int_0^1 e^t dt \\ &= -1 - e + 2(e-1) = e - 3. \end{aligned}$$

The reason that we defined this series now becomes apparent: $T_0 = y_0 e - x_0$ and $T_1 = y_1 e - x_1$. In general, it will be true that $T_n = y_n e - x_n$. We will now prove this fact.

It is clear that if T_n were to satisfy the same recurrence that the x_i and y_i do in (5.4.1), then the above statement holds by induction. (The initial conditions are correct, as needed.) So, we simplify T_n by integrating by parts twice in succession:

$$\begin{aligned}
T_n &= \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt \\
&= -\int_0^1 \frac{t^{n-1}(t-1)^n + t^n(t-1)^{n-1}}{(n-1)!}\ e^t dt \\
&= \int_0^1 \Big(\frac{t^{n-2}(t-1)^n}{(n-2)!} + n\frac{t^{n-1}(t-1)^{n-1}}{(n-1)!} \\
&\qquad\qquad + n\frac{t^{n-1}(t-1)^{n-1}}{(n-1)!} + \frac{t^n(t-1)^{n-2}}{(n-2)!}\Big)e^t dt \\
&= 2nT_{n-1} + \int_0^1 \frac{t^{n-2}(t-1)^{n-2}}{n-2!}(2t^2-2t+1)\ e^t dt \\
&= 2nT_{n-1} + 2\int_0^1 \frac{t^{n-1}(t-1)^{n-1}}{n-2!}\ e^t dt + \int_0^1 \frac{t^{n-2}(t-1)^{n-2}}{n-2!}\ e^t dt \\
&= 2nT_{n-1} + 2(n-1)T_{n-1} + T_{n-2} \\
&= 2(2n-1)T_{n-1} + T_{n-2},
\end{aligned}$$

which is the desired recurrence.

Therefore, $T_n = y_n e - x_n$. To conclude the proof, we consider the limit as n approaches infinity:

$$\lim_{n\to\infty} \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt = 0,$$

by inspection, and therefore

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \lim_{n\to\infty} (e - \frac{T_n}{y_n}) = e.$$

Therefore, the ratio x_n/y_n approaches e, and the continued fraction expansion $[2, 1, 2, 1, 1, 4, 1, 1, \ldots]$ does in fact converge to e.

5.4.4 Extensions of the Argument

The method of proof of this section generalizes to show that the continued fraction expansion of $e^{1/n}$ is

$$[1,\ (n-1),\ 1,\ 1,\ (3n-1),\ 1,\ 1,\ (5n-1),\ 1,\ 1,\ (7n-1), \ldots]$$

for all $n \in \mathbf{N}$ (see Exercise 5.6).

5.5 Quadratic Irrationals

The main result of this section is that the continued fraction expansion of a number is eventually repeating if and only if the number is a quadratic irrational. This can be viewed as an analog for continued fractions of the familiar fact that the decimal expansion of x is eventually repeating if and only if x is rational. The proof that continued fractions of quadratic irrationals eventually repeats is surprisingly difficult and involves an interesting finiteness argument. Section 5.5.2 emphasizes our striking ignorance about continued fractions of real roots of irreducible polynomials over $\mathbf{Q}$ of degree bigger than 2.

Definition 5.5.1 (Quadratic Irrational)**.** A *quadratic irrational* is a real number $\alpha \in \mathbf{R}$ that is irrational and satisfies a quadratic polynomial with coefficients in $\mathbf{Q}$.

Thus, for example, $(1+\sqrt{5})/2$ is a quadratic irrational. Recall that

$$\frac{1+\sqrt{5}}{2} = [1, 1, 1, \ldots].$$

The continued fraction of $\sqrt{2}$ is $[1, 2, 2, 2, 2, 2, \ldots]$, and the continued fraction of $\sqrt{389}$ is

$$[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38, \ldots].$$

Does the $[1, 2, 1, 1, 1, 1, 2, 1, 38]$ pattern repeat over and over again?

SAGE Example 5.5.2. We compute more terms of the continued fraction expansion of $\sqrt{389}$ using Sage:

```
sage: def cf_sqrt_d(d, bits):
...   x = sqrt(RealField(bits)(d))
...   return continued_fraction(x)
sage: cf_sqrt_d(389,50)
[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38]
sage: cf_sqrt_d(389,100)
[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38,
 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1,
 2, 1, 1]
```

5.5.1 Periodic Continued Fractions

Definition 5.5.3 (Periodic Continued Fraction)**.** A *periodic continued fraction* is a continued fraction $[a_0, a_1, \ldots, a_n, \ldots]$ such that

$$a_n = a_{n+h}$$

for some fixed positive integer h and all sufficiently large n. We call the minimal such h the *period of the continued fraction.*

Example 5.5.4. Consider the periodic continued fraction $[1,2,1,2,\ldots] = [\overline{1,2}]$. What does it converge to? We have

$$[\overline{1,2}] = 1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cdots}}}},$$

so if $\alpha = [\overline{1,2}]$ then

$$\alpha = 1 + \cfrac{1}{2 + \cfrac{1}{\alpha}} = 1 + \cfrac{1}{\cfrac{2\alpha+1}{\alpha}} = 1 + \frac{\alpha}{2\alpha+1} = \frac{3\alpha+1}{2\alpha+1}$$

Thus $2\alpha^2 - 2\alpha - 1 = 0$, so

$$\alpha = \frac{1+\sqrt{3}}{2}.$$

Theorem 5.5.5 (Periodic Characterization)**.** *An infinite simple continued fraction is periodic if and only if it represents a quadratic irrational.*

Proof. ($\Longrightarrow$) First suppose that

$$[a_0, a_1, \ldots, a_n, \overline{a_{n+1}, \ldots, a_{n+h}}]$$

is a periodic continued fraction. Set $\alpha = [a_{n+1}, a_{n+2}, \ldots]$. Then

$$\alpha = [a_{n+1}, \ldots, a_{n+h}, \alpha],$$

so by Proposition 5.2.5

$$\alpha = \frac{\alpha p_{n+h} + p_{n+h-1}}{\alpha q_{n+h} + q_{n+h-1}}.$$

Here we use that α is the last partial quotient. Thus, α satisfies a quadratic equation with coefficients in $\mathbf{Q}$. Computing as in Example 5.5.4 and rationalizing the denominators, and using that the a_i are all integers, shows that

$$\begin{aligned}[a_0, a_1, \ldots] &= [a_0, a_1, \ldots, a_n, \alpha] \\ &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{\alpha}}}\end{aligned}$$

is of the form $c + d\alpha$, with $c, d \in \mathbf{Q}$, so $[a_0, a_1, \ldots]$ also satisfies a quadratic polynomial over $\mathbf{Q}$.

The continued fraction procedure applied to the value of an infinite simple continued fraction yields that continued fraction back, so by Proposition 5.3.12, $\alpha \notin \mathbf{Q}$ because it is the value of an infinite continued fraction.

($\Longleftarrow$) Suppose $\alpha \in \mathbf{R}$ is an irrational number that satisfies a quadratic equation

$$a\alpha^2 + b\alpha + c = 0 \tag{5.5.1}$$

with $a, b, c \in \mathbf{Z}$ and $a \neq 0$. Let $[a_0, a_1, \ldots]$ be the continued fraction expansion of α. For each n, let

$$r_n = [a_n, a_{n+1}, \ldots],$$

so

$$\alpha = [a_0, a_1, \ldots, a_{n-1}, r_n].$$

We will prove periodicity by showing that the set of r_n's is finite. If we have shown finiteness, then there exists $n, h > 0$ such that $r_n = r_{n+h}$, so

$$\begin{aligned}
[a_0, \ldots, a_{n-1}, r_n] &= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, r_{n+h}] \\
&= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, r_n] \\
&= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, a_n, \ldots, a_{n+h-1}, r_{n+h}] \\
&= [a_0, \ldots, a_{n-1}, \overline{a_n, \ldots, a_{n+h-1}}].
\end{aligned}$$

It remains to show there are only finitely many distinct r_n. We have

$$\alpha = \frac{p_n}{q_n} = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}.$$

Substituting this expression for α into the quadratic equation (5.5.1), we see that

$$A_n r_n^2 + B_n r_n + C_n = 0,$$

where

$$\begin{aligned}
A_n &= a p_{n-1}^2 + b p_{n-1} q_{n-1} + c q_{n-1}^2, \\
B_n &= 2a p_{n-1} p_{n-2} + b(p_{n-1} q_{n-2} + p_{n-2} q_{n-1}) + 2c q_{n-1} q_{n-2}, \text{ and} \\
C_n &= a p_{n-2}^2 + b p_{n-2} q_{n-2} + c p_{n-2}^2.
\end{aligned}$$

Note that $A_n, B_n, C_n \in \mathbf{Z}$, that $C_n = A_{n-1}$, and that

$$B_n^2 - 4A_n C_n = (b^2 - 4ac)(p_{n-1} q_{n-2} - q_{n-1} p_{n-2})^2 = b^2 - 4ac.$$

Recall from the proof of Theorem 5.3.10 that

$$\left| \alpha - \frac{p_{n-1}}{q_{n-1}} \right| < \frac{1}{q_n q_{n-1}}.$$

Thus,

$$|\alpha q_{n-1} - p_{n-1}| < \frac{1}{q_n} < \frac{1}{q_{n-1}},$$

so

$$p_{n-1} = \alpha q_{n-1} + \frac{\delta}{q_{n-1}} \qquad \text{with } |\delta| < 1.$$

Hence,

$$\begin{aligned} A_n &= a\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right)^2 + b\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right) q_{n-1} + c q_{n-1}^2 \\ &= (a\alpha^2 + b\alpha + c) q_{n-1}^2 + 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta \\ &= 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta. \end{aligned}$$

Thus,

$$|A_n| = \left|2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta\right| < 2|a\alpha| + |a| + |b|.$$

We conclude that there are only finitely many possibilities for the integer A_n. Also,

$$|C_n| = |A_{n-1}| \quad \text{and} \quad |B_n| = \sqrt{b^2 - 4(ac - A_n C_n)},$$

so there are only finitely many triples (A_n, B_n, C_n), and hence only finitely many possibilities for r_n as n varies, which completes the proof. (The proof above closely follows [HW79, Thm. 177, pg.144–145].) □

5.5.2 Continued Fractions of Algebraic Numbers of Higher Degree

Definition 5.5.6 (Algebraic Number)**.** An *algebraic number* is a root of a polynomial $f \in \mathbf{Q}[x]$.

Open Problem 5.5.7. *Give a simple description of the complete continued fractions expansion of the algebraic number $\sqrt[3]{2}$. It begins*

$$[1, 3, 1, 5, 1, 1, 4, 1, 1, 8, 1, 14, 1, 10, 2, 1, 4, 12, 2, 3, 2, 1, 3, 4, 1, 1, 2, 14, \\ 3, 12, 1, 15, 3, 1, 4, 534, 1, 1, 5, 1, 1, \ldots]$$

The author does not see a pattern, and the 534 reduces his confidence that he will. Lang and Trotter (see [LT72]) analyzed many terms of the continued fraction of $\sqrt[3]{2}$ statistically, and their work suggests that $\sqrt[3]{2}$ has an "unusual" continued fraction; later work in [LT74] suggests that maybe it does not.

Khintchine (see [Khi63, pg. 59])

> No properties of the representing continued fractions, analogous to those which have just been proved, are known for algebraic numbers of higher degree [as of 1963]. [...] It is of interest to point out that up till the present time *no continued fraction development of an algebraic number of higher degree than the second is known* [emphasis added]. It is not even known if such a development has bounded elements. Generally speaking the problems associated with the continued fraction expansion of algebraic numbers of degree higher than the second are extremely difficult and virtually unstudied.

Richard Guy (see [Guy94, pg. 260])

> Is there an algebraic number of degree greater than two whose simple continued fraction has unbounded partial quotients? Does every such number have unbounded partial quotients?

Baum and Sweet [BS76] answered the analog of Richard Guy's question, but with algebraic numbers replaced by elements of a field K other than $\mathbf{Q}$. (The field K is $\mathbf{F}_2((1/x))$, the field of Laurent series in the variable $1/x$ over the finite field with two elements. An element of K is a polynomial in x plus a formal power series in $1/x$.) They found an α of degree 3 over K whose continued fraction has all terms of bounded degree, and other elements of various degrees greater than 2 over K whose continued fractions have terms of unbounded degree.

5.6 Recognizing Rational Numbers

Suppose that somehow you can compute approximations to some rational number, and want to figure what the rational number probably is. Computing the approximation to high enough precision to find a period in the decimal expansion is not a good approach, because the period can be huge (see below). A much better approach is to compute the simple continued fraction of the approximation, and truncate it before a large partial quotient a_n, then compute the value of the truncated continued fraction. This results in a rational number that has a relatively small numerator and denominator, and is close to the approximation of the rational number, since the tail end of the continued fraction is at most $1/a_n$.

We begin with a contrived example, which illustrates how to recognize a rational number. Let

$$x = 9495/3847 = 2.46815700545879906420587470756433584611385\ldots.$$

The continued fraction of the truncation 2.468157005458799064 is

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 328210621945, 2, 1, 1, 1, \ldots]$$

We have

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1] = \frac{9495}{3847}.$$

Notice that no repetition is evident in the digits of x given above, though we know that the decimal expansion of x must be eventually periodic, since all decimal expansions of rational numbers are eventually periodic. In fact, the length of the period of the decimal expansion of $1/3847$ is 3846, which is the order of 10 modulo 3847 (see Exercise 5.7).

For a slightly less contrived application of this idea, suppose $f(x) \in \mathbf{Z}[x]$ is a polynomial with integer coefficients, and we know for some reason that one root of f is a rational number. We can find that rational number, by using Newton's method to approximate each root, and continued fractions to decide whether each root is a rational number (we can substitute the value of the continued fraction approximation into f to see if it is actually a root). One could also use the well-known Rational Root Theorem, which asserts that any rational root n/d of f, with $n, d \in \mathbf{Z}$ coprime, has the property that n divides the constant term of f and d the leading coefficient of f. However, using that theorem to find n/d would require factoring the constant and leading terms of f, which could be completely impractical if they have a few hundred digits (see Section 1.1.3). In contrast, Newton's method and continued fractions should quickly find n/d, assuming the degree of f isn't too large.

For example, suppose $f = 3847x^2 - 14808904x + 36527265$. To apply Newton's method, let x_0 be a guess for a root of f. Iterate using the recurrence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Choosing $x_0 = 0$, approximations of the first two iterates are

$$x_1 = 2.466574501394566404103909378,$$

and

$$x_2 = 2.466815700480740192304316684 6.$$

The continued fraction of the approximations x_1 and x_2 are

$$[2, 2, 6, 1, 47, 2, 1, 4, 3, 1, 5, 8, 2, 3]$$

and

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 103, 8, 1, 2, 3, \ldots].$$

Truncating the continued fraction of x_2 before 103 gives

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1],$$

which evaluates to $9495/3847$, which is a rational root of f.

SAGE Example 5.6.1. We do the above calculation using SAGE. First we implement the Newton iteration:

```
sage: def newton_root(f, iterates=2, x0=0, prec=53):
...     x = RealField(prec)(x0)
...     R = PolynomialRing(ZZ,'x')
...     f = R(f)
...     g = f.derivative()
...     for i in range(iterates):
...         x = x - f(x)/g(x)
...     return x
```

Next we run the Newton iteration, and compute the continued fraction of the result:

```
sage: a = newton_root(3847*x^2 - 14808904*x + 36527265); a
2.46815700480740
sage: cf = continued_fraction(a); cf
[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 103, 8, 1, 2, 3, 1, 1]
```

We truncate the continued fraction and compute its value.

```
sage: c = cf[:12]; c
[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1]
sage: c.value()
9495/3847
```

Another computational application of continued fractions, which we can only hint at, is that there are functions in certain parts of advanced number theory (that are beyond the scope of this book) that take rational values at certain points, and which can only be computed efficiently via approximations; using continued fractions as illustrated above to evaluate such functions is crucial.

5.7 Sums of Two Squares

In this section, we apply continued fractions to prove the following theorem.

Theorem 5.7.1. *A positive integer n is a sum of two squares if and only if all prime factors of $p \mid n$ such that $p \equiv 3 \pmod 4$ have even exponent in the prime factorization of n.*

We first consider some examples. Notice that $5 = 1^2 + 2^2$ is a sum of two squares, but 7 is not a sum of two squares. Since 2001 is divisible by 3 (because $2 + 1$ is divisible by 3), but not by 9 (since $2 + 1$ is not), Theorem 5.7.1 implies that 2001 is not a sum of two squares. The theorem also implies that $2 \cdot 3^4 \cdot 5 \cdot 7^2 \cdot 13$ is a sum of two squares.

SAGE Example 5.7.2. We use Sage to write a short program that *naively* determines whether or not an integer n is a sum of two squares, and if so returns a, b such that $a^2 + b^2 = n$.

```
sage: def sum_of_two_squares_naive(n):
...    for i in range(int(sqrt(n))):
...        if is_square(n - i^2):
...            return i, (Integer(n-i^2)).sqrt()
...    return "%s is not a sum of two squares"%n
```

We next use our function in a couple of cases.

```
sage: sum_of_two_squares_naive(23)
'23 is not a sum of two squares'
sage: sum_of_two_squares_naive(389)
(10, 17)
sage: sum_of_two_squares_naive(2007)
'2007 is not a sum of two squares'
sage: sum_of_two_squares_naive(2008)
'2008 is not a sum of two squares'
sage: sum_of_two_squares_naive(2009)
(28, 35)
sage: 28^2 + 35^2
2009
sage: sum_of_two_squares_naive(2*3^4*5*7^2*13)
(189, 693)
```

Definition 5.7.3 (Primitive)**.** A representation $n = x^2 + y^2$ is *primitive* if x and y are coprime.

Lemma 5.7.4. *If n is divisible by a prime $p \equiv 3 \pmod 4$, then n has no primitive representations.*

Proof. Suppose n has a primitive representation, $n = x^2 + y^2$, and let p be any prime factor of n. Then

$$p \mid x^2 + y^2 \quad \text{and} \quad \gcd(x, y) = 1,$$

so $p \nmid x$ and $p \nmid y$. Since $\mathbf{Z}/p\mathbf{Z}$ is a field, we may divide by y^2 in the equation $x^2 + y^2 \equiv 0 \pmod p$ to see that $(x/y)^2 \equiv -1 \pmod p$. Thus the Legendre symbol $\left(\frac{-1}{p}\right)$ equals $+1$. However, by Proposition 4.2.1,

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$$

so $\left(\frac{-1}{p}\right) = 1$ if and only if $(p-1)/2$ is even, which is to say $p \equiv 1 \pmod 4$. $\square$

Proof of Theorem 5.7.1 ($\Longrightarrow$). Suppose that $p \equiv 3 \pmod 4$ is a prime, that $p^r \mid n$ but $p^{r+1} \nmid n$ with r odd, and that $n = x^2 + y^2$. Letting $d = \gcd(x, y)$, we have

$$x = dx', \quad y = dy', \quad \text{and} \quad n = d^2 n'$$

with $\gcd(x', y') = 1$ and

$$(x')^2 + (y')^2 = n'.$$

Because r is odd, $p \mid n'$, so Lemma 5.7.4 implies that $\gcd(x', y') > 1$, which is a contradiction. □

To prepare for our proof of the implication ($\Longleftarrow$) of Theorem 5.7.1, we reduce the problem to the case when n is prime. Write $n = n_1^2 n_2$, where n_2 has no prime factors $p \equiv 3 \pmod 4$. It suffices to show that n_2 is a sum of two squares, since

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2, \tag{5.7.1}$$

so a product of two numbers that are sums of two squares is also a sum of two squares. Since $2 = 1^2 + 1^2$ is a sum of two squares, it suffices to show that any prime $p \equiv 1 \pmod 4$ is a sum of two squares.

Lemma 5.7.5. *If $x \in \mathbf{R}$ and $n \in \mathbf{N}$, then there is a fraction $\frac{a}{b}$ in lowest terms such that $0 < b \leq n$ and*

$$\left| x - \frac{a}{b} \right| \leq \frac{1}{b(n+1)}.$$

Proof. Consider the continued fraction $[a_0, a_1, \ldots]$ of x. By Corollary 5.3.11, for each m

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}}.$$

Since $q_{m+1} \geq q_m + 1$ and $q_0 = 1$, either there exists an m such that $q_m \leq n < q_{m+1}$, or the continued fraction expansion of x is finite and n is larger than the denominator of the rational number x, in which case we take $\frac{a}{b} = x$ and are done. In the first case,

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}} \leq \frac{1}{q_m \cdot (n+1)},$$

so $\dfrac{a}{b} = \dfrac{p_m}{q_m}$ satisfies the conclusion of the lemma. □

Proof of Theorem 5.7.1 ($\Longleftarrow$). As discussed above, it suffices to prove that any prime $p \equiv 1 \pmod 4$ is a sum of two squares. Since $p \equiv 1 \pmod 4$,

$$(-1)^{(p-1)/2} = 1,$$

Proposition 4.2.1 implies that -1 is a square modulo p; i.e., there exists $r \in \mathbf{Z}$ such that $r^2 \equiv -1 \pmod{p}$. Lemma 5.7.5, with $n = \lfloor\sqrt{p}\rfloor$ and $x = -\frac{r}{p}$, implies that there are integers a, b such that $0 < b < \sqrt{p}$ and

$$\left|-\frac{r}{p} - \frac{a}{b}\right| \leq \frac{1}{b(n+1)} < \frac{1}{b\sqrt{p}}.$$

Letting $c = rb + pa$, we have that

$$|c| < \frac{pb}{b\sqrt{p}} = \frac{p}{\sqrt{p}} = \sqrt{p}$$

so

$$0 < b^2 + c^2 < 2p.$$

But $c \equiv rb \pmod{p}$, so

$$b^2 + c^2 \equiv b^2 + r^2b^2 \equiv b^2(1 + r^2) \equiv 0 \pmod{p}.$$

Thus $b^2 + c^2 = p$. □

Remark 5.7.6. Our proof of Theorem 5.7.1 leads to an efficient algorithm to compute a representation of any $p \equiv 1 \pmod{4}$ as a sum of two squares.

SAGE Example 5.7.7. We next use Sage and Theorem 5.7.1 to give an efficient algorithm for writing a prime $p \equiv 1 \pmod{4}$ as a sum of two squares. First we implement the algorithm that comes out of the proof of the theorem.

```
sage: def sum_of_two_squares(p):
...     p = Integer(p)
...     assert p%4 == 1, "p must be 1 modulo 4"
...     r = Mod(-1,p).sqrt().lift()
...     v = continued_fraction(-r/p)
...     n = floor(sqrt(p))
...     for x in v.convergents():
...         c = r*x.denominator() + p*x.numerator()
...         if -n <= c and c <= n:
...             return (abs(x.denominator()),abs(c))
```

Next we use the algorithm to write the first 10-digit prime $\equiv 1 \pmod{4}$ as a sum of two squares:

```
sage: p = next_prime(next_prime(10^10))
sage: sum_of_two_squares(p)
(55913, 82908)
```

The above calculation was essentially instantanoues. If instead we use the naive algorithm from before, it takes several seconds to write p as a sum of two squares.

```
sage: sum_of_two_squares_naive(p)
(55913, 82908)
```

5.8 Exercises

5.1 If $c_n = p_n/q_n$ is the nth convergent of $[a_0, a_1, \ldots, a_n]$ and $a_0 > 0$, show that

$$[a_n, a_{n-1}, \ldots, a_1, a_0] = \frac{p_n}{p_{n-1}}$$

and

$$[a_n, a_{n-1}, \ldots, a_2, a_1] = \frac{q_n}{q_{n-1}}.$$

(Hint: In the first case, notice that $\frac{p_n}{p_{n-1}} = a_n + \frac{p_{n-2}}{p_{n-1}} = a_n + \frac{1}{\frac{p_{n-1}}{p_{n-2}}}$.)

5.2 Show that every nonzero rational number can be represented in exactly two ways by a finite simple continued fraction. (For example, 2 can be represented by $[1, 1]$ and $[2]$, and $1/3$ by $[0, 3]$ and $[0, 2, 1]$.)

5.3 Evaluate the infinite continued fraction $[2, \overline{1, 2, 1}]$.

5.4 Determine the infinite continued fraction of $\frac{1+\sqrt{13}}{2}$.

5.5 Let $a_0 \in \mathbf{R}$ and $a_1, \ldots, a_n$ and b be positive real numbers. Prove that

$$[a_0, a_1, \ldots, a_n + b] < [a_0, a_1, \ldots, a_n]$$

if and only if n is odd.

5.6 (*) Extend the method presented in the text to show that the continued fraction expansion of $e^{1/k}$ is

$$[1,\ (k-1),\ 1,\ 1,\ (3k-1),\ 1,\ 1,\ (5k-1),\ 1,\ 1,\ (7k-1), \ldots]$$

for all $k \in \mathbf{N}$.

(a) Compute p_0, p_3, q_0, and q_3 for the above continued fraction. Your answers should be in terms of k.

(b) Condense three steps of the recurrence for the numerators and denominators of the above continued fraction. That is, produce a simple recurrence for r_{3n} in terms of r_{3n-3} and r_{3n-6} whose coefficients are polynomials in n and k.

(c) Define a sequence of real numbers by

$$T_n(k) = \frac{1}{k^n} \int_0^{1/k} \frac{(kt)^n (kt-1)^n}{n!} \ e^t dt.$$

i. Compute $T_0(k)$, and verify that it equals $q_0 e^{1/k} - p_0$.

ii. Compute $T_1(k)$, and verify that it equals $q_3 e^{1/k} - p_3$.

iii. Integrate $T_n(k)$ by parts twice in succession, as in Section 5.4, and verify that $T_n(k)$, $T_{n-1}(k)$, and $T_{n-2}(k)$ satisfy the recurrence produced in part 6b, for $n \geq 2$.

(d) Conclude that the continued fraction

$$[1,\ (k-1),\ 1,\ 1,\ (3k-1),\ 1,\ 1,\ (5k-1),\ 1,\ 1,\ (7k-1), \ldots]$$

represents $e^{1/k}$.

5.7 Let d be an integer that is coprime to 10. Prove that the decimal expansion of $\frac{1}{d}$ has a period equal to the order of 10 modulo d. (Hint: For every positive integer r, we have $\frac{1}{1-10^r} = \sum_{n \geq 1} 10^{-rn}$.)

5.8 Find a positive integer that has at least three different representations as the sum of two squares, disregarding signs and the order of the summands.

5.9 Show that if a natural number n is the sum of two two rational squares it is also the sum of two integer squares.

5.10 (*) Let p be an odd prime. Show that $p \equiv 1, 3 \pmod 8$ if and only if p can be written as $p = x^2 + 2y^2$ for some choice of integers x and y.

5.11 Prove that of any four consecutive integers, at least one is not representable as a sum of two squares.

6
Elliptic Curves

Elliptic curves are number theoretic objects that are central to both pure and applied number theory. Deep problems in number theory such as the congruent number problem—which integers are the area of a right triangle with rational side lengths?—translate naturally into questions about elliptic curves. Other questions, such as the famous Birch and Swinnerton-Dyer conjecture, describe mysterious structure that mathematicians expect elliptic curves to have. One can also associate finite abelian groups to elliptic curves, and in many cases these groups are well suited to the construction of cryptosystems. In particular, elliptic curves are widely believed to provide good security with smaller key sizes, something that is useful in many applications, for example, if we are going to print an encryption key on a postage stamp, it is helpful if the key is short! Morover, there is a way to use elliptic curves to factor integers, which plays a crucial role in sophisticated attacks on the RSA public-key cryptosystem of Section 3.3.

This chapter is a brief introduction to elliptic curves that builds on the ideas of Chapters 1–3 and introduces several deep theorems and ideas that we will not prove. In Section 6.1, we define elliptic curves and draw some pictures of them, and then in Section 6.2 we describe how to put a group structure on the set of points on an elliptic curve. Sections 6.3 and 6.4 are about how to apply elliptic curves to two cryptographic problems—constructing public-key cryptosystems and factoring integers. Finally, in Section 6.5, we consider elliptic curves over the rational numbers, and explain a deep connection between elliptic curves and a 1,000-year old unsolved problem.

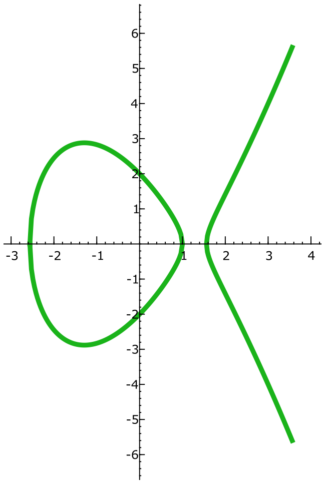


FIGURE 6.1. The elliptic curve $y^2 = x^3 - 5x + 4$ over $\mathbf{R}$

6.1 The Definition

Definition 6.1.1 (Elliptic Curve)**.** An *elliptic curve* over a field K is a curve defined by an equation of the form

$$y^2 = x^3 + ax + b,$$

where $a, b \in K$ and $-16(4a^3 + 27b^2) \neq 0$.

The condition that $-16(4a^3 + 27b^2) \neq 0$ implies that the curve has no "singular points," which will be essential for the applications we have in mind (see Exercise 6.1).

SAGE Example 6.1.2. We use the `EllipticCurve` command to create an elliptic curve over the rational field $\mathbf{Q}$ and draw the plot in Figure 6.1.

```
sage: E = EllipticCurve([-5, 4])
sage: E
Elliptic Curve defined by y^2  = x^3 - 5*x + 4
over Rational Field
sage: P = E.plot(thickness=4,rgbcolor=(0.1,0.7,0.1))
sage: P.show(figsize=[4,6])
```

We will use elliptic curves over finite fields to factor integers in Section 6.3 and to construct cryptosystems in Section 6.4. The following Sage code creates an elliptic curve over the finite field of order 37 and plots it, as illustrated in Figure 6.2.

```
sage: E = EllipticCurve(GF(37), [1,0])
sage: E
Elliptic Curve defined by y^2  = x^3 + x over
Finite Field of size 37
sage: E.plot(pointsize=45)
```

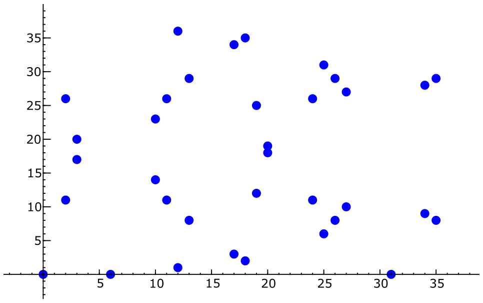


FIGURE 6.2. The elliptic curve $y^2 = x^3 + x$ over $\mathbf{Z}/37\mathbf{Z}$

In Section 6.2, we will put a natural abelian group structure on the set

$$E(K) = \{(x, y) \in K \times K : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\}$$

of K-rational points on an elliptic curve E over K. Here, $\mathcal{O}$ may be thought of as a point on E "at infinity." Figure 6.2 contains a plot of the points of $y^2 = x^3 + x$ over the finite field $\mathbf{Z}/37\mathbf{Z}$, though note that we do not explicitly draw the point at $\mathcal{O}$ at infinity.

Remark 6.1.3. If K has characteristic 2 (i.e., we have $1+1 = 0$ in K), then for any choice of a, b, the quantity $-16(4a^3 + 27b^2) \in K$ is 0, so according to Definition 6.1.1 there are no elliptic curves over K. There is a similar problem in characteristic 3. If we instead consider equations of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

we obtain a more general definition of elliptic curves, which correctly allows for elliptic curves in characteristics 2 and 3; these elliptic curves are popular in cryptography because arithmetic on them is often easier to efficiently implement on a computer.

6.2 The Group Structure on an Elliptic Curve

Let E be an elliptic curve over a field K, given by an equation $y^2 = x^3 + ax + b$. We begin by defining a binary operation $+$ on $E(K)$.

Algorithm 6.2.1 (Elliptic Curve Group Law)**.** Given $P_1, P_2 \in E(K)$, this algorithm computes a third point $R = P_1 + P_2 \in E(K)$.

1. [Is $P_i = \mathcal{O}$?] If $P_1 = \mathcal{O}$ set $R = P_2$ or if $P_2 = \mathcal{O}$ set $R = P_1$ and terminate. Otherwise write $(x_i, y_i) = P_i$.
2. [Negatives] If $x_1 = x_2$ and $y_1 = -y_2$, set $R = \mathcal{O}$ and terminate.
3. [Compute λ] Set $\lambda = \begin{cases} (3x_1^2 + a)/(2y_1) & \text{if } P_1 = P_2, \\ (y_1 - y_2)/(x_1 - x_2) & \text{otherwise.} \end{cases}$
4. [Compute Sum] Then $R = \big(\lambda^2 - x_1 - x_2, -\lambda x_3 - \nu\big)$, where $\nu = y_1 - \lambda x_1$ and $x_3 = \lambda^2 - x_1 - x_2$ is the x-coordinate of R.

Note that in Step 3, if $P_1 = P_2$, then $y_1 \neq 0$; otherwise, we would have terminated in the previous step.

Theorem 6.2.2. *The binary operation + defined in Algorithm 6.2.1 endows the set $E(K)$ with an abelian group structure, with identity $\mathcal{O}$.*

Before discussing why the theorem is true, we reinterpret + geometrically, so that it will be easier for us to visualize. We obtain the sum $P_1 + P_2$ by finding the third point P_3 of intersection between E and the line L determined by P_1 and P_2, then reflecting P_3 about the x-axis. (This description requires suitable interpretation in cases 1 and 2, and when $P_1 = P_2$.) This is illustrated in Figure 6.3, in which $(0, 2) + (1, 0) = (3, 4)$ on $y^2 = x^3 - 5x + 4$.

SAGE Example 6.2.3. We create the elliptic curve $y^2 = x^3 - 5x + 4$ in Sage, then add together $P = (1, 0)$ and $Q = (0, 2)$. We also compute $P + P$, which is the point $\mathcal{O}$ at infinity, which is represented in Sage by $(0 : 1 : 0)$, and compute the sum $P + Q + Q + Q + Q$, which is surprisingly large.

```
sage: E = EllipticCurve([-5,4])
sage: P = E([1,0]); Q = E([0,2])
sage: P + Q
(3 : 4 : 1)
sage: P + P
(0 : 1 : 0)
sage: P + Q + Q + Q + Q
(350497/351649 : 16920528/208527857 : 1)
```

To further clarify the above geometric interpretation of the group law, we prove the following proposition.

Proposition 6.2.4 (Geometric Group Law)**.** *Suppose $P_i = (x_i, y_i)$, $i = 1, 2$ are distinct points on an elliptic curve $y^2 = x^3 + ax + b$, and that $x_1 \neq x_2$. Let L be the unique line through P_1 and P_2. Then L intersects the graph of E at exactly one other point*

$$Q = \big(\lambda^2 - x_1 - x_2, \quad \lambda x_3 + \nu\big),$$

where $\lambda = (y_1 - y_2)/(x_1 - x_2)$ and $\nu = y_1 - \lambda x_1$.

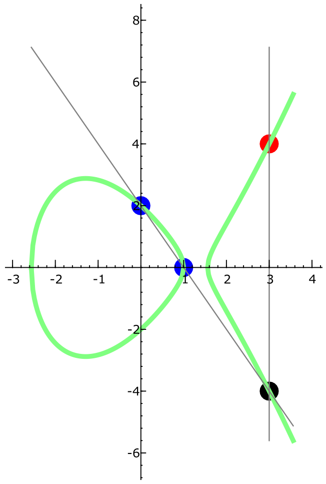


FIGURE 6.3. The Group Law: $(1,0)+(0,2)=(3,4)$ on $y^2=x^3-5x+4$

Proof. The line L through P_1, P_2 is $y=y_1+(x-x_1)\lambda$. Substituting this into $y^2=x^3+ax+b$, we get

$$(y_1+(x-x_1)\lambda)^2=x^3+ax+b.$$

Simplifying, we get $f(x)=x^3-\lambda^2x^2+\cdots=0$, where we omit the coefficients of x and the constant term since they will not be needed. Since P_1 and P_2 are in $L\cap E$, the polynomial f has x_1 and x_2 as roots. By Proposition 2.5.3, the polynomial f can have at most three roots. Writing $f=\prod(x-x_i)$ and equating terms, we see that $x_1+x_2+x_3=\lambda^2$. Thus, $x_3=\lambda^2-x_1-x_2$, as claimed. Also, from the equation for L we see that $y_3=y_1+(x_3-x_1)\lambda=\lambda x_3+\nu$, which completes the proof. □

To prove Theorem 6.2.2 means to show that $+$ satisfies the three axioms of an abelian group with $\mathcal{O}$ as identity element: existence of inverses, commutativity, and associativity. The existence of inverses follows immediately from the definition, since $(x,y)+(x,-y)=\mathcal{O}$. Commutativity is also clear from the definition of group law, since in Parts 1–3, the recipe is unchanged if we swap P_1 and P_2; in Part 4 swapping P_1 and P_2 does not change the line determined by P_1 and P_2, so by Proposition 6.2.4 it does not change the sum P_1+P_2.

It is more difficult to prove that $+$ satisfies the associative axiom, i.e., that $(P_1+P_2)+P_3=P_1+(P_2+P_3)$. This fact can be understood from at least three points of view. One is to reinterpret the group law geometrically (extending Proposition 6.2.4 to all cases), and thus transfer the problem to a question in plane geometry. This approach is beautifully explained with exactly the right level of detail in [ST92, §I.2]. Another approach is to use the formulas that define + to reduce associativity to checking specific algebraic identities; this is something that would be extremely tedious to do by hand, but can be done using a computer (also tedious). A third approach (see [Sil86] or [Har77]) is to develop a general theory of "divisors on algebraic curves," from which associativity of the group law falls out as a natural corollary. The third approach is the best, because it opens up many new vistas; however, we will not pursue it further because it is beyond the scope of this book.

SAGE Example 6.2.5. In the following Sage session, we use the formula from Algorithm 6.2.1 to verify that the group law holds for any choice of points P_1, P_2, P_3 on any elliptic curve over $\mathbf{Q}$ such that the points $P_1, P_2, P_3, P_1 + P_2, P_2 + P_3$ are all distinct and nonzero. We define a polynomial ring R in 8 variables.

```
sage: R.<x1,y1,x2,y2,x3,y3,a,b> = QQ[]
```

We define the relations the x_i will satisfy, and a quotient ring Q in which those relations are satisfied. (Quotients of polynomial rings are a generalization of the construction $\mathbf{Z}/n\mathbf{Z}$ that may be viewed as the quotient of the ring $\mathbf{Z}$ of integers by the relation that sets n to equal 0.)

```
sage: rels = [y1^2 - (x1^3 + a*x1 + b),
...           y2^2 - (x2^3 + a*x2 + b),
...           y3^2 - (x3^3 + a*x3 + b)]
...
sage: Q = R.quotient(rels)
```

We define the group operation, which assumes the points are distinct.

```
sage: def op(P1,P2):
...       x1,y1 = P1;  x2,y2 = P2
...       lam = (y1 - y2)/(x1 - x2); nu  = y1 - lam*x1
...       x3 = lam^2 - x1 - x2; y3 = -lam*x3 - nu
...       return (x3, y3)
```

We define three points, add them together via $P_1 + (P_2 + P_3)$ and $(P_1 + (P_2 + P_3))$, and observe that the results are the same modulo the relations.

```
sage: P1 = (x1,y1); P2 = (x2,y2); P3 = (x3,y3)
sage: Z = op(P1, op(P2,P3)); W = op(op(P1,P2),P3)
sage: (Q(Z[0].numerator()*W[0].denominator() -
...          Z[0].denominator()*W[0].numerator())) == 0
True
sage: (Q(Z[1].numerator()*W[1].denominator() -
...          Z[1].denominator()*W[1].numerator())) == 0
True
```

6.3 Integer Factorization Using Elliptic Curves

In 1987, Hendrik Lenstra published the landmark paper [Len87] that introduces and analyzes the Elliptic Curve Method (ECM), which is a powerful algorithm for factoring integers using elliptic curves. Lenstra's method is also described in [ST92, §IV.4], [Dav99, §VIII.5], and [Coh93, §10.3].

Lenstra's algorithm is well suited for finding "medium-sized" factors of an integer N, which today means between 10 to 40 decimal digits. The ECM method is not *directly* used for factoring RSA challenge numbers (see Section 1.1.3), but it is used on auxiliary numbers as a crucial step in the "number field sieve," which is the best known algorithm for hunting for such factorizations. Also, implementation of ECM typically requires little memory.

H. Lenstra

6.3.1 Pollard's $(p-1)$-Method

Lenstra's discovery of ECM was inspired by Pollard's $(p-1)$-method, which we describe in this section.

Definition 6.3.1 (Power Smooth)**.** Let B be a positive integer. If n is a positive integer with prime factorization $n = \prod p_i^{e_i}$, then n is *B-power smooth* if $p_i^{e_i} \leq B$ for all i.

For example, $30 = 2\cdot 3\cdot 5$ is B power smooth for $B = 5, 7$, but $150 = 2\cdot 3\cdot 5^2$ is not 5-power smooth (it is $B = 25$-power smooth).

We will use the following algorithm in both the Pollard $p-1$ and elliptic curve factorization methods.

Algorithm 6.3.2 (Least Common Multiple of First B Integers)**.** Given a positive integer B, this algorithm computes the least common multiple of the positive integers up to B.

1. [Sieve] Using, for example, the prime sieve (Algorithm 1.2.3), compute a list P of all primes $p \leq B$.
2. [Multiply] Compute and output the product $\prod_{p\in P} p^{\lfloor \log_p(B)\rfloor}$.

Proof. Set $m = \mathrm{lcm}(1, 2, \ldots, B)$. Then,

$$\mathrm{ord}_p(m) = \max(\{\mathrm{ord}_p(n) : 1 \leq n \leq B\}) = \mathrm{ord}_p(p^r),$$

where p^r is the largest power of p that satisfies $p^r \leq B$. Since $p^r \leq B < p^{r+1}$, we have $r = \lfloor \log_p(B)\rfloor$. □

SAGE Example 6.3.3. We implement Algorithm 6.3.2 in Sage and compute the least common multiple for $B = 100$ using both the above algorithm and a naive algorithm. We use `math.log` below so that $\log_p(B)$ is computed quickly using double precision numbers.

```
sage: def lcm_upto(B):
...       return prod([p^int(math.log(B)/math.log(p))
...                    for p in prime_range(B+1)])
sage: lcm_upto(10^2)
69720375229712477164533808935312303556800
sage: LCM([1..10^2])
69720375229712477164533808935312303556800
```

Algorithm 6.3.2 as implemented above in Sage takes about a second for $B = 10^6$.

Let N be a positive integer that we wish to factor. We use the Pollard $(p-1)$-method to look for a nontrivial factor of N as follows. First, we choose a positive integer B, usually with at most six digits. Suppose that there is a prime divisor p of N such that $p-1$ is B-power smooth. We try to find p using the following strategy. If $a > 1$ is an integer not divisible by p, then by Theorem 2.1.20,

$$a^{p-1} \equiv 1 \pmod{p}.$$

Let $m = \operatorname{lcm}(1, 2, 3, \ldots, B)$, and observe that our assumption that $p-1$ is B-power smooth implies that $p - 1 \mid m$, so

$$a^m \equiv 1 \pmod{p}.$$

Thus

$$p \mid \gcd(a^m - 1, N) > 1.$$

If $\gcd(a^m-1, N) < N$ also then $\gcd(a^m-1, N)$ is a nontrivial factor of N. If $\gcd(a^m - 1, N) = N$, then $a^m \equiv 1 \pmod{q^r}$ for every prime power divisor q^r of N. In this case, repeat the above steps but with a smaller choice of B or possibly a different choice of a. Also, it is a good idea to check from the start whether or not N is not a perfect power M^r and, if so, replace N by M. We formalize the algorithm as follows:

Algorithm 6.3.4 (Pollard $p-1$ Method)**.** Given a positive integer N and a bound B, this algorithm attempts to find a nontrivial factor g of N. (Each prime $p \mid g$ is likely to have the property that $p-1$ is B-power smooth.)

```
1. [Compute lcm] Use Algorithm 6.3.2 to compute $m = \operatorname{lcm}(1, 2, \ldots, B)$.
2. [Initialize] Set $a = 2$.
3. [Power and gcd] Compute $x = a^m - 1 \pmod{N}$ and $g = \gcd(x, N)$.
4. [Finished?] If $g \neq 1$ or $N$, output $g$ and terminate.
```

5. [Try Again?] If $a < 10$ (say), replace a by $a+1$ and go to step 3. Otherwise, terminate.

For fixed B, Algorithm 6.3.4 often splits N when N is divisible by a prime p such that $p-1$ is B-power smooth. Approximately 15 percent of primes p in the interval from 10^{15} and $10^{15}+10000$ are such that $p-1$ is 10^6 power smooth, so the Pollard method with $B = 10^6$ already fails nearly 85 percent of the time at finding 15-digit primes in this range (see also Exercise 6.10). We will not analyze Pollard's method further, since it was mentioned here only to set the stage for the elliptic curve factorization method.

The following examples illustrate the Pollard $(p-1)$-method.

Example 6.3.5. In this example, Pollard works perfectly. Let $N = 5917$. We try to use the Pollard $p-1$ method with $B = 5$ to split N. We have $m = \mathrm{lcm}(1,2,3,4,5) = 60$; taking $a = 2$, we have

$$2^{60} - 1 \equiv 3416 \pmod{5917}$$

and

$$\gcd(2^{60}-1, 5917) = \gcd(3416, 5917) = 61,$$

so 61 is a factor of 5917.

Example 6.3.6. In this example, we replace B with a larger integer. Let $N = 779167$. With $B = 5$ and $a = 2$, we have

$$2^{60} - 1 \equiv 710980 \pmod{779167},$$

and $\gcd(2^{60}-1, 779167) = 1$. With $B = 15$, we have

$$m = \mathrm{lcm}(1,2,\ldots,15) = 360360,$$

$$2^{360360} - 1 \equiv 584876 \pmod{779167},$$

and

$$\gcd(2^{360360}-1, N) = 2003,$$

so 2003 is a nontrivial factor of 779167.

Example 6.3.7. In this example, we replace B by a smaller integer. Let $N = 4331$. Suppose $B = 7$, so $m = \mathrm{lcm}(1,2,\ldots,7) = 420$,

$$2^{420} - 1 \equiv 0 \pmod{4331},$$

and $\gcd(2^{420}-1, 4331) = 4331$, so we do not obtain a factor of 4331. If we replace B by 5, Pollard's method works:

$$2^{60} - 1 \equiv 1464 \pmod{4331},$$

and $\gcd(2^{60}-1, 4331) = 61$, so we split 4331.

Example 6.3.8. In this example, $a = 2$ does not work, but $a = 3$ does. Let $N = 187$. Suppose $B = 15$, so $m = \mathrm{lcm}(1, 2, \ldots, 15) = 360360$,

$$2^{360360} - 1 \equiv 0 \pmod{187},$$

and $\gcd(2^{360360} - 1, 187) = 187$, so we do not obtain a factor of 187. If we replace $a = 2$ by $a = 3$, then Pollard's method works:

$$3^{360360} - 1 \equiv 66 \pmod{187},$$

and $\gcd(3^{360360} - 1, 187) = 11$. Thus $187 = 11 \cdot 17$.

SAGE Example 6.3.9. We implement the Pollard $(p - 1)$-method in Sage and use our implementation to do all of the above examples.

```
sage: def pollard(N, B=10^5, stop=10):
...       m = prod([p^int(math.log(B)/math.log(p))
...                 for p in prime_range(B+1)])
...       for a in [2..stop]:
...           x = (Mod(a,N)^m - 1).lift()
...           if x == 0: continue
...           g = gcd(x, N)
...           if g != 1 or g != N: return g
...       return 1
sage: pollard(5917,5)
61
sage: pollard(779167,5)
1
sage: pollard(779167,15)
2003
sage: pollard(4331,7)
1
sage: pollard(4331,5)
61
sage: pollard(187, 15, 2)
1
sage: pollard(187, 15)
11
```

6.3.2 Motivation for the Elliptic Curve Method

Fix a positive integer B. If $N = pq$ with p and q prime, and we assume that $p - 1$ and $q - 1$ are not B-power smooth, then the Pollard $(p - 1)$-method is unlikely to work. For example, let $B = 20$ and suppose that $N = 59 \cdot 101 = 5959$. Note that neither $59 - 1 = 2 \cdot 29$ nor $101 - 1 = 4 \cdot 25$ is B-power smooth. With $m = \mathrm{lcm}(1, 2, 3, \ldots, 20) = 232792560$, we have

$$2^m - 1 \equiv 5944 \pmod{N},$$

and $\gcd(2^m - 1, N) = 1$, so we do not find a factor of N.

As remarked above, the problem is that $p-1$ is not 20-power smooth for either $p = 59$ or $p = 101$. However, notice that $p - 2 = 3 \cdot 19$ is 20-power smooth. Lenstra's ECM replaces $(\mathbf{Z}/p\mathbf{Z})^*$, which has order $p - 1$, by the group of points on an elliptic curve E over $\mathbf{Z}/p\mathbf{Z}$. It is a theorem that

$$\#E(\mathbf{Z}/p\mathbf{Z}) = p + 1 \pm s$$

for some nonnegative integer $s < 2\sqrt{p}$ (see [Sil86, §V.1] for a proof). Also, every value of s subject to this bound occurs, as one can see using "complex multiplication theory." For example, if E is the elliptic curve

$$y^2 = x^3 + x + 54$$

over $\mathbf{Z}/59\mathbf{Z}$, then by enumerating points one sees that $E(\mathbf{Z}/59\mathbf{Z})$ is cyclic of order 57. The set of numbers $59 + 1 \pm s$ for $s \leq 15$ contains 14 numbers that are B-power smooth for $B = 20$, which illustrates that working with an elliptic curve gives us more flexibility. For example, $60 = 59 + 1 + 0$ is 5-power smooth and $70 = 59 + 1 + 10$ is 7-power smooth.

6.3.3 Lenstra's Elliptic Curve Factorization Method

Algorithm 6.3.10 (Elliptic Curve Factorization Method)**.** Given a positive integer N and a bound B, this algorithm attempts to find a nontrivial factor g of N or outputs "Fail."

1. [Compute lcm] Use Algorithm 6.3.2 to compute $m = \mathrm{lcm}(1, 2, \ldots, B)$.
2. [Choose Random Elliptic Curve] Choose a random $a \in \mathbf{Z}/N\mathbf{Z}$ such that $4a^3 + 27 \in (\mathbf{Z}/N\mathbf{Z})^*$. Then $P = (0, 1)$ is a point on the elliptic curve $y^2 = x^3 + ax + 1$ over $\mathbf{Z}/N\mathbf{Z}$.
3. [Compute Multiple] Attempt to compute mP using an elliptic curve analog of Algorithm 2.3.13. If at some point we cannot compute a sum of points because some denominator in Step 3 of Algorithm 6.2.1 is not coprime to N, we compute the greatest common divisor g of this denominator with N. If g is a nontrivial divisor, output it. If every denominator is coprime to N, output "Fail."

If Algorithm 6.3.10 fails for one random elliptic curve, there is an option that is unavailable with Pollard's $(p-1)$-method—we may repeat the above algorithm with a different elliptic curve. With Pollard's method we always work with the group $(\mathbf{Z}/N\mathbf{Z})^*$, but here we can try many groups $E(\mathbf{Z}/N\mathbf{Z})$ for many curves E. As mentioned above, the number of points on E over $\mathbf{Z}/p\mathbf{Z}$ is of the form $p + 1 - t$ for some t with $|t| < 2\sqrt{p}$; Algorithm 6.3.10 thus has a chance if $p + 1 - t$ is B-power smooth for some t with $|t| < 2\sqrt{p}$.

6.3.4 Examples

For simplicity, we use an elliptic curve of the form

$$y^2 = x^3 + ax + 1,$$

which has the point $P = (0, 1)$ already on it.

We factor $N = 5959$ using the elliptic curve method. Let

$$m = \mathrm{lcm}(1, 2, \ldots, 20) = 232792560 = 1101111000000010000111110000_2,$$

where x_2 means x is written in binary. First, we choose $a = 1201$ at random and consider $y^2 = x^3 + 1201x + 1$ over $\mathbf{Z}/5959\mathbf{Z}$. Using the formula for $P + P$ from Algorithm 6.2.1 we compute $2^i \cdot P = 2^i \cdot (0, 1)$ for $i \in B = \{4, 5, 6, 7, 8, 13, 21, 22, 23, 24, 26, 27\}$. Then $\sum_{i \in B} 2^i P = mP$. It turns out that during no step of this computation does a number not coprime to 5959 appear in any denominator, so we do not split N using $a = 1201$. Next, we try $a = 389$ and at some stage in the computation we add $P = (2051, 5273)$ and $Q = (637, 1292)$. When computing the group law explicitly, we try to compute $\lambda = (y_1 - y_2)/(x_1 - x_2)$ in $(\mathbf{Z}/5959\mathbf{Z})^*$, but we fail since $x_1 - x_2 = 1414$ and $\gcd(1414, 5959) = 101$. We thus find a nontrivial factor 101 of 5959.

SAGE Example 6.3.11. We implement elliptic curve factorization in Sage, then use it to do the above example and some other examples.

```
sage: def ecm(N, B=10^3, trials=10):
...       m = prod([p^int(math.log(B)/math.log(p))
...                 for p in prime_range(B+1)])
...       R = Integers(N)
...       # Make Sage think that R is a field:
...       R.is_field = lambda : True
...       for _ in range(trials):
...           while True:
...               a = R.random_element()
...               if gcd(4*a.lift()^3 + 27, N) == 1: break
...           try:
...               m * EllipticCurve([a, 1])([0,1])
...           except ZeroDivisionError, msg:
...               # msg: "Inverse of <int> does not exist"
...               return gcd(Integer(str(msg).split()[2]), N)
...       return 1
sage: set_random_seed(2)
sage: ecm(5959, B=20)
101
sage: ecm(next_prime(10^20)*next_prime(10^7), B=10^3)
10000019
```

6.3.5 A Heuristic Explanation

Let N be a positive integer and, for simplicity of exposition, assume that $N = p_1 \cdots p_r$ with the p_i distinct primes. It follows from Lemma 2.2.5 that there is a natural isomorphism

$$f : (\mathbf{Z}/N\mathbf{Z})^* \longrightarrow (\mathbf{Z}/p_1\mathbf{Z})^* \times \cdots \times (\mathbf{Z}/p_r\mathbf{Z})^*.$$

When using Pollard's method, we choose an $a \in (\mathbf{Z}/N\mathbf{Z})^*$, compute a^m, then compute $\gcd(a^m - 1, N)$. This gcd is divisible exactly by the primes p_i such that $a^m \equiv 1 \pmod{p_i}$. To reinterpret Pollard's method using the above isomorphism, let $(a_1, \ldots, a_r) = f(a)$. Then $(a_1^m, \ldots, a_r^m) = f(a^m)$, and the p_i that divide $\gcd(a^m - 1, N)$ are exactly the p_i such that $a_i^m = 1$. By Theorem 2.1.20, these p_i include the primes p_j such that $p_j - 1$ is B-power smooth, where $m = \operatorname{lcm}(1, \ldots, m)$.

We will not define $E(\mathbf{Z}/N\mathbf{Z})$ when N is composite, since this is not needed for the algorithm (where we assume that N is prime and hope for a contradiction). However, for the remainder of this paragraph, we pretend that $E(\mathbf{Z}/N\mathbf{Z})$ is meaningful and describe a heuristic connection between Lenstra and Pollard's methods. The significant difference between Pollard's method and the elliptic curve method is that the isomorphism f is replaced by an isomorphism (in quotes)

$$\text{“}g : E(\mathbf{Z}/N\mathbf{Z}) \to E(\mathbf{Z}/p_1\mathbf{Z}) \times \cdots \times E(\mathbf{Z}/p_r\mathbf{Z})\text{”}$$

where E is $y^2 = x^3 + ax + 1$, and the a of Pollard's method is replaced by $P = (0, 1)$. We put the isomorphism in quotes to emphasize that we have not defined $E(\mathbf{Z}/N\mathbf{Z})$. When carrying out the elliptic curve factorization algorithm, we attempt to compute mP, and if some components of $f(Q)$ are $\mathcal{O}$, for some point Q that appears during the computation, but others are nonzero, we find a nontrivial factor of N.

6.4 Elliptic Curve Cryptography

The idea to use elliptic curves in cryptography was independently proposed by Neil Koblitz and Victor Miller in the mid 1980s. In this section, we discuss an analog of Diffie-Hellman that uses an elliptic curve instead of $(\mathbf{Z}/p\mathbf{Z})^*$. We then discuss the ElGamal elliptic curve cryptosystem.

6.4.1 Elliptic Curve Analogs of Diffie-Hellman

The Diffie-Hellman key exchange from Section 3.2 works well on an elliptic curve with no serious modification. Michael and Nikita agree on a secret key as follows:

1. Michael and Nikita agree on a prime p, an elliptic curve E over $\mathbf{Z}/p\mathbf{Z}$, and a point $P \in E(\mathbf{Z}/p\mathbf{Z})$.
2. Michael secretly chooses a random m and sends mP.
3. Nikita secretly chooses a random n and sends nP.
4. The secret key is nmP, which both Michael and Nikita can compute.

Presumably, an adversary can not compute nmP without solving the discrete logarithm problem (see Problem 3.2.2 and Section 6.4.3 below) in $E(\mathbf{Z}/p\mathbf{Z})$. For well-chosen E, P, and p, experience suggests that the discrete logarithm problem in $E(\mathbf{Z}/p\mathbf{Z})$ is much more difficult than the discrete logarithm problem in $(\mathbf{Z}/p\mathbf{Z})^*$ (see Section 6.4.3 for more on the elliptic curve discrete log problem).

6.4.2 The ElGamal Cryptosystem and Digital Rights Management

This section is about the ElGamal cryptosystem, which works well on an elliptic curve. This section draws on a paper by a computer hacker named Beale Screamer who cracked a "Digital Rights Management" (DRM) system.

The elliptic curve used in the DRM is an elliptic curve over the finite field $k = \mathbf{Z}/p\mathbf{Z}$, where

$$p = 785963102379428822376694789446897396207498568951.$$

The number p in base 16 is

$$89\text{ABCDEF}012345672718281831415926141424\text{F}7,$$

which includes counting in hexadecimal, and digits of e, π, and $\sqrt{2}$. The elliptic curve E is

$$\begin{aligned} y^2 = x^3 &+ 317689081251325503476317476413827693272746955927x \\ &+ 79052896607878758718120572025718535432100651934. \end{aligned}$$

We have

$$\#E(k) = 785963102379428822376693024881714957612686157429,$$

and the group $E(k)$ is cyclic with generator

$$\begin{aligned} B = (&771507216262649826170648268565579889907769254176, \\ &390157510246556628525279459266514995562533196655). \end{aligned}$$

Our heroes Nikita and Michael share digital music when they are not out fighting terrorists. When Nikita installed the DRM software on her computer, it generated a private key

$$n = 670805031139910513517527207693060456300217054473,$$

which it hides in bits and pieces of files. In order for Nikita to play Juno Reactor's latest hit `juno.wma`, her web browser contacts a website that sells music. After Nikita sends her credit card number, that website allows Nikita to download a license file that allows her audio player to unlock and play `juno.wma`.

As we will see below, the license file was created using the ElGamal public-key cryptosystem in the group $E(k)$. Nikita can now use her license file to unlock `juno.wma`. However, when she shares both `juno.wma` and the license file with Michael, he is frustrated because even with the license, his computer still does not play `juno.wma`. This is because Michael's computer does not know Nikita's computer's private key (the integer n above), so Michael's computer can not decrypt the license file.

We now describe the ElGamal cryptosystem, which lends itself well to implementation in the group $E(\mathbf{Z}/p\mathbf{Z})$. To illustrate ElGamal, we describe how Nikita would set up an ElGamal cryptosystem that anyone could use to encrypt messages for her. Nikita chooses a prime p, an elliptic curve E over $\mathbf{Z}/p\mathbf{Z}$, and a point $B \in E(\mathbf{Z}/p\mathbf{Z})$, and publishes p, E, and B. She also chooses a random integer n, which she keeps secret, and publishes nB. Her public key is the four-tuple (p, E, B, nB).

Suppose Michael wishes to encrypt a message for Nikita. If the message is encoded as an element $P \in E(\mathbf{Z}/p\mathbf{Z})$, Michael computes a random integer r and the points rB and $P + r(nB)$ on $E(\mathbf{Z}/p\mathbf{Z})$. Then P is encrypted as the pair $(rB, P + r(nB))$. To decrypt the encrypted message, Nikita multiplies rB by her secret key n to find $n(rB) = r(nB)$, then subtracts this from $P + r(nB)$ to obtain

$$P = P + r(nB) - r(nB).$$

Remark 6.4.1. It also make sense to construct an ElGamal cryptosystem in the group $(\mathbf{Z}/p\mathbf{Z})^*$.

Returning to our story, Nikita's license file is an encrypted message to her. It contains the pair of points $(rB, P + r(nB))$, where

$$\begin{aligned} rB = (&1796710032183157463850266557330860449821944246 60,\\ &697834385359686368249301282675141830935176314718) \end{aligned}$$

and

$$\begin{aligned} P + r(nB) = (&1378510385482644673726451580930040003436391189 15,\\ &1108485892286762240572292302235808150242248756 99). \end{aligned}$$

When Nikita's computer plays `juno.wma`, it loads the secret key

$$n = 670805031139910513517527207693060456300217054473$$

into memory and computes

$$\begin{aligned} n(rB) = (&328901393518732637577115650601768681044040715701,\\ &586947838087815993601350565488788846203887988162). \end{aligned}$$

It then subtracts this from $P + r(nB)$ to obtain

$$\begin{aligned} P = (&14489646124220757767,\\ &669337780373284096274895136618194604469696830074). \end{aligned}$$

The x-coordinate 14489646124220757767 is the key that unlocks `juno.wma`.

If Nikita knew the private key n that her computer generated, she could compute P herself and unlock `juno.wma` and share her music with Michael. Beale Screamer found a weakness in the implementation of this system that allows Nikita to detetermine n, which is not a huge surprise since n is stored on her computer after all.

SAGE Example 6.4.2. We do the above examples in Sage:

```
sage: p = 785963102379428822376694789446897396207498568951
sage: E = EllipticCurve(GF(p), \
...    [317689081251325503476317476413827693272746955927,
...     79052896607878758718120572025718535432100651934])
sage: E.cardinality()
785963102379428822376693024881714957612686157429
sage: E.cardinality().is_prime()
True
sage: B = E([
...    771507216262649826170648268565579889907769254176,
...    390157510246556628525279459266514995562533196655])
sage: n=670805031139910513517527207693060456300217054473
sage: r=70674630913457179596452846564371866229568459543
sage: P = E([14489646124220757767,
...    669337780373284096274895136618194604469696830074])
sage: encrypt = (r*B, P + r*(n*B))
sage: encrypt[1] - n*encrypt[0] == P   # decrypting works
True
```

6.4.3 The Elliptic Curve Discrete Logarithm Problem

Problem 6.4.3 (Elliptic Curve Discrete Log Problem)**.** Suppose E is an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$ and $P \in E(\mathbf{Z}/p\mathbf{Z})$. Given a multiple Q of P, the *elliptic curve discrete log problem* is to find $n \in \mathbf{Z}$ such that $nP = Q$.

For example, let E be the elliptic curve given by $y^2 = x^3 + x + 1$ over the field $\mathbf{Z}/7\mathbf{Z}$. We have

$$E(\mathbf{Z}/7\mathbf{Z}) = \{\mathcal{O}, (2,2), (0,1), (0,6), (2,5)\}.$$

If $P = (2,2)$ and $Q = (0,6)$, then $3P = Q$, so $n = 3$ is a solution to the discrete logarithm problem.

If $E(\mathbf{Z}/p\mathbf{Z})$ has order p or $p \pm 1$, or is a product of reasonably small primes, then there are some methods for attacking the discrete log problem on E, which are beyond the scope of this book. It is therefore important to be able to compute $\#E(\mathbf{Z}/p\mathbf{Z})$ efficiently, in order to verify that the elliptic curve one wishes to use for a cryptosystem doesn't have any obvious vulnerabilities. The naive algorithm to compute $\#E(\mathbf{Z}/p\mathbf{Z})$ is to try each value of $x \in \mathbf{Z}/p\mathbf{Z}$ and count how often x^3+ax+b is a perfect square mod p, but this is of no use when p is large enough to be useful for cryptography. Fortunately, there is an algorithm due to Schoof, Elkies, and Atkin for computing $\#E(\mathbf{Z}/p\mathbf{Z})$ efficiently (polynomial time in the number of digits of p), but this algorithm is beyond the scope of this book.

In Section 3.2.1, we discussed the discrete log problem in $(\mathbf{Z}/p\mathbf{Z})^*$. There are general attacks called "index calculus attacks" on the discrete log problem in $(\mathbf{Z}/p\mathbf{Z})^*$ that are slow, but still faster than the known algorithms for solving the discrete log in a "general" group (one with no extra structure). For most elliptic curves, there is no known analog of index calculus attacks on the discrete log problem. At present, it appears that given p, the discrete log problem in $E(\mathbf{Z}/p\mathbf{Z})$ is much harder than the discrete log problem in the multiplicative group $(\mathbf{Z}/p\mathbf{Z})^*$. This suggests that by using an elliptic curve-based cryptosystem instead of one based on $(\mathbf{Z}/p\mathbf{Z})^*$, one gets equivalent security with much smaller numbers, which is one reason why building cryptosystems using elliptic curves is attractive to some cryptographers. For example, Certicom, a company that strongly supports elliptic curve cryptography, claims:

> "[Elliptic curve crypto] devices require less storage, less power, less memory, and less bandwidth than other systems. This allows you to implement cryptography in platforms that are constrained, such as wireless devices, handheld computers, smart cards, and thin-clients. It also provides a big win in situations where efficiency is important."

For an up-to-date list of elliptic curve discrete log challenge problems that Certicom sponsors, see [Cer]. For example, in April 2004, a specific cryptosystem was cracked that was based on an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$, where p has 109 bits. The first unsolved challenge problem involves an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$, where p has 131 bits, and the next challenge after that is one in which p has 163 bits. Certicom claims at [Cer] that the 163-bit challenge problem is computationally infeasible.

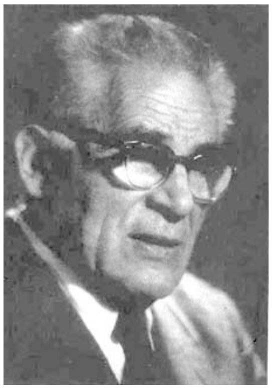


FIGURE 6.4. Louis J. Mordell

6.5 Elliptic Curves Over the Rational Numbers

Let E be an elliptic curve defined over $\mathbf{Q}$. The following is a deep theorem about the group $E(\mathbf{Q})$.

Theorem 6.5.1 (Mordell)**.** *The group $E(\mathbf{Q})$ is finitely generated. That is, there are points $P_1, \ldots, P_s \in E(\mathbf{Q})$ such that every element of $E(\mathbf{Q})$ is of the form $n_1P_1 + \cdots + n_sP_s$ for integers $n_1, \ldots n_s \in \mathbf{Z}$.*

Mordell's theorem implies that it makes sense to ask whether or not we can compute $E(\mathbf{Q})$, where by "compute" we mean find a finite set $P_1, \ldots, P_s$ of points on E that generate $E(\mathbf{Q})$ as an abelian group. There is a systematic approach to computing $E(\mathbf{Q})$ called "descent" (see, for example, [Cre97, Cre, Sil86]). It is widely believed that the method of descent will always succeed, but nobody has yet proved that it will. Proving that descent works for all curves is one of the central open problems in number theory, and is closely related to the Birch and Swinnerton-Dyer conjecture (one of the Clay Math Institute's million dollar prize problems). The crucial difficulty amounts to deciding whether or not certain explicitly given curves have any rational points on them or not (these are curves that have points over $\mathbf{R}$ and modulo n for all n).

The details of using descent to compute $E(\mathbf{Q})$ are beyond the scope of this book. In several places below, we will simply assert that $E(\mathbf{Q})$ has a certain structure or is generated by certain elements. In each case, we computed $E(\mathbf{Q})$ using a computer implementation of this method.

6.5.1 The Torsion Subgroup of $E(\mathbf{Q})$

For any abelian group G, let G_{tor} be the subgroup of elements of finite order. If E is an elliptic curve over $\mathbf{Q}$, then $E(\mathbf{Q})_{\text{tor}}$ is a subgroup of $E(\mathbf{Q})$, which must be finite because of Theorem 6.5.1 (see Exercise 6.6).

One can also prove that $E(\mathbf{Q})_{\mathrm{tor}}$ is finite by showing that there is a prime p and an injective reduction homomorphism $E(\mathbf{Q})_{\mathrm{tor}} \hookrightarrow E(\mathbf{Z}/p\mathbf{Z})$, then noting that $E(\mathbf{Z}/p\mathbf{Z})$ is finite. For example, if E is $y^2 = x^3 - 5x + 4$, then $E(\mathbf{Q})_{\mathrm{tor}} = \{\mathcal{O}, (1,0)\} \cong \mathbf{Z}/2\mathbf{Z}$.

The possibilities for $E(\mathbf{Q})_{\mathrm{tor}}$ are known.

Theorem 6.5.2 (Mazur, 1976)**.** *Let E be an elliptic curve over $\mathbf{Q}$. Then $E(\mathbf{Q})_{\mathrm{tor}}$ is isomorphic to one of the following 15 groups:*

$$\begin{aligned} \mathbf{Z}/n\mathbf{Z} &\qquad \textit{for } n \le 10 \textit{ or } n = 12,\\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2n &\qquad \textit{for } n \le 4. \end{aligned}$$

SAGE Example 6.5.3. We compute the structure of the torsion subgroups of some elliptic curves. In each case, the output of the function $T(a,b)$ below is a pair $c, d \in \mathbf{Z}$ (or integer c) such that the torsion subgroup of $y^3 = x^3 + ax + b$ is $\mathbf{Z}/c\mathbf{Z} \times \mathbf{Z}/d\mathbf{Z}$.

```
sage: T = lambda v: EllipticCurve(v
...           ).torsion_subgroup().invariants()
sage: T([-5,4])
[2]
sage: T([-43,166])
[7]
sage: T([-4,0])
[2, 2]
sage: T([-1386747, 368636886])
[8, 2]
```

6.5.2 The Rank of $E(\mathbf{Q})$

The quotient $E(\mathbf{Q})/E(\mathbf{Q})_{\mathrm{tor}}$ is a finitely generated free abelian group, so it is isomorphism to $\mathbf{Z}^r$ for some integer r, called the *rank* of $E(\mathbf{Q})$. For example, one can prove that if E is $y^2 = x^3 - 5x + 4$, then $E(\mathbf{Q})/E(\mathbf{Q})_{\mathrm{tor}}$ is generated by the point $(0,2)$.

SAGE Example 6.5.4. We use Sage to compute the ranks of some elliptic curves $y^2 = x^3 + ax + b$. The function $r(a,b)$ below returns the rank of this curve over $\mathbf{Q}$.

```
sage: r = lambda v: EllipticCurve(v).rank()
sage: r([-5,4])
1
sage: r([0,1])
0
sage: r([-3024, 46224])
2
sage: r([-112, 400])
```

```
3
sage: r([-102627, 12560670])
4
```

The following is a folklore conjecture, not associated with any particular mathematician:

Conjecture 6.5.5. *There are elliptic curves over* $\mathbf{Q}$ *of arbitrarily large rank.*

The world record is the following curve, whose rank is at least 28:

$$\begin{aligned} y^2 + &xy + y = x^3 - x^2 - \\ &20067762415575526585033208209338542750930230312178956502x + \\ &3448161179503055646703298569039072037485594435931918036126\ldots \\ &\ldots 6600829629193944873224342 9 \end{aligned}$$

It was discovered in May 2006 by Noam Elkies of Harvard University.

6.5.3 The Congruent Number Problem

Definition 6.5.6 (Congruent Number)**.** We call a nonzero rational number n a *congruent number* if $\pm n$ is the area of a right triangle with rational side lengths. Equivalently, n is a *congruent number* if the system of two equations

$$\begin{aligned} a^2 + b^2 &= c^2 \\ \frac{1}{2}ab &= n \end{aligned}$$

has a solution with $a, b, c \in \mathbf{Q}$.

For example, 6 is the area of the right triangle with side lengths 3, 4, and 5, so 6 is a congruent number. Less obvious is that 5 is also a congruent number; it is the area of the right triangle with side lengths 3/2, 20/3, and 41/6. It is nontrivial to prove that 1, 2, 3, and 4 are not congruent numbers. Here is a list of the integer congruent numbers up to 50:

$$5, 6, 7, 13, 14, 15, 20, 21, 22, 23, 24, 28, 29, 30, 31, 34, 37, 38, 39, 41, 45, 46, 47.$$

Every congruence class modulo 8 except 3 is represented in this list, which incorrectly suggests that if $n \equiv 3 \pmod 8$ then n is not a congruent number. Though no $n \leq 218$ with $n \equiv 3 \pmod 8$ is a congruent number, $n = 219$ is a congruent number congruent and $219 \equiv 3 \pmod 8$.

Deciding whether an integer n is a congruent number can be subtle, since the simplest triangle with area n can be very complicated. For example, as Zagier pointed out, the number 157 is a congruent number, and the "simplest" rational right triangle with area 157 has side lengths

$$a = \frac{6803298487826435051217540}{411340519227716149383203} \text{ and } b = \frac{411340519227716149383203}{21666555693714761309610}.$$

This solution would be difficult to find by a brute force search.

We call congruent numbers "congruent" because of the following proposition, which asserts that any congruent number is the common "congruence" between three perfect squares.

Proposition 6.5.7. *Suppose n is the area of a right triangle with rational side lengths a, b, c, with $a \leq b < c$. Let $A = (c/2)^2$. Then*

$$A - n, \quad A, \quad \text{and } A + n$$

are all perfect squares of rational numbers.

Proof. We have

$$\begin{aligned} a^2 + b^2 &= c^2 \\ \frac{1}{2}ab &= n \end{aligned}$$

Add or subtract 4 times the second equation to the first to get

$$\begin{aligned} a^2 \pm 2ab + b^2 &= c^2 \pm 4n \\ (a \pm b)^2 &= c^2 \pm 4n \\ \left(\frac{a \pm b}{2}\right)^2 &= \left(\frac{c}{2}\right)^2 \pm n \\ &= A \pm n \end{aligned}$$

□

The main motivating open problem related to congruent numbers is to give a systematic way to recognize them.

Open Problem 6.5.8. *Give an algorithm which, given n, outputs whether or not n is a congruent number.*

Fortunately, the vast theory developed about elliptic curves has something to say about the above problem. In order to understand this connection, we begin with an elementary algebraic proposition that establishes a link between elliptic curves and the congruent number problem.

Proposition 6.5.9 (Congruent numbers and elliptic curves)**.** *Let n be a rational number. There is a bijection between*

$$A = \left\{(a, b, c) \in \mathbf{Q}^3 \ : \ \frac{ab}{2} = n,\ a^2 + b^2 = c^2\right\}$$

and

$$B = \{(x,y) \in \mathbf{Q}^2 \,:\, y^2 = x^3 - n^2 x, \text{ with } y \neq 0\}$$

given explicitly by the maps

$$f(a,b,c) = \left(-\frac{nb}{a+c},\ \frac{2n^2}{a+c}\right)$$

and

$$g(x,y) = \left(\frac{n^2 - x^2}{y},\ -\frac{2xn}{y},\ \frac{n^2 + x^2}{y}\right).$$

The proof of this proposition is not deep, but involves substantial (elementary) algebra and we will not prove it in this book.

For $n \neq 0$, let E_n be the elliptic curve $y^2 = x^3 - n^2 x$.

Proposition 6.5.10 (Congruent number criterion)**.** *The rational number n is a congruent number if and only if there is a point $P = (x,y) \in E_n(\mathbf{Q})$ with $y \neq 0$.*

Proof. The number n is a congruent number if and only if the set A from Proposition 6.5.9 is nonempty. By the proposition A is nonempty if and only if B is nonempty. □

Example 6.5.11. Let $n = 5$. Then E_n is $y^2 = x^3 - 25x$, and we notice that $(-4,-6) \in E_n(\mathbf{Q})$. We next use the bijection of Proposition 6.5.9 to find the corresponding right triangle:

$$g(-4,-6) = \left(\frac{25-16}{-6}, -\frac{-40}{-6}, \frac{25+16}{-6}\right) = \left(-\frac{3}{2}, -\frac{20}{3}, -\frac{41}{6}\right).$$

Multiplying through by -1 yields the side lengths of a rational right triangle with area 5. *Are there any others?*

Observe that we can apply g to any point in $E_n(\mathbf{Q})$ with $y \neq 0$. Using the group law, we find that $2(-4,-6) = (1681/144, 62279/1728)$ and

$$g(2(-4,-6)) = \left(-\frac{1519}{492}, -\frac{4920}{1519}, \frac{3344161}{747348}\right).$$

This example foreshadows Theorem 6.5.14.

Example 6.5.12. Let $n = 1$, so E_1 is defined by $y^2 = x^3 - x$. Since 1 is not a congruent number, the elliptic curve E_1 has no point with $y \neq 0$. See Exercise 6.11.

SAGE Example 6.5.13. We implement the `cong` function in Sage, which returns a triple (a,b,c) whose entries are the sides of a rational right triangle of area n if one exists, and returns False if there are no such triangles.

```
sage: def cong(n):
...       G = EllipticCurve([-n^2,0]).gens()
...       if len(G) == 0: return False
...       x,y,_ = G[0]
...       return ((n^2-x^2)/y,-2*x*n/y,(n^2+x^2)/y)
sage: cong(6)
(3, 4, 5)
sage: cong(5)
(3/2, 20/3, 41/6)
sage: cong(1)
False
sage: cong(13)
(323/30, 780/323, 106921/9690)
sage: (323/30 * 780/323)/2
13
sage: (323/30)^2 + (780/323)^2 == (106921/9690)^2
True
```

Theorem 6.5.14 (Infinitely Many Triangles). *If n is a congruent number, then there are infinitely many distinct right triangles with rational side lengths and area n.*

We will not prove this theorem, except to note that one proves it by showing that $E_n(\mathbf{Q})_{\mathrm{tor}} = \{\mathcal{O}, (0,0), (n,0), (-n,0)\}$, so the elements of the set B in Proposition 6.5.9 all have infinite order. Hence, B is infinite so A is infinite.

Tunnell has proved that the Birch and Swinnerton-Dyer conjecture (alluded to above), implies the existence of an elementary way to decide whether or not an integer n is a congruent number. We state Tunnell's elementary way in the form of a conjecture.

Conjecture 6.5.15. *Let a, b, c denote integers. If n is an even square-free integer, then n is a congruent number if and only if*

$$\#\left\{(a,b,c) \in \mathbf{Z}^3 : 4a^2 + b^2 + 8c^2 = \frac{n}{2} : c \text{ is even}\right\}$$
$$= \#\left\{(a,b,c) : 4a^2 + b^2 + 8c^2 = \frac{n}{2} : c \text{ is odd}\right\}.$$

If n is odd and square free then n is a congruent number if and only if

$$\#\left\{(a,b,c) : 2a^2 + b^2 + 8c^2 = n : c \text{ is even}\right\}$$
$$= \#\left\{(a,b,c) : 2a^2 + b^2 + 8c^2 = n : c \text{ is odd}\right\}.$$

Enough of the Birch and Swinnerton-Dyer conjecture is known to prove one direction of Conjecture 6.5.15. In particular, it is a very deep theorem that if we do not have equality of the displayed cardinalities, then n is not a congruent number.

The even more difficult (and still open!) part of Conjecture 6.5.15 is the converse: If one has equality of the displayed cardinalities, prove that n is a congruent number. The difficulty in this direction, which appears to be very deep, is that we must somehow construct (or prove the existence of) elements of $E_n(\mathbf{Q})$. This has been accomplished in some cases due to the groundbreaking work of Gross and Zagier ([GZ86]) but much work remains to be done.

The excellent book [Kob84] is about congruent numbers and Conjecture 6.5.15, and we encourage the reader to consult it. The Birch and Swinnerton-Dyer conjecture is a Clay Math Institute million dollar millennium prize problem (see [Cla, Wil00]).

6.6 Exercises

6.1 Write down an equation $y^2 = x^3 + ax + b$ over a field K such that $-16(4a^3+27b^2) = 0$. Precisely what goes wrong when trying to endow the set $E(K) = \{(x, y) \in K \times K : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\}$ with a group structure?

6.2 One rational solution to the equation $y^2 = x^3 - 2$ is $(3, 5)$. Find a rational solution with $x \neq 3$ by drawing the tangent line to $(3, 5)$ and computing the second point of intersection.

6.3 Let E be the elliptic curve over the finite field $K = \mathbf{Z}/5\mathbf{Z}$ defined by the equation

$$y^2 = x^3 + x + 1.$$

(a) List all 9 elements of $E(K)$.

(b) What is the structure of $E(K)$, as a product of cyclic groups?

6.4 Let E be the elliptic curve defined by the equation $y^2 = x^3 + 1$. For each prime $p \geq 5$, let N_p be the cardinality of the group $E(\mathbf{Z}/p\mathbf{Z})$ of points on this curve having coordinates in $\mathbf{Z}/p\mathbf{Z}$. For example, we have that $N_5 = 6, N_7 = 12, N_{11} = 12, N_{13} = 12, N_{17} = 18, N_{19} = 12, , N_{23} = 24,$ and $N_{29} = 30$ (you do not have to prove this).

(a) For the set of primes satisfying $p \equiv 2 \pmod 3$, can you see a pattern for the values of N_p? Make a general conjecture for the value of N_p when $p \equiv 2 \pmod 3$.

(b) (*) Prove your conjecture.

6.5 Let E be an elliptic curve over the real numbers $\mathbf{R}$. Prove that $E(\mathbf{R})$ is not a finitely generated abelian group.

6.6 (*) Suppose G is a finitely generated abelian group. Prove that the subgroup G_{tor} of elements of finite order in G is finite.

6.7 Suppose $y^2 = x^3 + ax + b$ with $a, b \in \mathbf{Q}$ defines an elliptic curve. Show that there is another equation $Y^2 = X^3 + AX + B$ with $A, B \in \mathbf{Z}$ whose solutions are in bijection with the solutions to $y^2 = x^3 + ax + b$.

6.8 Suppose a, b, c are relatively prime integers with $a^2 + b^2 = c^2$. Then there exist integers x and y with $x > y$ such that $c = x^2 + y^2$ and either $a = x^2 - y^2$, $b = 2xy$ or $a = 2xy$, $b = x^2 - y^2$.

6.9 (*) Fermat's Last Theorem for exponent 4 asserts that any solution to the equation $x^4 + y^4 = z^4$ with $x, y, z \in \mathbf{Z}$ satisfies $xyz = 0$. Prove Fermat's Last Theorem for exponent 4, as follows.

(a) Show that if the equation $x^2 + y^4 = z^4$ has no integer solutions with $xyz \neq 0$, then Fermat's Last Theorem for exponent 4 is true.

(b) Prove that $x^2 + y^4 = z^4$ has no integer solutions with $xyz \neq 0$ as follows. Suppose $n^2 + k^4 = m^4$ is a solution with $m > 0$ minimal among all solutions. Show that there exists a solution with m smaller using Exercise 6.8 (consider two cases).

6.10 This problem requires a computer.

(a) Show that the set of numbers $59 + 1 \pm s$ for $s \leq 15$ contains 14 numbers that are B-power smooth for $B = 20$.

(b) Find the proportion of primes p in the interval from 10^{12} and $10^{12} + 1000$ such that $p - 1$ is $B = 10^5$ power smooth.

6.11 (*) Prove that 1 is not a congruent number by showing that the elliptic curve $y^2 = x^3 - x$ has no rational solutions except $(0, \pm 1)$ and $(0, 0)$, as follows:

(a) Write $y = \frac{p}{q}$ and $x = \frac{r}{s}$, where p, q, r, s are all positive integers and $\gcd(p, q) = \gcd(r, s) = 1$. Prove that $s \mid q$, so $q = sk$ for some $k \in \mathbf{Z}$.

(b) Prove that $s = k^2$, and substitute to see that $p^2 = r^3 - rk^4$.

(c) Prove that r is a perfect square by supposing that there is a prime ℓ such that $\mathrm{ord}_\ell(r)$ is odd, and analyzing ord_ℓ of both sides of $p^2 = r^3 - rk^4$.

(d) Write $r = m^2$, and substitute to see that $p^2 = m^6 - m^2k^4$. Prove that $m \mid p$.

(e) Divide through by m^2 and deduce a contradiction to Exercise 6.9.

Answers and Hints

• Chapter 1. Prime Numbers

2. They are $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97$.
3. Emulate the proof of Proposition 1.2.5.

• Chapter 2. The Ring of Integers Modulo n

2. They are 5, 13, 3, and 8.
3. For example, $x = 22$, $y = -39$.
4. Hint: Use the binomial theorem and prove that if $r \geq 1$, then p divides $\binom{p}{r}$.
7. For example, $S_1 = \{0, 1, 2, 3, 4, 5, 6\}$, $S_2 = \{1, 3, 5, 7, 9, 11, 13\}$, $S_3 = \{0, 2, 4, 6, 8, 10, 12\}$, and $S_4 = \{2, 3, 5, 7, 11, 13, 29\}$. In each we find S_i by listing the first seven numbers satisfying the ith condition, then adjust the last number if necessary so that the reductions will be distinct modulo 7.
8. An integer is divisible by 5 if and only if the last digits is 0 or 5. An integer is divisible by 9 if and only if the sum of the digits is divisible by 9. An integer is divisible by 11 if and only if the alternating sum of the digits is divisible by 11.
9. Hint for part (a): Use the divisibility rule you found in Exercise 1.8.

10. 71
11. 8
12. As explained on page 23, we know that $\mathbf{Z}/n\mathbf{Z}$ is a ring for any n. Thus to show that $\mathbf{Z}/p\mathbf{Z}$ is a field it suffices to show that every nonzero element $\overline{a} \in \mathbf{Z}/p\mathbf{Z}$ has an inverse. Lift a to an element $a \in \mathbf{Z}$, and set $b = p$ in Proposition 2.3.1. Because p is prime, $\gcd(a, p) = 1$, so there exists x, y such that $ax+py = 1$. Reducing this equality modulo p proves that $\overline{a}$ has an inverse $x \pmod{p}$. Alternatively, one could argue just like after Definition 2.1.16 that $\overline{a}^m = 1$ for some m, so some power of $\overline{a}$ is the inverse of $\overline{a}$.
13. 302
15. Only for $n = 1, 2$. If $n > 2$, then n is either divisible by an odd prime p or 4. If $4 \mid n$, then $2^e - 2^{e-1}$ divides $\varphi(n)$ for some $e \geq 2$, so $\varphi(n)$ is even. If an odd p divides n, then the even number $p^e - p^{e-1}$ divides $\varphi(n)$ for some $e \geq 1$.
16. The map ψ is a homomorphism since both reduction maps

$$\mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \quad \text{and} \quad \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$$

are homomorphisms. It is injective because if $a \in \mathbf{Z}$ is such that $\psi(a) = 0$, then $m \mid a$ and $n \mid a$, so $mn \mid a$ (since m and n are coprime), so $a \equiv 0 \pmod{mn}$. The cardinality of $\mathbf{Z}/mn\mathbf{Z}$ is mn and the cardinality of the product $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is also mn, so ψ must be an isomorphism. The units $(\mathbf{Z}/mn\mathbf{Z})^*$ are thus in bijection with the units $(\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*$.

For the second part of the exercise, let $g = \gcd(m, n)$ and set $a = mn/g$. Then $a \not\equiv 0 \pmod{mn}$, but $m \mid a$ and $n \mid a$, so $a \ker(\psi)$.

17. We express the question as a system of linear equations modulo various numbers, and use the Chinese remainder theorem. Let x be the number of books. The problem asserts that

$$\begin{aligned} x &\equiv 6 \pmod{7} \\ x &\equiv 2 \pmod{6} \\ x &\equiv 1 \pmod{5} \\ x &\equiv 0 \pmod{4} \end{aligned}$$

Applying CRT to the first pair of equations, we find that $x \equiv 20 \pmod{42}$. Applying CRT to this equation and the third, we find that $x \equiv 146 \pmod{210}$. Since 146 is not divisible by 4, we add multiples of 210 to 146 until we find the first x that is divisible by 4. The first multiple works, and we find that the aspiring mathematicians have 356 math books.

18. Note that $p = 3$ works, since $11 = 3^2 + 2$ is prime. Now suppose $p \neq 3$ is any prime such that p and $p^2 + 2$ are both prime. We must have $p \equiv 1 \pmod{3}$ or $p \equiv 2 \pmod{3}$. Then $p^2 \equiv 1 \pmod{3}$, so $p^2 + 2 \equiv 0 \pmod{3}$. Since $p^2 + 2$ is prime, we must have $p^2 + 2 = 3$, so $p = 1$, a contradiction as p is assumed prime.
19. For (a) $n = 1, 2$, see solution to Exercise 2.15. For (b), yes there are many such examples. For example, $m = 2$, $n = 4$.
20. By repeated application of multiplicativity and Equation (2.2.2) on page 31, we see that if $n = \prod_i p_i^{e_i}$ is the prime factorization of n, then

$$\varphi(n) = \prod_i (p_i^{e_i} - p_i^{e_i - 1}) = \prod_i p_i^{e_i - 1} \cdot \prod_i (p_i - 1).$$

23. 1, 6, 29, 34
24. Let $g = \gcd(12n+1, 30n+2)$. Then $g \mid 30n+2-2\cdot(12n+1) = 6n$. For the same reason, g also divides $12n + 1 - 2 \cdot (6n) = 1$, so $g = 1$, as claimed.
27. There is no primitive root modulo 8, since $(\mathbf{Z}/8\mathbf{Z})^*$ has order 4, but every element of $(\mathbf{Z}/8\mathbf{Z})^*$ has order 2. Prove that if ζ is a primitive root modulo 2^n, for $n \geq 3$, then the reduction of ζ mod 8 is a primitive root, a contradiction.
28. 2 is a primitive root modulo 125.
29. Let $\prod_{i=1}^m p_i^{e_i}$ be the prime factorization of n. Slightly generalizing Exercise 16, we see that

$$(\mathbf{Z}/n\mathbf{Z})^* \cong \prod (\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*.$$

Thus $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if the product $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ is cyclic. If $8 \mid n$, then there is no chance $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic, so assume $8 \nmid n$. Then by Exercise 2.28, each group $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ is itself cyclic. A product of cyclic groups is cyclic if and only the orders of the factors in the product are coprime (this follows from Exercise 2.16). Thus $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if the numbers $p_i(p_i - 1)$, for $i = 1, \ldots, m$ are pairwise coprime. Since $p_i - 1$ is even, there can be at most one odd prime in the factorization of n, and we see that $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if n is an odd prime power, twice an odd prime power, or $n = 4$.

• Chapter 3. Public-Key Cryptography

1. The best case is that each letter is A. Then the question is to find the largest n such that $1 + 27 + \cdots + 27^n \leq 10^{20}$. By computing

$\log_{27}(10^{20})$, we see that $27^{13} < 10^{20}$ and $27^{14} > 10^{20}$. Thus $n \leq 13$, and since $1+27+\cdots+27^{n-1} < 27^n$, and $2\cdot 27^{13} < 10^{20}$, it follows that $n = 13$.

2. This is not secure, since it is just equivalent to a "Ceaser Cipher," that is a permutation of the letters of the alphabet, which is well-known to be easily broken using a frequency analysis.

3. If we can compute the polynomial

$$f = (x-p)(x-q)(x-r) = x^3-(p+q+r)x^2+(pq+pr+qr)x-pqr,$$

then we can factor n by finding the roots of f, for example, using Newton's method (or Cardona's formula for the roots of a cubic). Because p, q, r, are distinct odd primes, we have

$$\varphi(n) = (p-1)(q-1)(r-1) = pqr - (pq+pr+qr) + p + q + r,$$

and

$$\sigma(n) = 1 + (p+q+r) + (pq+pr+qr) + pqr.$$

Since we know n, $\varphi(n)$, and $\sigma(n)$, we know

$$\begin{aligned} \sigma(n) - 1 - n &= (p+q+r) + (pq+pr+qr), \quad \text{and} \\ \varphi(n) - n &= (p+q+r) - (pq+pr+qr). \end{aligned}$$

We can thus compute both $p+q+r$ and $pq+pr+qr$, hence deduce f and find p, q, r.

• Chapter 4. Quadratic Reciprocity

1. They are all 1, −1, 0, and 1.

3. By Proposition 4.3.4, the value of $\left(\frac{3}{p}\right)$ depends only on the reduction $\pm p \pmod{12}$. List enough primes p such that $\pm p$ reduce to $1, 5, 7, 11$ modulo 12 and verify that the asserted formula holds for each of them.

7. Since $p = 2^{13} - 1$ is prime, there are either two solutions or no solutions to $x^2 \equiv 5 \pmod{p}$, and we can decide which using quadratic reciprocity. We have

$$\left(\frac{5}{p}\right) = (-1)^{(p-1)/2\cdot(5-1)/2}\left(\frac{p}{5}\right) = \left(\frac{p}{5}\right),$$

so there are two solutions if and only if $p = 2^{13}-1$ is ± 1 mod 5. In fact, $p \equiv 1 \pmod{5}$, so there are two solutions.

8. We have $4^{48} = 2^{96}$. By Euler's Theorem, $2^{96} = 1$, so $x = 1$.

9. For (a), take $a = 19$ and $n = 20$. We found this example using the Chinese remainder theorem applied to 4 (mod 5) and 3 (mod 4), and used that $\left(\frac{19}{20}\right) = \left(\frac{19}{5}\right) \cdot \left(\frac{19}{4}\right) = (-1)(-1) = 1$, yet 19 is not a square modulo either 5 or 4, so is certainly not a square modulo 20.

10. Hint: First reduce to the case that $6k - 1$ is prime, by using that if p and q are primes not of the form $6k - 1$, then neither is their product. If $p = 6k - 1$ divides $n^2 + n + 1$, it divides $4n^2 + 4n + 4 = (2n + 1)^2 + 3$, so -3 is a quadratic residue modulo p. Now use quadratic reciprocity to show that -3 is not a quadratic residue modulo p.

• Chapter 5. Continued Fractions

9. Suppose $n = x^2 + y^2$, with $x, y \in \mathbf{Q}$. Let d be such that $dx, dy \in \mathbf{Z}$. Then $d^2n = (dx)^2 + (dy)^2$ is a sum of two integer squares, so by Theorem 5.7.1, if $p \mid d^2n$ and $p \equiv 3 \pmod{4}$, then $\mathrm{ord}_p(d^2n)$ is even. We have $\mathrm{ord}_p(d^2n)$ is even if and only if $\mathrm{ord}_p(n)$ is even, so Theorem 5.7.1 implies that n is also a sum of two squares.

11. The squares modulo 8 are $0, 1, 4$, so a sum of two squares reduces modulo 8 to one of $0, 1, 2, 4$, or 5. Four consecutive integers that are sums of squares would reduce to four consecutive integers in the set $\{0, 1, 2, 4, 5\}$, which is impossible.

• Chapter 6. Elliptic Curves

2. The second point of intersection is $(129/100, 383/1000)$.

3. The group is cyclic of order 9, generated by $(4, 2)$. The elements of $E(K)$ are
$$\{\mathcal{O}, (4,2), (3,4), (2,4), (0,4), (0,1), (2,1), (3,1), (4,3)\}.$$

4. In part (a), the pattern is that $N_p = p + 1$. For part (b), a hint is that when $p \equiv 2 \pmod{3}$, the map $x \mapsto x^3$ on $(\mathbf{Z}/p\mathbf{Z})^*$ is an automorphism, so $x \mapsto x^3 + 1$ is a bijection. Now use what you learned about squares in $\mathbf{Z}/p\mathbf{Z}$ from Chapter 4.

5. For all sufficiently large real x, the equation $y^2 = x^3 + ax + b$ has a real solution y. Thus, the group $E(\mathbf{R})$ is not countable, since $\mathbf{R}$ is not countable. But any finitely generated group is countable.

6. In a course on abstract algebra, one often proves the nontrivial fact that every subgroup of a finitely generated abelian group is finitely generated. In particular, the torsion subgroup G_{tor} is

finitely generated. However, a finitely generated abelian torsion group is finite.

7. Hint: Multiply both sides of $y^2 = x^3 + ax + b$ by a power of a common denominator, and "absorb" powers into x and y.

8. Hint: see Exercise 4.6.

References

[ACD$^+$99] K. Aardal, S. Cavallar, B. Dodson, A. Lenstra, W. Lioen, P. L. Montgomery, B. Murphy, J. Gilchrist, G. Guillerm, P. Leyland, J. Marchand, F. Morain, A. Muffett, C.&C. Putnam, and P. Zimmermann, *Factorization of a 512-bit RSA key using the Number Field Sieve*, `http://www.loria.fr/~zimmerma/records/RSA155` (1999).

[AGP94] W. R. Alford, Andrew Granville, and Carl Pomerance, *There are infinitely many Carmichael numbers*, Ann. of Math. (2) **139** (1994), no. 3, 703–722. MR 95k:11114

[AKS02] M. Agrawal, N. Kayal, and N. Saxena, *PRIMES is in P*, to appear in Annals of Math., `http://www.cse.iitk.ac.in/users/manindra/primality.ps` (2002).

[BS76] Leonard E. Baum and Melvin M. Sweet, *Continued fractions of algebraic power series in characteristic* 2, Ann. of Math. (2) **103** (1976), no. 3, 593–610. MR 53 #13127

[Bur89] D. M. Burton, *Elementary Number Theory*, second ed., W. C. Brown Publishers, Dubuque, IA, 1989. MR 90e:11001

[Cal] C. Caldwell, *The Largest Known Primes,* `http://www.utm.edu/research/primes/largest.html`.

[Cer] Certicom, *The certicom ECC challenge,* `http://www.certicom.com/index.php?action=res,ecc_challenge`.

[Cla] Clay Mathematics Institute, *Millennium prize problems,* `http://www.claymath.org/millennium_prize_problems/`.

[Coh] H. Cohn, *A short proof of the continued fraction expansion of e,* `http://research.microsoft.com/~cohn/publications.html`.

[Coh93] H. Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics, vol. 138, Springer-Verlag, Berlin, 1993. MR 94i:11105

[Con97] John H. Conway, *The Sensual (Quadratic) Form*, Carus Mathematical Monographs, vol. 26, Mathematical Association of America, Washington, DC, 1997, With the assistance of Francis Y. C. Fung. MR 98k:11035

[CP01] R. Crandall and C. Pomerance, *Prime Numbers*, Springer-Verlag, New York, 2001, A computational perspective. MR 2002a:11007

[Cre] J. E. Cremona, `mwrank` *(computer software)*, `http://www.maths.nott.ac.uk/personal/jec/ftp/progs/`.

[Cre97] ______, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[Dav99] H. Davenport, *The Higher Arithmetic*, seventh ed., Cambridge University Press, Cambridge, 1999, An introduction to the theory of numbers, Chapter VIII by J. H. Davenport. MR 2000k:11002

[DH76] W. Diffie and M. E. Hellman, *New directions in cryptography*, IEEE Trans. Information Theory **IT-22** (1976), no. 6, 644–654. MR 55 #10141

[Eul85] Leonhard Euler, *An essay on continued fractions*, Math. Systems Theory **18** (1985), no. 4, 295–328, Translated from the Latin by B. F. Wyman and M. F. Wyman. MR 87d:01011b

[FT93] A. Fröhlich and M. J. Taylor, *Algebraic Number Theory*, Cambridge University Press, Cambridge, 1993. MR 94d:11078

[Guy94] R. K. Guy, *Unsolved Problems in Number Theory*, second ed., Springer-Verlag, New York, 1994, Unsolved Problems in Intuitive Mathematics, I. MR 96e:11002

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[Har77] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[Hoo67] C. Hooley, *On Artin's conjecture*, J. Reine Angew. Math. **225** (1967), 209–220. MR 34 #7445

[HW79] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, fifth ed., The Clarendon Press Oxford University Press, New York, 1979. MR 81i:10002

[IBM01] IBM, *IBM's Test-Tube Quantum Computer Makes History,* `http://www.research.ibm.com/resources/news/20011219_quantum.shtml`.

[IR90] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, second ed., Springer-Verlag, New York, 1990. MR 92e:11001

[Khi63] A. Ya. Khintchine, *Continued fractions*, Translated by Peter Wynn, P. Noordhoff Ltd., Groningen, 1963. MR 28 #5038

[Knu97] Donald E. Knuth, *The Art of Computer Programming*, third ed., Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1997, Volume 1: Fundamental algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[Knu98] ———, *The Art of Computer Programming. Vol. 2*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1998, Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing. MR 83i:68003

[Kob84] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Graduate Texts in Mathematics, vol. 97, Springer-Verlag, New York, 1984. MR 86c:11040

[Leh14] D. N. Lehmer, *List of Primes Numbers from* 1 *to* 10,006,721, Carnegie Institution Washington, D.C. (1914).

[Lem] F. Lemmermeyer, *Proofs of the Quadratic Reciprocity Law,* `http://www.rzuser.uni-heidelberg.de/~hb3/rchrono.html`.

[Len87] H. W. Lenstra, Jr., *Factoring integers with elliptic curves*, Ann. of Math. (2) **126** (1987), no. 3, 649–673. MR 89g:11125

[LL93] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The Development of the Number Field Sieve*, Lecture Notes in Mathematics, vol. 1554, Springer-Verlag, Berlin, 1993. MR 96m:11116

[LMG$^+$01] Vandersypen L. M., Steffen M., Breyta G., Yannoni C. S., Sherwood M. H., and Chuang I. L., *Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance*, Nature **414** (2001), no. 6866, 883–887.

[LT72] S. Lang and H. Trotter, *Continued fractions for some algebraic numbers*, J. Reine Angew. Math. **255** (1972), 112–134; addendum, ibid. **267** (1974), 219–220; MR **50** #2086. MR 46 #5258

[LT74] ______, *Addendum to: Continued fractions for some algebraic numbers (J. Reine Angew. Math.* **255** *(1972), 112–134)*, J. Reine Angew. Math. **267** (1974), 219–220. MR 50 #2086

[Mor93] P. Moree, *A note on Artin's conjecture*, Simon Stevin **67** (1993), no. 3-4, 255–257. MR 95e:11106

[MS08] B. Mazur and W. Stein, *What is Riemann's Hypothesis?*, 2008, In preparation.

[NZM91] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory of Numbers*, fifth ed., John Wiley & Sons Inc., New York, 1991. MR 91i:11001

[Old70] C. D. Olds, *The Simple Continued Fraction Expression of* e, Amer. Math. Monthly **77** (1970), 968–974.

[Per57] O. Perron, *Die Lehre von den Kettenbrüchen. Dritte, verbesserte und erweiterte Aufl. Bd. II. Analytisch-funktionentheoretische Kettenbrüche*, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1957. MR 19,25c

[RSA] RSA, *The New RSA Factoring Challenge,* `http://www.rsasecurity.com/rsalabs/challenges/factoring`.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Comm. ACM **21** (1978), no. 2, 120–126. MR 83m:94003

[Sag08] Sage, *Free Open Source Mathematical Software (Version 3.0.4)*, 2008, `http://www.sagemath.org`.

[Sch85] R. Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*, Mathematics of Computation **44** (1985), no. 170, 483–494.

[Sho97] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput. **26** (1997), no. 5, 1484–1509. MR 98i:11108

[Sho05] V. Shoup, *A Computational Introduction to Number Theory and Algebra*, Cambridge University Press, Cambridge, 2005. MR MR2151586 (2006g:11003)

[Sil86] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, New York, 1986. MR 87g:11070

[Sin99] S. Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, Doubleday, 1999.

[Slo] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

[ST92] J. H. Silverman and J. Tate, *Rational points on elliptic curves*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1992. MR 93g:11003

[Wal48] H. S. Wall, *Analytic Theory of Continued Fractions*, D. Van Nostrand Company, Inc., New York, N. Y., 1948. MR 10,32d

[Wei03] E. W. Weisstein, *RSA-576 Factored,* `http://mathworld.wolfram.com/news/2003-12-05/rsa/`.

[Wil00] A. J. Wiles, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[Zag75] D. Zagier, *The first 50 million prime numbers*, `http://modular.fas.harvard.edu/scans/papers/zagier/`.

Index

36 Fast computation of Hermite normal forms of random integer matrices, with C. Pernet

Journal of Number Theory ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt

Fast computation of Hermite normal forms of random integer matrices

Clément Pernet [a,1], William Stein [b,*,2]

[a] Grenoble université, INRIA-MOAIS, LIG; 51, avenue J. Kuntzmann; 38330 Montbonnot St-Martin, France
[b] University of Washington, Department of Mathematics, Box 354350, Seattle, WA, United States

ARTICLE INFO

Article history:
Received 21 April 2009
Available online xxxx
Communicated by Fernando Rodriguez Villegas

Keywords:
Hermite normal form
Exact linear algebra

ABSTRACT

This paper is about how to compute the Hermite normal form of a *random* integer matrix in practice. We propose significant improvements to the algorithm by Micciancio and Warinschi, and extend these techniques to the computation of the saturation of a matrix. We describe the fastest implementation for computing Hermite normal form for large matrices with large entries.

© 2010 Published by Elsevier Inc.

1. Introduction

This paper is about how to compute the Hermite normal form of a *random* integer matrix in practice. We describe the best known algorithm for random matrices, due to Micciancio and Warinschi [MW01] and explain some new ideas that make it practical. We also apply these techniques to give a new algorithm for computing the saturation of a module, and present timings.

In this paper we do not concern ourselves with nonrandom matrices, and instead refer the reader to [SL96,Sto98] for the state of the art for worse case complexity results. Our motivation for focusing on the random case is that it comes up frequently in algorithms for computing with modular forms.

Among the numerous notions of Hermite normal form, we use the following one, which is the closest to the familiar notion of reduced row echelon form.

Definition 1.1 *(Hermite normal form).* For any $n \times m$ integer matrix A the *Hermite normal form* (HNF) of A is the unique matrix $H = (h_{i,j})$ such that there is a unimodular $n \times n$ matrix U with $UA = H$, and such that H satisfies the following two conditions:

[*] Corresponding author.
E-mail address: wstein@gmail.com (W. Stein).

[1] Supported by the National Science Foundation under Grant No. 0713225.
[2] Supported by the National Science Foundation under Grant No. 0653968.

0022-314X/$ – see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.jnt.2010.01.017

- there exist a sequence of integers $j_1 < \cdots < j_n$ such that for all $0 \leqslant i \leqslant n$ we have $h_{i,j} = 0$ for all $j < j_i$ (row echelon structure),
- for $0 \leqslant k < i \leqslant n$ we have $0 \leqslant h_{k,j_i} < h_{i,j_i}$ (the pivot element is the greatest along its column and the coefficients above are nonnegative).

Thus the Hermite normal form is a generalization over $\mathbb{Z}$ of the reduced row echelon form of a matrix over $\mathbb{Q}$. Just as computation of echelon forms is a building block for many algorithms for computing with vector spaces, Hermite normal form is a building block for algorithms for computing with modules over $\mathbb{Z}$ (see, e.g., [Coh93, Chapter 2]).

Example 1.2. The HNF of the matrix

$$A = \begin{pmatrix} -5 & 8 & -3 & -9 & 5 & 5 \\ -2 & 8 & -2 & -2 & 8 & 5 \\ 7 & -5 & -8 & 4 & 3 & -4 \\ 1 & -1 & 6 & 0 & 8 & -3 \end{pmatrix}$$

is

$$H = \begin{pmatrix} 1 & 0 & 3 & 237 & -299 & 90 \\ 0 & 1 & 1 & 103 & -130 & 40 \\ 0 & 0 & 4 & 352 & -450 & 135 \\ 0 & 0 & 0 & 486 & -627 & 188 \end{pmatrix}.$$

Notice how the entries in the answer are quite large compared to the input.

Heuristic observations: For a random $n \times m$ matrix A with $n \leqslant m$, the number of digits of each entry of the rightmost $m - n + 1$ columns of H are similar in size to the determinant of the left $n \times n$ submatrix of A. For example, a random 250×250 matrix with entries in $[-2^{32}, 2^{32}]$ has HNF with entries in the last column all having about 2590 digits and determinant with about 2590 digits, but all other entries are likely to be very small (e.g., a single digit).

There are numerous algorithms for the computing HNF's, including [KB79,DKLET87,Bra89,MW01]. We describe an algorithm that is based on the heuristically fast algorithm by Micciancio and Warinschi [MW01], updated with several practical improvements.

In the rest of this paper, we mainly address computation of the HNF of a square nonsingular matrix A. We also briefly explain how to reduce the general case to the square case, discuss computation of saturation, and give timings. We give an outline of the algorithm in Section 2 and present more details in Sections 3, 5 and 6. The cases of more rows than columns and more columns than rows is discussed in the Section 7. In Section 9, we sketch the main features of our implementation in Sage, and compare the computation time for various class of matrices.

2. Outline of the algorithm when A is square

For the rest of this section, let $A = (a_{i,j})_{i,j=0,\ldots,n-1}$ be an $n \times n$ matrix with integer entries. There are two key ideas behind the algorithm of [MW01] for computing the HNF of A.

1. Every entry in the HNF H of a square matrix A is at most the absolute value of the determinant $\det(A)$, so one can compute H be working modulo the determinant of H. This idea was first introduced and developed in [DKLET87].
2. The determinant of A may of course still be extremely large. Micciancio and Warinschi's clever idea is to instead compute the Hermite form H' of a small-determinant matrix constructed

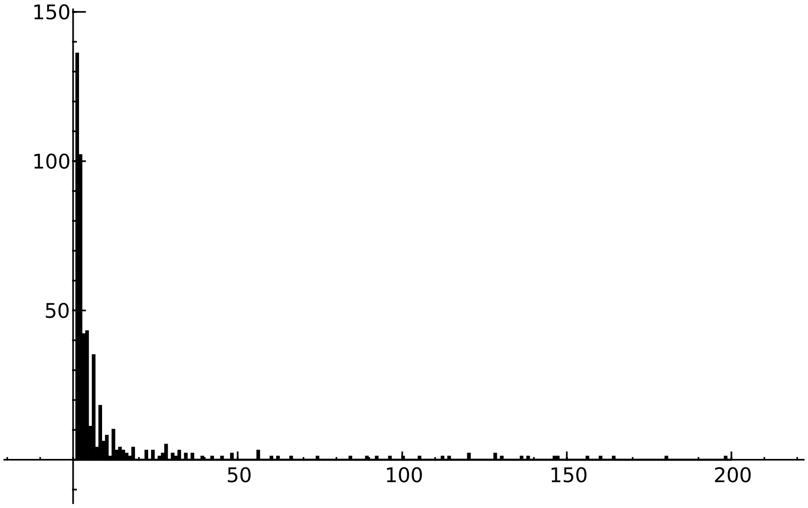


Fig. 2.1. Distribution of the determinants in (2.1), for 500 random matrices with $n = 100$ and entries uniformly chosen to satisfy $\log_2 \|A\| = 100$. Only 9 elements had a determinant larger than 200, and the largest one was 6816.

from A using the Euclidean algorithm and properties of determinants. Then we recover H from H' via three update steps.

We now explain the second key idea in more detail. Consider the following block decomposition of A:

$$A = \begin{bmatrix} B & b \\ c^T & a_{n-1,n} \\ d^T & a_{n,n} \end{bmatrix},$$

where B is the upper left $(n-2) \times (n-1)$ submatrix of A, and b, c, d are column vectors. Let $d_1 = \det\left(\begin{bmatrix} B \; c^T \end{bmatrix}\right)$ and $d_2 = \det\left(\begin{bmatrix} B \; d^T \end{bmatrix}\right)$. Use the extended Euclidean algorithm to find integers s, t such that

$$g = sd_1 + td_2,$$

where $g = \gcd(d_1, d_2)$.

Since the determinant is linear in row operations, we have

$$\det\left(\begin{bmatrix} B \\ sc^T + td^T \end{bmatrix}\right) = g. \tag{2.1}$$

For random matrices, g is likely to be very small. Fig. 2.1 illustrates the distribution of such gcd's, on a set of 500 random integer matrices of dimension 100 with 100-bit coefficients.

Algorithm 1 (on page 4) is essentially the algorithm of Micciancio and Warinschi. Our main improvement over their work is to greatly optimize Steps 3, 4 and 8. Step 8 is performed by a procedure they call `AddColumn` (see Algorithm 3 in Section 5 below), and steps 9 and 10 by a procedure they call `AddRow` (see Algorithm 4 in Section 6 below).

3. Double determinant computation

There are many algorithms for computing the determinant of an integer matrix A. One algorithm involves computing the Hadamard bound on $\det(A)$, then computing the determinant modulo p for

Algorithm 1: Hermite Normal Form [MW01]

```
    Data: $A$: an $n \times n$ nonsingular matrix over $\mathbb{Z}$
    Result: $H$: the Hermite normal form of $A$
1   begin
2       Write $A = \begin{bmatrix} B & b \\ c^T & a_{n-1,n} \\ d^T & a_{n,n} \end{bmatrix}$
3       Compute $d_1 = \det\left(\begin{bmatrix} B \\ c^T \end{bmatrix}\right)$
4       Compute $d_2 = \det\left(\begin{bmatrix} B \\ d^T \end{bmatrix}\right)$
5       Compute the extended gcd of $d_1$ and $d_2$: $g = sd_1 + td_2$
6       Let $C = \begin{bmatrix} B \\ sc^T + td^T \end{bmatrix}$
7       Compute $H_1$, the Hermite normal form of $C$, by working modulo $g$ as explained in Section 4
        below. (NOTE: In the unlikely case that $g = 0$ or $g$ is large, we compute $H_1$ using any HNF
        algorithm applied to $C$, e.g., by recursively applying the main algorithm of this paper to $C$.)
8       Obtain from $H_1$ the Hermite form $H_2$ of $\begin{bmatrix} B & b \\ sc^T + td^T & sa_{n-1,n} + ta_{n,n} \end{bmatrix}$
9       Obtain from $H_2$ the hermite form $H_3$ of $\begin{bmatrix} B & b \\ c^T & a_{n-1,n} \end{bmatrix}$
10      Obtain from $H_3$ the Hermite form $H$ of $\begin{bmatrix} B & b \\ c^T & a_{n-1,n} \\ d^T & a_{n,n} \end{bmatrix}$
11  end
```

sufficiently many p using an (asymptotically fast) Gaussian elimination algorithm, and finally using a Chinese remainder theorem reconstruction. This algorithm has bit complexity

$$\mathcal{O}\left(n^4\left(\log n + \log\|A\|\right) + n^3 \log^2 \|A\|\right),$$

or $\mathcal{O}(n^{\omega+1}(\log n + \log\|A\|))$ with fast matrix arithmetic (see [GG99, Chapter 5]).

Abbott, Bronstein and Mulders [ABM99] propose another determinant algorithm based on solving $Ax = v$ for a random integer vector v using an iterative p-adic solving algorithm (e.g., [Dix82,MC79]). In particular, by Cramer's rule the greatest common divisor of the denominators of the entries of x is a divisor d of $D = \det(A)$. The unknown integer D/d can be recovered by computing it modulo p for several primes and using the Chinese remainder theorem; usually D/d is very small, so this is fast. This approach has a similar worst case bit complexity: $\mathcal{O}(n^4 + n^3(\log n + \log\|A\|)^2)$ but a better average case complexity of $\mathcal{O}(n^3(\log^2 n + \log\|A\|)^2)$.

The computation time can also be improved by allowing early termination in the Chinese remainder algorithm: once a reconstruction stabilizes modulo several primes, the result is likely to remain the same with a certified probability, and one can avoid the remaining modular computations.

Further details on practical implementations for computing determinants of integer matrices can be found in [DU06].

Storjohann [Sto05] obtains the best known bit complexity for computing determinants using a Las Vegas algorithm. He obtains a complexity of $\mathcal{O}\tilde{}(n^\omega \log\|A\|)$, where ω is the exponent for matrix multiplication. However, no implementation of this algorithm is known that is better in practice than the p-adic lifting based method for practical problem sizes. Consequently, we based our implementation on this latter algorithm by [ABM99].

The computation of the two determinants (Steps 3 and 4) therefore involves the solving of two systems, with very similar matrices. We reduce it to only one system solution *in the generic case* using the following lemma. Since this is a bottleneck in the algorithm, this factor of two savings is huge in practice.

Please cite this article in press as: C. Pernet, W. Stein, Fast computation of Hermite normal forms of random integer matrices, J. Number Theory (2010), doi:10.1016/j.jnt.2010.01.017

Algorithm 2: Double determinant computation

```
Data: $B$: an $(n-1) \times n$ matrix over $\mathbb{Z}$
Data: $c, d$: two vectors in $\mathbb{Z}^n$
Result: $(d_1, d_2) = (\det(\begin{bmatrix} B^T & c \end{bmatrix}), \det(\begin{bmatrix} B^T & d \end{bmatrix}))$
begin
    Solve the system $\begin{bmatrix} B^T & c \end{bmatrix} x = d$ using Dixon's $p$-adic lifting
    Then $y_i = -x_i/x_n$, $y_n = 1/x_n$ solves $\begin{bmatrix} B^T & d \end{bmatrix} y = c$ by Lemma 3.1, unless $x_n = 0$, in which case
    we use the usual determinant algorithm to compute the determinants of the two matrices
    $u_1 = \mathrm{lcm}(\mathrm{denominators}(x))$
    $u_2 = \mathrm{lcm}(\mathrm{denominators}(y))$
    Compute Hadamard's bounds $h_1$ and $h_2$ on the determinants of $\begin{bmatrix} B^T & c \end{bmatrix}$ and $\begin{bmatrix} B^T & d \end{bmatrix}$
    Select a set of primes $(p_i)$ s.t. $\prod_i p_i > \max(h_1/u_1, h_2/u_2)$
    foreach $p_i$ do
        compute $B^T = LUP$, the LUP decomposition of $B^T \bmod p_i$
        $q = \prod_{i=1}^{n-1} U_{i,i} \bmod p_i$
        $x = L^{-1} c \bmod p_i$
        $y = L^{-1} d \bmod p_i$
        $v_1^{(i)} = q x_n \bmod p_i$
        $v_2^{(i)} = q y_n \bmod p_i$
    reconstruct $v_1$ and $v_2$ from $(v_1^{(i)})$ and $(v_2^{(i)})$ using CRT
    return $(d_1, d_2) = (u_1 v_1, u_2 v_2)$
end
```

Lemma 3.1. *Let A be an $n \times (n-1)$ matrix and c and d column vectors of degree n, and assume that the augmented matrices $[A|c]$ and $[A|d]$ are both invertible. Let $x = (x_i)$ be the solution of $[A|c]x = d$. If $x_n \neq 0$, then the solution $y = (y_i)$ to $[A|d]y = c$ is*

$$y = \left(-\frac{x_1}{x_n}, -\frac{x_2}{x_n}, \ldots, -\frac{x_{n-1}}{x_n}, \frac{1}{x_n}\right).$$

Proof. Write a_i for the ith column of A. The equation $[A|c]x = d$ is thus $(\sum_{i=1}^{n-1} a_i x_i) + c x_n = d$, so $(\sum_{i=1}^{n-1} a_i x_i) - d = -x_n c$. Dividing both sides by $-x_n$ yields $(\sum_{i=1}^{n-1}(-\frac{x_i}{x_n})a_i) + \frac{1}{x_n} d = c$, which proves the lemma. □

Example 3.2. Let $A = \begin{bmatrix} 1 & 2 \\ -4 & 3 \\ 2 & -5 \end{bmatrix}$, $c = (-1, 3, 5)^T$, and $d = (2, -3, 4)^T$. The solution to $[A|c]x = d$ is

$$x = \left(\frac{111}{68}, \frac{35}{68}, \frac{45}{68}\right).$$

Thus

$$y = \left(-\frac{x_1}{x_3}, -\frac{x_2}{x_3}, \frac{1}{x_3}\right) = \left(-\frac{37}{15}, -\frac{7}{9}, \frac{68}{45}\right).$$

Algorithm 2 (on page 5) describes how the two determinants are computed using Lemma 3.1.

4. Hermite form modulo g

Recall that C is a square nonsingular matrix with "small" determinant g. Step 7 of Algorithm 1 (on page 4) is to compute the HNF of C as explained in [DKLET87, §3]. There it is proved that since $g = \det(C)$, the Hermite normal form of $\left[\begin{smallmatrix} C \\ gI \end{smallmatrix}\right]$ is $\left[\begin{smallmatrix} H \\ 0 \end{smallmatrix}\right]$ where H is the Hermite normal form of C. Using this result, to compute H, we apply the standard row reduction Hermite normal form algorithm to C, always reducing all numbers modulo g. Conceptually, think of this as adding multiples of the rows of gI, which does not change the resulting Hermite form. At the end of this process we obtain a matrix $H = (h_{ij})$ with $0 \leqslant h_{ij} < g$ for all ij. There is one special case; since the product of the diagonal entries of the Hermite form of C is g, if the lower right entry of H is 0, then we replace it by g. Then the resulting matrix H is the Hermite normal form of C.

For additional discussion of the modular Hermite form algorithm, see [Coh93, §2.4, p. 71] which describes the algorithm in detail, including a discussion of our above remark about replacing 0 by g.

Example 4.1. Let $C = \left[\begin{smallmatrix} 5 & 26 \\ 2 & 11 \end{smallmatrix}\right]$. Then $g = \det(C) = 3$, and the reduction mod g of C is $\left[\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right]$. Subtracting the second row from the first yields $\left[\begin{smallmatrix} 2 & 2 \\ 0 & 0 \end{smallmatrix}\right]$, which is already reduced modulo 3. Then multiplying through the first row by -1 and reducing modulo 3 again, we obtain $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Then, as mentioned above, since the lower right entry is 0, we replace it by $g = 3$, obtaining the Hermite normal form $H = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 3 \end{smallmatrix}\right]$.

5. Add a column

Step 8 of Algorithm 1 is to find a column vector e such that

$$[\, H_1 \quad e \,] = U \begin{bmatrix} B & b \\ sc^T + td^T & a_{n-1,n} \end{bmatrix} \tag{5.1}$$

is in Hermite form, for a unimodular matrix U.

By hypothesis $C = \left[\begin{smallmatrix} B \\ sc^T + td^T \end{smallmatrix}\right]$ is invertible, so from (5.1), one gets

$$\begin{aligned} e &= U \begin{bmatrix} b \\ a_{n-1,n-1} \end{bmatrix} \\ &= H_1 \begin{bmatrix} B \\ sc^T + td^T \end{bmatrix}^{-1} \begin{bmatrix} b \\ a_{n-1,n-1} \end{bmatrix}. \end{aligned}$$

In [MW01], the column e is computed using multi-modular computations and a tight bound on the size of the entries of e. We instead use the p-adic lifting algorithm of [Dix82,MC79] to solve the system

$$\begin{bmatrix} B \\ sc^T + td^T \end{bmatrix} x = \begin{bmatrix} b \\ a_{n-1,n-1} \end{bmatrix}.$$

However, the last row $sc^T + td^T$ typically has much larger coefficients than the rest of the matrix, thus unduly penalizing the complexity of finding a solution. Our key idea is to replace the row $sc^T + td^T$ by a random row u that has small entries such that the resulting matrix is still invertible, find the solution y of this modified system, then recover x as follows. Let $\{k\}$ be a basis of the 1-dimensional kernel of B. Then the sought for solution of the original system is

$$x = y + \alpha k,$$

```
Algorithm 3: AddColumn
Data: $B = \begin{bmatrix} B_1 & b_2 \\ b_3^T & b_4 \end{bmatrix}$: an $n \times n$ matrix over $\mathbb{Z}$, where $B_1$ is $(n-1) \times (n-1)$ and $b_2, b_3$ are vectors
Data: $H_1$: the Hermite normal form of $\begin{bmatrix} B_1 \\ b_3^T \end{bmatrix}$
Result: $H$: the Hermite normal form of $B$
begin
    Pick a random vector $u$ such that $|u_i| \leqslant \|B\|\ \forall i$
    Solve $\begin{bmatrix} B_1 \\ u \end{bmatrix} y = \begin{bmatrix} b_2 \\ b_4 \end{bmatrix}$
    Compute a kernel basis vector $k$ of $B_1$
    $\alpha = b_4 - \frac{b_3^T \cdot y}{b_3^T \cdot k}$
    $x = y + \alpha k$
    $e = H_1 x$
    return $[H_1\ e]$
end
```

where α satisfies

$$\left(sc^T + td^T\right) \cdot (y + \alpha k) = a_{n-1,n-1}.$$

By linearity of the dot product, we have

$$\alpha = \frac{a_{n-1,n-1} - (sc^T + td^T) \cdot y}{(sc^T + td^T) \cdot k}.$$

Note that if $(sc^T + td^T) \cdot k = 0$, then $Ck = 0$, which would contradict our assumption that $C = \begin{bmatrix} B \\ sc^T + td^T \end{bmatrix}$ is invertible.

6. Add a row

Steps 9 and 10 of Algorithm 1 consist of adding a new row to the current Hermite form and updating it to obtain a new matrix in Hermite form.

The principle is to eliminate the new row with all existing pivots and update the already computed parts when necessary. Algorithm 4 (on page 8) describes this in more detail.

7. The nonsquare case

In the case where the matrix is rectangular, with dimensions $m \times n$, we reduce to the case of a square nonsingular matrix as follows: first compute the column and row rank profile (pivot columns and subset of independent rows) of A modulo a random word-size prime. With high probability, the matrix A has the same column and row rank profile over $\mathbb{Q}$, so we can now apply Algorithm 1 to the square nonsingular $r \times r$ matrix obtained by picking the row and column rank profile submatrix of A over $\mathbb{Z}$.

The additional rows and columns are then incorporated as follows:

additional columns: use Algorithm 3 (`AddColumn`) with a block of column vectors instead of just one column. If this fails, then we computed the rank profile incorrectly, in which case we start over with a different random prime.

additional rows: use Algorithm 4 (`AddRow`) for each additional row.

```
Algorithm 4: AddRow
Data: $A$: an $m \times n$ matrix in Hermite normal form
Data: $b$: a vector of degree $n$
Result: $H$: the Hermite normal form of $\begin{bmatrix} A \\ b \end{bmatrix}$
begin
    forall pivots $a_{i,j_i}$ of $A$ do
        if $b_{j_i} = 0$ then
            continue
        if $A_{i,j_i} | b_{j_i}$ then
            $b := b - b_{j_i}/A_{i,j_i} A_{i,1\dots n}$
        else
            /* Extended gcd based elimination */
            $(g, s, t) = \mathrm{XGCD}(a_{i,j_i}, b_{j_i})$ ;        /* so $g = sa_{i,j_i} + tb_{j_i}$ */
            $A_{i,1\dots n} := sA_{i,1\dots n} + tb_{j_i}$
            $b := b_{j_i}/gA_{i,1\dots n} - A_{i,j_i}/gb$
            for $k = 1$ to $i - 1$ do
                /* Reduces row $k$ with row $i$ */
                $A_{k,1\dots n} := A_{k,1\dots n} - \lfloor A_{k,j_i}/A_{i,j_i} \rfloor A_{i,1\dots n}$
    if $b \neq 0$ then
        let $j$ be the index of the first nonzero element of $b$
        insert $b^T$ between rows $i$ and $i + 1$ such that $j_i < j < j_{i+1}$
    Return $H = \begin{bmatrix} A \\ b \end{bmatrix}$
end
```

8. Saturation

If M is a submodule of $\mathbb{Z}^n$ for some n, then the saturation of M is $\mathbb{Z}^n \cap (\mathbb{Q}M)$, i.e., the intersection with $\mathbb{Z}^n$ of the $\mathbb{Q}$-span of any basis of M. For example, if M has rank n, then the saturation of M just equals $\mathbb{Z}^n$. Also, kernels of homomorphisms of free $\mathbb{Z}$-modules are saturated. Saturation comes up in many number theoretic algorithms, e.g., saturation is an important step in computing a basis over $\mathbb{Z}$ for the space of q-expansions of cuspidal modular forms of given weight and level, and comes up in explicit computation with homology of modular curves using modular symbols.

There is a well-known connection between saturation and Hermite form. If A is a basis matrix for M, and H is the Hermite form of the transpose of A with any 0 rows at the bottom deleted (so H is square), then $H^{-1}A$ is a matrix whose rows are a basis for the saturation of M. Thus computation of a saturation of a matrix reduces to computation of one Hermite form and solving a system $HX = A$.

If A is sufficiently random, then the Hermite form matrix H has a very large last column and all other entries are small, so we exploit the trick in Section 6 and instead solve a much easier system.

9. Implementation

Our implementation of the algorithms described in this paper are included in Sage [Ste]. This implementation relies on `IML` [SC] for the solution of integer systems using p-adic lifting, and on `LinBox` [Lin] for the computation of determinants modulo p (the `IML` and `LinBox` libraries are both part of Sage). Our implementation is primarily optimized for the square case.

We illustrate computing a Hermite normal form and saturation in Sage.

```
sage: A = matrix(ZZ,3,5,[-1,2,5,65,2,4,-1,-3,1,-2,-1,-2,1,-1,1])
sage: A
[-1  2  5 65  2]
```

Please cite this article in press as: C. Pernet, W. Stein, Fast computation of Hermite normal forms of random integer matrices, J. Number Theory (2010), doi:10.1016/j.jnt.2010.01.017

```
[ 4 -1 -3  1 -2]
[-1 -2  1 -1  1]
sage: A.hermite_form()
[  1   0  17 259   7]
[  0   1  31 453  13]
[  0   0  40 582  17]
sage: A.saturation()
[-1  2  5 65  2]
[ 4 -1 -3  1 -2]
[-1 -2  1 -1  1]
```

There are implementations of Hermite normal form algorithms in NTL [Sho], PARI [PAR], GAP [GAP], Maple and Mathematica. The algorithm in this paper is asymptotically better than these standard implementations.

Acknowledgments

We thank Allan Steel for providing us with notes from a talk he gave on his implementation of [MW01]. Steel's implementation was much faster than any other available (e.g., PARI, NTL, Mathematica, Maple, and GAP), which was the motivation for this paper. We would also like to thank Burcin Erocal for implementing mod n computation of Hermite form in Sage, Robert Bradshaw for help benchmarking our implementation, Arne Storjohann for helpful conversations about the nonrandom case, and Andrew Crites and Michael Goff for discussions.

References

[ABM99] J. Abbott, M. Bronstein, T. Mulders, Fast deterministic computation of determinants of dense matrices, in: International Symposium on Symbolic and Algebraic Computation, ACM Press, 1999.

[Bra89] R.J. Bradford, Hermite normal forms for integer matrices, in: European Conference on Computer Algebra (EUROCAL '87), Springer-Verlag, Berlin–Heidelberg–New York, 1989, pp. 315–316.

[Coh93] H. Cohen, A Course in Computational Algebraic Number Theory, Springer-Verlag, Berlin, 1993. MR 94i:11105.

[Dix82] John D. Dixon, Exact solution of linear equations using p-adic expansions, Numer. Math. 40 (1982) 137–141.

[DKLET87] P.D. Domich, R. Kannan, L.E. Trotter Jr., Hermite normal form computation using modulo determinant arithmetic, Math. Oper. Res. 12 (1) (1987) 50–59.

[DU06] Jean-Guillaume Dumas, Anna Urbanska, An introspective algorithm for the integer determinant, in: Proc. Transgressive Computing, Grenade, Spain, 2006, pp. 185–202.

[GAP] GAP, Groups, algorithms, programming—a system for computational discrete algebra, http://www-gap.mcs.st-and.ac.uk/.

[GG99] Joachim von zur Gathen, Jürgen Gerhard, Modern Computer Algebra, Cambridge University Press, New York, NY, USA, 1999.

[KB79] Ravindran Kannan, Achim Bachem, Polynomial algorithms for computing the smith and hermite normal forms of an integer matrix, SIAM J. Comput. 8 (4) (1979) 499–507.

[Lin] The LinBox Group, LinBox: Exact linear algebra with dense and blackbox matrices (Version 1.1.6), http://www.linalg.org.

[MC79] R. Moenck, J. Carter, Approximate algorithms to derive exact solutions to systems of linear equations, in: Edward W. Ng (Ed.), Proceedings of the International Symposium on Symbolic and Algebraic Manipulation (EUROSAM '79), Marseille, France, in: Lecture Notes in Comput. Sci., vol. 72, Springer-Verlag, 1979, pp. 65–73.

[MW01] Daniele Micciancio, Bogdan Warinschi, A linear space algorithm for computing the Hermite Normal Form, in: International Symposium on Symbolic and Algebraic Computation, ACM Press, 2001, pp. 231–236.

[PAR] PARI, A computer algebra system designed for fast computations in number theory, http://pari.math.u-bordeaux.fr/.

[SC] Arne Storjohann, Zhuliang Chen, Integer Matrix Library (Version 1.0.2), http://www.cs.uwaterloo.ca/~z4chen/iml.html.

[Sho] V. Shoup, NTL: Number theory library, http://www.shoup.net/ntl/.

[SL96] Arne Storjohann, George Labahn, Asymptotically fast computation of Hermite normal forms of integer matrices, in: Proc. International Symp. on Symbolic and Algebraic Computation: ISSAC '96, ACM Press, 1996, pp. 259–266.

[Ste] William Stein, Sage: Open Source Mathematical Software (Version 3.2.3), The Sage Group, http://www.sagemath.org.

[Sto98] Arne Storjohann, Computing Hermite and Smith normal forms of triangular integer matrices, Linear Algebra Appl. 282 (1998) 25–45.

[Sto05] Arne Storjohann, The shifted number system for fast linear algebra on integer matrices, J. Complexity 21 (4) (2005) 609–650.

Please cite this article in press as: C. Pernet, W. Stein, Fast computation of Hermite normal forms of random integer matrices, J. Number Theory (2010), doi:10.1016/j.jnt.2010.01.017

37 Toward a Generalization of the Gross-Zagier Conjecture

OXFORD UNIVERSITY PRESS LTD JOURNAL **00** (0000), 1–17
doi:10.1093/OUP Journal/XXX000

Toward a Generalization of the Gross-Zagier Conjecture

William Stein[1]*
[1] *Department of Mathematics, University of Washington, Web:* `http://wstein.org`*, Email:* `wstein@uw.edu`

Abstract: We review some of Kolyvagin's results and conjectures about elliptic curves, then make a new conjecture that slightly refines Kolyvagin's conjectures. We introduce a definition of finite index subgroups $W_p \subset E(K)$, one for each prime p that is inert in a fixed imaginary quadratic field K. These subgroups generalize the group $\mathbb{Z}y_K$ generated by the Heegner point $y_K \in E(K)$ in the case $r_{\rm an} = 1$. For any curve with $r_{\rm an} \geq 1$, we give a description of W_p, which is conditional on truth of the Birch and Swinnerton-Dyer conjecture and our conjectural refinement of Kolyvagin's conjecture. We then deduce the following conditional theorem, up to an explicit finite set of primes: (a) the set of indexes $[E(K) : W_p]$ is finite, and (b) the subgroups W_p with $[E(K) : W_p]$ maximal satisfy a higher-rank generalization of the Gross-Zagier formula. We also investigate a higher-rank generalization of a conjecture of Gross-Zagier.

KEY WORDS number theory; Birch and Swinnerton-Dyer Conjecture; Euler systems; Heegner points; Kolyvagin's conjecture; computational number theory; Gross-Zagier theorem

Received

1 Introduction

Let E be an elliptic curve defined over $\mathbb{Q}$. The order of vanishing $r_{\rm an}$ at $s = 1$ of the Hasse-Weil L-series $L(E/\mathbb{Q}, s)$ of E is defined because E is modular (see [BCDT01, Wil95]). The Birch and Swinnerton-Dyer (BSD) rank conjecture [Bir65] asserts that $r_{\rm an}$ is equal to the algebraic rank $r_{\rm alg}$ of $E(\mathbb{Q})$. The BSD formula then gives a conjectural formula for the leading coefficient of the Taylor expansion about $s = 1$ of $L(E/\mathbb{Q}, s)$; this formula resembles the analytic class number formula. The BSD rank conjecture is known for curves with $r_{\rm an} \leq 1$, but there has been relatively little progress toward the BSD rank conjecture when $r_{\rm an} \geq 2$.

In the late 1980s, Kolyvagin wrote several landmark papers that combined the Gross-Zagier theorem [GZ86] about heights of Heegner points over quadratic imaginary fields K, a theorem [BFH90] about nonvanishing of special values of twists of L-functions, and relations involving Hecke operators between Heegner points over ring class fields of K to prove that if $r_{\rm an} \leq 1$, then the BSD rank conjecture is true for E. Kolyvagin wrote [Kol91a] on the case of general rank, in which he computes the elementary invariants of the Selmer groups of any elliptic curve E of any rank in terms of properties of Heegner points, assuming a certain nontriviality hypothesis. It was until recently unclear whether or not this hypothesis was ever satisfied for any curve with $r_{\rm an} \geq 2$. Fortunately, this hypothesis has now been confirmed numerically (with high probability) in one case of a rank 2 curve [JLS08].

We review some of Kolyvagin's results and conjectures from [Kol91a], then make a new conjecture that refines Kolyvagin's conjectures. Using reduction modulo p of Heegner points, we introduce a definition of finite index subgroups $W_p \subset E(K)$, one for each prime p that is inert in K. Let $y_K \in E(K)$ be the associated Heegner point as in Equation (1) below. Then these subgroups W_p generalize the group $\mathbb{Z}y_K$ in the case $r_{\rm an} = 1$. For any $r_{\rm an} \geq 1$, we give a description of W_p, which is conditional on truth of the BSD conjecture and our conjectural refinement of Kolyvagin's conjecture. We then deduce the following conditional theorem (see Theorems 7.5 and 7.7), up to an explicit finite set of primes: (a) the set of indexes $[E(K) : W_p]$ is finite, and (b) the subgroups W_p with $[E(K) : W_p]$ maximal satisfy a higher-rank generalization of the Gross-Zagier formula (see (5) below). We also give numerical data and a new conjecture about the existence of Gross-Zagier subgroups.

We leave open far more questions than we answer, and we intend to follow up on these questions in subsequent papers. For example, perhaps the definition of the groups W_p can be refined and generalized in various ways, and results similar to those in this paper proved about them. It would be interesting to find a practical algorithm that can provably compute the groups W_p for a particular p, assuming that $E(K)$ has already been computed. We also hope to find a higher-rank analogue of the Gross-Zagier formula over the Hilbert class field of K, involving the Petersson inner product, modular forms, and Rankin-Selberg convolutions $L_{\mathcal{A}}(f, s)$, as in [GZ86], which is consistent with the results we prove about the groups W_p in this paper. It would also be

*Supported by NSF grant 0555776.

Copyright © 0000 Oxford University Press

valuable to give proofs of the results of [Kol91a] building on [McC91] instead of [Kol91b], possibly using results from the present paper.

We briefly outline the structure of this paper. In the first few sections, we state the BSD conjecture and Gross-Zagier formula, define Kolyvagin points, state Kolyvagin's conjectures, and then define certain finite index subgroups W_p of $E(K)$. In the rest of the paper, we study reduction mod p, conditionally deduce the structure of W_p, and give some numerical examples.

More precisely, we do the following. In Section 2 we state the full Birch and Swinnerton-Dyer conjecture over an imaginary quadratic field K, and state a generalized Gross-Zagier formula for elliptic curves of any rank. In Section 3, we introduce the Kolyvagin points P_λ on E over ring class fields of K, and deduce some key properities of these points. We state Kolvagin's conjectures from [Kol91a] along with some of their consequences in Section 4. We also state a conjecture that refines Kolyvagin's conjectures and also refines a conjecture of Gross-Zagier. In Section 5 we use reductions of Kolyvagin points to define, for every prime p that is inert in K, a finite index subgroup W_p of $E(K)$. Section 6 lays some general foundations for our later determination of the structure of W_p by studying the image of a fixed $Q \in E(K)$ in $E(\mathbb{F}_{p^2})/(p+1)$. Section 7 presents a conditional proof that (up to primes not in the set $B(E)$) maximal index subgroups exist and that they satisfy our generalized Gross-Zagier formula. Finally, in Section 8 we numerically investigate the existence of Gross-Zagier subgroups of $E(K)$, and give evidence for a higher-rank generalization of a conjecture of Gross-Zagier.

Acknowledgement: We thank R. Bradshaw, K. Buzzard, C. Citro, J. Coates, C. Cornut, M. Flach, R. Greenberg, B. Gross, D. Jetchev, K. Lauter, B. Mazur, R. Miller, and Tonghai Yang for helpful conversations. We thank Amod Agashe and Andrei Jorza for carefully reading a draft of the paper and providing many helpful comments, and we thank the anonymous referee for much helpful feedback.

> "It is always good to try to prove true theorems."
>
> – Bryan Birch

1.1 Notation and Conventions

Let A be an abelian group. Let $A_{\rm tor}$ be the subgroup of elements of A of finite order and let $A_{/\,{\rm tor}} = A/A_{\rm tor}$ denote the quotient of A by its torsion subgroup. Let $A[n]$ be the subgroup of elements of A of order n, and for any prime ℓ, let $A(\ell)$ be the subgroup of elements of ℓ-power order. For $z \in A$, let $e = \mathrm{ord}_\ell(z)$ be the largest integer e such that $z = \ell^e y$ for some $y \in A$, or $\mathrm{ord}_\ell(z) = \infty$ if the set of e is unbounded. If $a_1, \ldots, a_n$ are elements of an additive or multiplicative group A, we let $\langle a_1, \ldots, a_n\rangle$ denote the subgroup of A generated by the a_i.

Throughout this paper, E denotes an elliptic curve defined over $\mathbb{Q}$ of conductor N, and K is a quadratic imaginary field with $D = \mathrm{disc}(K)$ coprime to N that satisfies the *Heegner hypothesis*—each prime dividing N splits in K. We fix an ideal $\mathcal{N}$ in $\mathcal{O}_K$ such that $\mathcal{O}_K/\mathcal{N}$ is cyclic of order N. Let H be the Hilbert class field of K, let $\pi : X_0(N) \to E$ be a fixed choice of modular parametrization (see Section 3 below), and let

$$y_K = \mathrm{Tr}_{H/K}(\pi((\mathbb{C}/\mathcal{O}_K, \mathcal{N}^{-1}/\mathcal{O}_K))) \in E(K) \tag{1}$$

be the Heegner point associated to K.

Let c denote the Manin constant of E (see Section 2), and c_q the Tamagawa numbers of E at primes $q \mid N$. Unless otherwise stated, everywhere in this paper p denotes a prime that is inert in K.

2 Gross-Zagier Subgroups

In this section, we fix our notation and conventions, and define the Manin constant. Then we recall the statement of the full Birch and Swinnerton-Dyer conjecture over an imaginary quadratic field K. We give a new definition of *Gross-Zagier subgroups* of $E(K)$ and prove that they all satisfy a Gross-Zagier style formula. When $r_{\rm an} = 1$, we prove that $\mathbb{Z}y_K$ is the unique Gross-Zagier subgroup, up to torsion.

Let E be an elliptic curve over $\mathbb{Q}$ and let K be a quadratic imaginary field that satisfies the Heegner hypothesis – so K has discriminant $D < -4$, each prime dividing the conductor N of E splits in K, and $\gcd(D, N) = 1$. Let $\mathcal{O}_K$ be the ring of integers of K. Let E^D denote the quadratic twist of E by D. Throughout this paper, except briefly in Section 8, we *always* assume that

$$r_{\rm an}(E/\mathbb{Q}) > r_{\rm an}(E^D/\mathbb{Q}) \leq 1 \tag{2}$$

Recall that under the Heegner hypothesis the sign of the functional equation of

$$L(E/K, s) = L(E/\mathbb{Q}, s) \cdot L(E^D/\mathbb{Q}, s)$$

is -1, so the sign in the functional equations for $L(E/\mathbb{Q},s)$ and $L(E^D/\mathbb{Q},s)$ are different, hence

$$\operatorname{ord}_{s=1} L(E/\mathbb{Q},s) \not\equiv \operatorname{ord}_{s=1} L(E^D/\mathbb{Q},s) \pmod 2.$$

Proposition 2.1. *Suppose E is an elliptic curve with $r_{\mathrm{an}}(E/\mathbb{Q}) > 0$. Then there exist infinitely many D satisfying the Heegner hypothesis with*

$$r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1.$$

Proof. The main theorem of [BFH90] implies the existence of infinitely many D with $r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$. Since $r_{\mathrm{an}}(E/\mathbb{Q}) > 0$ and $r_{\mathrm{an}}(E/\mathbb{Q}) \not\equiv r_{\mathrm{an}}(E^D/\mathbb{Q}) \pmod 2$, the inequality $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q})$ also holds.

Let $\omega = 2\pi i c f(z) dz$ be the pullback of a minimal invariant differential on E, where $f(z) \in S_2(\Gamma_0(N))$ is a cuspidal newform, and c is the Manin constant of E (see [ARS06]). For each prime $q \mid N$, let c_q be the Tamagawa number of E at q. Set $r = r_{\mathrm{an}}(E/K) = \operatorname{ord}_{s=1} L(E/K,s)$, which is defined since every elliptic curve over $\mathbb{Q}$ is modular. Let $\|\omega\|^2 = \int_{E(\mathbb{C})} \omega \wedge \overline{i\omega} = 2 \cdot \operatorname{Vol}(\mathbb{C}/\Lambda)$. The Shafarevich-Tate group of E over a number field M is

$$Ш(E/M) = \ker\left(\mathrm{H}^1(M,E) \to \bigoplus_v H^1(M_v,E)\right).$$

The following is a formulation of the Birch and Swinnerton-Dyer conjecture [GZ86, pg. 311] over K.

Conjecture 2.2 (Birch and Swinnerton-Dyer)**.** *The Mordell-Weil group $E(K)$ has rank $r = \operatorname{ord}_{s=1} L(E/K,s)$, the Shafarevich-Tate group $Ш(E/K)$ is finite, and*

$$\frac{L^{(r)}(E/K,1)}{r!} = \frac{\#Ш(E/K) \cdot \|\omega\|^2 \cdot \operatorname{Reg}(E/K) \cdot \left(\prod_{q|N} c_q\right)^2}{\#E(K)_{\mathrm{tor}}^2 \cdot \sqrt{|D|}}. \tag{3}$$

Let $Ш_{\mathrm{an}}$ be the order of $Ш(E/K)$ that is predicted by Conjecture 2.2. The existence of the Cassels-Tate pairing implies that if $Ш(E/K)$ is finite, then $\#Ш(E/K)$ is a perfect square, so Conjecture 2.2 implies that $\sqrt{Ш_{\mathrm{an}}}$ is an integer. Recall from Section 1.1 that $A_{/\mathrm{tor}} = A/A_{\mathrm{tor}}$.

Definition 2.3 (Gross-Zagier subgroup)**.** *A* Gross-Zagier subgroup *$W \subset E(K)$ is a torsion-free subgroup such that $E^D(\mathbb{Q}) \subset W + E(K)_{\mathrm{tor}}$, the quotient $E(K)_{/\mathrm{tor}}/W$ is cyclic, and*

$$[E(K) : W] = c \cdot \prod c_q \cdot \sqrt{Ш_{\mathrm{an}}}. \tag{4}$$

For any set S of primes, we say that a subgroup $W \subset E(K)$ is a *Gross-Zagier subgroup up to primes not in S* if W has no p-torsion for $p \notin S$ and all the conditions of Definition 2.3 holds up to primes not in S.

We will numerically investigate the existence of Gross-Zagier subgroups in Section 8, assuming that Conjecture 2.2 is true. Even the existence of Gross-Zagier subgroups of every $E(K)$ is far from clear, since if they exist, then $\#E(K)_{\mathrm{tor}}$ divides $c \prod c_q \cdot \sqrt{Ш_{\mathrm{an}}}$. In fact, we will give an example of an $E(K)$ that does not have any Gross-Zagier subgroups (this example does not satisfy (2)).

In the following proposition we do not assume the Conjecture 2.2. Thus $Ш_{\mathrm{an}}$ *a priori* could just be some meaningless transcendental number. Also, for any subgroup $H \subset E(K)$, we write $\operatorname{Reg}(H)$ for the absolute value of the determinant of the height pairing matrix on any basis for H modulo torsion.

Proposition 2.4. *If W is a Gross-Zagier subgroup, then W satisfies the* generalized Gross-Zagier formula*:*

$$\frac{L^{(r)}(E/K,1)}{r!} = \frac{\|\omega\|^2}{c^2 \cdot \sqrt{|D|}} \cdot \operatorname{Reg}(W). \tag{5}$$

More generally, a torsion-free subgroup $W \subset E(K)$ satisfies the generalized Gross-Zagier formula if and only if it has index $c \cdot \prod c_q \cdot \sqrt{Ш_{\mathrm{an}}}$ in $E(K)$.

Proof. The BSD formula (3) with $\#Ш(E/K)$ replaced by $Ш_{\mathrm{an}}$ implies that (5) holds if and only if

$$\frac{\|\omega\|^2}{c^2 \cdot \sqrt{|D|}} \cdot \operatorname{Reg}(W) = \frac{Ш_{\mathrm{an}} \cdot \|\omega\|^2 \cdot \operatorname{Reg}(E/K) \cdot \left(\prod_{p|N} c_q\right)^2}{\#E(K)_{\mathrm{tor}}^2 \cdot \sqrt{|D|}}. \tag{6}$$

Our hypotheses that $[E(K) : W]$ is finite and that W is torsion free imply that

$$[E(K) : W]^2 = \frac{\operatorname{Reg}(W) \cdot \#E(K)_{\mathrm{tor}}^2}{\operatorname{Reg}(E/K)} \tag{7}$$

Manipulate (6) by cancelling everything in common on both sides and putting the regulators and torsion on the left, and everything else on the right. The substitution (7) then shows that $[E(K) : W]^2 = c^2 \cdot \left(\prod_{q|N} c_q\right)^2 \cdot Ш_{\mathrm{an}}$ if and only if (5) holds. Taking square roots proves the proposition. □

Corollary 2.5. *Let $y_K \in E(K)$ be the Heegner point after fixing a choice of ideal $\mathcal{N}$ as in Equation* (1)*, and assume that E has analytic rank* 1*. Then the Gross-Zagier subgroups of $E(K)$ are the cyclic groups $\langle y_K + P\rangle$, for all $P \in E(K)_{\mathrm{tor}}$.*

Proof. By [Kol88], $E(K)$ is of rank 1, and by Proposition 2.4 the Gross-Zagier formula [GZ86, Thm. 2.1, pg. 311] implies that $[E(K) : \langle y_K\rangle] = c \prod c_q \sqrt{\text{Ш}_{\mathrm{an}}}$ (see also, [GZ86, Conj. 2.2, pg. 311]). Since $E(K)_{/\,\mathrm{tor}}$ is free of rank 1 and $\langle y_K\rangle$ is torsion free, $E(K)_{/\,\mathrm{tor}}/\langle y_K\rangle$ is cyclic, so $\langle y_K\rangle$ is a Gross-Zagier subgroup. The same argument proves this with y_K replaced by $y_K + P$ for any $P \in E(K)_{\mathrm{tor}}$, since y_K and $y_K + P$ have the same height. If W is any Gross-Zagier subgroup, then since $E(K)$ has rank one we must have $W \equiv \langle y_K\rangle \pmod{E(K)_{\mathrm{tor}}}$, so $W = \langle y_K + P\rangle$ for some $P \in E(K)_{\mathrm{tor}}$.

3 Heegner and Kolyvagin Points

In this section, we define certain subsets $\Lambda^k_{\ell^n} \subset \mathbb{Z}$ of positive square-free integers. For each integer $\lambda \in \Lambda^k_{\ell^n}$, we consider the corresponding ring class field K_λ, and we define elements $I_\lambda, J_\lambda \in \mathbb{Z}[\mathrm{Gal}(K_\lambda/K)]$. We then apply these group ring elements to the Heegner points $y_\lambda \in E(K_\lambda)$ to obtain the Kolyvagin points $P_\lambda \in E(K_\lambda)$. Finally, we prove the $\mathrm{Gal}(K_\lambda/K)$-equivariance of the equivalence class $P_\lambda + \ell^n E(K)$ in $E(K)/\ell^n E(K)$.

For any integer m, let $a_m = a_m(E)$ be the mth coefficient of the L-series $\sum a_m/m^s$ attached to E. Let ℓ be any prime and n any positive integer. For any nonnegative integer k, let $\Lambda^k_{\ell^n}$ be the set of squarefree positive integers $\lambda = p_1 \dots p_k$ coprime to N, where each p_i is inert in K and

$$a_{p_i} \equiv p_i + 1 \equiv 0 \pmod{\ell^n}.$$

When $k = 0$, we set $\Lambda^0_{\ell^n} = \{1\}$. The Chebotarev density theorem implies that $\Lambda^k_{\ell^n}$ is infinite for any $k \geq 1$.

Recall from Section 1.1 that we fixed an ideal $\mathcal{N}$ in $\mathcal{O}_K$ such that $\mathcal{O}_K/\mathcal{N}$ is cyclic of order N, and let $\mathcal{O}_\lambda = \mathbb{Z} + \lambda\mathcal{O}_K$ be the order in $\mathcal{O}_K$ of conductor λ. Let $X_0(N)$ be the compact modular curve defined over $\mathbb{Q}$ that classifies isomorphism classes of elliptic curves equipped with a cyclic subgroup of order N. Fix a choice of minimal modular parametrization $\pi : X_0(N) \to E$, which exists by the modularity theorem [BCDT01, Wil95]. For each $\lambda \in \Lambda^k_{\ell^n}$, the *Heegner point*

$$x_\lambda = [(\mathbb{C}/\mathcal{O}_\lambda, (\mathcal{N} \cap \mathcal{O}_\lambda)^{-1}/\mathcal{O}_\lambda)] \in X_0(N)(K_\lambda)$$

is defined over the ring class field K_λ of K of conductor λ.

Definition 3.1 (Heegner point)**.** *The Heegner point y_λ associated to $\lambda \in \Lambda^k_{\ell^n}$ is*

$$y_\lambda = \pi(x_\lambda) \in E(K_\lambda).$$

We emphasize that that y_λ depends on the choice of modular parametrization π_E and the ideal $\mathcal{N}$ in $\mathcal{O}_K$ with $\mathcal{O}_K/\mathcal{N} = \mathbb{Z}/N\mathbb{Z}$. However, once we fix that data, the Heegner points for all λ are defined.

For $\lambda \in \Lambda^k_{\ell^n}$, let $G_\lambda = \mathrm{Gal}(K_\lambda/K_1)$ and note that we have a canonical isomorphism

$$G_\lambda \cong \prod_{p\mid\lambda} G_p,$$

where the group $G_p = \mathrm{Gal}(K_p/K_1) = \langle t_p\rangle$ is cyclic of order $p+1$, with some (non-canonical) choice t_p of generator. Let

$$I_p = \sum_{i=1}^{p} i t_p^i \in \mathbb{Z}[G_p] \quad \text{and} \quad I_\lambda = \prod_{p\mid\lambda} I_p \in \mathbb{Z}[G_\lambda].$$

Let R be a set of representatives in $\mathrm{Gal}(K_\lambda/K)$ for the quotient group $\mathrm{Gal}(K_\lambda/K)/\mathrm{Gal}(K_\lambda/K_1) \cong \mathrm{Gal}(K_1/K)$, and let

$$J_\lambda = \sum_{g\in R} g \in \mathbb{Z}[G_\lambda].$$

Definition 3.2 (Kolyvagin Point)**.** *The Kolyvagin point P_λ associated to $\lambda \in \Lambda^k_{\ell_n}$ is*

$$P_\lambda = J_\lambda I_\lambda y_\lambda \in E(K_\lambda).$$

Note that $P_1 = y_K \in E(K)$.

Let $R = \mathrm{End}(E/\mathbb{C})$ and let $B(E)$ be the set of *odd* primes ℓ that do not divide $\mathrm{disc}(R)$ and such that the ℓ-adic representation $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_R(\mathrm{Tate}_\ell(E))$ is surjective. By a theorem of Serre [Ser72], the set $B(E)$ contains all but finitely many primes (see [GJP+09] for algorithms to bound $B(E)$). Let T_p be the pth Hecke operator on the Jacobian $J_0(N)$ of $X_0(N)$, and for each prime $p \mid \lambda$, let Tr_p be the trace $J_0(N)(K_\lambda) \to J_0(N)(K_{\lambda/p})$.

Proposition 3.3. *The points y_λ form an Euler system, in the sense that if $\lambda = p\lambda'$ for a prime p and $\lambda \in \Lambda_\ell$, then $y_\lambda = \mathrm{Frob}_\wp(y_{\lambda'}) \pmod{\wp}$ for all primes $\wp$ of K_λ over p, and $\mathrm{Tr}_p(x_\lambda) = T_p(x_{\lambda'})$ in $J_0(N)$.*

Proof. See [Gro91, Prop. 3.7].

Proposition 3.4. *We have*

$$[I_\lambda y_\lambda] \in (E(K_\lambda)/\ell^n E(K_\lambda))^{G_\lambda} \qquad \textit{and} \qquad [P_\lambda] \in (E(K_\lambda)/\ell^n E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)}$$

Proof. Though standard (see, e.g., [Gro91, Prop. 3.6]) this proposition plays a key role in Section 5, so we give a proof here for the convenience of the reader. The first statement implies the second, since $[P_\lambda]$ is the $\mathrm{Gal}(K_1/K)$ trace of $[I_\lambda y_\lambda]$. It remains to prove the first inclusion. For this, it suffices to show that $[I_\lambda y_\lambda]$ is fixed by t_p for all primes $p \mid \lambda$, as these elements generate G_λ. We will prove this by showing that $(t_p - 1)I_\lambda y_\lambda$ lies in $\ell^n E(K_\lambda)$.

Write $\lambda = p\lambda'$. We have

$$(t_p - 1)I_p = (t_p - 1) \cdot \left(\sum_{i=1}^{p} i t_p^i\right) = p + 1 - \mathrm{Tr}_p, \tag{8}$$

where as above $\mathrm{Tr}_p = \mathrm{Tr}_{K_\lambda/K_{\lambda'}}$. Note that this is the only place in the proof where we use the explicit definition of I_p as $\sum_{i=1}^{p} i t_p^i$, and in fact we could instead replace I_p by any element I of $\mathbb{Z}[G_\lambda]$ such that

$$(t_p - 1)I = p + 1 - \mathrm{Tr}_p,$$

but doing so does not seem to lead to anything interesting. Note that the Euler system relation (see Proposition 3.3) and our hypothesis that $a_p \equiv 0 \pmod{\ell^n}$ together imply that

$$\mathrm{Tr}_p I_{\lambda'} y_\lambda = I_{\lambda'} \mathrm{Tr}_p y_\lambda = I_{\lambda'} a_p y_{\lambda'} \in \ell^n E(K_\lambda).$$

We have

$$(t_p - 1)I_\lambda = (t_p - 1)I_p I_{\lambda'} = (p + 1 - \mathrm{Tr}_p)I_{\lambda'}$$

in $\mathbb{Z}[G_\lambda]$, so since $p + 1 \equiv 0 \pmod{\ell^n}$

$$(t_p - 1)I_\lambda y_\lambda = (p + 1)I_{\lambda'} y_\lambda - \mathrm{Tr}_p I_{\lambda'} y_\lambda \in \ell^n E(K_\lambda).$$

□

4 Kolyvagin's Conjectures and their Consequences

For any prime ℓ and positive integer n, let

$$\Lambda_{\ell^n} = \bigcup_{\text{all } k \geq 0} \Lambda_{\ell^n}^k$$

be the set of square-free positive integers λ such that $\ell^n \mid \gcd(a_p, p+1)$ for each $p \mid \lambda$. In this section, we define maps $n, m : \Lambda_\ell \to \mathbb{Z} \cup \{\infty\}$ that measure ℓ-divisibility properties of λ and P_λ for all $\lambda \in \Lambda_\ell$. We state Kolyvagin's "Conjecture A" that there exists λ with $m(\lambda) \neq \infty$, then state Kolyvagin's structure theorem, which describes the structure of $\mathrm{Sel}^{(\ell^b)}(E/K)$, for b sufficiently large, in terms of the maps n and m. Finally, we state Kolyvagin's stronger "Conjecture D", which basically asserts that if f is the smallest nonnegative integer such that $m(\lambda) \neq \infty$ for some $\lambda \in \Lambda_\ell^f$, then for sufficiently large k the cohomology classes τ_{λ,ℓ^n} with $\lambda \in \Lambda_{\ell^n+k}^f$ generate a subgroup of $\mathrm{Sel}^{(\ell^n)}(E/K)$ that equals the image of a subgroup V of $E(K)$. To motivate Conjecture 4.9, we prove that it implies that $\mathrm{rank}(E(\mathbb{Q})) = f + 1$ and $Ш(E/K)(\ell)$ is finite for each $\ell \in B(E)$ and determine the structure of V (see Proposition 4.11).

Recall that we defined ord_ℓ in Section 1.1. Define two set-theoretic maps

$$n, m : \Lambda_\ell \to \mathbb{Z} \cup \{\infty\}$$

by

$$n(\lambda) = \max\{e : \lambda \in \Lambda_{\ell^e}\} \qquad \text{and} \qquad m(\lambda) = \mathrm{ord}_\ell([P_\lambda]),$$

where $[P_\lambda]$ denotes the equivalence class of P_λ in $E(K_\lambda)/\ell^{n(\lambda)}E(K_\lambda)$. For each integer $k \geq 0$, let

$$m_{\ell,k} = \min(m(\Lambda_\ell^k)) \qquad \text{and} \qquad m_\ell = \min(m(\Lambda_\ell)) = \min(\{m_{\ell,k} : k \geq 0\}).$$

Also, let

$$f_\ell = \min\{k : m_{\ell,k} < \infty\} \leq \infty, \tag{9}$$

where we let $f_\ell = \infty$ if $m_\ell = \infty$.

Kolyvagin proves [Kol91b, Thm. C] that $m_{\ell,0} \geq m_{\ell,1} \geq m_{\ell,2} \geq \dots$.

Conjecture 4.1 (Kolyvagin's Conjecture A_ℓ)**.** *$m_\ell < \infty$. Equivalently, there exists $\lambda \in \Lambda_\ell$ such that $[P_\lambda] \neq 0$.*

See [JLS08] for the first computational evidence for Conjecture 4.1. For example, for a specific rank 2 elliptic curve, that paper shows that $m_3 = m_{3,1} = 0$ and $f_3 = 1$, assuming that the numerical computation of a certain Heegner point y_λ was done to sufficient precision. (If the computation were not done to sufficient precision it is highly likely that we would haved detected this.)

Conjecture 4.1 is quite powerful, as the following theorem shows. For an abelian group A of odd order with an action of complex conjugation, let A^+ denote the $+1$ eigenspace for conjugation and A^- the minus eigenspace, so $A = A^+ \oplus A^-$. As always, we continue to assume our minimality hypothesis that

$$r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1.$$

Theorem 4.2 (Kolyvagin)**.** *Let $\ell \in B(E)$, suppose Conjecture 4.1 is true for ℓ, and let $f = f_\ell$. For every k, let $b_k = \ell^{m_{\ell,k} - m_{\ell,k+1}}$. Then for every $n \geq m_{\ell,f}$, we have*

$$\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q}) = \mathrm{Sel}^{(\ell^n)}(E/K)^+ \approx (\mathbb{Z}/\ell^n\mathbb{Z})^{f+1} \oplus (\mathbb{Z}/b_{f+1}\mathbb{Z})^2 \oplus (\mathbb{Z}/b_{f+3}\mathbb{Z})^2 \oplus \cdots$$

and

$$\mathrm{Sel}^{(\ell^n)}(E^D/\mathbb{Q}) = \mathrm{Sel}^{(\ell^n)}(E/K)^- \approx (\mathbb{Z}/\ell^n\mathbb{Z})^{h} \oplus (\mathbb{Z}/b_{f}\mathbb{Z})^2 \oplus (\mathbb{Z}/b_{f+2}\mathbb{Z})^2 \oplus \cdots$$

where $h = \mathrm{rank}(E^D(\mathbb{Q})) \leq 1$.

Proof. The leftmost equality in the above two equations is true because ℓ is odd, and Theorem 1 of [Kol91a] implies both of the rightmost equalities, but possibly with $\mathrm{Sel}^{(\ell^n)}(E/K)^+$ and $\mathrm{Sel}^{(\ell^n)}(E/K)^-$ swapped and a different value for h. Theorem 1 of [Kol91a] is proved by inductively constructing cohomology classes with good properties with respect to certain localization homomorphisms. To finish the proof, we establish that these two Selmer groups are not swapped and that $h = \mathrm{rank}(E^D(\mathbb{Q}))$.

First note that by [Kol88, BFH90], our hypothesis that $r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$ implies that $r_{\mathrm{an}}(E^D/\mathbb{Q}) = \mathrm{rank}(E^D(\mathbb{Q}))$ and $Ш(E^D/\mathbb{Q})$ is finite.

If $f = 0$, then the Heegner point y_K has infinite order, so by [GZ86] we have $r_{\mathrm{an}}(E/K) = 1$ and by [Kol88], $E(K)$ has rank 1 and $Ш(E/K)$ is finite. By our minimality hypothesis, we have $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q})$, so $r_{\mathrm{an}}(E/\mathbb{Q}) = \mathrm{rank}(E(\mathbb{Q})) = 1$ and $r_{\mathrm{an}}(E^D/\mathbb{Q}) = \mathrm{rank}(E^D(\mathbb{Q})) = 0$. Thus the two displayed Selmer groups $\mathrm{Sel}^{(\ell^n)}(E/K)^\pm$ are in the claimed order. Moreover, $h = 0 = \mathrm{rank}(E^D(\mathbb{Q}))$.

Next assume $f > 0$. Then one of the two Selmer groups contained $(\mathbb{Z}/\ell^n\mathbb{Z})^{f+1}$ for arbitrarily large n. Since we know that $Ш(E^D/\mathbb{Q})$ is finite and $\mathrm{rank}(E^D(\mathbb{Q})) \leq 1$ but $f + 1 \geq 2$, the Selmer group that contains $(\mathbb{Z}/\ell^n\mathbb{Z})^{f+1}$ must be $\mathrm{Sel}^{(\ell^n)}(E/K)^+$. Thus again we see that the two displayed Selmer groups are in the claimed order. Also, again $h = \mathrm{rank}(E^D(\mathbb{Q}))$ follows.

□

Remark 4.3. *Suppose the hypotheses of Theorem 4.2 are satisfied. Then comparing the conclusion about the choice of signs in Theorem 4.2 with the statement of Theorem 1 in [Kol91a] shows that $f + 1 \equiv r_{\mathrm{an}}(E/\mathbb{Q}) \pmod 2$, which implies the parity conjecture for the Selmer group of E at ℓ.*

Proposition 4.4. *Let $\ell \in B(E)$. Then $f_\ell = \mathrm{rank}(E(\mathbb{Q})) - 1$ if and only if $Ш(E/\mathbb{Q})(\ell)$ is finite and Conjecture 4.1 holds for ℓ.*

Proof. First suppose $f_\ell = \mathrm{rank}(E(\mathbb{Q})) - 1$. Then $f_\ell \neq \infty$, so Conjecture 4.1 holds. To prove that $\text{Ш}(E/\mathbb{Q})(\ell)$ is finite, use Theorem 4.2 and that by our rank hypothesis the image of $E(\mathbb{Q})$ in $\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q})$ is $(\mathbb{Z}/\ell^n\mathbb{Z})^{f+1}$. Thus $\text{Ш}(E/\mathbb{Q})[\ell^n]$ is a quotient of the ℓ^n-torsion subgroup of the finite group $(\mathbb{Z}/b_{f+1}\mathbb{Z})^2 \oplus (\mathbb{Z}/b_{f+3}\mathbb{Z})^2 \oplus \cdots$, so $\text{Ш}(E/\mathbb{Q})(\ell)$ is finite.

Conversely, suppose the ℓ-primary group $\text{Ш}(E/\mathbb{Q})(\ell)$ is finite and that Conjecture 4.1 holds. Let b be a positive integer such that $\ell^b\text{Ш}(E/\mathbb{Q})(\ell) = 0$. Then the map $\mathrm{Sel}^{(\ell^b)}(E/\mathbb{Q}) \to \text{Ш}(E/\mathbb{Q})(\ell)$ is surjective, and for every integer $n \geq b$, the map $\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q})[\ell^b] \to \text{Ш}(E/\mathbb{Q})(\ell)$ is also surjective, since $\mathrm{Sel}^{(\ell^b)}(E/\mathbb{Q}) \to \mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q})[\ell^b]$. The image of $\ell^b\,\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q})$ in $\text{Ш}(E/\mathbb{Q})(\ell)$ is trivial. Since $\ell \in B(E)$ we have $E(\mathbb{Q})_{\mathrm{tor}}[\ell] = 0$, so exactness of the sequence

$$0 \to E(\mathbb{Q})/\ell^n E(\mathbb{Q}) \to \mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q}) \to \text{Ш}(E/\mathbb{Q})(\ell) \to 0$$

implies that $\ell^b\,\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q}) \approx \ell^b(\mathbb{Z}/\ell^n\mathbb{Z})^r$, where $r = \mathrm{rank}(E(\mathbb{Q}))$. On the other hand, if we also choose $\ell^b \geq b_{f+1}$, then Theorem 4.2 implies that $\ell^b\,\mathrm{Sel}^{(\ell^n)}(E/\mathbb{Q}) \approx \ell^b(\mathbb{Z}/\ell^n\mathbb{Z})^{f+1}$. We conclude that $r = f+1$.

□

Kolyvagin's other conjectures involve $\mathrm{H}^1(K, E[\ell^\infty]) = \varinjlim_m \mathrm{H}^1(K, E[\ell^m])$.

Lemma 4.5. *Suppose $E(K)[\ell] = 0$. Then for every $m \geq 1$, the natural map $\mathrm{H}^1(K, E[\ell^m]) \to \mathrm{H}^1(K, E[\ell^\infty])$ is injective.*

Proof. This lemma is of course very well known, but we give a proof for completeness. It suffices to show that for any pair a, b of nonnegative integers that the map

$$\mathrm{H}^1(K, E[\ell^a]) \to \mathrm{H}^1(K, E[\ell^{a+b}]) \tag{10}$$

is injective. Taking Galois cohomology of $0 \to E[\ell^a] \to E[\ell^{a+b}] \to E[\ell^{a+b}]/E[\ell^a] \to 0$ we see that $\mathrm{H}^0(K, E[\ell^{a+b}]/E[\ell^a])$ surjects onto the kernel of (10). We have an exact sequence of Galois modules

$$0 \to E[\ell^a] \to E[\ell^{a+b}] \xrightarrow{\ell^a} E[\ell^b] \to 0,$$

so $\mathrm{H}^0(K, E[\ell^{a+b}]/E[\ell^a]) \cong \mathrm{H}^0(K, E[\ell^b]) = E(K)[\ell^b] = 0$, since $E(K)[\ell] = 0$.

□

We now define Galois cohomology classes associated to the Kolyvagin points P_λ. For $\lambda \in \Lambda_{\ell^n}$ with $\ell \in B(E)$, let $\tau_{\lambda,\ell^n} \in \mathrm{H}^1(K, E[\ell^n])$ be the image of P_λ under the map

$$(E(K_\lambda)/\ell^n E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)} \hookrightarrow \mathrm{H}^1(K_\lambda, E[\ell^n])^{\mathrm{Gal}(K_\lambda/K)} \cong \mathrm{H}^1(K, E[\ell^n]),$$

where the last map is an isomorphism because $\ell \in B(E)$ (see, e.g., [Gro91, §4]). Kolyvagin also remarks that one can define Galois cohomology classes τ_{λ,ℓ^n} for $\ell \notin B(E)$ and all $\lambda \in \Lambda_{\ell^{k_0+n}}$, where k_0 is the smallest nonnegative even integer such that $\ell^{k_0/2}E(\mathbf{K})(\ell) = 0$ and $\mathbf{K}$ is the compositum of all K_λ for $\lambda \in \Lambda$. Of course, for all $\ell \in B(E)$ we have $k_0 = 0$.

Let τ'_{λ,ℓ^n} be the image in $H^1(K, E[\ell^\infty])$ of τ_{λ,ℓ^n} (note that for the moment we are not assuming that $\ell \in B(E)$, so the natural map $\mathrm{H}^1(K, E[\ell^m]) \to \mathrm{H}^1(K, E[\ell^\infty])$ need not be injective). For any integers $a \geq 0$, $k \geq k_0$ and $n \geq 1$, let

$$V^a_{k,\ell^n} = \langle \tau'_{\lambda,\ell^n} : \lambda \in \Lambda^a_{\ell^{n+k}} \rangle \subset H^1(K, E[\ell^\infty])$$

Since $\Lambda^a_{\ell^{n+k+1}} \subset \Lambda^a_{\ell^{n+k}}$, we have

$$V^a_{0,\ell^n} \supset V^a_{1,\ell^n} \supset V^a_{2,\ell^n} \supset \cdots .$$

We have $\ell\tau_{\lambda,\ell^{n+1}} = \tau_{\lambda,\ell^n}$, because the following diagram commutes, with $G = \mathrm{Gal}(K_\lambda/K)$:

$$\begin{tikzcd}
(E(K_\lambda)/\ell^{k+n+1}E(K_\lambda))^G \arrow[r, hook] & \mathrm{H}^1(K_\lambda, E[\ell^{k+n+1}])^G \\
(E(K_\lambda)/\ell^{k+n}E(K_\lambda))^G \arrow[r, hook] \arrow[u, "{[\ell]}"] & \mathrm{H}^1(K_\lambda, E[\ell^{k+n}])^G \arrow[u]
\end{tikzcd}$$

Thus $\ell V^a_{k,\ell^{n+1}} \subset V^a_{k,\ell^n}$.

We say $\{\tau_{\lambda,\ell^n}\}$ is a *strong nonzero system* if there exists $a \geq 0$ such that for all $k \geq k_0$ there exists n such that $V^a_{k,\ell^n} \neq 0$. In other words, if one continues the grid of subgroups of $\mathrm{H}^1(K, E[\ell^\infty])$ below infinitely far to the right and up in the obvious way, then it is *not* the case that sufficiently far to the right every single group is 0.

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
V^a_{0,\ell^3} & V^a_{1,\ell^3} \arrow[l] & V^a_{2,\ell^3} \arrow[l] & \cdots \arrow[l] \\
V^a_{0,\ell^2} & V^a_{1,\ell^2} \arrow[l] & V^a_{2,\ell^2} \arrow[l] & \cdots \arrow[l] \\
V^a_{0,\ell} & V^a_{1,\ell} \arrow[l] & V^a_{2,\ell} \arrow[l] & \cdots \arrow[l]
\end{tikzcd}$$

Conjecture 4.6 (Kolyvagin's Conjecture B_ℓ)**.** *$\{\tau_{\lambda,\ell^n}\}$ is a strong nonzero system.*

Remark 4.7. *Kolyvagin remarks [Kol91a, pg. 258] that if $\ell \in B(E)$, then $\{\tau_{\lambda,\ell^n}\}$ is a strong nonzero system if and only if there exists n such that $V^a_{0,\ell^n} \neq 0$. By Lemma 4.5, this is the case if and only if some τ is nonzero. So for $\ell \in B(E)$, Conjectures 4.1 is true if and only if Conjecture 4.6 is true.*

The following conjecture is motivated by Theorem 4.2 and the conjecture that $Ш(E/K)$ is finite.

Conjecture 4.8 (Kolyvagin's Conjecture C)**.** *The set of primes ℓ such that $m_\ell \neq 0$ is finite.*

Let $r_{\mathrm{an}} = \mathrm{ord}_{s=1} L(E,s)$, and let $\varepsilon = (-1)^{r_{\mathrm{an}}-1}$. For any module A with an action of complex conjugation σ, and $\nu \in \{0,1\}$, let $A^\nu = (1-(-1)^\nu \varepsilon\sigma)A$.

Conjecture 4.9 (Kolyvagin's Conjecture D_ℓ)**.** *There exists $\nu \in \{0,1\}$ and a subgroup $V \subset (E(K)/E(K)_{\mathrm{tor}})^\nu$ such that $1 \leq \mathrm{rank}(V) \equiv \nu \pmod 2$ and for all $n \geq 1$ and all sufficiently large k, one has*

$$V^a_{k,\ell^n} \equiv V \pmod{\ell^n(E(K)_{/\,\mathrm{tor}})},$$

where $a = \mathrm{rank}(V) - 1$.

The following conjecture is the natural generalization to higher rank of the hypothesis when $r_{\mathrm{an}}(E/\mathbb{Q}) = 1$ that the Hegner point y_K has infinite order.

Conjecture 4.10 (Kolyvagin's Conjecture D)**.** *There exists a single subgroup V of $E(K)$ such that Conjecture 4.9 holds simultaneously for all ℓ with that V.*

Conjecture 4.9 has numerous consequences. Much of the following proposition is implicitly stated without any proofs in [Kol91a, pg. 258–259], so we give complete proofs below.

Proposition 4.11. *Assume our running minimality hypothesis that $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$. Suppose Conjecture 4.9 is true for $\ell \in B(E)$ and let $f = f_\ell$. Then*

1. *$(E(K)/E(K)_{\mathrm{tor}})^\nu = (E(K)/E(K)_{\mathrm{tor}})^+$,*
2. *$a = f$,*
3. *$\mathrm{rank}(E(\mathbb{Q})) = f + 1$,*
4. *$Ш(E/K)(\ell)$ is finite,*
5. *$r_{\mathrm{an}}(E/\mathbb{Q}) \equiv \mathrm{rank}(E(\mathbb{Q})) \pmod 2$, and*
6. *$V \otimes \mathbb{Z}_\ell = \ell^{m_f} E(\mathbb{Q}) \otimes \mathbb{Z}_\ell$.*

Proof. By Conjecture 4.9, there exists $\nu \in \{0,1\}$ and a subgroup $V \subset (E(K)/E(K)_{\mathrm{tor}})^\nu$ such that $1 \leq \mathrm{rank}(V) \equiv \nu \pmod 2$ and for all $n > 0$ and all sufficiently large k we have

$$V^a_{k,\ell^n} \equiv V \pmod{\ell^n E(K)_{\mathrm{tor}}},$$

where $a = \mathrm{rank}(V) - 1$.

If $\mathrm{rank}(V) = 1$, then $a = 0$, so $V^0_{k,\ell^n} \neq 0$ for some k, so since f is the smallest integer such that $V^f_{k,\ell^n} \neq 0$, this implies that $f = 0$ giving Part 2; thus the Heegner point y_K has infinite order and $r_{\mathrm{an}}(E/K) = 1$. Since $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q})$, we have $r_{\mathrm{an}}(E/\mathbb{Q}) = 1$ and $r_{\mathrm{an}}(E^D/\mathbb{Q}) = 0$, so Parts 1,3,4, 5 follows. Finally, Part 6 follows since V^0_{k,ℓ^n} is just the image of the Heegner point y_K under the connecting homomorphism, and $\mathrm{ord}_\ell(y_K) = m_f$.

Next assume that $\mathrm{rank}(V) > 1$. By our minimality hypothesis, $r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$, so $\mathrm{rank}(E^D(\mathbb{Q})) \leq 1$, hence $V \not\subset (E(K)/E(K)_{\mathrm{tor}})^-$, so $V \subset (E(K)/E(K)_{\mathrm{tor}})^+$, which proves Part 1. We have $f \leq a$ since $V^a_{k,\ell^n} \neq 0$ for some $k \geq 0$. Also, since $f < \infty$, Theorem 4.2 implies that $\mathrm{rank}(E(\mathbb{Q})) \leq f + 1$. Since $\mathrm{rank}((E(K)/E(K)_{\mathrm{tor}})^+) = \mathrm{rank}(E(\mathbb{Q}))$, we have

$$a + 1 = \mathrm{rank}(V) \leq \mathrm{rank}(E(\mathbb{Q})) \leq f + 1 \leq a + 1.$$

We conclude that the above inequalities are equalities, so $a = f$ which proves Part 2, and $\mathrm{rank}(E(\mathbb{Q})) = f + 1$, which proves Part 3. Also because $\mathrm{rank}(E(\mathbb{Q})) = f + 1$, Theorem 4.2 implies that $Ш(E/\mathbb{Q})(\ell)$ is finite, so since $Ш(E^D/\mathbb{Q}))$ is also finite, Part 4 is true. Considering the definition of the A^ν before the statement of Conjecture 4.9, we see that $1 - (-1)^\nu(-1)^{r_{\mathrm{an}}-1}\sigma = 1 + \sigma$, so $\nu \equiv r_{\mathrm{an}} \pmod 2$. Since part of Conjecture 4.9 is that $\mathrm{rank}(V) \equiv \nu \pmod 2$, and we proved that $\mathrm{rank}(V) = \mathrm{rank}(E(\mathbb{Q}))$, we conclude that $r_{\mathrm{an}} \equiv \mathrm{rank}(E(\mathbb{Q}))$ (mod 2), which is Part 5. By [Kol91a, Thm. 3], for all $k \geq m_f$ the subgroup $V^f_{k,\ell^n} \subset \mathrm{H}^1(K, E[\ell^\infty])$ contains $(\ell^{m_f}\mathbb{Z}/\ell^n\mathbb{Z})^{f+1} = \delta(\ell^{m_f}E(\mathbb{Q}))$, so $\ell^{m_f}E(\mathbb{Q}) \otimes \mathbb{Z}_\ell \subset V \otimes \mathbb{Z}_\ell$. On the other hand, by definition of m_f, every cohomology class τ_{λ,ℓ^n} is contained in $\ell^{m_f}\,\mathrm{H}^1(K, E[\ell^n])$. Thus $\delta(V) \subset \ell^{m_f}\,\mathrm{H}^1(K, E[\ell^\infty])$, so $V \subset \ell^{m_f}E(\mathbb{Q})$. This proves Part 6.

□

Recall from Section 2 that c is the Manin constant of E and the c_q are the Tamagawa numbers of E. We make the following new refinement of Kolyvagin's Conjecture 4.8.

Conjecture 4.12. *We have* $m_\ell = \mathrm{ord}_\ell(c \cdot \prod_{q|N} c_q)$.

Theorem 7.5 and Theorem 7.7 below serve as our motivation to make Conjecture 4.12. In particular, Kolyvagin proved that at primes $\ell \in B(E)$, Conjecture 4.12 is equivalent to [GZ86, Conj 2.2, pg 311] in the special case when E has analytic rank 1 over K.

5 Mod p Kolyvagin Points and Kolyvagin Subgroups

As always, we assume E is an elliptic curve over $\mathbb{Q}$, that K is a quadratic imaginary field satisfying the Heegner hypothesis, and p is a prime that is inert in K. The Heegner hypothesis implies that the primes of bad reduction for E split in K, so p must be a prime of good reduction. For each such prime, we define a finite-index subgroup W_p of $E(K)$. We do this by extending Kolyvagin's construction of points P_λ to obtain a new well-defined construction of elements of the quotient group

$$E(\mathbb{F}_p)/(p+1) = E(\mathbb{F}_p)/(p+1)E(\mathbb{F}_p)$$

for any inert prime p. Thus this section takes Kolyvagin's definition of points P_λ one step further to define elements of $E(\mathbb{F}_p)/(p+1)$. We first compute the structure of the odd part of the group $E(\mathbb{F}_p)/(p+1)$ for any good prime p. We then use properties of splitting of primes in certain ring class fields to define the canonical reduction $R_{p,\lambda} \in E(\mathbb{F}_p)/(p+1)$ of the Kolyvagin points P_λ, and consider the subgroup X_p of $E(\mathbb{F}_p)/(p+1)$ generated by the $R_{p,\lambda}$ for certain λ. We then define W_p to be the inverse image of X_p and finish with some results about the structure of W_p.

If A is a finite abelian group, the *odd part* of A is the subgroup of A of all elements of odd order, and if n is an integer, the odd part of n is $n/2^{\mathrm{ord}_2(n)}$.

Lemma 5.1. *The odd part of* $E(\mathbb{F}_p)/(p+1)$ *is cyclic of order the odd part of* $\gcd(p+1, a_p)$.

Proof. Suppose ℓ is an odd prime divisor of $\#(E(\mathbb{F}_p)/(p+1))$. If the ℓ-primary subgroup of $E(\mathbb{F}_p)/(p+1)$ is not cyclic, then since $\ell \neq p$ we have $E(\mathbb{F}_p)[\ell] \approx (\mathbb{Z}/\ell\mathbb{Z})^2$. The Weil pairing induces an isomorphism of Galois modules $\bigwedge^2 E[\ell] \cong \mu_\ell$ and $E[\ell] \subset E(\mathbb{F}_p)$, so $\mu_\ell \subset \mathbb{F}_p^*$, hence $\ell \mid (p-1)$. Since ℓ divides $\#(E(\mathbb{F}_p)/(p+1))$ and ℓ is prime, we have $\ell \mid (p+1)$, so $\ell \mid \gcd(p-1, p+1) = 2$, a contradiction, since ℓ is odd.

The group $E(\mathbb{F}_p)$ has order $p + 1 - a_p$, and we just proved above that $E(\mathbb{F}_p)(\ell)$ is cyclic for any odd prime divisor ℓ of $p+1$. Thus the quotient ℓ-primary group $(E(\mathbb{F}_p)/(p+1))(\ell) = (E(\mathbb{F}_p)(\ell))/(p+1)$ has order ℓ^m, where

$$m = \mathrm{ord}_\ell(\gcd(p+1, \#E(\mathbb{F}_p))) = \mathrm{ord}_\ell(\gcd(p+1, p+1-a_p)) = \mathrm{ord}_\ell(\gcd(p+1, a_p)).$$

Taking the product over all odd primes ℓ, shows that the odd part of $E(\mathbb{F}_p)/(p+1)$ has order the odd part of $\gcd(p+1, a_p)$. □

Remark 5.2. 1. *Lemma 5.1 is true even if p is a good prime that is not inert in K (in fact, the lemma and proof have nothing to do with K).*

2. *Lemma 5.1 is false if we do not restrict to odd parts. For example, if E is $y^2 = x^3 - x$ and $p = 3$, then $E(\mathbb{F}_3) \approx (\mathbb{Z}/2\mathbb{Z})^2$, so $E(\mathbb{F}_3)/4 \approx (\mathbb{Z}/2\mathbb{Z})^2$ is not cyclic.*
3. *For every prime ℓ, there exists infinitely many primes p such that $E(\mathbb{F}_p)(\ell)$ is not cyclic. Indeed, by the Chebotarev density theorem there are infinitely many p that split completely in the field $\mathbb{Q}(E[\ell])$, and for these p we have $(\mathbb{Z}/\ell\mathbb{Z})^2 \subset E(\mathbb{F}_p)$.*

Lemma 5.3. *If p is inert in K and does not divide λ, then the prime ideal $p\mathcal{O}_K$ of K splits completely in K_λ. In particular, if $p \in \Lambda^1_{\ell^n}$ and $\lambda \in \Lambda_{\ell^n}$ with $p \nmid \lambda$, then $p\mathcal{O}_K$ splits completely in K_λ.*

Proof. (Compare line -3 on page 103 of [Kol91b].) Since p is inert, the ideal $p\mathcal{O}_K$ is a prime principal ideal of $\mathcal{O}_K$, hence splits completely in the Hilbert class field K_1. As explained in [Gro91, pg. 238], class field theory identifies $\mathrm{Gal}(K_\lambda/K_1)$ with $C = (\mathcal{O}_K/\lambda\mathcal{O}_K)^*/(\mathbb{Z}/\lambda\mathbb{Z})^*$. The image of p is trivial in C, so the Frobenius element attached to $p\mathcal{O}_K$ is trivial, hence $p\mathcal{O}_K$ splits completely in the ring of integers of K_λ, as claimed. □

Define the reduction map $E(K) \to E(\mathbb{F}_{p^2})$ by reducing the Néron model $\mathcal{E}$ of E over $\mathcal{O}_K$ modulo $p\mathcal{O}_K$, and using the natural maps $E(K) \cong \mathcal{E}(\mathcal{O}_K) \to \mathcal{E}_{\mathbb{F}_{p^2}}(\mathbb{F}_{p^2}) \cong E(\mathbb{F}_{p^2})$. Let $\pi_p : E(K) \to E(\mathbb{F}_p)/(p+1)$ be the composition of reduction modulo the prime ideal $p\mathcal{O}_K$ with $\mathrm{Tr}_{\mathbb{F}_{p^2}/\mathbb{F}_p} : E(\mathbb{F}_{p^2}) \to E(\mathbb{F}_p)$ followed by quotienting out by the subgroup $(p+1)E(\mathbb{F}_p)$. Fix a *choice* $\wp$ of prime ideal of K_λ over $p\mathcal{O}_K$. Extend π_p to a map $\pi_\wp : E(K_\lambda) \to E(\mathbb{F}_p)/(p+1)$ by quotienting out by $\wp$, as illustrated in the following diagram:

$$\begin{tikzcd}
E(K_\lambda) \arrow[drr, "\text{mod } \wp"] & & \\
E(K) \arrow[u] \arrow[r, "\cong"] \arrow[ddrr, dotted, "\pi_p"'] & \mathcal{E}(\mathcal{O}_K) \arrow[r] & E(\mathbb{F}_{p^2}) \arrow[d, "\text{trace}"] \\
& & E(\mathbb{F}_p) \arrow[d] \\
& & E(\mathbb{F}_p)/(p+1)
\end{tikzcd}$$

For each $\ell \mid (p+1)$, let $v_\ell = \mathrm{ord}_\ell(\gcd(a_p, p+1))$, and define $\pi_{\wp,\ell} : E(K_\lambda) \to (E(\mathbb{F}_p)/(p+1))(\ell)$ by

$$\pi_{\wp,\ell}(S) = \pi_\wp\left(\frac{p+1}{\ell^{v_\ell}} S\right).$$

We now study how the homomorphism $\pi_{\wp,\ell}$ depends on our choice of prime of $\wp$ over $p\mathcal{O}_K$.

Proposition 5.4. *The map $\pi_{\wp,\ell}$ induces a well-defined (independent of choice of $\wp$) homomorphism*

$$\vartheta : (E(K_\lambda)/\ell^{v_\ell}E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)} \to E(\mathbb{F}_p)/(p+1).$$

Proof. Let $[S] \in (E(K_\lambda)/\ell^{v_\ell}E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)}$ with $S \in E(K_\lambda)$. If $\wp'$ is another prime of K_λ over $p\mathcal{O}_K$, then because the Galois group acts transitively on the primes over a given prime, there is $\sigma \in \mathrm{Gal}(K_\lambda/K)$ such that $\pi_{\wp',\ell}(S) = \pi_{\wp,\ell}(\sigma(S))$. Since $[S]$ is $\mathrm{Gal}(K_\lambda/K)$-equivariant, we have $\sigma(S) = S + \ell^{v_\ell} \cdot Q$, for some $Q \in E(K_\lambda)$, so

$$\begin{aligned}
\vartheta([\sigma(S)]) &= \pi_{\wp,\ell}(\sigma(S)) \\
&= \pi_\wp\left(\frac{p+1}{\ell^{v_\ell}}\sigma(S)\right) \\
&= \pi_\wp\left(\frac{p+1}{\ell^{v_\ell}}S\right) + \pi_\wp((p+1)Q) \\
&= \pi_{\wp,\ell}(S) + 0 = \vartheta([S]),
\end{aligned}$$

where $\pi_\wp((p+1)Q) = (p+1)\pi_\wp(Q)$ is 0, since the group $E(\mathbb{F}_p)/(p+1)$ is killed by $p+1$. □

By Proposition 3.4, $[P_\lambda]$ is in the domain of the homomorphism ϑ of Proposition 5.4.

Definition 5.5 (Mod p Kolyvagin Point)**.** *The* mod p Kolyvagin point *associated to $p \in \Lambda^1_{\ell^n}$ and $\lambda \in \Lambda_{\ell^n}$ is*

$$R_{p,\lambda} = \vartheta([P_\lambda]) \in E(\mathbb{F}_p)/(p+1),$$

where ϑ is as in Proposition 5.4.

As above, let $v_\ell = \mathrm{ord}_\ell(\gcd(a_p, p+1))$. For each $k \geq 0$, let

$$X_{k,p} = \left\langle R_{p,\lambda} : \lambda \in \bigcup_\ell \Lambda^{f_\ell}_{\ell^{v_\ell+k}} \right\rangle \subset E(\mathbb{F}_p)/(p+1) \tag{11}$$

be the subgroup generated by all mod p Kolyvagin points associated to λ that are a product of f_ℓ primes, where f_ℓ is from Equation (9). Note that the subscript of Λ in (11) is $\ell^{v_\ell+k}$, and we take the union over *all* ℓ thus obtaining a subgroup $X_{k,p}$ that need not be ℓ-primary for any ℓ, despite $R_{p,\lambda}$ being ℓ-primary. Let

$$X_p = \bigcap_{k \geq 0} X_{k,p}.$$

Let $W_{k,p}$ be the inverse image of $X_{k,p}$ under the map π_p:

$$W_{k,p} = \pi_p^{-1}(X_{k,p}) \subset E(K),$$

and

$$W_p = \pi_p^{-1}(X_p) \subset E(K).$$

Since $E(\mathbb{F}_p)/(p+1)$ is finite, $W_{k,p}$ and W_p have finite index in $E(K)$; also, by Lemma 5.1, the odd part of this index divides $\gcd(p+1, a_p)$.

Remark 5.6. *Note that $E^D(\mathbb{Q})$ is in the kernel of the trace map, hence in the kernel of π_p, so $E^D(\mathbb{Q}) \subset W_p$. Thus it is possible that W_p contains torsion, hence W_p in general need not be a Gross-Zagier subgroup as in Definition 2.3. In a future paper, we intend to give a more refined definition of a sequence of groups W_p^a, for each $a \geq 0$, which better accounts for torsion. We would then search for a Gross-Zagier style formula for each group W_p^a for $a \leq f+1$, in order to more closely relate $r_{\mathrm{an}}(E/\mathbb{Q})$ to $f+1$.*

6 Controlling the Reduction Map

The main result of this section is a proof that under certain hypothesis, if a point Q has infinite order and n is a positive integer, then there are infinitely many primes p such that the image of Q in $E(\mathbb{F}_{p^2})/(p+1)$ has order divisible by n. We prove this using Galois cohomology and by converting a condition on ℓ-divisibility of points into a Chebotarev condition. We will use this result later to study the maximum index $[E(K) : W_p]$ that can occur and prove a generalized Gross-Zagier formula for such W_p.

Let E, K, etc., be as above, and let $\ell \in B(E)$, where $B(E)$ is the set of primes defined on page 5. Suppose $Q \in E(K)$ has infinite order, and let n be an odd positive integer. Suppose that for each prime $\ell \mid n$, the set of cardinalities $\{\# \mathrm{H}^1(K(E[\ell^j])/K, E[\ell^j]) : j \geq 1\}$ is bounded. This hypothesis is satisfied if $\ell \in B(E)$, since then $\mathrm{H}^1(K(E[\ell^j])/K, E[\ell^j]) = 0$ for all j (see [Gro91, pg. 241] and [GJP$^+$09, Prop. 5.2]).

Proposition 6.1. *Let Q and n be as above. Let S be the set of primes p such that p is inert in K, p splits completely in $K(E[n])/K$, and the image of Q in $E(\mathbb{F}_{p^2})/(p+1)E(\mathbb{F}_{p^2})$ has order divisible by n. Then S has positive (Dirichlet) density.*

Proof. Let $m = \prod \ell_i^{e_i}$ with ℓ_i the distinct primes that divide n, and e_i any positive integers, which we will fix later in the argument. Fix any i, and let $L = K(E[\prod_{j \neq i} \ell_j])$, which is a Galois extension of K. Define homomorphisms Ψ_i, f, g, and h as in the following commutative diagram:

$$\begin{tikzcd}
E(K(E[m]))/\ell_i^{e_i} E(K(E[m])) \arrow[r, hook] & \mathrm{H}^1(K(E[m]), E[\ell_i^{e_i}]) \\
 & \mathrm{H}^1(L(E[\ell_i^{e_i}]), E[\ell_i^{e_i}]) \arrow[u, "f"'] \\
 & \mathrm{H}^1(K(E[\ell_i^{e_i}]), E[\ell_i^{e_i}]) \arrow[u, "h"'] \\
E(K)/\ell_i^{e_i} E(K) \arrow[uuu] \arrow[r, hook] \arrow[uuur, dotted, "\psi_i"] & \mathrm{H}^1(K, E[\ell_i^{e_i}]) \arrow[u, "g"']
\end{tikzcd}$$

The horizontal maps above are induced by the short exact sequence coming from multiplication by $\ell_i^{e_i}$, and the vertical maps on the right are the restriction maps. The diagram commutes so the order of the image of Q in $E(K(E[m]))/\ell_i^{e_i}E(K(E[m]))$ is the same as the order of $\Psi_i(Q)$.

By hypothesis and the inflation restriction sequence the cardinality of $\ker(g)$ is bounded independently of i and e_i. Also, $[L : K]$ depends only on the set of prime divisors ℓ_i of n, not their exponents, so

$$\#\ker(h) = \#\mathrm{H}^1(L(E[\ell_i^{e_i}])/K(E[\ell_i^{e_i}]), E[\ell_i^{e_i}]) = \#\mathrm{Hom}(\mathrm{Gal}(L(E[\ell_i^{e_i}])/K(E[\ell_i^{e_i}])), E[\ell_i^{e_i}])$$

is also bounded independent of e_i, because every homomorphism has image in the fixed subset $E[\ell_i^d]$, where d is the exponent of the group $\mathrm{Gal}(L/K)$. Finally, the map f is injective, since

$$\ker(f) \cong \mathrm{H}^1(K(E[m])/L(E[\ell_i^{e_i}]), E[\ell_i^{e_i}])$$

and $\#\mathrm{Gal}(K(E[m])/L(E[\ell_i^{e_i}]))$ is divisible only by the primes ℓ_j for $j \neq i$ and these are all coprime to $\#E[\ell_i^{e_i}] = \ell_i^{2e_i}$. We conclude that there is an integer b such that $\#\ker(\Psi_i) \leq \ell_i^b$, and this bound holds no matter how we increase the numbers e_i and e_j (for all j).

The above proof that $\ker(\Psi_i)$ is uniformly bounded is completely general. See Remark 6.3 for a sketch of an alternative proof of this bound in the special case when $\ell \in B(E)$ for all $\ell \mid n$, which is the only case we will use in this paper.

Because $\ker(\Psi_i)$ is uniformly bounded independent of our choice of e_i, for each i, we can choose e_i large enough so that $\Psi_i(Q)$ has order divisible by $\ell_i^{\mathrm{ord}_{\ell_i}(n)}$. Then for each i, let d_i be maximal such that $\ell_i^{d_i}$ divides Q in $E(K(E[m]))$. Note that $d_i < e_i$ for each i, since $\Psi_i(Q) \neq 0$ and $\Psi(Q)$ is an element of a group that is killed by $\ell_i^{e_i}$. Since $m = \prod \ell_i^{e_i}$, we have $\ell_i^{d_i+1} \mid m$, so

$$M_i = K\left(E[m],\ \frac{1}{\ell_i^{d_i+1}}Q\right)$$

does not depend on the choice of $\ell_i^{d_i+1}$th root of Q, is a Galois extension of $K(E[m])$, and $[M_i : K(E[m])]$ is a nontrivial power of ℓ_i. Thus the M_i for all i are linearly disjoint as extensions of $K(E[m])$.

Let M be the compositum of the fields M_i defined above. Since the M_i are linearly disjoint nontrivial extensions of $K(E[m])$, there exists an automorphism $\sigma \in \mathrm{Gal}(M/\mathbb{Q})$ such that $\sigma|_{K(E([m]))}$ is complex conjugation, and $\sigma|_{M_i}$ has order divisible by ℓ_i for each i. By the Chebotarev density theorem, there is a positive density of primes $p \in \mathbb{Z}$ that are unramified in M and have Frobenius the class of σ. Such primes are inert in K since complex conjugation acts nontrivially on K, split completely in $K(E[m])/K$ since complex conjugation has order 2, and each prime over p in $K(E[m])$ does not split completely in any of the extensions $M_i/K(E[m])$ since $[\mathrm{Frob}_p]|_{M_i} = \sigma|_{M_i}$ has order divisible by $\ell_i > 2$. Note that this is the only place in the argument where we use that n is odd.

Let p be any prime as in the previous paragraph. We have

$$E(\mathbb{F}_{p^2})/\ell_i^{e_i}E(\mathbb{F}_{p^2}) \cong (\mathbb{Z}/\ell_i^{e_i}\mathbb{Z})^2$$

since $p\mathcal{O}_K$ splits completely in $K(E[m])$ and $\ell_i^{e_i} \mid m$. Also, the Frobenius condition implies that the primes of M_i over $p\mathcal{O}_K$ do not have residue class degree 1, so since M_i is generated by any choice of $\frac{1}{\ell^{d_i+1}}Q$, the reduction $\overline{Q}$ of Q modulo any prime over $p\mathcal{O}_K$ is not divisible by $\ell_i^{d_i+1}$ in $E(\mathbb{F}_{p^2})$. Note that $\ell_i^{d_i}$ divides $\overline{Q}$, because the prime $p\mathcal{O}_K$ splits completely in $K(E[m])/K$ and $\ell_i^{d_i}$ divides Q in $K(E[m])$, so d_i is the largest integer such that $\ell_i^{d_i}$ divides the image of $\overline{Q}$ in $E(\mathbb{F}_{p^2})$. We conclude that for each i the image of Q in $E(\mathbb{F}_{p^2})/\ell_i^{e_i}E(\mathbb{F}_{p^2})$ has order the same as the order of $\Psi_i(Q)$.

By hypothesis, $e_i \geq \mathrm{ord}_{\ell_i}(n)$ and $\Psi_i(Q)$ has order divisible by $\ell_i^{\mathrm{ord}_{\ell_i}(n)}$ for each i, so the image of Q in $E(\mathbb{F}_{p^2})/mE(\mathbb{F}_{p^2})$ has order divisible by n. For any such p, we also have that the characteristic polynomial of the class of Frob_p in $\mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q})$ acting on $E[m]$ is $x^2 - a_px + p \pmod{m}$. On the other hand, since $[\mathrm{Frob}_p]$ on $E[m]$ is the class of complex conjugation and complex conjugation acts nontrivially (since m is odd) hence has characteristic polynomial $x^2 - 1$, we have $x^2 - a_px + p \equiv x^2 - 1 \pmod{m}$. Thus $m \mid (p+1)$, so the image of Q in $E(\mathbb{F}_{p^2})/(p+1)E(\mathbb{F}_{p^2})$ also has order divisible by n, which completes the proof. □

Remark 6.2. *Proposition 6.1 is analogous to the statement that if $x, n \in \mathbb{Z}$ with $\gcd(n, x) = 1$ and $\mathbb{Q}(\zeta_n, \sqrt[n]{x})$ is an extension of $\mathbb{Q}(\zeta_n)$ of degree n, then there exist a positive density of primes p such that the multiplicative order of x modulo p is divisible by n. The proof of this statement resembles the proof of Proposition 6.1, except we work with the field $\mathbb{Q}(\zeta_n, \sqrt[n]{x})$. The idea of the proof of Proposition 6.1 is well-known to experts who study questions such as the Lang-Trotter conjecture about reduction of points on elliptic curves.*

Remark 6.3. *If for every prime $\ell \mid n$ we have $\ell \in B(E)$, we can alternatively use that $K(E[\ell_1^\infty])$ and $K(E[\ell_2^\infty])$ are linearly disjoint for distinct odd primes ℓ_1 and ℓ_2 in $B(E)$ to give a different proof that the maps Ψ_i have uniformly bounded kernel in Proposition 6.1. In that case we have that* $\mathrm{Gal}(K(E[n])/K) \approx \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$, *so*

$$\ker\Big(\mathrm{H}^1(K, E[n]) \to \mathrm{H}^1(K(E[n]), E[n])\Big) \cong \mathrm{H}^1(K(E[n])/K, E[n]) = \mathrm{H}^1(\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z}), (\mathbb{Z}/n\mathbb{Z})^2) = 0,$$

where the last group is 0 *by a standard group cohomology argument (see, e.g., [Ste02, §5.1]). This implies that the maps Ψ_i are all injective. The linear disjointness of $K(E[\ell_1^\infty])$ and $K(E[\ell_2^\infty])$ for the distinct* odd *primes ℓ_1 and ℓ_2 follows by a Galois theory argument using the structure of* $\mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$. *We thank R. Greenberg for this observation.*

7 Maximal Index Subgroups W_p

As above, we assume that E is an elliptic curve over $\mathbb{Q}$ with positive analytic rank and that $K = \mathbb{Q}(\sqrt{D})$ is a quadratic imaginary field that satisfies the Heegner hypothesis and the minimality hypothesis that $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$.

Recall that for each inert prime p of K we defined a subgroup $X_p \subset E(\mathbb{F}_p)/(p+1)$ in Equation (11) of Section 5. This was a group got by reducing Kolyvagin points associated to all primes ℓ modulo a choice of prime over p. In this section, for all $\ell \in B(E)$ we conditionally compute, in terms of $m_{\ell,f}$, the ℓ-primary part $X_p(\ell)$ of this subgroup $X_p \subset E(\mathbb{F}_p)/(p+1)$. We relate our refinement of Kolyvagin's conjectures to the generalized Gross-Zagier formula (5). We also conditionally compute X_p in terms of $c \cdot \prod c_q \cdot \sqrt{\#\text{Ш}(E/K)}$ using Theorem 4.2. We apply our description of X_p to prove that, up to primes not in $B(E)$, the subgroups W_p with $[E(K) : W_p]$ maximal are all Gross-Zagier subgroups of $E(K)$.

Proposition 7.1. *Conjecture 4.9 implies that for every $\ell \in B(E)$,*

$$X_p(\ell) = \frac{p+1}{\ell^{v_\ell}} \cdot \pi_p(\ell^{m_{\ell,f}} E(\mathbb{Q})),$$

where $v_\ell = \mathrm{ord}_\ell(p+1)$.

Proof. Let Φ be the composite homomorphism

$$(E(K_\lambda)/\ell^{v_\ell}E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)} \hookrightarrow \mathrm{H}^1(K_\lambda, E[\ell^{v_\ell}])^{\mathrm{Gal}(K_\lambda/K)} \cong \mathrm{H}^1(K, E[\ell^{v_\ell}]),$$

and let $\delta : E(K) \to \mathrm{H}^1(K, E[\ell^{v_\ell}])$. We are assuming Conjecture 4.9, so we may apply Proposition 4.11 Part 6 (taking into account Lemma 4.5), to see that for all k sufficiently large we have $\delta(\ell^{m_{\ell,f}}E(\mathbb{Q})) = V^f_{k,\ell^{v_\ell}}$. Thus

$$\delta(\ell^{m_{\ell,f}}E(\mathbb{Q})) = \langle \Phi([P_\lambda]) : \lambda \in \Lambda^f_{\ell^{v_\ell+k}} \rangle.$$

Let $i : E(K) \to (E(K_\lambda)/\ell^{v_\ell}E(K_\lambda))^{\mathrm{Gal}(K_\lambda/K)}$. For any $Q \in E(K)$ we have $\frac{p+1}{\ell^{v_\ell}} \cdot \pi_p(Q) = \vartheta(i(Q))$ where ϑ is as in Proposition 5.4. Since $\delta = \Phi \circ i$ and Φ is injective, the group $X_{k,p}(\ell)$ generated by all $\vartheta([P_\lambda])$ is equal to $\vartheta(i(\ell^{m_{\ell,f}}E(\mathbb{Q})))$. Since this is true for all sufficiently large k, the proposition follows for X_p. □

Theorem 7.5 below generalizes [Kol91b, Thm. E] to arbitrary rank. To prove it we first prove some lemmas and make a definition.

Lemma 7.2. *Suppose A is a nonzero finitely generated free abelian group and $\varphi : A \to \mathbb{Z}/d\mathbb{Z}$ is a surjective homomorphism. For every nonzero integer c we have $[A : \varphi^{-1}(\varphi(cA))] = \gcd(c, d)$.*

Proof. Let $B = \varphi^{-1}(\varphi(cA))$. We have $\varphi(\ker(\varphi)) = 0 \subset \varphi(cA)$, so $\ker(\varphi) \subset B$. Since $\ker(\varphi) \subset B$, the isomorphism $A/\ker(\varphi) \cong \mathbb{Z}/d\mathbb{Z}$ induces an isomorphism $A/B \cong (\mathbb{Z}/d\mathbb{Z})/\varphi(B)$. But φ is surjective, so $\varphi(B) = \varphi(\varphi^{-1}(\varphi(cA))) = \varphi(cA) = c\varphi(A) = c(\mathbb{Z}/d\mathbb{Z})$, so $A/B \cong (\mathbb{Z}/d\mathbb{Z})/(c(\mathbb{Z}/d\mathbb{Z})) \cong \mathbb{Z}/\gcd(d,c)\mathbb{Z}$. □

Recall (see page 5) that $B(E)$ is a set of primes that have certain good properties for E. Below, for any integer n we either let $n' = \ell^{\mathrm{ord}_\ell(n)}$ be the ℓ-part of n or the maximal divisor of n divisible only by primes in $B(E)$, depending on whether we are considering the first or second part of the following lemma.

Lemma 7.3. *Assume $E(\mathbb{Q})$ has positive rank and let t be a positive integer.*

1. *If $\ell \in B(E)$ is such that $X_p(\ell) = \frac{p+1}{\ell^{v_\ell}} \cdot \pi_p(tE(\mathbb{Q}))$ for all inert primes p, then*

$$\max\{\mathrm{ord}_\ell([E(K) : W_p]) : \textit{ all inert } p\} = \mathrm{ord}_\ell(t).$$

2. If for all $\ell \in B(E)$ we have $X_p(\ell) = \frac{p+1}{\ell^{v_\ell}} \cdot \pi_p(tE(\mathbb{Q}))$ for all inert primes p, then

$$\max\{[E(K) : W_p]' : \text{ all inert } p\} = t'.$$

Proof. Let p be any inert prime, and recall that p is a prime of good reduction, since all bad primes split in K. By Lemma 5.1, the odd part of the image of $\pi_p : E(K) \to E(\mathbb{F}_p)/(p+1)$ is a cyclic group $\mathbb{Z}/n\mathbb{Z}$ for some integer n. Since $\pi_p(E^D(\mathbb{Q})) = 0$ (see Remark 5.6), we have

$$\pi_p(tE(\mathbb{Q}))' = \pi_p(tE(\mathbb{Q}) + tE^D(\mathbb{Q}))' = \pi_p(tE(K))',$$

so by Proposition 7.6, $W_p' = \pi_p^{-1}(X_p)' = \pi_p^{-1}(\pi_p(tE(K))')$. Thus Lemma 7.2 implies that $[E(K)' : W_p']$ is $\gcd(t,n)'$. This proves that set of indexes $[E(K)' : W_p']$ all divide t'.

We show the maximum equals t' by proving that there is a positive density of primes p such that the n above is divisible by t'. By hypothesis, there is a point $P \in E(\mathbb{Q})$ of infinite order. By Proposition 6.1, there exists a positive density of primes p that are inert in K such that $\pi_p(P) \in E(\mathbb{F}_p)/(p+1)$ has order divisible by t'. For such p, the n above is thus divisible by t', so $\gcd(t,n)' = t'$, which completes the proof. □

Let

$$w_\ell = \sup(\{\mathrm{ord}_\ell([E(K) : W_p]) : \text{ all inert } p\}) \le \infty. \tag{12}$$

Lemma 7.4. *Suppose $\ell \in B(E)$, that Conjecture 4.9 is true for E, and assume that p is an inert prime such that $\mathrm{ord}_\ell([E(K) : W_p])$ is maximal in the sense that it equals w_ℓ. Then $m_{\ell,f} = \mathrm{ord}_\ell([E(K) : W_p])$.*

Proof. We are assuming that Conjecture 4.9 is true, so Proposition 7.1 applies and gives an explicit formula for $X_p(\ell)$. Namely, we may take $t = m_{\ell,f}$ in Lemma 7.3. Also, by Conjecture 4.9 (and Proposition 4.11) we have $E(\mathbb{Q})$ has rank at least 1. The lemma then follows from Lemma 7.3. □

Theorem 7.5. *Suppose $\ell \in B(E)$, that Conjectures 2.2 and 4.9 are true for E, and that p is an inert prime such that $w_\ell = \mathrm{ord}_\ell([E(K) : W_p])$, where w_ℓ is as in (12) above. Then W_p satisfies the generalized Gross-Zagier formula (5) up to a rational factor that is coprime to ℓ if and only if Conjecture 4.12 is true for ℓ.*

Proof. We are assuming Conjecture 4.9, which implies Conjecture 4.1, so we may apply Theorem 4.2, which has Conjecture 4.1 as a hypothesis. Let b_k be as in Theorem 4.2 for our given prime ℓ. Theorem 4.2 implies that

$$\begin{aligned}\#Ш(E/K)(\ell) &= \#((\mathbb{Z}/b_f\mathbb{Z})^2 \oplus (\mathbb{Z}/b_{f+1}\mathbb{Z})^2 \oplus (\mathbb{Z}/b_{f+2}\mathbb{Z})^2 \oplus \cdots) \\ &= (b_f \cdot b_{f+1} \cdots)^2 \\ &= \ell^{2(m_{\ell,f} - m_{\ell,f+1} + m_{\ell,f+1} - m_{\ell,f+2} + m_{\ell,f+2} - \cdots)} \cdots \\ &= \ell^{2(m_{\ell,f} - m_\ell)},\end{aligned}$$

so $m_{\ell,f} - m_\ell = \mathrm{ord}_\ell(\sqrt{\#Ш(E/K)(\ell)})$.

We will now show that the generalized Gross-Zagier formula (5) holds up to a rational factor that is coprime to ℓ if and only if Conjecture 4.12 that $m_\ell = \mathrm{ord}_\ell(c\prod c_q)$ is true for ℓ. We will repeatedly use Lemma 7.4 that $m_{\ell,f} = \mathrm{ord}_\ell([E(K) : W_p])$.

First, suppose that the generalized Gross-Zagier formula (5) holds up to a rational factor that is coprime to ℓ. Proposition 2.4 combined with Conjecture 2.2 (that $Ш_{\mathrm{an}} = \#Ш$), implies that this hypothesis means that $\mathrm{ord}_\ell([E(K) : W_p]) = \mathrm{ord}_\ell\left(c\prod c_q \cdot \sqrt{\#Ш(E/K)(\ell)}\right)$. Thus:

$$\begin{aligned}m_{\ell,f} &= \mathrm{ord}_\ell([E(K) : W_p]) \\ &= \mathrm{ord}_\ell\left(c\prod c_q \cdot \sqrt{\#Ш(E/K)(\ell)}\right) \\ &= \mathrm{ord}_\ell\left(c\prod c_q\right) + \mathrm{ord}_\ell(\sqrt{\#Ш(E/K)(\ell)}) \\ &= \mathrm{ord}_\ell\left(c\prod c_q\right) + m_{\ell,f} - m_\ell,\end{aligned}$$

where in the last equality we use the formula for $\#Ш(E/K)(\ell)$ that we derived above using Theorem 4.2. Subtracting $m_{\ell,f}$ from both sides shows that $m_\ell = \mathrm{ord}_\ell(c\prod c_q)$.

Conversely, suppose that $m_\ell = \mathrm{ord}_\ell(c\prod c_q)$. From Theorem 4.2 we have

$$m_{\ell,f} - m_\ell = \mathrm{ord}_\ell(\sqrt{\#Ш(E/K)(\ell)}),$$

so

$$\mathrm{ord}_\ell([E(K) : W_p]) = m_{\ell,f} = \mathrm{ord}_\ell\left(c\prod c_q\right) + m_{\ell,f} - m_\ell = \mathrm{ord}_\ell\left(c\prod c_q \cdot \sqrt{\#Ш(E/K)(\ell)}\right).$$

Proposition 2.4 then implies that W_p satisfies the generalized Gross-Zagier formula up to a rational factor coprime to ℓ. □

For any integer n, let n' denote the maximal divisor of n that is divisible only by primes in $B(E)$, and for any abelian group A, let $A' = A \otimes \mathbb{Z}[1/b]$, where b is the product of the finitely many primes not in $B(E)$. Let

$$T = c \cdot \prod_{q|N} c_q \cdot \sqrt{\#\text{Ш}(E/K)}.$$

Proposition 7.6. *Conjectures 4.9 and 4.12 together imply that* $X_p' = \pi_p(TE(\mathbb{Q}))'$.

Proof. Using the calculation in the first paragraph of the proof of Theorem 7.5 along with Conjecture 4.12 combined with Theorem 4.2, shows that for every $\ell \in B(E)$, we have

$$m_{\ell,f} = \text{ord}_\ell(T).$$

Since the integers $T/\ell^{m_{\ell,f}}$ and $(p+1)/\ell^{v_\ell}$, for $v_\ell = \text{ord}_\ell(p+1)$, both act as automorphisms on any ℓ-primary group,

$$\begin{aligned}\pi_p(TE(\mathbb{Q}))(\ell) &= \left(\frac{T}{\ell^{m_{\ell,f}}} \cdot \pi_p(\ell^{m_{\ell,f}} E(\mathbb{Q}))\right)(\ell)\\ &= \pi_p(\ell^{m_{\ell,f}} E(\mathbb{Q}))(\ell)\\ &= \frac{p+1}{\ell^{v_\ell}} \cdot \pi_p(\ell^{m_{\ell,f}} E(\mathbb{Q})) = X_p(\ell),\end{aligned}$$

where the last equality uses Proposition 7.1 (which assumes that Conjecture 4.9 is true). We conclude that $X_p' = \pi(TE(\mathbb{Q}))'$.

Theorem 7.7 is a partial converse to Theorem 7.5.

Theorem 7.7. *Assume that $E(\mathbb{Q})$ has positive rank. Then Conjectures 4.9 and 4.12 together imply that the maximum index $[E(K)' : W_p']$ over all inert p is $(c \cdot \prod c_q \cdot \sqrt{\#\text{Ш}(E/K)})'$.*

Proof. The conjectures we're assuming allow us to use Proposition 7.6 and hence take $t = T$ in Lemma 7.3. This proves the theorem. □

Conclusion: By Proposition 2.4, if W_p' has maximal index in $E(K)'$, then imply that we have an equality

$$\frac{L^{(r)}(E,1)}{r!} = \frac{\|\omega\|^2}{c \cdot \sqrt{|D|}} \cdot \text{Reg}(W_p),$$

up to powers of primes not in $B(E)$. Thus the W_p' of maximal index satisfy this generalized Gross-Zagier formula.

Conjecture 7.8. *If $W \subset E(K)$ is* any *Gross-Zagier subgroup of index ℓ^{w_ℓ}, then there exists an inert prime p such that W_p' equals W'.*

8 Existence of Gross-Zagier Subgroups

Let E, K, etc., be as in Section 1.1, and let

$$t = c \cdot \prod_{q|N} c_q \cdot \sqrt{\#\text{Ш}(E/K)_{\text{an}}}.$$

In this section we investigate the analogue of the conjectures on pages 311–312 of [GZ86]. In particular, the existence of any Gross-Zagier subgroup for $E(K)$ combined with the BSD conjecture implies that $\#E(K)_{\text{tor}} \mid t$. The main theorem of [GZ86] thus led Gross-Zagier to make the following conjecture.

Conjecture 8.1 (Gross-Zagier)**.** *If $E(K)$ has rank 1, then the integer t is divisible by $\#E(\mathbb{Q})_{\text{tor}}$.*

Proposition 8.2. *Assume the BSD formula. If there exists any subgroup W of $E(K)$ such that the generalized Gross-Zagier formula (5) holds for W, then $\#E(K)_{\text{tor}} \mid t$. Note that we do not assume W is torsion free.*

Proof. Let W be such a subgroup. Arguing as in the proof of Proposition 2.4, we see that

$$\#E(K)_{\text{tor}}^2 \cdot (\text{Reg}(W)/\text{Reg}(E/K)) = c^2 \cdot \left(\prod_{q|N} c_q\right)^2 \cdot \text{Ш}_{\text{an}} = t^2.$$

The quotient $\text{Reg}(W)/\text{Reg}(E/K)$ is a square integer, so taking square roots of both sides yields the claim.

□

Because of Proposition 8.2, we view the divisibility $\#E(K)_{\mathrm{tor}} \mid t$ as a sort of *"litmus test"* for whether there could be a generalization of the Gross-Zagier formula in general. First, we observe that the most naive generalization of Conjecture 8.1 to higher rank is *false* (!), as the following example shows.

Example 8.3. *Let E be the curve 65a of rank 1 over $\mathbb{Q}$ given by $y^2 + xy = x^3 - x$ and let $D = -56$. Then $\#\text{Ш}_{\mathrm{an}}(E/K) = \prod c_q = c = 1$, so $t = 1$, but $\#E(\mathbb{Q})_{\mathrm{tor}} = 2$. Here $E^D(\mathbb{Q})$ has rank 2, so $\mathrm{rank}(E(K)) = 3$, and the rank hypothesis of Conjecture 8.1 is not satisfied.*

Proposition 8.4. *Suppose $\mathrm{rank}(E(\mathbb{Q})) > 0$ and that t is a positive integer. Then there exists a Gross-Zagier subgroup $W \subset E(K)$ if and only if $\#E(K)_{\mathrm{tor}} \mid t$.*

Proof. Suppose $W \subset E(K)$ is a Gross-Zagier subgroup. Then $[E(K) : W] = t$. By hypothesis W is torsion free, so $E(K)_{\mathrm{tor}} \hookrightarrow E(K)/W$, so $\#E(K)_{\mathrm{tor}} \mid \#(E(K)/W) = t$.

Conversely, suppose that $\#E(K)_{\mathrm{tor}} \mid t$, and note that by hypothesis $E(\mathbb{Q})$ has positive rank. The group $E(K)/(E^D(\mathbb{Q}) + E(K)_{\mathrm{tor}})$ is thus a finitely generated infinite abelian group, so has subgroups of all index. In particular, it has a subgroup W' such that the quotient by W' is cyclic of order $t/\#E(K)_{\mathrm{tor}}$. Let $\tilde{W}$ be the inverse image of W' in $E(K)$, so $E(K)_{\mathrm{tor}}, E^D(\mathbb{Q}) \subset \tilde{W}$, and $[E(K) : \tilde{W}] = t/\#E(K)_{\mathrm{tor}}$. Since $\tilde{W}$ is finitely generated, there exists a torsion free subgroup $W \subset \tilde{W}$ such that $W \oplus E(K)_{\mathrm{tor}} = \tilde{W}$. Then

$$[E(K) : W] = \#E(K)_{\mathrm{tor}} \cdot [E(K) : \tilde{W}] = \#E(K)_{\mathrm{tor}} \cdot \frac{t}{\#E(K)_{\mathrm{tor}}} = t.$$

□

Elsewehere in this paper, for technical reasons in order to apply Kolyvagin's theorems, we made a minimality hypothesis on $r_{\mathrm{an}}(E^D/\mathbb{Q})$, and based on extensive numerical data, we conjecture that this is the right hypothesis to guarantee the existence of Gross-Zagier subgroups $W \subset E(K)$.

Conjecture 8.5. *If $r_{\mathrm{an}}(E/\mathbb{Q}) > r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$, then $\#E(K)_{\mathrm{tor}} \mid t$. In particular, there exists a Gross-Zagier subgroup $W \subset E(K)$.*

We obtain evidence for Conjecture 8.5 using Sage[†] [S$^+$09, Creb, PAR], Cremona's tables [Crea], Proposition 8.4, and assuming the Birch and Swinnerton-Dyer conjecture. More precisely, we check that Conjecture 8.5 is "probably true" for every elliptic curve of rank ≥ 2 and conductor $\leq 130,000$ and the first three D that satisfy the Heegner hypothesis, except possibly for the triples $(E, D, \#E(K)_{\mathrm{tor}})$ in Table 1 where the computation of the conjectural order of $\#\text{Ш}(E/K)$ took too long.

Table 1: All triples up to conductor 130,000 where we did not yet verify Conjecture 8.5

$(8320e1, -191, 2), (9842d1, -223, 3), (9842d1, -255, 3), (9842d1, -447, 3), (74655j1, -251, 3),$
$(87680a1, -119, 2), (87680a1, -151, 2), (87680b1, -119, 2), (87680b1, -151, 2), (89465a1, -51, 2),$
$(89465a1, -59, 2), (89465a1, -71, 2), (95545b1, -191, 2), (95545b1, -219, 2), (104585b1, -139, 2),$
$(104585b1, -179, 2), (104585b1, -191, 2), (114260a1, -231, 2), (114260a1, -239, 2), (114260a1, -431, 2),$
$(122486a1, -103, 3), (122486a1, -55, 3), (122486a1, -87, 3), (126672r1, -335, 2), (126672r1, -647, 2),$
$(126672r1, -719, 2), (129940a1, -111, 2), (129940a1, -71, 2), (129940a1, -79, 2)$

In our computations, we considered the first three Heegner D, *without* making the condition $r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$. The conjecture is *false* without the hypothesis that $r_{\mathrm{an}}(E^D/\mathbb{Q}) \leq 1$, as Example 8.3 above shows. Moreover, we found two further similar examples in which, however, E has rank 2 and E^D has rank 3. First, for the curve E with Cremona label 20672m1, equation $y^2 = x^3 - 431x - 3444$ and $D = -127$, we have $\mathrm{rank}(E(\mathbb{Q})) = 2$, $\mathrm{rank}(E^D(\mathbb{Q})) = 3$, and $\#E(K)_{\mathrm{tor}} = 2$, but $t = 1$. A second example is E given by 18560c1 and $D = -151$, in which again $\mathrm{rank}(E(\mathbb{Q})) = 2$, $\mathrm{rank}(E^D(\mathbb{Q})) = 3$, $\#E(K)_{\mathrm{tor}} = 2$, but $t = 1$.

This was a large computation that relies on a range of nontrivial computer code, which we carried out as follows. First we computed $\#E(K)_{\mathrm{tor}}$ for each of the 78,420 elliptic curve of conductor $\leq$130,000 with rank ≥ 2 and the first three Heegner D. We then determined whether $\#E(K)_{\mathrm{tor}}$ divides $c \cdot \prod c_q$. Since we are verifying that something divides $c \cdot \prod c_q$, there is no loss at all in assuming Manin's conjecture that $c = 1$ for the optimal quotient of $X_0(N)$. We then computed the Manin constant c for non-optimal curves by finding a shortest isogeny path from the optimal curve in the isogeny graph of E (there is unfortunately a small possibility of error in computation of the isogeny graph, due to numerical precision used in the implementation). We found only 37 remaining curves E of rank ≥ 2 such that $\#E(K)_{\mathrm{tor}} \nmid c \cdot \prod c_q$, and $37 \cdot 3 = 111$ corresponding pairs (E, D). It

[†] Running on hardware purchased using National Science Foundation Grant No. DMS-0821725.

turns out that all of these curves are optimal hence have $c = 1$. For each of these pairs (E, D) we attempted to compute $\#Ш(E/K)_{\mathrm{an}}$ using Conjecture 2.2 and some results of [GJP+09], and the computation finished in all but 29 cases. The main difficulty was computing $\mathrm{Reg}(E/K)$ in terms of $\mathrm{Reg}(E/\mathbb{Q})$ and $\mathrm{Reg}(E^D/\mathbb{Q})$ by saturating the sum of $E(\mathbb{Q})$ and $E^D(\mathbb{Q})$ in $E(K)$. Computing $E^D(\mathbb{Q})$ was sometimes very difficult, since E^D has huge conductor and rank 1, and this sometimes took as long as a day when it completed. For more details, the reader is urged to read the source code of the Sage command `heegner_sha_an` in Sage-3.4.1 and later.

References

[ARS06] A. Agashe, K. A. Ribet, and W. A. Stein, *The Manin Constant*, JPAM Coates Volume (2006), `http://wstein.org/papers/ars-manin/`.

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[BFH90] Daniel Bump, Solomon Friedberg, and Jeffrey Hoffstein, *Nonvanishing theorems for L-functions of modular forms and their derivatives*, Invent. Math. **102** (1990), no. 3, 543–618. MR MR1074487 (92a:11058)

[Bir65] B. J. Birch, *Conjectures concerning elliptic curves*, Proceedings of Symposia in Pure Mathematics, VIII, Amer. Math. Soc., Providence, R.I., 1965, pp. 106–112. MR 30 #4759

[Crea] J. E. Cremona, *Elliptic Curves Data*, `http://www.warwick.ac.uk/~masgaj/ftp/data/`.

[Creb] ———, `mwrank` *(computer software)*,
`http://www.maths.nott.ac.uk/personal/jec/ftp/progs/`.

[GJP+09] G. Grigorov, A. Jorza, S. Patrikis, C. Patrascu, and W. Stein, *Verification of the Birch and Swinnerton-Dyer Conjecture for Specific Elliptic Curves*, To appear in Mathematics of Computation (2009).

[Gro91] B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 235–256.

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[JLS08] D. Jetchev, K. Lauter, and W. Stein, *Explicit heegner points: Kolyvagin's conjecture and nontrivial elements in the Shafarevich-Tate group.*

[Kol88] V. A. Kolyvagin, *Finiteness of* $E(\mathbf{Q})$ *and* $Ш(E, \mathbf{Q})$ *for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671. MR 89m:11056

[Kol91a] V. A. Kolyvagin, *On the structure of Selmer groups*, Math. Ann. **291** (1991), no. 2, 253–259. MR 93e:11073

[Kol91b] ———, *On the structure of Shafarevich-Tate groups*, Algebraic geometry (Chicago, IL, 1989), Springer, Berlin, 1991, pp. 94–121.

[McC91] W. G. McCallum, *Kolyvagin's work on Shafarevich-Tate groups*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, pp. 295–316. MR 92m:11062

[PAR] PARI, *A computer algebra system designed for fast computations in number theory*, `http://pari.math.u-bordeaux.fr/`.

[S+09] W. A. Stein et al., *Sage Mathematics Software (Version 3.3)*, The Sage Development Team, 2009, `http://www.sagemath.org`.

[Ser72] J-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331.

[Ste02] W. A. Stein, *There are genus one curves over* **Q** *of every odd index*, J. Reine Angew. Math. **547** (2002), 139–147. MR 2003c:11059

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

38 The Modular Degree, Congruence Primes and Multiplicity One, with A. Agashe and K. Ribet

The Modular Degree, Congruence Primes and Multiplicity One

Amod Agashe Kenneth A. Ribet William A. Stein

Abstract.

The modular degree and congruence number are two fundamental invariants of an elliptic curve over the rational field. Frey and Müller have asked whether these invariants coincide. Although this question has a negative answer, we prove a theorem about the relation between the two invariants: one divides the other, and the ratio is divisible only by primes whose squares divide the conductor of the elliptic curve. We discuss the ratio even in the case where the square of a prime does divide the conductor, and we study analogues of the two invariants for modular abelian varieties of arbitrary dimension.

1 Introduction

Let E be an elliptic curve over $\mathbf{Q}$. By [BCDT01], we may view E as an abelian variety quotient over $\mathbf{Q}$ of the modular Jacobian $J_0(N)$, where N is the conductor of E. After possibly replacing E by an isogenous curve, we may assume that the kernel of the map $J_0(N) \to E$ is connected, i.e., that E is an *optimal quotient* of $J_0(N)$.

Let $f_E = \sum a_n q^n \in S_2(\Gamma_0(N))$ be the newform attached to E. The *congruence number* r_E of E is the largest integer such that there is an element $g = \sum b_n q^n \in S_2(\Gamma_0(N))$ with integer Fourier coefficients b_n that is orthogonal to f_E with respect to the Peterson innner product, and congruent to f_E modulo r_E (i.e., $a_n \equiv b_n \pmod{r_E}$ for all n). The *modular degree* m_E is the degree of the composite map $X_0(N) \to J_0(N) \to E$, where we map $X_0(N)$ to $J_0(N)$ by sending $P \in X_0(N)$ to $[P] - [\infty] \in J_0(N)$.

Section 2 is about relations between r_E and m_E. For example, $m_E \mid r_E$. In [FM99, Q. 4.4], Frey and Müller asked whether $r_E = m_E$. We give examples in which $r_E \neq m_E$, then conjecture that for any prime p, $\mathrm{ord}_p(r_E/m_E) \leq \frac{1}{2}\mathrm{ord}_p(N)$. We prove this conjecture when $\mathrm{ord}_p(N) \leq 1$.

In Section 3, we consider analogues of congruence primes and the modular degree for optimal quotients that are not necessarily elliptic curves; these are quotients of $J_0(N)$ and $J_1(N)$ of any dimension associated to ideals of the relevant Hecke algebras. In Section 4 we prove the main theorem of this paper, and in Section 5 we give some new examples of failure of multiplicity one motivated by the arguments in Section 4.

Acknowledgment. The authors are grateful to A. Abbes, R. Coleman, B. Conrad, J. Cremona, H. Lenstra, E. de Shalit, B. Edixhoven, L. Merel, and R. Taylor for several discussions and advice regarding this paper.

2 Congruence Primes and the Modular Degree

Let N be a positive integer and let $X_0(N)$ be the modular curve over $\mathbf{Q}$ that classifies isomorphism classes of elliptic curves with a cyclic subgroup of order N. The Hecke algebra $\mathbf{T}$ of level N is the subring of the ring of endomorphisms of $J_0(N) = \mathrm{Jac}(X_0(N))$ generated by the Hecke operators T_n for all $n \geq 1$. Let f be a newform of weight 2 for $\Gamma_0(N)$ with integer Fourier coefficients, and let I_f be kernel of the homomorphism $\mathbf{T} \to \mathbf{Z}[\ldots, a_n(f), \ldots]$ that sends T_n to a_n. Then the quotient $E = J_0(N)/I_f J_0(N)$ is an elliptic curve over $\mathbf{Q}$. We call E the *optimal quotient* associated to f. Composing the embedding $X_0(N) \hookrightarrow J_0(N)$ that sends ∞ to 0 with the quotient map $J_0(N) \to E$, we obtain a surjective morphism of curves $\phi_{E} : X_0(N) \to E$.

Definition 2.1. The *modular degree m_{E}* of E is the degree of ϕ_{E}.

Congruence primes have been studied by Doi, Hida, Ribet, Mazur and others (see, e.g., [Rib83, §1]), and played an important role in Wiles's work [Wil95] on Fermat's last theorem. Frey and Mai-Murty have observed that an appropriate asymptotic bound on the modular degree is equivalent to the *abc*-conjecture (see [Fre97, p.544] and [Mur99, p.180]). Thus, results that relate congruence primes and the modular degree are of great interest.

Theorem 2.2. *Let E be an elliptic curve over $\mathbf{Q}$ of conductor N, with modular degree m_{E} and congruence number r_{E}. Then $m_{E} \mid r_{E}$ and if $\mathrm{ord}_p(N) \leq 1$ then $\mathrm{ord}_p(r_{E}) = \mathrm{ord}_p(m_{E})$.*

We will prove a generalization of Theorem 2.2 in Section 4 below.

The divisibility $m_{E} \mid r_{E}$ was first discussed in [Zag85, Th. 3], where it is attributed to the second author (Ribet); however in [Zag85] the divisibility was mistakenly written in the opposite direction. For some other expositions of the proof, see [AU96, Lem 3.2] and [CK04]. We generalize this divisibility in Proposition 4.5. The second part of Theorem 2.2, i.e., that if $\mathrm{ord}_p(N) \leq 1$ then $\mathrm{ord}_p(r_{E}) = \mathrm{ord}_p(m_{E})$, follows from the more general Theorem 3.7 below. Note that [AU96, Prop. 3.3–3.4] implies the weaker statement that if $p \nmid N$ then $\mathrm{ord}_p(r_{E}) = \mathrm{ord}_p(m_{E})$, since [AU96, Prop. 3.3] implies

$$\mathrm{ord}_p(r_{E}) - \mathrm{ord}_p(m_{E}) = \mathrm{ord}_p(\#\mathcal{C}) - \mathrm{ord}_p(c_{E}) - \mathrm{ord}_p(\#\mathcal{D}),$$

Table 1: Differing Modular Degree and Congruence Number

Curve	m_E	r_E
54B1	2	6
64A1	2	4
72A1	4	8
80A1	4	8
88A1	8	16
92B1	6	12
96A1	4	8
96B1	4	8
99A1	4	12
108A1	6	18
112A1	8	16
112B1	4	8
112C1	8	16
120A1	8	16
124A1	6	12
126A1	8	24
128A1	4	32
128B1	8	32
128C1	4	32
128D1	8	32
135A1	12	36
144A1	4	8
144B1	8	16

and by [AU96, Prop. 3.4] $\mathrm{ord}_p(\#\mathcal{C}) = 0$. (Here c_E is the Manin constant of E, which is an integer by results of Edixhoven and Katz-Mazur; see e.g., [ARS06] for more details.)

Frey and Müller [FM99, Ques. 4.4] asked whether $r_E = m_E$ in general. After implementing an algorithm to compute r_E in Magma [BCP97], we quickly found that the answer is no. The counterexamples at conductor $N \leq 144$ are given in Table 1, where the curve is given using the notation of [Cre97]:

For example, the elliptic curve 54B1, given by the equation $y^2 + xy + y = x^3 - x^2 + x - 1$, has $r_E = 6$ and $m_E = 2$. To see explicitly that $3 \mid r_E$, observe that the newform corresponding to E is $f = q + q^2 + q^4 - 3q^5 - q^7 + \cdots$ and the newform corresponding to $X_0(27)$ if $g = q - 2q^4 - q^7 + \cdots$, so $g(q) + g(q^2)$ appears to be congruent to f modulo 3. To prove this congruence, we checked it for 18 Fourier coefficients, where the sufficiency of precision to degree 18 was determined using [Stu87].

In our computations, there appears to be no absolute bound on the p that occur. For example, for the curve 242B1 of conductor $N = 2 \cdot 11^2$ we have[1]

$$m_E = 2^4 \neq r_E = 2^4 \cdot 11.$$

We propose the following replacement for Question 4.4 of [FM99]:

Conjecture 2.3. *Let E be an optimal elliptic curve of conductor N and p be any prime. Then*

$$\mathrm{ord}_p\left(\frac{r_E}{m_E}\right) \leq \frac{1}{2}\,\mathrm{ord}_p(N).$$

We verified Conjecture 2.3 using Magma for every optimal elliptic curve quotient of $J_0(N)$, with $N \leq 539$.

If $p \geq 5$ then $\mathrm{ord}_p(N) \leq 2$, so a special case of the conjecture is

$$\mathrm{ord}_p\left(\frac{r_E}{m_E}\right) \leq 1 \qquad \text{for any } p \geq 5.$$

[1] The curve 242a1 in "modern notation."

Remark 2.4. It is often productive to parametrize elliptic curves by $X_1(N)$ instead of $X_0(N)$ (see, e.g., [Ste89] and [Vat05]). Suppose E is an optimal quotient of $X_1(N)$, let m'_E be the degree of the modular parametrization, and let r'_E be the $\Gamma_1(N)$-congruence number, which is defined as above but with $S_2(\Gamma_0(N))$ replaced by $S_2(\Gamma_1(N))$. For the optimal quotient of $X_1(N)$ isogenous to 54B1, we find using Magma that $m'_E = 18$ and $r'_E = 6$. Thus the equality $m'_E = r'_E$ fails, and the analogous divisibility $m'_E \mid r'_E$ no longer holds. Also, for a curve of conductor 38 we have $m'_E = 18$ and $r'_E = 6$, so equality need not hold even if the level is square free. We hope to investigate this in a future paper.

3 Modular abelian varieties of arbitrary dimension

For $N \geq 4$, let Γ be a fixed choice of either $\Gamma_0(N)$ or $\Gamma_1(N)$, let X be the modular curve over $\mathbf{Q}$ associated to Γ, and let J be the Jacobian of X. Let I be a *saturated* ideal of the corresponding Hecke algebra $\mathbf{T} \subset \mathrm{End}(J)$, so $\mathbf{T}/I$ is torsion free. Then $A = A_I = J/IJ$ is an optimal quotient of J since IJ is an abelian subvariety.

Definition 3.1. If $f = \sum a_n(f)q^n \in S_2(\Gamma)$ and $I_f = \ker(\mathbf{T} \to \mathbf{Z}[\ldots, a_n(f), \ldots])$, then $A = A_f = J/I_f J$ is the *newform quotient* associated to f. It is an abelian variety over $\mathbf{Q}$ of dimension equal to the degree of the field $\mathbf{Q}(\ldots, a_n(f), \ldots)$.

In this section, we generalize the notions of the congruence number and the modular degree to quotients $A = A_I$, and state a theorem relating the two numbers, which we prove in Sections 4.1–4.2.

Let ϕ_2 denote the quotient map $J \to A$. By Poincare reducibility over $\mathbf{Q}$ there is a unique abelian subvariety $A^\vee$ of J that projects isogenously to the quotient A (equivalently, which has finite intersection with $\ker(\phi_2)$), and so by Hecke equivariance of $J \to A$ it follows that $A^\vee$ is $\mathbf{T}$-stable. Let ϕ be the composite isogeny

$$\phi : A^\vee \xrightarrow{\phi_1} J \xrightarrow{\phi_2} A.$$

Remark 3.2. Note that $A^\vee$ is the dual abelian variety of A. More generally, if C is any abelian variety, let $C^\vee$ denote the dual of C. There is a canonical principal polarization $J \cong J^\vee$, and dualizing ϕ_2, we obtain a map $\phi_2^\vee : A^\vee \to J^\vee$, which we compose with $\theta^{-1} : J^\vee \cong J$ to obtain a map $\phi_1 : A^\vee \to J$. Note also that φ is a polarization (induced by pullback of the theta divisor).

The *exponent* of a finite group G is the smallest positive integer n such that every element of G has order dividing n.

Definition 3.3. The *modular exponent* of A is the exponent of the kernel of the isogeny ϕ, and the *modular number* of A is the degree of ϕ.

We denote the modular exponent of A by $\tilde{n}_{A}$ and the modular number by n_{A}. When A is an elliptic curve, the modular exponent is equal to the modular degree of A, and the modular number is the square of the modular degree (see, e.g., [AU96, p. 278]).

If R is a subring of $\mathbf{C}$, let $S_2(R) = S_2(\Gamma; R)$ denote the subgroup of $S_2(\Gamma)$ consisting of cups forms whose Fourier expansions at the cusp ∞ have coefficients in R. (Note that Γ is fixed for this whole section.) Let $S_2(\Gamma;\mathbf{Z})[I]^{\perp}$ denote the orthogonal complement of $S_2(\Gamma;\mathbf{Z})[I]$ in $S_2(\Gamma;\mathbf{Z})$ with respect to the Petersson inner product.

The following is well known, but we had difficulty finding a good reference.

Proposition 3.4. *The group $S_2(\Gamma;\mathbf{Z})$ is of finite rank as a $\mathbf{Z}$-module.*

Proof. Using the standard pairing between $\mathbf{T}$ and $S_2(\Gamma,\mathbf{Z})$ (see also [Rib83, Theorem 2.2]) we see that $S_2(\Gamma,\mathbf{Z}) \cong \mathrm{Hom}(\mathbf{T},\mathbf{Z})$. Thus $S_2(\Gamma,\mathbf{Z})$ is finitely generated over $\mathbf{Z}$ if and only if $\mathbf{T}$ is finitely generated over $\mathbf{Z}$. But the action of $\mathbf{T}$ on $\mathrm{H}_1(J,\mathbf{Z})$ is a faithful representation that embeds $\mathbf{T}$ into $\mathrm{Mat}_{2d}(\mathbf{Z}) \cong \mathbf{Z}^{(2d)^2}$. But $\mathbf{Z}$ is Noetherian, so $\mathbf{T}$ is finitely generated over $\mathbf{Z}$. □

Definition 3.5. The exponent of the quotient group

$$\frac{S_2(\Gamma;\mathbf{Z})}{S_2(\Gamma;\mathbf{Z})[I] + S_2(\Gamma;\mathbf{Z})[I]^{\perp}} \tag{1}$$

is the *congruence exponent* $\tilde{r}_{A}$ of A and its order is the *congruence number* r_{A}.

Remark 3.6. Note that $S_2(\Gamma,\mathbf{Z}) \otimes_{\mathbf{Z}} R = S_2(\Gamma,R)$; see, e.g., the discussion in [DI95, §12]. Thus the analogue of Definition 3.5 with $\mathbf{Z}$ replaced by an algebraic integer ring (or even $\overline{\mathbf{Z}}$) gives a torsion module whose annihilator ideal meets $\mathbf{Z}$ in the ideal generated by the congruence exponent.

Our definition of r_{A} generalizes the definition in Section 2 when A is an elliptic curve (see [AU96, p. 276]), and the following generalizes Theorem 2.2:

Theorem 3.7. *If $f \in S_2(\mathbf{C})$ is a newform, then*

(a) We have $\tilde{n}_{A_f} \mid \tilde{r}_{A_f}$, and

(b) If $p^2 \nmid N$, then $\mathrm{ord}_p(\tilde{r}_{A_f}) = \mathrm{ord}_p(\tilde{n}_{A_f})$.

Remark 3.8. When A_f is an elliptic curve, Theorem 3.7 implies that the modular degree divides the congruence number (since for an elliptic curve the modular degree and modular exponent are the same), i.e., $\sqrt{n_{A_f}} \mid r_{A_f}$. In general, the divisibility $n_{A_f} \mid r_{A_f}^2$ need not hold. For example, there is a newform of degree 24 in $S_2(\Gamma_0(431))$ such that

$$n_{A_f} = (2^{11}\cdot 6947)^2 \nmid r_{A_f}^2 = (2^{10}\cdot 6947)^2.$$

Note that 431 is prime and mod 2 multiplicity one fails for $J_0(431)$ (see [Kil02]).

4 Proof of the Main Theorem

In this section we prove Theorem 3.7. We continue using the notation introduced so far.

4.1 Proof of Theorem 3.7 (a)

We begin with a remark about compatibilities. In general, the polarization of J induced by the theta divisor need not be Hecke equivariant, because if T is a Hecke operator on J, then on $J^\vee$ it acts as $W_N T W_N$, where W_N is the Atkin-Lehner involution (see e.g., [DI95, Rem. 10.2.2]). However, on J^{new} the action of the Hecke operators commutes with that of W_N, so if the quotient map $J \to A$ factors through J^{new}, then the Hecke action on $A^\vee$ induced by the embedding $A^\vee \to J^\vee$ and the action on $A^\vee$ induced by $\phi_1 : A^\vee \to J$ are the same. Hence $A^\vee$ is isomorphic to $\phi_1(A^\vee)$ as a $\mathbf{T}$-module.

Recall that f is a newform, $I_f = \mathrm{Ann}_{\mathbf{T}}(f)$, and $J = J_0(N)$. Let $B = I_f J$, so that $A^\vee + B = J$, and $J/B \cong A$. The following lemma is well known, but we prove it here for the convenience of the reader.

Lemma 4.1. $\mathrm{Hom}_{\mathbf{Q}}(A^\vee, B) = 0$.

Proof. Pick a prime ℓ. Then $\overline{\mathbf{Q}}_\ell \otimes V_\ell(J)^{\mathrm{ss}}$ as a $\overline{\mathbf{Q}}_\ell[G_{\mathbf{Q}}]$-module is a direct sum of copies of the representations ρ_g as g ranges through all normalized eigenforms of weight 2 and level N with coefficients in $\overline{\mathbf{Q}}$; by a well-known result of the second author, these representations are absolutely irreducible. Now since f is a newform and $A^\vee \to A$ is an isogeny, $\overline{\mathbf{Q}}_\ell \otimes V_\ell(A^\vee)^{\mathrm{ss}}$ is a direct sum of copies of $\rho_{\sigma(f)}$ as σ ranges over all embeddings of K_f into $\overline{\mathbf{Q}}$. Thus, by the analytic theory of multiplicity one (see [Li75, Cor. 3, pg. 300]), the Galois modules $V_\ell(A^\vee)$ and $V_\ell(B) = V_\ell(J)/V_\ell(A^\vee)$ share no common Jordan-Hölder factors even when coefficients are extended to $\overline{\mathbf{Q}}_\ell$, so $\mathrm{Hom}_{\mathbf{Q}}(A', B) = 0$. □

Let $\mathbf{T}_1$ be the image of $\mathbf{T}$ in $\mathrm{End}(A^\vee)$, and let $\mathbf{T}_2$ be the image of $\mathbf{T}$ in $\mathrm{End}(B)$. We have the following commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & \mathbf{T} \arrow[r] \arrow[d] & \mathbf{T}_1 \oplus \mathbf{T}_2 \arrow[r] \arrow[d] & \dfrac{\mathbf{T}_1 \oplus \mathbf{T}_2}{\mathbf{T}} \arrow[r] \arrow[d, dotted] & 0 \\
0 \arrow[r] & \mathrm{End}(J) \arrow[r] & \mathrm{End}(A^\vee) \oplus \mathrm{End}(B) \arrow[r] & \dfrac{\mathrm{End}(A^\vee) \oplus \mathrm{End}(B)}{\mathrm{End}(J)} \arrow[r] & 0.
\end{tikzcd} \tag{2}$$

Let

$$e = (1, 0) \in \mathbf{T}_1 \oplus \mathbf{T}_2,$$

and let e_1 and e_2 denote the images of e in the groups $(\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}$ and $(\mathrm{End}(A^\vee) \oplus \mathrm{End}(B))/\mathrm{End}(J)$, respectively. It follows from Lemma 4.1 that the two quotient groups on the right hand side of (2) are finite, so e_1 and e_2

have finite order. Note that because e_2 is the image of e_1, the order of e_2 is a divisor of the order of e_1.

The *denominator* of any $\varphi \in \mathrm{End}(J) \otimes \mathbf{Q}$ is the smallest positive integer n such that $n\varphi \in \mathrm{End}(J)$.

Let $\pi_{A^\vee}, \pi_B \in \mathrm{End}(J) \otimes \mathbf{Q}$ be projection onto $A^\vee$ and B, respectively. Note that the denominator of $\pi_{A^\vee}$ equals the denominator of π_B, since $\pi_{A^\vee} + \pi_B = 1_J$, so that $\pi_B = 1_J - \pi_{A^\vee}$.

Lemma 4.2. *The element $e_2 \in (\mathrm{End}(A^\vee) \oplus \mathrm{End}(B))/\mathrm{End}(J)$ defined above has order $\tilde{n}_A$.*

Proof. Let n be the order of e_2, so n is the denominator of $\pi_{A^\vee}$, which, as mentioned above, is also the denominator of π_B. We want to show that n is equal to $\tilde{n}_A$, the exponent of $A^\vee \cap B$.

Let $i_{A^\vee}$ and i_B be the embeddings of $A^\vee$ and B into J, respectively. Then

$$\varphi = (n\pi_{A^\vee}, n\pi_B) \in \mathrm{Hom}(J, A^\vee \times B)$$

and $\varphi \circ (i_{A^\vee} + i_B) = [n]_{A^\vee \times B}$. We have an exact sequence

$$0 \to A^\vee \cap B \xrightarrow{x \mapsto (x,-x)} A^\vee \times B \xrightarrow{i_{A^\vee} + i_B} J \to 0.$$

Let Δ be the image of $A^\vee \cap B$. Then by exactness,

$$[n]\Delta = (\varphi \circ (i_{A^\vee} + i_B))(\Delta) = \varphi \circ ((i_{A^\vee} + i_B)(\Delta)) = \varphi(\{0\}) = \{0\},$$

so n is a multiple of the exponent $\tilde{n}_A$ of $A^\vee \cap B$.

To show the opposite divisibility, consider the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & A^\vee \cap B \arrow[r, "x \mapsto (x,-x)"] \arrow[d, "{[\tilde{n}_A]}"] & A^\vee \times B \arrow[r] \arrow[d, "{([\tilde{n}_A],0)}"] & J \arrow[r] \arrow[d, dotted, "\psi"] & 0 \\
0 \arrow[r] & A^\vee \cap B \arrow[r, "x \mapsto (x,-x)"] & A^\vee \times B \arrow[r] & J \arrow[r] & 0,
\end{tikzcd}$$

where the middle vertical map is $(a, b) \mapsto (\tilde{n}_A a, 0)$ and the map ψ exists because $[\tilde{n}_A](A^\vee \cap B) = 0$. But $\psi = \tilde{n}_A \pi_{A^\vee}$ in $\mathrm{End}(J) \otimes \mathbf{Q}$. This shows that $\tilde{n}_A \pi_{A^\vee} \in \mathrm{End}(J)$, i.e., that $\tilde{n}_A$ is a multiple of the denominator n of $\pi_{A^\vee}$.

□

Let $\mathrm{Ext}^1 = \mathrm{Ext}^1_{\mathbf{Z}}$ denote the first Ext functor in the category of $\mathbf{Z}$-modules.

Lemma 4.3. *The group $(\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}$ is isomorphic to the quotient (1) in Definition 3.5, so $r_A = \#((\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T})$ and $\tilde{r}_A$ is the exponent of $(\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}$. More precisely, $\mathrm{Ext}^1((\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}, \mathbf{Z})$ is isomorphic as a $\mathbf{T}$-module to the quotient (1).*

Proof. Apply the $\mathrm{Hom}(-, \mathbf{Z})$ functor to the first row of (2) to obtain a three-term exact sequence

$$0 \to \mathrm{Hom}(\mathbf{T}_1 \oplus \mathbf{T}_2, \mathbf{Z}) \to \mathrm{Hom}(\mathbf{T}, \mathbf{Z}) \to \mathrm{Ext}^1((\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}, \mathbf{Z}) \to 0. \qquad (3)$$

There is a $\mathbf{T}$-equivariant bilinear pairing $\mathbf{T} \times S_2(\mathbf{Z}) \to \mathbf{Z}$ given by $(t, g) \mapsto a_1(t(g))$, which is perfect by [AU96, Lemma 2.1] (see also [Rib83, Theorem 2.2]). Using this pairing, we transform (3) into an exact sequence

$$0 \to S_2(\mathbf{Z})[I_f] \oplus S_2(\Gamma; \mathbf{Z})[I_f]^{\perp} \to S_2(\mathbf{Z}) \to \mathrm{Ext}^1((\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}, \mathbf{Z}) \to 0$$

of $\mathbf{T}$-modules. Here we use that $\mathrm{Hom}(\mathbf{T}_2, \mathbf{Z})$ is the unique saturated Hecke-stable complement of $S_2(\mathbf{Z})[I_f]$ in $S_2(\mathbf{Z})$, hence must equal $S_2(\mathbf{Z})[I_f]^{\perp}$. Finally note that if G is any finite abelian group, then $\mathrm{Ext}^1(G, \mathbf{Z}) \approx G$ as groups, which gives the desired result. □

LEMMA 4.4. *The element $e_1 \in (\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}$ has order $\tilde{r}_A$.*

Proof. By Lemma 4.3, the lemma is equivalent to the assertion that the order r of e_1 equals the exponent of $M = (\mathbf{T}_1 \oplus \mathbf{T}_2)/\mathbf{T}$. Since e_1 is an element of M, the exponent of M is divisible by r.

To obtain the reverse divisibility, consider any element x of M. Let $(a, b) \in \mathbf{T}_1 \oplus \mathbf{T}_2$ be such that its image in M is x. By definition of e_1 and r, we have $(r, 0) \in \mathbf{T}$, and since $1 = (1, 1) \in \mathbf{T}$, we also have $(0, r) \in \mathbf{T}$. Thus $(\mathbf{T}r, 0)$ and $(0, \mathbf{T}r)$ are both subsets of $\mathbf{T}$ (i.e., in the image of $\mathbf{T}$ under the map $\mathbf{T} \to \mathbf{T}_1 \oplus \mathbf{T}_2$), so $r(a, b) = (ra, rb) = (ra, 0) + (0, rb) \in \mathbf{T}$. This implies that the order of x divides r. Since this is true for every $x \in M$, we conclude that the exponent of M divides r. □

PROPOSITION 4.5. *If $f \in S_2(\mathbf{C})$ is a newform, then $\tilde{n}_{A_f} \mid \tilde{r}_{A_f}$.*

Proof. Since e_2 is the image of e_1 under the right-most vertical homomorphism in (2), the order of e_2 divides that of e_1. Now apply Lemmas 4.2 and 4.4. □

This finishes the proof of the first statement in Theorem 3.7.

4.2 PROOF OF THEOREM 3.7 (B)

Let $\mathbf{T}'$ be the saturation of $\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots]$ in $\mathrm{End}(J_0(N))$, i.e., the set of elements of $\mathrm{End}(J_0(N)) \otimes \mathbf{Q}$ some positive multiple of which lie in $\mathbf{T}$. The quotient $\mathbf{T}'/\mathbf{T}$ is a finitely generated abelian group because both $\mathbf{T}$ and $\mathrm{End}(J_0(N))$ are finitely generated over $\mathbf{Z}$. Since $\mathbf{T}'/\mathbf{T}$ is also a torsion group, it is finite.

In Section 4.2.1, we will give some conditions under which $\mathbf{T}$ and $\mathbf{T}'$ agree locally at maximal ideal of $\mathbf{T}$. In Section 4.2.2, we will explain how the ratio of the congruence number to the modular degree is closely related to the order of $\mathbf{T}'/\mathbf{T}$, and finally deduce that this ratio is 1 (for quotients associated to newforms) locally at a prime p such that $p^2 \nmid N$.

4.2.1 Multiplicity One

Fixt an integer N and a prime $p \mid N$. Suppose for a moment that N is prime, so $p = N$. In [Maz77], Mazur proves that $\mathbf{T} = \mathbf{T}'$; he combines this result with the equality

$$\mathbf{T} \otimes \mathbf{Q} = \mathrm{End}(J_0(p)) \otimes \mathbf{Q},$$

to deduce that $\mathbf{T} = \mathrm{End}(J_0(p))$. This result, combined with Ribet's result [Rib75] or [Rib81] to the effect that $\mathbf{T} \otimes \mathbf{Q} = (\mathrm{End}_{\overline{\mathbf{Q}}} J_0(N)) \otimes \mathbf{Q}$, shows that $\mathbf{T}$ is the full ring of endomorphisms of $J_0(N)$ over $\overline{\mathbf{Q}}$. When N is no longer necessarily prime, the method of [Maz77] shows that $\mathbf{T}$ and $\mathbf{T}'$ agree locally at a maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ that satisfies a simple condition involving differentials form mod ℓ, where ℓ is the residue characteristic of $\mathfrak{m}$.

For the sake of completeness, we state and prove a lemma that can be easily extracted from [Maz77]. Let m be the largest square dividing N and let $R = \mathbf{Z}[\frac{1}{m}]$. Let $X_0(N)_R$ denote the minimal regular model of $X_0(N)$ over R. Let $\Omega = \Omega_{X_0(N)/R}$ denote the sheaf of regular differentials on $X_0(N)_R$, as in [Maz78, §2(e)]. If ℓ is a prime such that $\ell^2 \nmid N$, then $X_0(N)_{\mathbf{F}_\ell}$ denotes the special fiber of $X_0(N)_R$ at the prime ℓ.

Lemma 4.6 (Mazur). *Let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ of residue characteristic ℓ such that $\ell^2 \nmid N$. Suppose that*

$$\dim_{\mathbf{T}/\mathfrak{m}} \mathrm{H}^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}] \leq 1.$$

Then $\mathbf{T}$ and $\mathbf{T}'$ agree locally at $\mathfrak{m}$.

Proof. Let M denote the group $H^1(X_0(N)_R, \mathcal{O}_{X_0(N)})$, where $\mathcal{O}_{X_0(N)}$ is the structure sheaf of $X_0(N)$. As explained in [Maz77, p. 95], we have an action of $\mathrm{End}_{\mathbf{Q}} J_0(N)$ on M, and the action of $\mathbf{T}$ on M via the inclusion $\mathbf{T} \subseteq \mathrm{End}_{\mathbf{Q}} J_0(N)$ is faithful, so likewise for the action by $\mathbf{T}'$. Hence we have an injection $\phi : \mathbf{T}' \hookrightarrow \mathrm{End}_{\mathbf{T}} M$. Suppose $\mathfrak{m}$ is a maximal ideal of $\mathbf{T}$ that satisfies the hypotheses of the lemma. To prove that $\mathbf{T}_{\mathfrak{m}} = \mathbf{T}'_{\mathfrak{m}}$ it suffices to prove the following claim:

Claim: The map $\phi|_{\mathbf{T}}$ is surjective locally at $\mathfrak{m}$.

Proof. By Nakayama's lemma, to show that M is generated as a single element over $\mathbf{T}$ locally at $\mathfrak{m}$, it suffices to check that the dimension of the $\mathbf{T}/\mathfrak{m}$-vector space $M/\mathfrak{m}M$ is at most one. Since $\ell^2 \nmid N$, $M/\mathfrak{m}M$ is dual to $H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}]$ (see, e.g., [Maz78, §2]). Since we are assuming that $\dim_{\mathbf{T}/\mathfrak{m}} H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}] \leq 1$, we have $\dim_{\mathbf{T}/\mathfrak{m}}(M/\mathfrak{m}M) \leq 1$, which proves the claim. □

□

If $\mathfrak{m}$ is a maximal ideal of the Hecke algebra $\mathbf{T}$ of residue characteristic ℓ, we say that $\mathfrak{m}$ satisfies *multiplicity one for differentials* if

$$\dim(\mathrm{H}^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}]) \leq 1.$$

By Lemma 4.6, multiplicity one for $\mathrm{H}^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}]$ implies that $\mathbf{T}$ and $\mathbf{T}'$ agree at $\mathfrak{m}$.

There is quite a bit of literature on the question of multiplicity 1 for $\mathrm{H}^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}]$. The easiest case is that ℓ is prime to the level N:

LEMMA 4.7 (Mazur). *If* $\mathfrak{m}$ *is a maximal ideal of* $\mathbf{T}$ *of residue characteristic* ℓ *such that* $\ell \nmid N$*, then*

$$\dim_{\mathbf{T}/\mathfrak{m}} \mathrm{H}^0(X_0(N)_{\mathbf{F}_\ell}, \Omega)[\mathfrak{m}] \leq 1.$$

Proof. Mazur deduces this lemma from injectivity of the q-expansion map. The reader may find the following alternative approach to part of the argument easier to follow than the one on p. 95 of [Maz77]. We have an $\mathbf{F}_\ell$-vector space that embeds in $\mathbf{F}_\ell[[q]]$, for example a space V of differentials that is killed by a maximal ideal $\mathfrak{m}$. This space is a $\mathbf{T}/\mathfrak{m}$-vector space, and we want to see that its dimension over $\mathbf{T}/\mathfrak{m}$ is at most 1. Mazur invokes tensor products and eigenvectors; alternatively, we note that V embeds in $\mathrm{Hom}_{\mathbf{F}_\ell}(\mathbf{T}/\mathfrak{m}, \mathbf{F}_\ell)$ via the standard duality that sends $v \in V$ to the linear form whose value on a Hecke operator T is the qth coefficient of $v|T$. The group $\mathrm{Hom}_{\mathbf{F}_\ell}(\mathbf{T}/\mathfrak{m}, \mathbf{F}_\ell)$ has the same size as $\mathbf{T}/\mathfrak{m}$, which completes the argument because $\mathrm{Hom}_{\mathbf{F}_\ell}(\mathbf{T}/\mathfrak{m}, \mathbf{F}_\ell)$ has dimension 1 as a $\mathbf{T}/\mathfrak{m}$-vector space. □

In the context of Mazur's paper, where the level N is prime, we see from Lemma 4.7 that $\mathbf{T}$ and $\mathbf{T}'$ agree away from N. Locally at N, Mazur proved that $\mathbf{T} = \mathbf{T}'$ by an analogue of the arguments that he used away from N; see Chapter II of [Maz77] (and especially Prop. 9.4 and 9.5 of that chapter) as well as [MR91], where these arguments are taken up in a context where the level is no longer necessarily prime (and where one works locally at a prime whose square does not divide the level). Thus in the prime level case, $\mathbf{T} = \mathbf{T}'$, as we asserted above.

Now let p be a prime such that $p \parallel N$, and let $M = N/p$. The question of multiplicity 1 at p for $\mathrm{H}^0(X_0(pM)_{\mathbf{F}_p}, \Omega)[\mathfrak{m}]$ is discussed in [MR91], where the authors establish multiplicity 1 for maximal ideals $\mathfrak{m} \mid p$ for which the associated mod p Galois representation is irreducible and *not* p-old. (A representation of level pM is p-old if it arises from $S_2(\Gamma_0(M))$.)

If $\mathfrak{m}$ is a maximal ideal of $\mathbf{T}$ of residue characteristic ℓ, then we say that $\mathfrak{m}$ is ordinary if $T_\ell \notin \mathfrak{m}$ (note that T_ℓ is often denoted U_ℓ if $\ell \mid N$). For our purposes, the following lemma is convenient:

LEMMA 4.8 (Wiles). *If* $\mathfrak{m}$ *is an ordinary maximal ideal of* $\mathbf{T}$ *of characteristic* p*, then*

$$\dim_{\mathbf{T}/\mathfrak{m}} \mathrm{H}^0(X_0(pM)_{\mathbf{F}_p}, \Omega)[\mathfrak{m}] \leq 1.$$

This is essentially Lemma 2.2 in [Wil95, pg. 485]; we make a few comments about how it applies on our situation:

1. Wiles considers $X_1(M,p)$ instead of $X_0(pM)$, which means that he is using $\Gamma_1(M)$-structure instead of $\Gamma_0(M)$-structure. This surely has no relevance to the issue at hand.

2. Wiles assumes (on page 480) that p is an odd prime, but again this assumption is not relevant to our question.

3. The condition that $\mathfrak{m}$ is ordinary does not appear explicitly in the statement of Lemma 2.2 in [Wil95]; instead it is a reigning assumption in the context of his discussion.

4. We see by example that Wiles's "ordinary" assumption is less stringent than the assumption in [MR91]; note that [MR91] rule out cases where $\mathfrak{m}$ is both old and new at p, whereas Wiles is happy to include such cases. (On the other hand, Wiles's assumption is certainly nonempty, since it rules out maximal ideals $\mathfrak{m}$ that arise from non-ordinary (old) forms of level M. Here is an example with $p = 2$ and $M = 11$, so $N = 22$: There is a unique newform $f = \sum a_n q^n$ of level 11, and $\mathbf{T} = \mathbf{Z}[T_2] \subset \mathrm{End}(J_0(22))$, where $T_2^2 - a_2 T_2 + 2 = 0$. Since $a_2 = -2$, we have $\mathbf{T} \cong \mathbf{Z}[\sqrt{-1}]$. We can choose the square root of -1 to be $T_2 + 1$. Then T_2 is a generator of the unique maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ with residue characteristic 2, and this maximal ideal is not ordinary.)

We now summarize the conclusions we can make from the lemmas so far:

Proposition 4.9. *The modules $\mathbf{T}$ and $\mathbf{T}'$ agree locally at each maximal ideal $\mathfrak{m}$ that is either prime to N or that satisfies the following supplemental hypothesis: the residue characteristic of $\mathfrak{m}$ divides N only to the first power and $\mathfrak{m}$ is ordinary.*

Proof. This follows easily from Lemmas 4.6, 4.7, and 4.8. □

In Mazur's original context, where the level N is prime, we have $T_N^2 = 1$ because there are no forms of level 1. Accordingly, each $\mathfrak{m}$ dividing N is ordinary, and we recover Mazur's equality $\mathbf{T} = \mathbf{T}'$ in this special case.

4.2.2 Degrees and Congruences

Let $e \in \mathbf{T} \otimes \mathbf{Q}$ be as in Section 4.1, and let p, N, M be as before Lemma 4.8. The image of e in $J_0(pM)$ is the $\mathbf{T}$-stable abelian subvariety denoted $A^\vee$ in Section 4.1, but since we shall now exclusively work with this subvariety rather than the corresponding optimal quotient of $J_0(pM)$ (which was denoted A earlier), we will now write A to denote the image of e (without risk of confusion). We also write B to denote the unique $\mathbf{T}$-stable abelian subvariety of $J_0(pM)$ complementary to A.

For $t \in \mathbf{T}$, let t_A be the restriction of t to A, and let t_B be the image of t in $\mathrm{End}(B)$. Let $\mathbf{T}_A$ be the subgroup of $\mathrm{End}(A)$ consisting of the various t_A, and define $\mathbf{T}_B$ similarly. As before, we obtain an injection $j : \mathbf{T} \hookrightarrow \mathbf{T}_A \times \mathbf{T}_B$ with finite cokernel. Because j is an injection, we refer to the maps $\pi_A : \mathbf{T} \to \mathbf{T}_A$ and $\pi_B : \mathbf{T} \to \mathbf{T}_B$, given by $t \mapsto t_A$ and $t \mapsto t_B$, respectively, as "projections".

Definition 4.10. The *congruence ideal* associated with the projector e is $I = \pi_A(\ker(\pi_B)) \subset \mathbf{T}_A$.

Viewing $\mathbf{T}_A$ as $\mathbf{T}_A \times \{0\}$, we may view $\mathbf{T}_A$ as a subgroup of $\mathbf{T} \otimes \mathbf{Q} \cong (\mathbf{T}_A \times \mathbf{T}_B) \otimes \mathbf{Q}$. Also, we may view $\mathbf{T}$ as embedded in $\mathbf{T}_A \times \mathbf{T}_B$, via the map j.

Lemma 4.11. *We have* $I = \mathbf{T}_A \cap \mathbf{T}$.

A larger ideal of $\mathbf{T}_A$ is $J = \mathrm{Ann}_{\mathbf{T}_A}(A \cap B)$; it consists of restrictions to A of Hecke operators that vanish on $A \cap B$.

Lemma 4.12. *We have* $I \subset J$.

Proof. The image in $\mathbf{T}_A$ of an operator that vanishes on B also vanishes on $A \cap B$. □

Lemma 4.13. *We have* $J = \mathbf{T}_A \cap \mathrm{End}(J_0(pM)) = \mathbf{T}_A \cap \mathbf{T}'$.

Proof. This is elementary; it is an analogue of Lemma 4.11. □

Proposition 4.14. *There is a natural inclusion* $J/I \hookrightarrow \mathbf{T}'/\mathbf{T}$ *of* $\mathbf{T}$*-modules.*

Proof. Consider the map $\mathbf{T} \to \mathbf{T} \otimes \mathbf{Q}$ given by $t \mapsto te$. This homomorphism factors through $\mathbf{T}_A$ and yields an injection $\iota_A : \mathbf{T}_A \hookrightarrow \mathbf{T} \otimes \mathbf{Q}$. Symmetrically, we also obtain $\iota_B : \mathbf{T}_B \hookrightarrow \mathbf{T} \otimes \mathbf{Q}$. The map $(t_A, t_B) \mapsto \iota_A(t_A) + \iota_B(t_B)$ is an injection $\mathbf{T}_A \times \mathbf{T}_B \hookrightarrow \mathbf{T} \otimes \mathbf{Q}$. The composite of this map with the inclusion $j : \mathbf{T} \hookrightarrow \mathbf{T}_A \times \mathbf{T}_B$ defined above is the natural map $\mathbf{T} \hookrightarrow \mathbf{T} \otimes \mathbf{Q}$. We thus have a sequence of inclusions

$$\mathbf{T} \hookrightarrow \mathbf{T}_A \times \mathbf{T}_B \hookrightarrow \mathbf{T} \otimes \mathbf{Q} \subset \mathrm{End}(J_0(pM)) \otimes \mathbf{Q}.$$

By Lemma 4.11 and Lemma 4.13, we have $I = \mathbf{T}_A \cap \mathbf{T}$ and $J = \mathbf{T}_A \cap \mathbf{T}'$. Thus $I = J \cap \mathbf{T}$, where the intersection is taken inside $\mathbf{T}'$. Thus

$$J/I = J/(J \cap \mathbf{T}) \cong (J + \mathbf{T})/\mathbf{T} \hookrightarrow \mathbf{T}'/\mathbf{T}.$$

□

Corollary 4.15. *If* $\mathfrak{m}$ *is a maximal ideal not in* $\mathrm{Supp}_{\mathbf{T}}(\mathbf{T}'/\mathbf{T})$*, then* $\mathfrak{m}$ *is not in the support of* J/I*, i.e., if* $\mathbf{T}$ *and* $\mathbf{T}'$ *agree locally at* $\mathfrak{m}$*, then* I *and* J *also agree locally at* $\mathfrak{m}$*.*

Note that the Hecke algebra $\mathbf{T}$ acts on J/I through its quotient $\mathbf{T}_A$, since the action of $\mathbf{T}$ on I and on J factors through this quotient.

Now we specialize to the case where A is ordinary at p, in the sense that the image of T_p in $\mathbf{T}_A$, which we denote $T_{p,A}$, is invertible modulo every maximal ideal of $\mathbf{T}_A$ that divides p. (This case occurs when A is a subvariety of the p-new subvariety of $J_0(pM)$, since the square of $T_{p,A}$ is the identity.)

If $\mathfrak{m} \mid p$ is a maximal ideal of $\mathbf{T}$ that arises by pullback from a maximal ideal of $\mathbf{T}_A$, then $\mathfrak{m}$ is ordinary in the sense used above. When A is ordinary at p, it follows from Proposition 4.9 and Corollary 4.15 that $I = J$ locally at p. The reason is simple: regarding I and J as $\mathbf{T}_A$-modules, we realize that we need to test that $I = J$ at maximal ideals of $\mathbf{T}_A$ that divide p. These ideals correspond to maximal ideals $\mathfrak{m} \mid p$ of $\mathbf{T}$ that are automatically ordinary, so we have $I = J$ locally at $\mathfrak{m}$ because of Proposition 4.9. By Proposition 4.9, we have $\mathbf{T} = \mathbf{T}'$ locally at primes away from the level pM. Thus we conclude that $I = J$ locally at all primes $\ell \nmid pM$ and also at p, a prime that divides the level pM exactly once.

Suppose, finally, that A is the abelian variety associated to a newform f of level pM. The ideal $I \subset \mathbf{T}_A$ measures congruences between f and the space of forms in $S_2(\Gamma_0(pM))$ that are orthogonal to the space generated by f. Also, $A \cap B$ is the kernel in A of the map "multiplication by the modular element e". In this case, the inclusion $I \subset J$ corresponds to the divisibility $\tilde{n}_A \mid \tilde{r}_A$, and we have equality at primes at which $I = J$ locally. We conclude that the congruence exponent and the modular exponent agree both at p and at primes not dividing pM, which completes our proof of Theorem 3.7(b).

Remark 4.16. The ring

$$R = \mathrm{End}(J_0(pM)) \cap (\mathbf{T}_A \times \mathbf{T}_B)$$

is often of interest, where the intersection is taken in $\mathrm{End}(J_0(pM)) \otimes \mathbf{Q}$. We proved above that there is a natural inclusion $J/I \hookrightarrow \mathbf{T}'/\mathbf{T}$. This inclusion yields an isomorphism $J/I \xrightarrow{\sim} R/\mathbf{T}$. Indeed, if (t_A, u_B) is an endomorphism of $J_0(pM)$, where $t, u \in \mathbf{T}$, then $(t_A, u_B) - u = (t_A, 0)$ is an element of J. The ideals I and J are equal to the extent that the rings $\mathbf{T}$ and R coincide. Even when $\mathbf{T}'$ is bigger than $\mathbf{T}$, its subring R may be not far from $\mathbf{T}$.

5 Failure of Multiplicity One

In this section, we discuss examples of failure of multiplicity one (in two different but related senses). The notion of multiplicity one, originally due to Mazur [Maz77], has played an important role in several places (e.g., in Wiles's proof of Fermat's last theorem [Wil95]). This notion is closely related to Gorensteinness of certain Hecke algebras (e.g., see [Til97]). Kilford [Kil02] found examples of failure of Gorensteinness (and multiplicity one) at the prime 2 for certain prime levels. Motivated by the arguments in Section 4, in this section we give examples of failure of multiplicity one for primes (including odd primes) whose square divides the level.

5.1 Multiplicity One for Differentials

In connection with the arguments in Section 4, especially Lemmas 4.6 and 4.8, it is of interest to compute the index $[\mathbf{T}' : \mathbf{T}]$ for various N. We can compute this index in Magma, e.g., the following commands compute the index for $N = 54$: “`J := JZero(54); T := HeckeAlgebra(J); Index(Saturation(T), T);`” We obtain Table 2, where the first column contains N and the second column contains $[\mathbf{T}' : \mathbf{T}]$:

Let $\mathfrak{m}$ be a maximal ideal of the Hecke algebra $\mathbf{T} \subset \mathrm{End}(J_0(N))$ of residue characteristic p. Recall that we say that $\mathfrak{m}$ satisfies *multiplicity one for differentials* if $\dim(\mathrm{H}^0(X_0(N)_{\mathbf{F}_p}, \Omega)[\mathfrak{m}]) \leq 1$.

In each case in which $[\mathbf{T}' : \mathbf{T}] \neq 1$, Lemma 4.6 implies that there is some maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ such that $\dim(\mathrm{H}^0(X_0(N)_{\mathbf{F}_p}, \Omega)[\mathfrak{m}]) > 1$, which is an example of failure of multiplicity one for differentials.

In Table 2, whenever $p \mid [\mathbf{T}' : \mathbf{T}]$, then $p^2 \mid 2N$. This is a consequence of Proposition 4.9, which moreover asserts that when 2 exactly divides N and $2 \mid [\mathbf{T}' : \mathbf{T}]$ then there is a non-ordinary (old) maximal ideal of characteristic 2 in the support of $\mathbf{T}'/\mathbf{T}$.

Moreover, notice that Theorem 3.7(b) (whose proof is in Section 4.2) follows formally from two key facts: that A_f is new and that multiplicity one for differentials holds for ordinary maximal ideals with residue characteristic $p \mid\mid N$ and for all maximal ideals with residue characteristic $p \nmid N$. The conclusion of Theorem 3.7(b) does not hold for the counterexamples in Section 2 (e.g., for 54B1), which are all new elliptic curves, so multiplicity one for differentials does not hold for certain maximal ideals that arise from the new quotient of the Hecke algebra. Note that in all examples we have $p \mid (r/m)$ with $p^2 \mid N$, which raises the question: are there non-ordinary counterexamples with $p \mid\mid N$?

5.2 Multiplicity One for Jacobians

We say that a maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ satisfies *multiplicity one* if $J_0(N)[\mathfrak{m}]$ is of dimension two over $\mathbf{T}/\mathfrak{m}$. We sometimes use the phrase “multiplicity one for $J_0(N)$” in order to distinguish this notion from the notion of multiplicity one for differentials.

Proposition 5.1. *Suppose E is an optimal elliptic curve over $\mathbf{Q}$ of conductor N and p is a prime such that $p \mid r_E$ but $p \nmid m_E$. Let $\mathfrak{m}$ be the annihilator in $\mathbf{T}$ of $E[p]$. Then multiplicity one fails for $\mathfrak{m}$, i.e., $\dim_{\mathbf{T}/\mathfrak{m}} J_0(N)[\mathfrak{m}] > 2$.*

Proof. Using the principal polarization $E \cong E^\vee$ we view E as an abelian subvariety of $J = J_0(N)$ and consider the complementary $\mathbf{T}$-stable abelian subvariety A of E (thus A is the kernel of the modular parametrization map $J \to E$). In this setup, $J = E + A$, and the intersection of E and A is $E[m_E]$. Here we use that the composite map $E \simeq E^\vee \to J^\vee \to J \to E$ is a polarization, and hence is multiplication by a positive integer m_E. Because $p \nmid m_E$, we have $E[p] \cap A = 0$. On the other hand, let $\mathfrak{m}$ be the annihilator of $E[p]$ inside $\mathbf{T}$.

Table 2: The Index $[\mathbf{T}' : \mathbf{T}]$

11	1	51	1	91	1	131	1	171	9
12	1	52	1	92	16	132	8	172	8
13	1	53	1	93	1	133	1	173	1
14	1	54	3	94	4	134	1	174	4
15	1	55	1	95	1	135	27	175	5
16	1	56	2	96	8	136	16	176	512
17	1	57	1	97	1	137	1	177	1
18	1	58	1	98	1	138	4	178	1
19	1	59	1	99	9	139	1	179	1
20	1	60	2	100	1	140	8	180	72
21	1	61	1	101	1	141	1	181	1
22	1	62	2	102	1	142	8	182	1
23	1	63	1	103	1	143	1	183	1
24	1	64	2	104	4	144	32	184	1024
25	1	65	1	105	1	145	1	185	1
26	1	66	1	106	1	146	1	186	4
27	1	67	1	107	1	147	7	187	1
28	1	68	2	108	54	148	4	188	256
29	1	69	1	109	1	149	1	189	243
30	1	70	1	110	2	150	5	190	8
31	1	71	1	111	1	151	1	191	1
32	1	72	2	112	8	152	32	192	4096
33	1	73	1	113	1	153	9	193	1
34	1	74	1	114	1	154	1	194	1
35	1	75	1	115	1	155	1	195	1
36	1	76	2	116	4	156	32	196	14
37	1	77	1	117	1	157	1	197	1
38	1	78	2	118	2	158	4	198	81
39	1	79	1	119	1	159	1	199	1
40	1	80	4	120	32	160	256	200	80
41	1	81	1	121	1	161	1	201	1
42	1	82	1	122	1	162	81	202	1
43	1	83	1	123	1	163	1	203	1
44	2	84	2	124	16	164	8	204	32
45	1	85	1	125	25	165	1	205	1
46	2	86	1	126	18	166	2	206	4
47	1	87	1	127	1	167	1	207	81
48	1	88	8	128	64	168	128	208	256
49	1	89	1	129	1	169	13	209	1
50	1	90	1	130	1	170	1	210	2

Then $J[\mathfrak{m}]$ contains $E[p]$ and also $A[\mathfrak{m}]$, and because p is a congruence prime, the submodule $A[\mathfrak{m}] \subset J[\mathfrak{m}]$ is nonzero. Thus the sum $E[p] + A[\mathfrak{m}]$ is a direct sum and is larger than $E[p]$, which is of dimension 2 over $\mathbf{T}/\mathfrak{m} = \mathbf{Z}/p\mathbf{Z}$. Hence the dimension of $J[\mathfrak{m}]$ over $\mathbf{T}/\mathfrak{m}$ is bigger than 2, as claimed. □

Proposition 5.1 implies that any example in which simultaneously $p \nmid m_E$ and $\mathrm{ord}_p(r_E) \neq \mathrm{ord}_p(m_E)$ produces an example in which multiplicity one for $J_0(N)$ fails. For example, for the curve 54B1 and $p = 3$, we have $\mathrm{ord}_3(r_E) = 1$ but $\mathrm{ord}_3(m_E) = 0$, so multiplicity one at 3 fails for $J_0(54)$.

References

[ARS06] A. Agashe, K. Ribet and W. Stein, *The Manin Constant*, QJPAM, Coates Volume (2006), to appear.

[AU96] A. Abbes and E. Ullmo, *À propos de la conjecture de Manin pour les courbes elliptiques modulaires*, Compositio Math. 103 (1996), no. 3, 269–286.

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. 24 (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993).

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. 14 (2001), no. 4, 843–939 (electronic).

[CK04] Alina Carmen Cojocaru and Ernst Kani, *The modular degree and the congruence number of a weight 2 cusp form*, Acta Arith. 114 (2004), no. 2, 159–167.

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997,
`http://www.maths.nott.ac.uk/personal/jec/book/`.

[CSS97] G. Cornell, J. H. Silverman, and G. Stevens (eds.), *Modular forms and Fermat's last theorem (boston,ma, 1995)*, New York, Springer-Verlag, 1997, Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[Fre97] G. Frey, *On ternary equations of Fermat type and relations with elliptic curves*, Modular forms and Fermat's last theorem (Boston, MA, 1995), Springer, New York, 1997, pp. 527–548.

[FM99] G. Frey and M. Müller, *Arithmetic of modular curves and applications*, Algorithmic algebra and number theory (Heidelberg, 1997), Springer, Berlin, 1999, pp. 11–48.

[Har77] R. Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[Kil02] L. J. P. Kilford, *Some non-Gorenstein Hecke algebras attached to spaces of modular forms*, J. Number Theory 97 (2002), no. 1, 157–164.

[Lan83] S. Lang, *Abelian varieties*, Springer-Verlag, New York, 1983, Reprint of the 1959 original.

[Li75] W-C. Li, *Newforms and functional equations*, Math. Ann. 212 (1975), 285–315.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. 44 (1978), no. 2, 129–162.

[MR91] B. Mazur and K. A. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque (1991), no. 196–197, 6, 215–255 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[Mum70] D. Mumford, *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, No. 5, Published for the Tata Institute of Fundamental Research, Bombay, 1970.

[Mur99] M. R. Murty, *Bounds for congruence primes*, Automorphic forms, automorphic representations, and arithmetic (Fort Worth, TX, 1996), Amer. Math. Soc., Providence, RI, 1999, pp. 177–192.

[Rib75] K. A. Ribet, *Endomorphisms of semi-stable abelian varieties over number fields*, Ann. Math. (2) 101 (1975), 555–562.

[Rib81] K. A. Ribet, *Endomorphism algebras of abelian varieties attached to newforms of weight* 2, Seminar on Number Theory, Paris 1979–80, Progr. Math., vol. 12, Birkhäuser Boston, Mass., 1981, pp. 263–276.

[Rib83] K. A. Ribet, *Mod p Hecke operators and congruences between modular forms*, Invent. Math. 71 (1983), no. 1, 193–205.

[Ste89] G. Stevens, *Stickelberger elements and modular parametrizations of elliptic curves*, Invent. Math. 98 (1989), no. 1, 75–106.

[Stu87] J. Sturm, *On the congruence of modular forms*, Number theory (New York, 1984–1985), Springer, Berlin, 1987, pp. 275–280.

[Til97] J. Tilouine, *Hecke algebras and the Gorenstein property*, Modular forms and Fermat's last theorem (Boston, MA, 1995), Springer, New York, 1997, pp. 327–342.

[Vat05] V. Vatsal, *Multiplicative subgroups of $J_0(N)$ and applications to elliptic curves*, J. Inst. Math. Jussieu 4 (2005), no. 2, 281–316.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) 141 (1995), no. 3, 443–551.

[Zag85] D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. 28 (1985), no. 3, 372–384.

39 The Sage Project: Unifying Free Mathematical Software to Create a Viable Alternative to Magma, Maple, Mathematica and Matlab, with B. Erocal

The Sage Project:
Unifying Free Mathematical Software to Create a Viable Alternative to Magma, Maple, Mathematica and MATLAB

Burçin Eröcal[1] and William Stein[2]

[1] Research Institute for Symbolic Computation
Johannes Kepler University,
Linz, Austria
Supported by FWF grants P20347 and DK W1214.
burcin@erocal.org
[2] Department of Mathematics
University of Washington
wstein@uw.edu
Supported by NSF grant DMS-0757627 and DMS-0555776.

Abstract. Sage is a free, open source, self-contained distribution of mathematical software, including a large library that provides a unified interface to the components of this distribution. This library also builds on the components of Sage to implement novel algorithms covering a broad range of mathematical functionality from algebraic combinatorics to number theory and arithmetic geometry.

Keywords: Python, Cython, Sage, Open Source, Interfaces

1 Introduction

In order to use mathematical software for exploration, we often push the boundaries of available computing resources and continuously try to improve our implementations and algorithms. Most mathematical algorithms require basic building blocks, such as multiprecision numbers, fast polynomial arithmetic, exact or numeric linear algebra, or more advanced algorithms such as Gröbner basis computation or integer factorization. Though implementing some of these basic foundations from scratch can be a good exercise, the resulting code may be slow and buggy. Instead, one can build on existing optimized implementations of these basic components, either by using a general computer algebra system, such as Magma, Maple, Mathematica or MATLAB, or by making use of the many high quality open source libraries that provide the desired functionality. These two approaches both have significant drawbacks. This paper is about Sage,[3] which provides an alternative approach to this problem.

[3] http://www.sagemath.org

Having to rely on a closed propriety system can be frustrating, since it is difficult to gain access to the source code of the software, either to correct a bug or include a simple optimization in an algorithm. Sometimes this is by design:

> “Indeed, in almost all practical uses of Mathematica, issues about how Mathematica works inside turn out to be largely irrelevant. You might think that knowing how Mathematica works inside would be necessary [...]” (See [Wol].)

Even if we manage to contact the developers, and they find time to make the changes we request, it might still take months or years before these changes are made available in a new release.

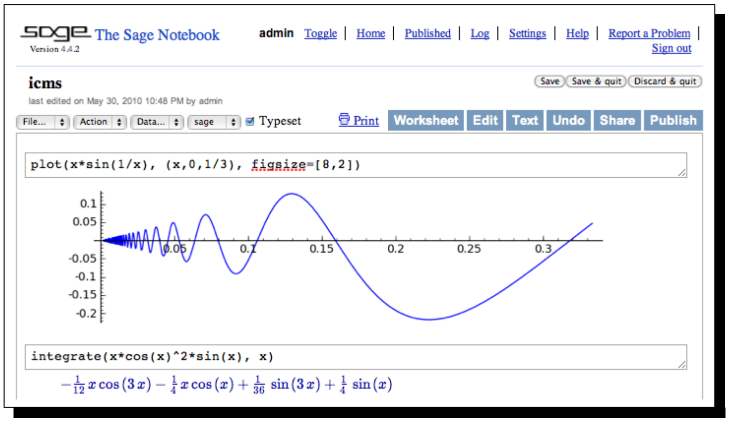
Fundamental questions of correctness, reproducibility and scientific value arise when building a mathematical research program on top of proprietary software (see, e.g., [SJ07]). There are many published refereed papers containing results that rely on computations performed in Magma, Maple, or Mathematica.[4] In some cases, a specific version of Magma is the only software that can carry out the computation. This is not the infrastructure on which we want to build the future of mathematical research.

In sharp contrast, open source libraries provide a great deal of flexibility, since anyone can see and modify the source code as they wish. However, functionality is often segmented into different specialized libraries and advanced algorithms are hidden behind custom interpreted languages. One often runs into trouble trying to install dependencies before being able use an open source software package. Also, converting the output of one package to the input format of another package can present numerous difficulties and introduce subtle errors.

Sage, which started in 2005 (see [SJ05]), attacks this problem by providing:

1. a *self-contained distribution* of mathematical software that installs from source easily, with the only dependency being compiler tools,
2. *unified interfaces* to other mathematical software to make it easier to use all these programs together, and
3. a *new library* that builds on the included software packages and implements a broad range of mathematical functionality.

The rest of this paper goes into more detail about Sage. In Section 1.1, we describe the Sage graphical user interface. Section 1.2 is about the Sage development process, Sage days workshops, mailing lists, and documentation. The subject of Section 2 is the sophisticated way in which Sage is built out of a wide range of open source libraries and software. In Section 2.1 we explain how we use Python and Cython as the glue that binds the compendium of software included in Sage into a unified whole. We then delve deeper into Python, Cython and the Sage preparser in Section 2.2, and illustrate some applications to mathematics in Section 2.3. Sage is actively used for research, and in Section 3 we describe some capabilities of Sage in advanced areas of mathematics.


sage The Sage Notebook
Version 4.4.2
admin Toggle | Home | Published | Log | Settings | Help | Report a Problem | Sign out
Save Save & quit Discard & quit
File... Action Data... sage ☑ Typeset
Print Worksheet Edit Text Undo Share Publish
0.1
0.05
-0.05
-0.1
-0.15
-0.2
0.05
0.1
0.15
0.2
0.25
0.3

icms

last edited on May 30, 2010 10:48 PM by admin

```
plot(x*sin(1/x), (x,0,1/3), figsize=[8,2])
```

```
integrate(x*cos(x)^2*sin(x), x)
```

$-\frac{1}{12}\,x\cos\left(3\,x\right) - \frac{1}{4}\,x\cos\left(x\right) + \frac{1}{36}\,\sin\left(3\,x\right) + \frac{1}{4}\,\sin\left(x\right)$

Fig. 1. The Sage Notebook

1.1 The Notebook

As illustrated in Figure 1, the graphical user interface for Sage is a web application, inspired by Google Documents [Goo], which provides convenient access to all capabilities of Sage, including 3D graphics. In single user mode, Sage works like a regular application whose main window happens to be your web browser. In multiuser mode, this architecture allows users to easily set up servers for accessing their work over the Internet as well as sharing and collaborating with colleagues. One can try the Sage notebook by visiting `www.sagenb.org`, where there are over 30,000 user accounts and over 2,000 published worksheets.

Users also download Sage to run it directly on their computers. We track all downloads from `www.sagemath.org`, though there are several other high-profile sites that provide mirrors of our binaries. Recently, people download about 6,000 copies of Sage per month directly from the Sage website.

1.2 The Sage Development Process

There are over 200 developers from across the world who have contributed to the Sage project. People often contribute because they write code using Sage as part of a research project, and in this process find and fix bugs, speed up parts of Sage, or want the code portion of their research to be peer reviewed. Each contribution to Sage is first posted to the Sage Trac server `trac.sagemath.org`; it is then peer reviewed, and finally added to Sage after all issues have been sorted out and all requirements are met. Nothing about this process is anonymous; every step of the peer review process is recorded indefinitely for all to see.

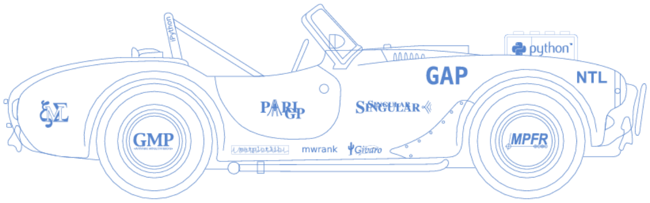
[4] Including by the second author of this paper, e.g., [CES03]!

The Sage Developer's Guide begins with an easy-to-follow tutorial that guides developers through each step involved in contributing code to Sage. Swift feedback is available through the `sage-devel` mailing list, and the `#sage-devel` IRC chat room on `irc.freenode.net` (see `www.sagemath.org/development.html`).

Much development of Sage has taken place at the Sage Days workshops. There have been two dozen Sage Days [Sagb] and many more are planned. These are essential to sustaining the momentum of the Sage project and also help ensure that developers work together toward a common goal, rather than competing with each other and fragmenting our limited community resources.

A major goal is ensuring that there will be many Sage Days workshops for the next couple of years. The topics will depend on funding, but will likely include numerical computation, large-scale bug fixing, L-functions and modular forms, function fields, symbolic computation, topology, and combinatorics. The combination of experienced developers with a group of enthusiastic mathematicians at each of these workshops has rapidly increased the developer community, and we hope that it will continue to do so.

2 Building the Car...



With the motto "building the car instead of reinventing the wheel," Sage brings together numerous open source software packages (see Table 1 and [Saga]).

Many applications of Sage require using these libraries together. Sage handles the conversion of data behind the scenes, automatically using the best tool for the job, and allows the user to concentrate on the problem at hand.

In the following example, which we explain in detail below, Sage uses the FLINT library [HH] for univariate polynomials over the ring $\mathbb{Z}$ of integers, whereas Singular [DGPS10] is used for multivariate polynomials. The option to use the NTL library [Sho] for univariate polynomials is still available, if the user so chooses.

```
sage: R.<x> = ZZ[]
sage: type(R.an_element())
<type 'sage.rings...Polynomial_integer_dense_flint'>
sage: R.<x,y> = ZZ[]
sage: type(R.an_element())
<type 'sage.rings...MPolynomial_libsingular'>
sage: R = PolynomialRing(ZZ, 'x', implementation='NTL')
sage: type(R.an_element())
<type 'sage.rings...Polynomial_integer_dense_ntl'>
```

Table 1. Packages Included With Every Copy of Sage-4.4.2

atlas	gap	libgcrypt	palp	scipy_sandbox
blas	gd	libgpg_error	pari	scons
boehm_gc	gdmodule	libm4ri	pexpect	setuptools
boost	genus2reduction	libpng	pil	singular
cddlib	gfan	linbox	polybori	sphinx
cliquer	ghmm	matplotlib	pycrypto	sqlalchemy
cvxopt	givaro	maxima	pygments	sqlite
cython	gnutls	mercurial	pynac	symmetrica
docutils	gsl	moin	python	sympow
ecl	iconv	mpfi	python_gnutls	sympy
eclib	iml	mpfr	r	tachyon
ecm	ipython	mpir	ratpoints	termcap
f2c	jinja	mpmath	readline	twisted
flint	jinja2	networkx	rubiks	weave
flintqs	lapack	ntl	sagenb	zlib
fortran	lcalc	numpy	sagetex	zn_poly
freetype	libfplll	opencdk	scipy	zodb3

The first line in the example above constructs the univariate polynomial ring `R` $= \mathbb{Z}[x]$, and assigns the variable `x` to be the generator of this ring. Note that $\mathbb{Z}$ is represented by `ZZ` in Sage. The expression `R.<x> = ZZ[]` is not valid Python, but can be used in Sage code as a shorthand as explained in Section 2.2. The next line asks the ring `R` for an element, using the `an_element` function, then uses the builtin Python function `type` to query its type. We learn that it is an instance of the class `Polynomial_integer_dense_flint`. Similarly line 4 constructs `R` $= \mathbb{Z}[x, y]$ and line 7 defines `R` $= \mathbb{Z}[x]$, but this time using the `PolynomialRing` constructor explicitly and specifying that we want the underlying implementation to use the NTL library.

Often these interfaces are used under the hood, without the user having to know anything about the corresponding systems. Nonetheless, there are easy ways to find out what is used by inspecting the source code, and users are strongly encouraged to cite components they use in published papers. The following example illustrates another way to get a list of components used when a specific command is run.

```
sage: from sage.misc.citation import get_systems
sage: get_systems(’integrate(x^2, x)’)
[’ginac’, ’Maxima’]
sage: R.<x,y,z> = QQ[]
sage: I = R.ideal(x^2+y^2, z^2+y)
sage: get_systems(’I.primary_decomposition()’)
[’Singular’]
```

2.1 Interfaces

Sage makes it possible to use a wide range of mathematical software packages together by providing a unified interface that handles data conversion automatically. The complexity and functionality of these interfaces varies greatly, from simple text-based interfaces that call external software for an individual computation, to using a library as the basis for an arithmetic type. The interfaces can also run code from libraries written in the interpreted language of another program. Table 2 lists the interfaces provided by Sage.

Table 2. Sage Interfaces to the above Mathematical Software

Pexpect	axiom, ecm, fricas, frobby, gap, g2red, gfan, gnuplot, gp, kash, lie, lisp, macaulay2, magma, maple, mathematica, matlab, maxima, mupad, mwrank, octave, phc, polymake, povray, qepcad, qsieve, r, rubik, scilab, singular, tachyon
C Library	eclib, fplll, gap (in progress), iml, linbox, maxima, ratpoints, r (via rpy2), singular, symmetrica
C Library arithmetic	flint, mpir, ntl, pari, polybori, pynac, singular

The above interfaces are the result of many years writing Python and Cython [BBS] code to adapt Singular [DGPS10], GAP [L$^+$], Maxima [D$^+$], Pari [PAR], GiNaC/Pynac [B$^+$], NTL [Sho], FLINT [HH], and many other libraries, so that they can be used smoothly and efficiently in a unified way from Python [Ros]. Some of these programs were originally designed to be used only through their own interpreter and made into a library by Sage developers. For example libSingular was created by Martin Albrecht in order to use the fast multivariate polynomial arithmetic in Singular from Sage. The libSingular interface is now used by other projects, including Macaulay2 [GS] and GFan [Jen].

There are other approaches to linking mathematical software together. The recent paper [LHK$^+$] reports on the state of the art using OpenMath. Sage takes a dramatically different approach to this problem. Instead of using a general string-based XML protocol to communicate with other mathematical software, Sage interfaces are tailor made to the specific software and problem at hand. This results in far more efficient and flexible interfaces. The main disadvantage compared to OpenMath is that the interfaces all go through Sage.

Having access to many programs which can perform the same computation, without having to worry about data conversion, also makes it easier to double check results. For example, below we first use Maxima, an open source symbolic computation package distributed with Sage, to integrate a function, then perform the same computation using Maple and Mathematica.

```
sage: var('x')
sage: integrate(sin(x^2), x)
1/8*((I - 1)*sqrt(2)*erf((1/2*I - 1/2)*sqrt(2)*x) + \
  (I + 1)*sqrt(2)*erf((1/2*I + 1/2)*sqrt(2)*x))*sqrt(pi)
sage: maple(sin(x^2)).integrate(x)
1/2*2^(1/2)*Pi^(1/2)*FresnelS(2^(1/2)/Pi^(1/2)*x)
sage: mathematica(sin(x^2)).Integrate(x)
Sqrt[Pi/2]*FresnelS[Sqrt[2/Pi]*x]
```

The most common type of interface, called a pexpect interface, communicates with another command line program by reading and writing strings to a text console, as if another user was in front of the terminal. Even though these are relatively simple to develop, the overhead of having to print and parse strings to represent the data makes this process potentially cumbersome and inefficient. This is the default method of communication with most high level mathematics software, including commercial and open source programs, such as Maple, Mathematica, Magma, KASH or GAP.

Sage provides a framework to represent elements over these interfaces, perform arithmetic with them or apply functions to the given object, as well as using a file to pass the data if the string representation is too big. The following demonstrates arithmetic with GAP elements.

```
sage: a = gap('22')
sage: a*a
484
```

It is also possible to use pexpect interfaces over remote consoles. In the following code, we connect to the localhost as a different user and call Mathematica functions. Note that the interface can handle indexing vectors as well.

```
sage: mma = Mathematica(server="rmma60@localhost")
sage: mma("2+2")
4
sage: t = mma("Cos[x]")
sage: t.Integrate('x')
Sin[x]
sage: t = mma('{0,1,2,3}')
sage: t[2]
1
```

Sage also includes specialized libraries that are linked directly from compiled code written in Cython. These are used to handle specific problems, such as the characteristic polynomial computation in the example below.

```
sage: M = Matrix(GF(5), 10, 10)
sage: M.randomize()
sage: M.charpoly(algorithm='linbox')
x^10 + 4*x^9 + 4*x^7 + 3*x^4 + 3*x^3 + 3*x^2 + 4*x + 3
```

Many basic arithmetic types also use Cython to directly utilize data structures from efficient arithmetic libraries, such as MPIR or FLINT. An example of this can be seen at the beginning of this section, where elements of the ring $\mathbb{Z}[x]$ are represented by the class Polynomial_integer_dense_flint.

The Singular interface is one of the most advanced included in Sage. Singular has a large library of code written in its own language. Previously the only way to access these functions, which include algorithms for Gröbner basis and primary decomposition, was to call Singular through a pexpect interface, passing data back and forth using strings. Recently, due to work of Michael Brickenstein and Martin Albrecht, Sage acquired the ability to call these functions directly.

In the example below, we import the function primdecSY from primdec.lib, and call it the same way we would call a Python function. The interface handles the conversion of the data to Singular's format and back. Since Sage already uses Singular data structures directly to represent multivariate polynomials and ideals over multivariate polynomial rings, there are no conversion costs. It is only a matter of passing the right pointer.

```
sage: pr = sage.libs.singular.ff.primdec__lib.primdecSY
sage: R.<x,y,z> = QQ[]
sage: p = z^2+1; q = z^3+2
sage: I = R.ideal([p*q^2,y-z^2])
sage: pr(I)
[[[z^2 - y, y^3 + 4*y*z + 4], \
    [z^2 - y, y*z + 2, y^2 + 2*z]], \
    [[y + 1, z^2 + 1], [y + 1, z^2 + 1]]]
```

Efforts are under way to extend these capabilities to other programs, for example to GAP which provides Sage's underlying group theory functionality. Up to now, GAP was only available through its interpreter, through a pexpect interface that was written by Steve Linton. As the following example demonstrates, the performance of this interface is far from ideal.[5]

```
sage: b = gap('10')
sage: b*b
100
sage: timeit('b*b')
625 loops, best of 3: 289 microseconds per loop
```

The code snippet above constructs the element b in GAP using the pexpect interface, and measures the time it takes to square b. Compare these numbers to the following example, which uses the library interface to GAP, recently developed by the second author (but *not* included in Sage yet).

```
sage: import sage.libs.gap.gap as g
sage: a = g.libgap('10'); a
10
sage: type(a)
<type 'sage.libs.gap.gap.GapElement'>
sage: a*a
100
sage: timeit('a*a')
625 loops, best of 3: 229 nanoseconds per loop
```

The library interface is about 1,000 times faster than the pexpect interface.

[5] All timings in this paper were performed on an 2.66GHz Intel Xeon X7460 based computer.

2.2 Python - a mainstream language

In line with the principle of not reinventing the wheel, Sage is built on the mainstream programming language Python, both as the main development language and the user language. This frees the Sage developers, who are mainly mathematicians, from the troubles of language design, and gives access to an immense array of general purpose Python libraries and tools.

Python is an interpreted language with a clear, easy to read and learn syntax. Since it is dynamically typed, it is ideal for rapid prototyping, providing an environment to easily test new ideas and algorithms.

A fast interpreter In the following Singular session, we first declare the ring `r` $= \mathbb{Q}[x, y, z]$ and the polynomial `f` $\in$ `r`, then measures the time to square `f` repeatedly, 10,000 times.

```
singular: int t = timer; ring r = 0,(x,y,z), dp;
singular: def f = y^2*z^2-x^2*y^3-x*z^3+x^3*y*z;
singular: int j; def g = f;
singular: for (j = 1; j <= 10^5; j++) { g = f*f; }
singular: (timer-t), system("--ticks-per-sec");
990 1000
```

The elapsed time is 990 milliseconds. Next we use Sage to do the same computation, using the same Singular data structures directly, but without going through the interpreter.

```
sage: R.<x,y,z> = QQ[]
sage: f = y^2*z^2 - x^2*y^3 - x*z^3 + x^3*y*z;  type(f)
<type 'sage.rings.polynomial...MPolynomial_libsingular'>
sage: timeit('for j in xrange(10^5): g = f*f')
5 loops, best of 3: 91.8 ms per loop
```

Sage takes only 91.8 milliseconds for the same operation. This difference is because the Python interpreter is more efficient at performing `for` loops.

Cython - compiled extensions Python alone is too slow to implement a serious mathematical software system. Fortunately, Cython [BBS] makes it easy to optimize parts of your program or access existing C/C++ libraries. It can translate Python code with annotations containing static type information to C/C++ code, which is then compiled as a Python extension module.

Many of the basic arithmetic types in Sage are provided by Cython wrappers of C libraries, such as FLINT for univariate polynomials over $\mathbb{Z}$, Singular for multivariate polynomials, and Pynac for symbolic expressions.

The code segment below defines a Python function to add integers from 0 to N and times the execution of this function with the argument `10^7`.

```
sage: def mysum(N):
....:     s = int(0)
....:     for k in xrange(1,N): s += k
....:     return s
....:
```

```
sage: time mysum(10^7)
CPU times: user 0.52 s, sys: 0.00 s, total: 0.52 s
49999995000000
```

Here is the same function, but the loop index `k` is declared to be a C integer and the accumulator `s` is a C `long long`.

```
sage: cython("""
....: def mysum_cython(N):
....:     cdef int k
....:     cdef long long s = 0
....:     for k in xrange(N): s += k
....:     return s
....: """)
sage: time mysum_cython(10^7)
CPU times: user 0.01 s, sys: 0.00 s, total: 0.01 s
49999995000000L
```

The code is compiled and linked to the interpreter on the fly, and the function `mysum_cython` is available immediately. Note that the run time for the Cython function is 60 times faster than the Python equivalent.

Cython also handles the conversion of Python types to C types automatically. In the following example, we call the C function `sinl` using Cython to wrap it in a Python function named `sin_c_wrap`.

```
sage: cython("""
....: cdef extern from "math.h":
....:         long double sinl(long double)
....: def sin_c_wrap(a):
....:         return sinl(a)
....: """)
sage: sin_c_wrap(3.14)
0.0015926529164868282
sage: sin_c_wrap(1)
0.8414709848078965
sage: sin_c_wrap(1r)
0.8414709848078965
```

Note that the conversion of Sage types in the first two calls to `sin_c_wrap` or the Python type integer in the last call is performed transparently by Cython.

The Preparser While Python has many advantages as a programming and glue language, it also has some undesirable features. Sage hides these problems by using a preparser to change the commands passed to Python in an interactive session (or when running a script with the `.sage` extension). In order to maintain compatibility with Python, changes performed by the preparser are kept to a minimum. Moreover, the Sage library code is not preparsed, and is written in Cython or Python directly.

Python, like C and many other programming languages, performs integer floor division. This means typing `1/2` results in 0, not the rational number 1/2. Sage wraps all numeric literals entered in the command line or the notebook with its own type declarations, which behave as expected with respect to arithmetic and have the advantage that they are backed by efficient multiprecision arithmetic libraries such as MPIR [H+] and MPFR [Z+], which are thousands of times faster than Python for large integer arithmetic.

To call the preparser directly on a given string, use the preparse function.

```
sage: preparse("1/2")
'Integer(1)/Integer(2)'
sage: preparse("1.5")
"RealNumber('1.5')"
```

Adding a trailing r after a number indicates that the preparser should leave that as the “raw” literal. The following illustrates division with Python integers.

```
sage: preparse("1r/2r")
'1/2'
sage: 1r/2r
0
```

Here is the result of performing the same division in Sage.

```
sage: 1/2
1/2
sage: type(1/2)
<type 'sage.rings.rational.Rational'>
sage: (1/2).parent()
Rational Field
```

The preparser also changes the ^ sign to the exponentiation operator ** and provides a shorthand to create new mathematical domains and name their generator in one command.

```
sage: preparse("2^3")
'Integer(2)**Integer(3)'
sage: preparse("R.<x,y> = ZZ[]")
"R = ZZ['x, y']; (x, y,) = R._first_ngens(2)"
```

2.3 Algebraic, Symbolic and Numerical Tools

Sage combines algebraic, symbolic and numerical computation tools under one roof, enabling users to choose the tool that best suits the problem. This combination also makes Sage more accessible to a wide audience—scientists, engineers, pure mathematicians and mathematics teachers can all use the same platform for scientific computation.

While not concentrating on only one of these domains might seem to divide development resources unnecessarily, it actually results in a better overall experience for everyone, since users do not have to come up with makeshift solutions to compensate for the lack of functionality from a different field. Moreover, because Sage is a distributed mostly-volunteer open source project, widening our focus results in substantially more developer resources.

Algebraic Tools: The Coercion System An algebraic framework, similar to that of Magma or Axiom, provides access to efficient data structures and specialized algorithms associated to particular mathematical domains. The Python language allows classes to define how arithmetic operations like `+` and `*` will be handled, in a similar way to how C++ allows overloading of operators. However, the built-in support for overloading in Python is too simple to support operations with a range of objects in a mathematical type hierarchy.

Sage abstracts the process of deciding what an arithmetic operation means, or equivalently, in which domain the operation should be performed, in a framework called the *coercion system*, which was developed and implemented by Robert Bradshaw, David Roe, and many others. Implementations of new mathematical objects only need to define which other domains have a natural embedding to their domain. When performing arithmetic with objects, the coercion system will find a common domain where both arguments can be canonically mapped, perform the necessary type conversions automatically, thus allowing the implementation to only handle the case where both objects have the same parent.

In the following example, the variable `t` is an element of $\mathbb{Z}$ whereas `u` is in $\mathbb{Q}$. In order to perform the addition, the coercion system first deduces that the result should be in $\mathbb{Q}$ from the fact that `t` can be converted to the domain of `u`, namely $\mathbb{Q}$, but canonical conversion in the other direction is not possible. Then the addition is performed with both operands having the same domain $\mathbb{Q}$.

```
sage: t = 1
sage: t.parent()
Integer Ring
sage: u = 1/2
sage: u.parent()
Rational Field
sage: v = t + u; v
3/2
sage: v.parent()
Rational Field
```

Similarly, in the following example, the common domain $\mathbb{Q}[x]$ is found for arguments from $\mathbb{Z}[x]$ and $\mathbb{Q}$. Note that in this case, the result is not in the domain of either of the operands.

```
sage: R.<x> = ZZ[]
sage: r = x + 1/2
sage: r.parent()
Univariate Polynomial Ring in x over Rational Field
sage: 5*r
5*x + 5/2
```

Algebraic Tools: The Category Framework Another abstraction to make implementing mathematical structures easier is the *category framework*, whose development was spearheaded by Nicolas Thiéry and Florent Hivert. Similar in spirit to the mathematical programming facilities developed in Axiom and encapsulated in Aldor, the category framework uses Python's dynamic class creation capabilities to combine functions relevant for a mathematical object, inherited through a mathematical hierarchy, into a class at run time.

This process greatly simplifies the troubles of having to combine object-oriented programming concepts with mathematical structural concerns, while keeping efficiency in mind. Efficient implementations can keep the inheritance hierarchy imposed by the data structures, while generic methods to compute basic properties are implemented in the *category* and automatically attached to the element classes when they are needed.

Symbolic Tools The symbolic subsystem of Sage provides an environment similar to Maple or Mathematica, where the input is treated only as an expression without any concern about the underlying mathematical structure.

Sage uses Pynac [ES], a hybrid C++ and Cython library built on top of GiNaC [B$^+$], to work with symbolic expressions. High level symbolic calculus problems including symbolic integration, solution of differential equations and Laplace transforms are solved using Maxima behind the scenes.

Here is an example of how to use the symbolic computation facilities in Sage. Note that in contrast to other symbolic software such as Maple, variables must be declared before they are used.

```
sage: x,y,z = var(’x,y,z’)
sage: sin(x).diff(x)
cos(x)
sage: psi(x).series(x,4)
(-1)*x^(-1) + (-euler_gamma) + (1/6*pi^2)*x + \
    (-zeta(3))*x^2 + (1/90*pi^4)*x^3 + Order(x^4)
sage: w = SR.wild()  # wildcard for symbolic substitutions
sage: ((x^2+y^2+z^2)*zeta(x)).subs({w^2:5})
15*zeta(x)
```

Numerical Tools In addition to code for symbolic computation, the standard numerical Python packages NumPy, SciPy, and Matplotlib are included in Sage, along with the numerical libraries cvxopt, GSL, Mpmath, and R.

For numerical applications, Robert Bradshaw and Carl Witty developed a compiler for Sage that converts symbolic expressions into an internal format suitable for blazingly fast floating point evaluation.

```
sage: f(x,y) = sqrt(x^2 + y^2)
sage: a = float(2)
sage: timeit(’float(f(a,a))’)
625 loops, best of 3: 216 microseconds per loop
sage: g = fast_float(f)
sage: timeit(’float(g(a,a))’)
625 loops, best of 3: 0.406 microseconds per loop
```

The `fast_float` feature is automatically used by the `minimize` command.

```
sage: minimize(f, (a,a))
(-5.65756135618e-05, -5.65756135618e-05)
```

Performance is typically within a factor of two from what one gets using a direct implementation in C or Fortran.

3 Afterword

In this article, we have showed that Sage is a powerful platform for developing sophisticated mathematical software. Sage is actively used in research mathematics, and people use Sage to develop state-of-the-art algorithms. Sage is particularly strong in number theory, algebraic combinatorics, and graph theory. For further examples, see the 53 published articles, 11 Ph.D. theses, 10 books, and 30 preprints at `www.sagemath.org/library-publications.html`.

For example, Sage has extensive functionality for computations related to the Birch and Swinnerton-Dyer conjecture. In addition to Mordell-Weil group computations using [Cre] and point counting over large finite fields using the SEA package in [PAR], there is much novel elliptic curve code written directly for Sage. This includes the fastest known algorithm for computation of p-adic heights [Har07,MST06], and code for computing p-adic L-series of elliptic curves at ordinary, supersingular, and split multiplicative primes. Sage combines these capabilities to compute explicitly bounds on Shafarevich-Tate groups of elliptic curves [SW10]. Sage also has code for computation with modular forms, modular abelian varieties, and ideal class groups in quaternion algebras.

The MuPAD-combinat project, which was started by Florent Hivert and Nicolas M. Thiéry in 2000, built the world's preeminent system for algebraic combinatorics on top of MuPAD (see [Des06] and [HT05]). Page 54 of [HT05]: "They [MuPAD] also have promised to release the code source of the library under a well known open-source license, some day." In 2008, MuPAD was instead purchased by MathWorks (makers of MATLAB), so MuPAD is no longer available as a separate product, and will probably never be open source. Instead it now suddenly costs $3000 (commercial) or $700 (academic).

As a result, the MuPAD-combinat group has spent several years reimplementing everything in Sage (see [T$^+$] for the current status). The MuPAD-combinat group was not taken by surprise by the failure of MuPAD, but instead were concerned from the beginning by the inherent risk in building their research program on top of MuPAD. In fact, they decided to switch to Sage two months before the bad news hit, and have made tremendous progress porting:

> "It has been such a relief during the last two years not to have this Damocles sword on our head!"
>
> – Nicolas Thiéry

References

B$^+$. Christian Bauer et al., **Ginac: is not a CAS**, `http://www.ginac.de/`.

BBS. Stefan Behnel, Robert Bradshaw, and Dag Seljebotn, **Cython: C-Extensions for Python**, `http://www.cython.org/`.

CES03. B. Conrad, S. Edixhoven, and W. A. Stein, $J_1(p)$ **Has Connected Fibers**, Documenta Mathematica **8** (2003), 331–408.

Cre. J. E. Cremona, `mwrank` **(computer software)**, `http://www.warwick.ac.uk/staff/J.E.Cremona/mwrank/`.

D$^+$. Robert Dodier et al., **Maxima: A Computer Algebra System**, `http://maxima.sourceforge.net/`.

Des06. Francois Descouens, **Making research on symmetric functions with MuPAD-Combinat**, Mathematical software—ICMS 2006, Lecture Notes in Comput. Sci., vol. 4151, Springer, Berlin, 2006, pp. 407–418.

DGPS10. W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann, SINGULAR **3-1-1 — A computer algebra system for polynomial computations**, `http://www.singular.uni-kl.de`.

ES. Burcin Erocal and William Stein, **Pynac – symbolic computation with python objects**, `http://pynac.sagemath.org/`.

Goo. Google, **Google Documents**, `http://docs.google.com/`.

GS. Daniel R. Grayson and Michael E. Stillman, **Macaulay2, a software system for research in algebraic geometry**, Available at http://www.math.uiuc.edu/Macaulay2/.

H$^+$. Bill Hart et al., **MPIR: Multiprecision Integers and Rationals**, `http://www.mpir.org/`.

Har07. David Harvey, **Efficient computation of p-adic heights**, `http://arxiv.org/abs/0708.3404`.

HH. Bill Hart and David Harvey, **Flint: Fast library for number theory**, `http://www.flintlib.org/`.

HT05. Florent Hivert and Nicolas M. Thiéry, **MuPAD-Combinat, an open-source package for research in algebraic combinatorics**, Sém. Lothar. Combin. **51** (2004/05), Art. B51z, 70 pp. (electronic), `http://www.emis.de/journals/SLC/wpapers/s51thiery.html`.

Jen. Anders Jensen, **Gfan: software for computing Gröbner fans and tropical varieties**, `http://www.math.tu-berlin.de/~jensen/software/gfan/gfan.html`.

L$^+$. Steve Linton et al., **Gap: Groups, algorithms and programming**, `http://www.gap-system.org/`.

LHK$^+$. S. Linton, K. Hammond, A. Konovalov, et al., **Easy Composition of Symbolic Computation Software: A New Lingua Franca for Symbolic Computation**, `www.win.tue.nl/~droozemo/site/pubs/1004ISSAC2010.pdf`.

MST06. Barry Mazur, William Stein, and John Tate, **Computation of p-adic heights and log convergence**, Doc. Math. (2006), no. Extra Vol., 577–614 (electronic). MR MR2290599 (2007i:11089)

PAR. PARI, **A computer algebra system designed for fast computations in number theory**, `http://pari.math.u-bordeaux.fr/`.

Ros. Guido van Rossum, **Python**, `http://www.python.org`.

Saga. Sage, **Components**, `http://sagemath.org/links-components.html`.

Sagb. ———, **Sage days workshops**, `http://wiki.sagemath.org/Workshops`.

Sho. V. Shoup, **NTL: Number theory library**, `http://www.shoup.net/ntl/`.

SJ05. William Stein and David Joyner, **Open source mathematical software**, ACM SIGSAM Bulletin **39** (2005).

SJ07. ———, **Open source mathematical software**, Notices Amer. Math. Soc. (2007), `http://www.ams.org/notices/200710/tx071001279p.pdf`.

SW10. W. Stein and C. Wuthrich, **Computations About Tate-Shafarevich Groups Using Iwasawa Theory**, In preparation (2010), `http://wstein.org/papers/shark/`.

T$^+$. N. Thiery et al., **Sage Combinat Roadmap**, `http://trac.sagemath.org/sage_trac/wiki/SageCombinatRoadMap`.

Wol. Wolfram, **Why you do not usually need to know about internals**, `http://reference.wolfram.com/mathematica/tutorial/WhyYouDoNotUsuallyNeedToKnowAboutInternals.html`.

Z$^+$. Paul Zimmerman et al., **The MPFR Library**, `http://www.mpfr.org/`.

40 Heegner Points and the Arithmetic of Elliptic Curves over Ring Class Extensions, with R. Bradshaw

Heegner Points and the Arithmetic of Elliptic Curves Over Ring Class Extensions

Robert Bradshaw and William Stein[1a,b]

[a] *Google, Seattle*
[b] *University of Washington*

Abstract

Let E be an elliptic curve over $\mathbb{Q}$ and let K be a quadratic imaginary field that satisfies the Heegner hypothesis. We study the arithmetic of E over ring class extensions of K, with particular focus on the case when E has analytic rank at least 2 over $\mathbb{Q}$. We also point out an issue in the literature regarding generalizing the Gross-Zagier formula, and offer a conjecturally correct formula.

Keywords: elliptic curve, Gross-Zagier formula, Birch and Swinnerton-Dyer conjecture, Shafarevich-Tate groups
2000 MSC: 11G05

1. Introduction

Let E be an elliptic curve over $\mathbb{Q}$. By [Wil95, BCDT01], $L(E,s)$ extends to an entire function on $\mathbb{C}$, so $r_{\mathrm{an}}(E/\mathbb{Q}) = \mathrm{ord}_{s=1} L(E,s)$ is defined. Let $r_{\mathrm{alg}}(E/\mathbb{Q}) = \mathrm{rank}(E(\mathbb{Q}))$.

Conjecture 1 (Birch and Swinnerton-Dyer (see [Wil00]))**.** *We have*

$$r_{\mathrm{an}}(E/\mathbb{Q}) = r_{\mathrm{alg}}(E/\mathbb{Q}).$$

Let K be a quadratic imaginary field such that all primes dividing the conductor N of E split in K, and let $u = \#\mathcal{O}_K^\times/2$, which is 1 unless $K = \mathbb{Q}(\sqrt{-1})$ or $\mathbb{Q}(\sqrt{-3})$. For each squarefree product c of primes that are inert

[1] This work was supported by NSF grants DMS-0757627, DMS-0653968 and the Mathematical Sciences Research Institute.

in K, let K_c denote the ring class field of conductor c, which is an abelian extension of K ramified exactly at primes dividing c. Moroever, K_1 is the Hilbert class field of K, and (see [Gro91, §3])

$$\mathrm{Gal}(K_c/K_1) \cong (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times.$$

Heegner points are certain points in $E(K_c)$ that are constructed using complex multiplication and a fixed choice of modular parametrization $\phi_E : X_0(N) \to E$ of minimal degree. In this paper, we study the subgroup of $E(K_c)$ generated by Galois conjugates of Heegner points, and relate it to $\#Ш(E/K_c)$.

Our motivation for this paper is that the subgroup W of any Mordell-Weil group generated by Heegner points typically fits into an analogue of the BSD conjecture, but with the "difficult" factors such as the Shafarevich-Tate group and Tamagawa numbers removed (see [Ste10b]). Thus according to the BSD formula (see Conjecture 12 below), we expect that the index of W in its saturation (or the closely related index of $E(K) + W$ in $E(K_c)$) in the Mordell-Weil group is related to the order of Ш and Tamagawa numbers. In Theorem 13 below, which is conditional on the BSD formula (see Conjecture 12 below), we compute this index in terms of other invariants of E. Intriguingly, in order for our result to satisfy certain consistency checks, we discover that the previously published explicit generalizations of the Gross-Zagier formula to ring class fields appear to be wrong, e.g., they do not properly take into account either the conductor of the ring class character or the degree of the ring class field.

Our hypothesis that every prime dividing N splits in K implies that there is a factorization of the ideal $N\mathcal{O}_K$ as $\mathcal{N}\bar{\mathcal{N}}$ with $\mathcal{O}_K/\mathcal{N} \cong \mathbb{Z}/N\mathbb{Z}$. Fix an embedding $K \hookrightarrow \mathbb{C}$ and view $\mathcal{O}_K$ as a lattice in $\mathbb{C}$, so $\mathbb{C}/\mathcal{O}_K$ is a CM elliptic curve, and $\mathcal{N}^{-1}/\mathcal{O}_K$ defines a cyclic subgroup of order N. Let $X_0(N)$ be the standard modular curve whose affine points over $\mathbb{C}$ parameterize isomorphism classes of pairs (F, C), where F is an elliptic curve over $\mathbb{C}$ and C is a cyclic subgroup of F of order N. Let x_1 be the point in $X_0(N)(K_1)$ defined by the isomorphism class of $(\mathbb{C}/\mathcal{O}_K, \mathcal{N}^{-1}/\mathcal{O}_K)$. Using the modular parameterization $\phi_E : X_0(N) \to E$, we obtain a point $y_1 = \phi_E(x_1) \in E(K_1)$. Let $y_K = \mathrm{Tr}_{K_1/K}(y_1)$ be the trace of y_1. After fixing our choice of ϕ_E, the point y_K is well defined up to sign, since making a different choice of $\mathcal{N}$ replaces y_K by its image under an Atkin-Lehner involution, as explained in [Wat06, §2] or [Coh07, Thm. 8.7.7], and Atkin-Lehner acts as ±1 on E.

In addition to their central importance to explicit computation of rational points on elliptic curves, Heegner points play an essential role in results toward Conjecture 1 (see, e.g., [Gro91]):

Theorem 2 (Gross-Zagier, Kolyvagin, et al.)**.** *Let $E/\mathbb{Q}$ be an elliptic curve with $r_{\mathrm{an}}(E/\mathbb{Q}) \leq 1$. Then $r_{\mathrm{an}}(E/\mathbb{Q}) = r_{\mathrm{alg}}(E/\mathbb{Q})$ and $Ш(E/\mathbb{Q})$ is finite.*

The proof that $Ш(E/\mathbb{Q})$ is finite also yields an explicit computable upper bound on the p-part of $\#Ш(E/\mathbb{Q})$ (see [GJP$^+$09, Thm. 3.4]) at primes p where $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p])$ has sufficiently large image (see [Cha05, GJP$^+$09, Jet08, SW11]). The bound is in terms of $[E(K) : \mathbb{Z}y_K]$, for any choice of K. This bound plays an essential role in verifying the full BSD formula (Conjecture 12) for specific elliptic curves, as in [GJP$^+$09, Mil10, MS10].

If M is any number field, let $\hat{h}_M$ denote the Néron-Tate canonical height on $E(M)$ over M. If S is an extension of M and $P \in E(M)$, then $\hat{h}_S(P) = [S : M] \cdot \hat{h}_M(P)$ (see [Sil92, Prop. VIII.5.4]). Following [GZ86, §I.6 and §V.2], we have

$$\|\omega_E\|^2 = \frac{8\pi^2 \cdot (f,f) \cdot c_E^2}{\deg(\phi_E)}, \tag{1}$$

where ω_E is a minimal differential on E, c_E is the Manin constant, $\deg(\phi_E)$ is the modular degree, f is the newform corresponding to E, and (f,f) is the Petersson inner product of f with itself (see also [GJP$^+$09, §3]).

Remark 3. We assume that $c_E = 1$ in the rest of this paper. As explained in [ARS06] this should be a harmless assumption, and conjecturally amounts to working with the optimal elliptic curve isogenous to E.

The following theorem is in [GZ86, §V.2, pg. 311]:

Theorem 4 (Gross-Zagier)**.** *We have*

$$L'(E/K,1) = \frac{\|\omega_E\|^2}{u^2 \cdot \sqrt{|D_K|}} \cdot \hat{h}_K(y_K).$$

Let E be an elliptic curve over $\mathbb{Q}$ and assume that $r_{\mathrm{an}}(E/K) = 1$. The subgroup of $E(K)$ generated by the Heegner point plays an essential role in the proof of Theorem 2. One uses the nontorsion point $y_K = \mathrm{Tr}_{K_1/K}(y_1)$ to bound the rank of $E(K)$ from below. There are also higher Heegner points $y_c = \phi_E(x_c)$ (see Section 2) that are used to construct elements of various Selmer groups associated to E, which one then uses to bound the rank of $E(K)$ from above.

Assume $L'(E/K,1) \neq 0$. Then, as explained in [Ste10b, §2], the Gross-Zagier formula and the BSD formula for $L'(E/K,1)$ together imply that

$$[E(K) : \mathbb{Z}y_K]^2 = \#Ш(E/K) \cdot \prod c_{v,K},$$

where the $c_{v,K}$ are the Tamagawa numbers of E/K. Note that since each prime divisor $p \mid N$ splits in K, the product of the Tamagawa numbers of E/K is the square of $\prod_{p|N} c_p$, where the c_p are the Tamagawa numbers of $E/\mathbb{Q}$. See the proof of Proposition 14 for related remarks, and [Ste10b, Prop. 2.4] for a discussion of what happens when E has rank ≥ 2.

In Section 2, we recall the definition of Heegner points over ring class fields, set up some notation involving characters and corresponding idempotent projectors, and discuss generalization of the Gross-Zagier formula to higher Heegner points. In Section 3, we introduce the subgroup W of $E(K_c)$ generated by Galois conjugates of Heegner points and describe a theorem of Bertolini-Darmon that allows us to deduce conditions under which $W+E(K)$ has finite index in $E(K_c)$. In Section 4, we use a generalization of the Gross-Zagier formula to derive a formula for $\mathrm{Reg}(W)$, then use the BSD formula to compute the index of $W+E(K)$ in $E(K_c)$. We also compute the index of W in its saturation. Section 5 gives an example that illustrates the results of Section 4. Finally, Section 6 suggests some avenues for future investigation.

2. Higher Heegner Points

Fix a positive squarefree integer c whose prime divisors are inert in K and coprime to N. Let $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$ and $\mathcal{N}_c = \mathcal{N} \cap \mathcal{O}_c$. Then the pair $(\mathbb{C}/\mathcal{O}_c, \mathcal{N}_c^{-1}/\mathcal{O}_c)$ defines a CM elliptic curve equipped with a cyclic subgroup of order N, and the isomorphism class of this pair defines a point $x_c \in X_0(N)(K_c)$. We use the modular parameterization ϕ_E to map x_c to a point $y_c = \phi_E(x_c) \in E(K_c)$.

Let $G = \mathrm{Gal}(K_c/K)$ and let

$$h_c = [K_c : K] = \#\,\mathrm{Cl}(\mathcal{O}_c) = \#G$$

be the class number of the order $\mathcal{O}_c$. For any character $\chi : G \to \mathbb{C}^\times$, let e_χ be the idempotent

$$e_\chi = \frac{1}{h_c}\sum_{\sigma \in G} \chi^{-1}(\sigma)\sigma \in \mathbb{C}[G],$$

which projects to the χ-isotypical component of any module. Note that if $\sigma \in G$, then $\sigma e_\chi = \chi(\sigma) e_\chi$; also, $1 = \sum_{\chi : G \to \mathbb{C}^\times} e_\chi$.

Following [Gro84, (10.1)], we extend the Néron-Tate height pairing $\langle\ ,\ \rangle_{K_c}$ on $E(K_c)$ defined by h_{K_c} to a Hermitian inner product on the complex vector space $V = E(K_c) \otimes_{\mathbb{Z}} \mathbb{C}$ by letting

$$\langle \alpha P,\, \beta Q \rangle = \alpha \overline{\beta} \langle P,\, Q \rangle_{K_c} \tag{2}$$

and extending linearly. We also view V as a $\mathbb{C}[G]$-module by making $\sigma \in G$ act by $\sigma(P \otimes \alpha) = \sigma(P) \otimes \alpha$. Since E is defined over $\mathbb{Q}$, the height pairing on V is $\mathrm{Gal}(K_c/\mathbb{Q})$-equivariant (see [Sil92, Lem. VIII.5.10]), in the sense that for any $\sigma \in \mathrm{Gal}(K_c/\mathbb{Q})$ and $P, Q \in E(K_c)$, we have $\langle \sigma(P), \sigma(Q) \rangle = \langle P, Q \rangle$.

Lemma 5. *The χ eigenspaces of V are orthogonal with respect to the height pairing.*

Proof. This is standard, but for the convenience of the reader we give a proof. If χ, χ' are two characters of G, then for any $P, Q \in E(K_c)$ and $\sigma \in G$, we have

$$\begin{aligned} \langle e_\chi P,\, e_{\chi'} Q \rangle &= \langle \sigma(e_\chi P),\, \sigma(e_{\chi'} Q) \rangle \\ &= \langle \chi(\sigma) e_\chi P,\, \chi'(\sigma) e_{\chi'} Q \rangle \\ &= \chi(\sigma) \chi'(\sigma)^{-1} \langle e_\chi P,\, e_{\chi'} Q \rangle. \end{aligned}$$

Thus if $\langle e_\chi P,\, e_{\chi'} Q \rangle \neq 0$ for some P, Q, then $\chi(\sigma)\chi'(\sigma)^{-1} = 1$ for all σ, hence $\chi = \chi'$. □

We next explain how the heights $\hat{h}_{K_c}(e_\chi y_c)$ are related to the special values of certain L-functions. Let $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ be the newform corresponding to E, let χ be a character of G, and let $L(f, \chi, s)$ be the Rankin-Selberg L-series $L(f \otimes g_\chi, s)$, as described in [Gro84, §III]. According to [Gro84, Prop. 21.2], the sign in the functional equation for $L(f, \chi, s)$ is -1, so $L(f, \chi, s)$ vanishes to odd order at $s = 1$. In [Zha01a, Thm. 1.2.1], Zhang proves a generalization of the Gross-Zagier formula (Theorem 4 above) that relates the height of $e_\chi y_c$ to $L'(f, \chi, 1)$. Unfortunately, the literature on this formula is inconsistent. For nontrival χ, [JLS09, §A.2] asserts that Zhang's theorem implies that

$$L'(f, \chi, 1) = \frac{4(f, f)}{u^2 \sqrt{|D_K|}} \cdot \hat{h}_{K_c}(e_\chi y_c). \tag{3}$$

The earlier paper [Hay95, Thm. 2] conjectures that the formula is

$$L'(f,\chi,1) = \frac{8\pi^2(f,f)}{u^2\sqrt{|D_K|}} \cdot \hat{h}_{K_c}(e_\chi y_c). \tag{4}$$

However, somewhat bizarrely, immediately after stating the above, [Hay95] then states that the formula is instead

$$L'(f,\chi,1) = \frac{h_c \cdot 8\pi^2(f,f)}{u^2\sqrt{|D_K|}} \cdot \hat{h}_{K_c}(e_\chi y_c). \tag{5}$$

which is closer to what we expect (see Conjecture 6).

Consistency checks with the BSD formula (see Proposition 14 and the discussion on page 14 right after the proof of Theorem 13) very strongly suggest that Equations (3), (4) and (5) are all incorrect. Zhang remarks at the end of Section 1 of [Zha04], "I would like to thank N. Vatsal and H. Xue for pointing out many inaccuracies in our previous paper [Zha01a]," and in an email to the authors: "You are right that my formula cited in your paper is not accurate. A correct version is in my paper [Zha04]."

Instead, we propose the following closely related formula, which also features the *conductor* of the character $\chi : \mathrm{Gal}(K_c/K) \to \mathbb{C}^\times$, which is the smallest integer divisor $c' \mid c$ such that χ factors through the natural quotient map $\mathrm{Gal}(K_c/K) \to \mathrm{Gal}(K_{c'}/K)$.

Conjecture 6. *If χ is a nontrivial character of G, then*

$$L'(f,\chi,1) = \frac{h_c \cdot 8\pi^2(f,f)}{\mathrm{cond}(\chi) \cdot u^2 \cdot \sqrt{|D_K|}} \cdot \hat{h}_{K_c}(e_\chi y_c).$$

Remark 7. Zhang has explained to us that one can deduce the above conjecture from his [Zha04, Thm 6.1]. Zhang and his students intend to give the details in a future paper.

3. The Heegner Point Subgroup

In this section we state a theorem of Bertolini-Darmon, and use it to understand when $W + E(K)$ generates a finite index subgroup of $E(K_c)$. We also give equivalent conditions under which W and $E(K)$ are orthogonal.

Let E and K be as above. We continue to fix an integer c whose prime divisors are inert in K and coprime to N, and let a_c be the cth Fourier coefficient of the newform attached to the elliptic curve E. Consider the subgroup $W = \mathbb{Z}[G]y_c$ of $E(K_c)$ spanned by the G-conjugates of y_c.

Recall from Section 2 the vector space $V = E(K_c) \otimes_{\mathbb{Z}} \mathbb{C}$, which is a finite-dimensional $\mathbb{C}[G]$-module equipped with a G-invariant bilinear Hermitian height pairing (2). For any character χ of G, let $V^\chi = e_\chi V$ be the subspace of V on which G acts via χ. Because $1 = \sum_\chi e_\chi$, we have

$$V = \bigoplus_{\chi : G \to \mathbb{C}^\times} V^\chi,$$

and Lemma 5 asserts that the V^χ are mutually orthogonal. Let $y_{c,\chi} = e_\chi(y_c) \in V^\chi$.

Theorem 8 (Bertolini-Darmon [BD90])**.** *If $y_{c,\chi} \neq 0$ then $V^\chi = \mathbb{C}y_{c,\chi}$.*

Remark 9. The converse of Theorem 8 is the assertion that if $y_{c,\chi} = 0$ then $V^\chi \neq \mathbb{C}y_{c,\chi} = 0$. As explained in [BD90], this is consistent with a natural refinement of the BSD rank conjecture (Conjecture 1), which asserts that V^χ has odd rank (see also [YZZ10, Conj. 1.4.1]). It is a difficult open problem to come up with any way to construct points in V^χ when $\mathbb{C}y_{c,\chi} = 0$.

Proposition 10. *If for* all *nontrivial characters χ of G we have $L'(f, \chi, 1) \neq 0$, then the index $[E(K_c) : W + E(K)]$ is finite.*

Proof. By tensoring with $\mathbb{C}$, we see that the claim is equivalent to showing that the $\mathbb{C}$ span of $W + E(K)$ is V. Let χ_1 denote the trivial character. Then

$$V = \bigoplus_{\chi : G \to \mathbb{C}^\times} V^\chi = V^{\chi_1} \oplus \bigoplus_{\chi \neq \chi_1} V^\chi.$$

We have $V^{\chi_1} = E(K) \otimes \mathbb{C}$. Theorem 8 and our hypothesis that $L'(f, \chi, 1) \neq 0$ for all nontrivial χ imply that $W \otimes \mathbb{C} = \oplus_{\chi \neq \chi_1} V^\chi$, □

As explained in [Gro84, §6] and [Gro91, Prop. 3.7], we have $\mathrm{Tr}_{K_c/K}(y_c) = a_c y_K$, which motivates the appearance of $a_c y_K$ in the following proposition.

Proposition 11. *The following are equivalent:*

1. *The two subgroups W and $E(K)$ of $E(K_c)$ are mutually orthogonal.*
2. *The point $a_c y_K$ is torsion.*
3. $a_c = 0$ *or* $r_{\mathrm{an}}(E/K) > 1$.

Proof. To prove that 1 implies 2, suppose that W is orthogonal to $E(K)$. The height pairing on $E(K_c)$ is 0 only on torsion points, so $W \cap E(K)$ is a torsion group. But $a_c y_K = \mathrm{Tr}_{K_c/K}(y_c) \in W \cap E(K)$, so $a_c y_K$ is torsion, as claimed.

To prove that 2 implies 1, assume that $a_c y_K$ is torsion. Choose $P \in E(K)$ and $Q \in W$. For any $\sigma \in G$, we have

$$\mathrm{Tr}_{K_c/K}(\sigma(y_c)) = \sigma(\mathrm{Tr}_{K_c/K}(y_c)) = \sigma(a_c y_K) = a_c y_K \in E(K)_{\mathrm{tor}}. \tag{6}$$

Since Q is a linear combination of $\sigma(y_c)$ for various σ, Equation (6) implies that $\mathrm{Tr}_{K_c/K}(Q)$ is torsion. The height pairing is Galois equivariant, so for all $\sigma \in G$, we have $\langle P, Q\rangle = \langle \sigma P, \sigma Q\rangle = \langle P, \sigma Q\rangle$. Thus

$$\langle P, Q\rangle = \frac{1}{h_c}\sum_{\sigma\in G}\langle P, \sigma Q\rangle = \frac{1}{h_c}\langle P, \mathrm{Tr}_{K_c/K} Q\rangle = 0.$$

Finally we observe that 2 and 3 are equivalent. If $a_c = 0$ then $a_c y_K = 0$. If $r_{\mathrm{an}}(E/K) > 1$, then Theorem 4 implies that y_K is torsion. Conversely, suppose $a_c y_K$ is torsion. If $a_c \neq 0$, then y_K is also torsion, so Theorem 4 implies that $r_{\mathrm{an}}(E/K) > 1$. □

4. Regulators and Indexes

In this section we study the index $[E(K_c) : W + E(K)]$, and under certain hypotheses, conjecturally relate it to various arithmetic invariants of E. In particular, we prove Theorem 13, which is a conjectural formula for the index $[E(K_c)_{/\mathrm{tor}} : (E(K) + W)_{/\mathrm{tor}}]$ under any of the equivalent hypotheses of Proposition 11.

If H is any subgroup of a Mordell-Weil group $E(M)$, let $\mathrm{Reg}_M(H)$ be the absolute value of the determinant of the height pairing $\langle\, , \rangle_M$ on a basis of H. We emphasize here that we use the height relative to M and not the absolute height on $E(\overline{\mathbb{Q}})$.

Theorem 13 below is conditional on the BSD formula over number fields.

Conjecture 12 (Birch and Swinnerton-Dyer Formula)**.** *If E is an elliptic curve of rank r over a number field F then*

$$\frac{L^{(r)}(E/F, 1)}{r!} = \frac{\Omega_{E/F} \cdot \mathrm{Reg}_F(E(F)) \cdot \#Ш(E/F) \cdot \prod_v c_{v,F}}{\sqrt{|D_F|} \cdot \#E(F)_{\mathrm{tor}}^2},$$

where $D_F \in \mathbb{Z}$ is the discriminant of F, and the other quantities are as in [Lan91, III, §5].

If E is defined over $\mathbb{Q}$ and F is totally imaginary, as it is in our application in which $F = K$ or $F = K_c$, we have $\Omega_{E/F} = \|\omega_E\|^{[F:\mathbb{Q}]}$, where $\|\omega_E\|$ is as in Equation (1) (see also [GZ86, §6]).

Much of the rest of this section is devoted to proving the following theorem.

Theorem 13. *Assume Conjectures 6 and 12 for E, that $\mathrm{ord}_{s=1} L(E/K, \chi, s) = 1$ for each nontrivial ring class character χ of conductor dividing c, and that $a_c y_K$ is torsion. Let $r = r_{\mathrm{an}}(E/K) = \mathrm{ord}_{s=1} L(E/K, s)$ and assume that $r = \mathrm{rank}(E(K))$, as predicted by Conjecture 1. Then*

$$[E(K_c)_{/\mathrm{tor}} : (E(K)+W)_{/\mathrm{tor}}]^2 = \frac{\#\text{Ш}(E/K_c)}{\#\text{Ш}(E/K)} \cdot \frac{\prod_w c_{w,K_c}}{\prod_v c_{v,K}} \cdot \frac{\#E(K)_{\mathrm{tor}}^2}{\#E(K_c)_{\mathrm{tor}}^2} \cdot h_c^{r-1} \cdot u^{2h_c}.$$

Because of the the Cassels-Tate pairing, we expect that $\#\text{Ш}(E/K_c)$ and $\#\text{Ш}(E/K)$ are both perfect squares (see, e.g., [Ste04, Thm. 1.2]). The following Proposition is thus an important consistency check for Theorem 13.

Proposition 14. *Theorem 13 predicts that $\frac{\#\text{Ш}(E/K_c)}{\#\text{Ш}(E/K)}$ is a perfect square.*

Proof. We check that each factor, except the quotient of Shafarevich-Tate groups appearing in the theorem, is a perfect square, especially the Tamagawa number factors. Each prime of bad reduction for E splits in K, and for the two primes v and v' over a split prime p of $\mathbb{Q}$, we have $c_{v,K} = c_{v',K}$, so

$$\prod_v c_{v,K} = \left(\prod_{p|N} c_{p,\mathbb{Q}}\right)^2.$$

The extension K_c/K is unramified at each prime of bad reduction for E, and the formation of Néron models commutes with unramified base change (see [BLR90, §1.2, Prop. 2]), so for each prime v of K and each prime w of K_c with $w \mid v$, we have $c_{w,K_c} = c_{v,K}$. Let g_v be the number of primes of K_c over the prime v of K. Then

$$\prod_{w \text{ of } K_c} c_{w,K_c} = \prod_{v \text{ of } K} c_{v,K}^{g_v} = \prod_{p|N} c_{p,\mathbb{Q}}^{2g_v} = \left(\prod_{p|N} c_{p,\mathbb{Q}}^{g_v}\right)^2.$$

Finally, the factor h_c^{r-1} is a perfect square since the sign of the functional equation for $L(E/K, s)$ is odd, so r is odd. □

Lemma 15. *With hypotheses as in Theorem 13, $L(E/K_c, s)$ vanishes to order exactly $r + h_c - 1$ and*

$$\frac{L^{(r+h_c-1)}(E/K_c, 1)}{(r + h_c - 1)!} = \frac{L^{(r)}(E/K, 1)}{r!} \cdot \prod_{\chi \neq \chi_1} L'(E/K, \chi, 1). \tag{7}$$

Proof. The L-function of E over K_c factors as

$$L(E/K_c, s) = \prod_{\chi} L(f, \chi, s) = L(E/K, s) \cdot \prod_{\chi \neq \chi_1} L(f, \chi, s),$$

where the first product is over characters $\chi : G \to \mathbb{C}^\times$, and χ_1 is the trivial character. This implies the order of vanishing statement. The leading coefficient of the product of power series is the product of the leading coefficients of those series, which gives the formula for the leading coefficient. □

In using Conjecture 12 to deduce Theorem 13, we will make use of an explicit formula for the discriminant D_{K_c}.

Lemma 16. *We have*

$$D_{K_c} = D_K^{h_c} \cdot \prod_{p \mid c} p^{\frac{2 \cdot p \cdot h_c}{p+1}}$$

Proof. Consider a prime divisor $p \mid c$, and write $c = pc'$. The prime $p\mathcal{O}_K$ above p splits completely in $K_{c'}/K$ (as explained in [Ste10b, Lem. 5.3]). Going from $K_{c'}$ to K_c, the primes above $p\mathcal{O}_K$ are totally ramified, with ramification index $[K_c : K_{c'}] = [K_p : K_1] = p+1$. Combining this information for all $p \mid c$ and applying [FT93, Thm. 26, Ch. III], implies that the different $\delta_{K_c/K}$ is $\prod_{p|c} \prod_{\mathfrak{p}|p} \mathfrak{p}^p$. Let $\mathfrak{p}$ be any prime of K_c over p. As explained above, since p is inert in $K/\mathbb{Q}$, the prime $p\mathcal{O}_K$ splits completely in $K_{c'}/K$, then totally and tamely ramifies in K_c/K_c', so $\mathrm{norm}_{K_c/\mathbb{Q}}(\mathfrak{p}) = p^2$, and the number of primes $\mathfrak{p}$ over a given p is $h_c/(p + 1)$. The different ideal is multiplicative in towers, and the discriminant is the norm of the different, so

$$\begin{aligned} D_{K_c} &= \mathrm{norm}_{K_c/\mathbb{Q}}(\delta_{K_c/\mathbb{Q}}) \\ &= \mathrm{norm}_{K_c/\mathbb{Q}}(\delta_{K/\mathbb{Q}} \cdot \delta_{K_c/K}) \\ &= \mathrm{norm}_{K_c/\mathbb{Q}}(\delta_{K/\mathbb{Q}}) \cdot \prod_{p|c} \prod_{\mathfrak{p}|p} \mathrm{norm}_{K_c/\mathbb{Q}}(\mathfrak{p})^p \\ &= D_K^{h_c} \cdot \prod_{p|c} p^{\frac{2h_c p}{p+1}}. \end{aligned}$$

□

The product of prime divisors of c in Lemma 16 can be expressed in terms of conductors as follows:

Lemma 17. *We have*

$$D_{K_c} = D_K^{h_c} \cdot \prod_{\chi \neq \chi_1} \operatorname{cond}(\chi)^2. \tag{8}$$

Proof. Consider the set of characters $\chi : G \to \mathbb{C}^\times$. A character χ has conductor not divisible by p precisely if it factors through $\operatorname{Gal}(K_{c'}/K)$, so the number of characters χ with conductor not divisible by p is the number of characters of $\operatorname{Gal}(K_{c'}/K)$, which is $\#\operatorname{Gal}(K_{c'}/K) = h_c/(p+1)$. Thus the number of characters with conductor divisible by p is $h_c - h_c/(p+1)$. As $\operatorname{cond}(\chi) \mid c$ we have

$$\prod_{\chi \neq \chi_1} \operatorname{cond}(\chi) = \prod_{p \mid c} p^{h_c - h_c/(p+1)} = \prod_{p \mid c} p^{h_c p/(p+1)},$$

which, combined with Lemma 16, implies the claimed formula. □

We will use the following lemma in computing a certain regulator in the proof of Proposition 19 below.

Lemma 18. *Let $M_m(a,b)$ be the $m \times m$ matrix with $a+b$ along the diagonal and all other entries equal to b. Then* $\det M_m(a,b) = (a+mb)a^{m-1}$.

Proof. The case for $m = 1, 2$ is clear. For $m > 2$, first consider the determinant of the matrix $M'_m(a,b)$ of size $m \times m$ whose entries are all b except for the first upper off diagonal whose entries are all $a+b$ (see Equation (9) below). We claim that $\det M'_m(a,b) = (-a)^{m-1}b$. For $m = 1, 2$ this is clear. For larger m we perform a row operation (subtract row 2 from row 1) and expand by minors, as follows:

$$\det M'_m(a,b) = \begin{vmatrix} b & a+b & \cdots & b \\ b & b & \ddots & \vdots \\ \vdots & & \ddots & a+b \\ b & \cdots & b & b \end{vmatrix} = \begin{vmatrix} 0 & a & \cdots & 0 \\ b & b & \ddots & \vdots \\ \vdots & & \ddots & a+b \\ b & \cdots & b & b \end{vmatrix} \tag{9}$$

$$= -a \cdot \det M'_{m-1}(a,b) = -a(-a)^{m-2}b = (-a)^{m-1} \cdot b. \tag{10}$$

Using this formula for $\det M'_m(a,b)$ allows us to compute $\det M_m(a,b)$ as follows, where in the first step we subtract the last row from the first row:

$$\det M_m(a,b) = \begin{vmatrix} a+b & b & \cdots & b \\ b & a+b & & \vdots \\ \vdots & & \ddots & b \\ b & \cdots & b & a+b \end{vmatrix} = \begin{vmatrix} a & 0 & \cdots & -a \\ b & a+b & & \vdots \\ \vdots & & \ddots & b \\ b & \cdots & b & a+b \end{vmatrix}$$
$$= a \cdot \det M_{m-1}(a,b) \;+\; (-1)^m(-a)\det M'_{m-1}(a,b)$$
$$= (a+mb)\cdot a^{m-1}.$$

□

Proposition 19. *With hypotheses as in Theorem 13 (but without assuming any conjectures!), we have*

$$\mathrm{Reg}_{K_c}(W) = h_c^{h_c-2} \cdot \prod_{\chi\neq\chi_1} \hat{h}_{K_c}(y_{c,\chi}).$$

Proof. In this proof we will work everywhere with the images of points in $V = E(K_c)\otimes\mathbb{C}$, which should not cause confusion.

The hypotheses imply that for each nontrivial character χ, the point $y_{c,\chi}$ has infinite order. Lemma 5 asserts that the $y_{c,\chi}$ are mutually orthogonal, so there is a lattice Λ in $W\otimes\mathbb{C}$ with basis the $y_{c,\chi}$, which has rank $h_c - 1$ (the number of nontrivial characters χ). Because the $y_{c,\chi}$ are all nonzero and orthogonal, we have

$$\mathrm{Reg}_{K_c}(\Lambda) = \prod_{\chi\neq\chi_1} \hat{h}_{K_c}(y_{c,\chi}).$$

By Proposition 10, the elements $(y_c^\sigma)_{1\neq\sigma\in G}$ are independent and nonzero, so they form a basis for their $\mathbb{Z}$-span $W_{/\mathrm{tor}}$ in V. Let M be the $(h_c-1)\times(h_c-1)$ change of basis matrix with respect to these two bases. More precisely, if for any fixed basis of V, we let B_Λ be the matrix with rows our chosen basis for Λ and B_W the matrix with rows our basis for W, then $B_\Lambda = M\cdot B_W$. We have $\mathrm{Reg}_{K_c}(\Lambda) = \det(M)^2\cdot\mathrm{Reg}_{K_c}(W)$, so to compute $\mathrm{Reg}_{K_c}(W)$, it suffices to compute $\det(M)^2$. By definition of e_χ and using that $\mathrm{Tr}_{K_c/K}(y_c) = 0$ (in V) we have

$$y_{c,\chi} = \frac{1}{h_c}\sum_{\sigma\in G}\chi^{-1}(\sigma)y_c^\sigma = \frac{1}{h_c}\sum_{1\neq\sigma\in G}(\chi^{-1}(\sigma)-1)y_c^\sigma,$$

from which we read off the rows of the matrix M. For any two rows M_{χ_i}, M_{χ_j} of M,

$$\begin{aligned} M_{\chi_i} \cdot M_{\chi_j} &= \frac{1}{h_c^2} \sum_{1 \neq \sigma \in G} (\chi_i^{-1}(\sigma) - 1)(\chi_j^{-1}(\sigma) - 1) \\ &= \frac{1}{h_c^2} \sum_{\sigma \in G} (\chi_i^{-1}(\sigma) - 1)(\chi_j^{-1}(\sigma) - 1) \\ &= \frac{1}{h_c^2} \sum_{\sigma \in G} (\chi_i \chi_j)^{-1}(\sigma) - \chi_i^{-1}(\sigma) - \chi_j^{-1}(\sigma) + 1 = \begin{cases} \frac{2}{h_c} & \text{if } \chi_i = \chi_j^{-1}, \\ \frac{1}{h_c} & \text{otherwise.} \end{cases} \end{aligned}$$

Thus

$$(\det M)^2 = \det MM^T = \det(M_{\chi_i} \cdot M_{\chi_j})_{i,j} = \pm \begin{vmatrix} \frac{2}{h_c} & \frac{1}{h_c} & \cdots & \frac{1}{h_c} \\ \frac{1}{h_c} & \frac{2}{h_c} & & \vdots \\ \vdots & & \ddots & \frac{1}{h_c} \\ \frac{1}{h_c} & \cdots & \frac{1}{h_c} & \frac{2}{h_c} \end{vmatrix},$$

where the columns in the final matrix have been permuted so we have $2/h_c$ down the diagonal and $1/h_c$ everywhere else, which only affects the determinant up to sign. To evaluate this determinant we use Lemma 18 with $a = b = 1/h_c$ and $m = h_c - 1$ and obtain

$$\det(M)^2 = \left(\frac{1}{h_c} + (h_c - 1) \cdot \frac{1}{h_c} \right) \cdot \left(\frac{1}{h_c} \right)^{h_c - 2} = 1/h_c^{h_c - 2}.$$

Thus

$$\mathrm{Reg}_{K_c}(W) = (\det M)^{-2} \cdot \mathrm{Reg}_{K_c}(\Lambda) = h_c^{h_c - 2} \cdot \prod_{\chi \neq \chi_1} \hat{h}_{K_c}(y_{c,\chi}).$$

□

Proof of Theorem 13. Apply Conjecture 12 to the left hand side of Equation (7), and to the first factor on the right hand side, and Conjecture 6 to the remaining factors on the right hand side, to get

$$\frac{\|\omega_f\|^{2h_c} \cdot \mathrm{Reg}_{K_c}(E(K_c)) \cdot \#\text{Ш}(E/K_c) \cdot \prod c_{w,K_c}}{\sqrt{|D_{K_c}|} \cdot \#E(K_c)_{\mathrm{tor}}^2}$$

$$= \frac{\|\omega_f\|^2 \cdot \mathrm{Reg}_K(E(K)) \cdot \#Ш(E/K) \cdot \prod c_{v,K}}{\sqrt{|D_K|} \cdot \#E(K)_{\mathrm{tor}}^2} \cdot \prod_{\chi \neq \chi_1} \frac{h_c \cdot \|\omega_f\|^2}{\mathrm{cond}(\chi) \cdot u^2 \cdot \sqrt{|D_K|}} \cdot \hat{h}_{K_c}(y_{c,\chi}).$$

Cancelling $\|\omega_f\|^{2h_c}$ from both sides, and rearranging factors gives

$$u^{2h_c} \cdot \frac{\sqrt{|D_K|^{h_c}} \cdot \prod_{\chi \neq \chi_1} \mathrm{cond}(\chi)}{\sqrt{|D_{K_c}|}} \cdot \frac{\prod c_{w,K_c}}{\prod c_{v,K}} \cdot \frac{\#Ш(E/K_c)}{\#Ш(E/K)}$$

$$= \frac{\mathrm{Reg}_K(E(K)) \cdot h_c^{h_c-1} \cdot \prod_{\chi \neq \chi_1} \hat{h}_{K_c}(y_{c,\chi})}{\mathrm{Reg}_{K_c}(E(K_c))} \cdot \frac{\#E(K_c)_{\mathrm{tor}}^2}{\#E(K)_{\mathrm{tor}}^2}. \tag{11}$$

We have $r = \mathrm{rank}(E(K))$, because we are assuming Conjecture 1 for E/K, and Proposition 11 implies that W and $E(K)$ are orthogonal, so

$$\mathrm{Reg}_{K_c}(E(K)+W) = \mathrm{Reg}_{K_c}(E(K)) \cdot \mathrm{Reg}_{K_c}(W) = h_c^r \cdot \mathrm{Reg}_K(E(K)) \cdot \mathrm{Reg}_{K_c}(W). \tag{12}$$

Combining Equation (12) with Proposition 19 yields

$$\mathrm{Reg}_K(E(K)) \cdot h_c^{h_c-1} \cdot \prod_{\chi \neq \chi_1} \hat{h}_{K_c}(y_{c,\chi}) = \mathrm{Reg}_K(E(K)) \cdot h_c \cdot \mathrm{Reg}_{K_c}(W)$$

$$= h_c^{1-r} \cdot \mathrm{Reg}_{K_c}(E(K) + W).$$

Taking square roots of the absolute value of both sides of the formula in Lemma 17 and simplify Equation (11) using the above, we obtain

$$u^{2h_c} \cdot \frac{\prod c_{w,K_c}}{\prod c_{v,K}} \cdot \frac{\#Ш(E/K_c)}{\#Ш(E/K)} = h_c^{1-r} \cdot \frac{\mathrm{Reg}_{K_c}(E(K)+W)}{\mathrm{Reg}_{K_c}(E(K_c))} \cdot \frac{\#E(K_c)_{\mathrm{tor}}^2}{\#E(K)_{\mathrm{tor}}^2}$$

$$= h_c^{1-r} \cdot [E(K_c)_{/\mathrm{tor}} : (E(K)+W)_{/\mathrm{tor}}]^2 \cdot \frac{\#E(K_c)_{\mathrm{tor}}^2}{\#E(K)_{\mathrm{tor}}^2}.$$

Solving for $[E(K_c)_{/\mathrm{tor}} : (E(K)+W)_{/\mathrm{tor}}]^2$ then yields the claimed formula in Theorem 13. □

If we remove the cond(χ) factor from Conjecture 6, then rederive Theorem 13 as in the proof above, the one change is that in Equation (11), instead of having

$$\frac{\sqrt{|D_K|^{h_c}} \cdot \prod_{\chi \neq \chi_1} \mathrm{cond}(\chi)}{\sqrt{|D_{K_c}|}} = 1$$

we get an extra factor of

$$\frac{\sqrt{|D_K|^{h_c}}}{\sqrt{|D_{K_c}|}}$$

next to u^{2h_c}. According to Lemma 16, we have

$$\frac{\sqrt{|D_{K_c}|}}{\sqrt{|D_K|^{h_c}}} = \prod_{p|c} p^{\frac{ph_c}{p+1}}.$$

In the special case when $c = p$ is an odd prime and K has class number 1, this simplifies to

$$\frac{\sqrt{|D_{K_c}|}}{\sqrt{|D_K|^{h_c}}} = p^{\frac{p(p+1)}{p+1}} = p^p,$$

which is never a perfect square, which leads to a contradiction (see Proposition 14).

5. An Example

Suppose E is the elliptic curve 389a given by $y^2 + y = x^3 + x^2 - 2x$, which has rank 2 and conductor 389. The field $K = \mathbb{Q}(\sqrt{-7})$ satisfies the Heegner hypothesis, $c = 5$ is inert in K, and $u = 1$. Since K has class number 1, we have $h_c = c + 1 = 6$. According to [JLS09], the field K_c is obtained by adjoining a root of

$$z^6 + 1750z^5 - 26551875z^4 - 570237500z^3 + 202540106562500z^2$$
$$- 2921132756718750000z + 1345371129783105468 75$$

to K, and we find by computer calculation (or Lemma 16) that

$$D_{K_c} = 5^{10} \cdot 7^6 = (-7)^6 5^{(2\cdot 5\cdot 6)/(5+1)}.$$

All of the p-adic Galois representations associated to E are surjective, so $E(K_c)_{\mathrm{tor}} = 0$. The BSD conjecture and a computation using [S$^+$11] implies that $\text{Ш}(E/K) = 1$, and we find by computation that $r = r_{\mathrm{an}}(E/K) = 3$. The Tamagawa numbers of E at 389 is 1. Assuming the hypothesis of Theorem 13 are satisfied, we have

$$[E(K_5) : E(K) + W]^2 = \#\text{Ш}(E/K_5) \cdot 6^2. \tag{13}$$

Let σ be a choice of generator for $G = \mathrm{Gal}(K_5/K)$. As explained in [JLS09, Ste10a], the Kolyvagin class $\tau \in \mathrm{H}^1(K, E[3])$ associated to y_5 is nonzero and $\text{Ш}(E/K)[3] = 0$, so there is some nonzero $P \in E(K)/3E(K)$ such that $[P] \mapsto [P_5] \in E(K_5)/3E(K_5)$, where $P_5 = \sum i\sigma^i(y_5) \in W$. Thus $P - P_5 = 3Q \in 3E(K_5)$, where $Q \in E(K_5)$ but $Q \notin E(K) + W$. Hence $3 \mid [E(K_5) : E(K) + W]$, as predicted by Equation (13).

6. Ideas for Future Work

It would be of interest to compute the relevant L-functions in this paper for several specific examples, using the methods of Dokchitser [Dok04] or Rubinstein. In addition, one could explicitly compute the Mordell-Weil group $E(K_c)$ in some examples. It would also be of interest to find explicit examples that illustrate the situation discussed in Remark 9, in which $\mathrm{ord}_{s=1} L(E, \chi, s) \geq 3$, since we are currently not aware of any such examples.

Regarding generalizations, it would be natural to fully treat the case when $r = 1$, so that W has finite index in $E(K_c)$. It would also be good to extend the results of this paper to modular abelian varieties A_f attached to newforms in $S_2(\Gamma_0(N))$. Another possible generalization would be to quadratic imaginary fields that do not satisfy the Heegner hypothesis, so the modular curve $X_0(N)$ is replaced by a Shimura curve (see, e.g., the extensive work of Bertolini and Darmon). In another direction, one could likely generalize our results to elliptic curves (or abelian varieties) over totally real fields, following the program initiated by Zhang in [Zha01b].

Assume that for all nontrivial χ we have $\mathrm{ord}_{s=1} L(E, \chi, s) = 1$. Under this hypothesis, it would be of great interest to prove the divisibility

$$\frac{\#\text{Ш}(E/K_c)}{\#\text{Ш}(E/K)} \;\Big|\; [E(K_c) : E(K) + W]^2,$$

at least away from an explicit finite list of primes. This might make it possible to compute $\text{Ш}(E/K_c)/\text{Ш}(E/K)$ for a specific elliptic curve. This would be a generalization of the explicit upper bounds on $\#\text{Ш}(E/K)$ from [GJP$^+$09, Thm. 3.4]. The cryptic [Ber10, Remark 5.23(1)] is relevant, because it claims one can prove at least finiteness of $\text{Ш}(E/K_c)(\chi)$, in the Shimura curve case, though warns "The original methods of Kolyvagin, based on the Gross-Zagier formula, allow to prove a similar statement only when χ is quadratic." This should be contrasted with [YZZ10, §1.6, Thm. C], where it is claimed that under our hypothesis Tian-Zhang have in fact proved that $\text{Ш}(E/K_c)(\chi)$ is finite, using the original method of Kolyvagin based on their generalization of the Gross-Zagier formula.

References

[ARS06] Amod Agashe, Kenneth Ribet, and William A. Stein, *The Manin constant*, Pure Appl. Math. Q. **2** (2006), no. 2, part 2, 617–636, `http://wstein.org/papers/ars-manin/`. MR 2251484 (2007c:11076)

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic), `http://math.stanford.edu/~conrad/papers/tswfinal.pdf`. MR 2002d:11058

[BD90] Massimo Bertolini and Henri Darmon, *Kolyvagin's descent and Mordell-Weil groups over ring class fields*, J. Reine Angew. Math. **412** (1990), 63–74, `http://www.math.mcgill.ca/darmon/pub/Articles/Research/04.Kolyvagin-descent/paper.pdf`. MR 1079001 (91j:11048)

[Ber10] Massimo Bertolini, *Report on the Birch and Swinnerton-Dyer conjecture*, Milan J. Math. **78** (2010), 153–178, `http://newrobin.mat.unimi.it/users/mbertoli/report.bsd.pdf`. MR 2684777

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR 91i:14034

[Cha05] Byungchul Cha, *Vanishing of some cohomology groups and bounds for the Shafarevich-Tate groups of elliptic curves*, J. Number Theory. **111** (2005), 154–178, `http://wstein.org/papers/bib/cha-vanishing_of_some_cohomology_groups_and_bounds_for_the_Shafarevich-Tate_groups_of_elliptic_curves.pdf`.

[Coh07] Henri Cohen, *Number theory. Vol. I. Tools and Diophantine equations*, Graduate Texts in Mathematics, vol. 239, Springer, New York, 2007. MR 2312337 (2008e:11001)

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149, `http://arxiv.org/abs/math/0207280`. MR 2068888 (2005f:11128)

[FT93] A. Fröhlich and M. J. Taylor, *Algebraic number theory*, Cambridge Studies in Advanced Mathematics, vol. 27, Cambridge University Press, Cambridge, 1993. MR 1215934 (94d:11078)

[GJP$^+$09] G. Grigorov, A. Jorza, S. Patrikis, C. Tarnita, and W. Stein, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. **78** (2009), 2397–2425, `http://wstein.org/papers/bsdalg/`.

[Gro84] Benedict H. Gross, *Heegner points on $X_0(N)$*, Modular forms (Durham, 1983), Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., Horwood, Chichester, 1984, `http://wstein.org/papers/bib/gross-heegner_points_on_X0N.pdf`, pp. 87–105. MR 803364 (87f:11036b)

[Gro91] B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, `http://wstein.org/papers/bib/gross-kolyvagins_work_on_modular_elliptic_curves.pdf`, pp. 235–256.

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320, `http://wstein.org/papers/bib/Gross-Zagier_Heegner_points_and_derivatives_of_Lseries.pdf`. MR 87j:11057

[Hay95] Yoshiki Hayashi, *The Rankins L-function and Heegner points for general discriminants.*, Proc. Japan Acad. Ser. A Math. Sci. (1995), no. 71(2), 30–32, `http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.pja/1195510808`.

[Jet08] Dimitar Jetchev, *Global divisibility of Heegner points and Tamagawa numbers*, Compos. Math. **144** (2008), no. 4, 811–826, `http://arxiv.org/abs/math/0703431`. MR 2441246 (2010b:11072)

[JLS09] Dimitar Jetchev, Kristin Lauter, and William Stein, *Explicit Heegner points: Kolyvagin's conjecture and non-trivial elements in the Shafarevich-Tate group*, J. Number Theory **129** (2009), no. 2, 284–302, `http://wstein.org/papers/kolyconj/`. MR 2473878 (2009m:11080)

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR 93a:11048

[Mil10] Robert L. Miller, *Proving the Birch and Swinnerton-Dyer conjecture for specific elliptic curves of analytic rank zero and one*, `http://arxiv.org/abs/1010.2431`, 2010.

[MS10] Robert L. Miller and Michael Stoll, *Explicit isogeny descent on elliptic curves*, `http://arxiv.org/abs/1010.3334`, 2010.

[S$^+$11] W. A. Stein et al., *Sage Mathematics Software (Version 4.6.2)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[Ste04] W. A. Stein, *Shafarevich-Tate Groups of Nonsquare Order*, Modular Curves and Abelian Varieties, Progress of Mathematics (2004), 277–289, `http://wstein.org/papers/nonsquaresha/`.

[Ste10a] William Stein, *Heegner points on rank two elliptic curves*, `http://wstein.org/papers/kolyconj2/`.

[Ste10b] ______, *Toward a Generalization of the Gross-Zagier Conjecture*, Internat. Math. Res. Notices (2010), `http://wstein.org/papers/stein-ggz/`.

[SW11] William Stein and Christian Wuthrich, *Computations About Tate-Shafarevich Groups Using Iwasawa Theory*, in preparation (2011), `http://wstein.org/papers/shark/`.

[Wat06] Mark Watkins, *Some remarks on Heegner point computations*, Preprint (2006), `http://arxiv.org/abs/math/0506325`.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551, `http://users.tpg.com.au/nanahcub/flt.pdf`.

[Wil00] ———, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[YZZ10] X. Yuan, S. Zhang, and W. Zhang, *Gross-Zagier formula*, `http://www.math.columbia.edu/~yxy/preprints/GZ.pdf`.

[Zha01a] Shou-Wu Zhang, *Gross-Zagier formula for* GL_2, Asian J. Math. **5** (2001), no. 2, 183–290, `http://intlpress.com/AJM/p/2001/5_2/AJM-5-2-183-290.pdf`. MR 1868935 (2003k:11101)

[Zha01b] ———, *Heights of Heegner points on Shimura curves*, Ann. of Math. (2) **153** (2001), no. 1, 27–147. MR 1826411 (2002g:11081)

[Zha04] ———, *Gross-Zagier formula for* GL(2)*. II*, Heegner points and Rankin L-series, Math. Sci. Res. Inst. Publ., vol. 49, Cambridge Univ. Press, Cambridge, 2004, `http://www.math.columbia.edu/~szhang/papers/gzes.pdf`, pp. 191–214. MR 2083213

41 Kolyvagin's Conjecture for Some Specific Higher Rank Elliptic Curves

Kolyvagin's Conjecture for Specific Higher Rank Elliptic Curves

William Stein*

March 26, 2011

Abstract

We study Heegner points and Kolyvagin classes for elliptic curves over $\mathbb{Q}$, with special focus on curves that have analytic rank at least 2. We reinterpret Kolyvagin's "derived classes" construction in the context of divisors on modular curves directly in characteristic ℓ, and prove compatibility and multiplicity one results. We use these results to give the first complete algorithm for explicitly computing (certain) Kolyvagin classes, and thus verify a conjecture of Kolyvagin for some specific elliptic curves.

1 Introduction

A *higher rank* elliptic curve is an elliptic curve E over $\mathbb{Q}$ of analytic rank at least 2. Let K be a quadratic imaginary field such that each prime dividing the conductor of E splits in K. This paper is about the Galois cohomology classes $\tau_{c,p^n} \in \mathrm{H}^1(K, E[p^n])$ defined by Kolyvagin (see, e.g., [Kol88a, Gro91, McC91]). Our main motivation is the explicit study of these classes on higher rank elliptic curves, inspired by the results of [Ste10, BS11] and open conjectures of Kolyvagin (see [Kol91, ÇW08]). In particular, consider Conjecture A of [Kol91, pg. 255]:

Conjecture 1.1 (Kolyvagin)**.** *For each prime p, there is some n and squarefree product $c = \prod p_i$ of primes that are inert in K with $p^n \mid \gcd(a_{p_i}, p_i + 1)$ such that $\tau_{c,p^n} \neq 0$.*

For elliptic curves with analytic rank ≤ 1 over K, this conjecture with $c = 1$ follows from [GZ86], but for higher rank curves the conjecture is wide open, and we have only computational data.

The goal of this paper is to shed some light on Conjecture 1.1 by making it more explicit and computing many examples, as follows. Let p^n and c be as in Conjecture 1.1 We adapt Kolyvagin's construction to define elements in $E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z})$, then give an algorithm to compute these elements in many cases. When one of these elements is nonzero, the corresponding Kolyvagin cohomology class τ_{c,p^n} is also nonzero, which allows us to verify, in several specific examples, Conjecture 1.1. This is significant because until now this conjecture had not been verified in even a single case. In particular, we present a powerful and fairly general approach to explicitly computing information about particular classes $\tau_{c,p^n} \in \mathrm{H}^1(K, A_f[p^n])$ for a modular abelian varieties A_f, squarefree integer c and prime power p^n. Thus, our results provide further motivation and much needed tools for studying Heegner points in the context of higher rank elliptic curves

*The work was supported by NSF grant 0555776 and the Clay Mathematics Institute.

and modular abelian varieties. Moreover, we provide new algorithms for computing with Selmer groups of elliptic curves, which exploit different methods than explicit n-descent for small n (see [Cre97, §3.5] and [CFO$^+$08]) or explicit Iwasawa theory as in [SW11].

Our approach is inspired by groundbreaking work of Cornut, Vatsal, Gross, Jetchev-Kane, and Mazur (see [JK10, Cor02, Vat02]), in which they establish nontriviality results about Heegner points. Our new idea is simple: *use rational quaternion algebras to give an explicit description of the Kolyvagin derived classes construction modulo an auxiliary prime ℓ that is inert in the quadratic imaginary field K* (see Section 6). Many of the objects we use play a central role in the work of Cornut mentioned above. We hope that some of our techniques may also be useful for exploring and refining other ideas related to extra structure on higher rank elliptic curves arising from Heegner points.

The Birch and Swinnerton-Dyer conjectural rank formula (see Conjecture 3.1 below) asserts that $\mathrm{ord}_{s=1} L(E,s) = \mathrm{rank}(E(\mathbb{Q}))$. This conjecture is a theorem when E is an elliptic curve over $\mathbb{Q}$ of analytic rank ≤ 1 (see [BCDT01, GZ86, Kol88b] and Theorem 3.2 below). In sharp contrast, when E is a higher rank curve, the BSD conjecture remains shrouded in mystery, as do potential generalizations of the Gross-Zagier formula (see, e.g., [Ste10]). Unfortunately, the many exciting generalizations of the Gross-Zagier formula to other settings (see [BY09, Zha01, Zha04, YZZ11]) so far seem to yield little new insight in the higher rank case. As explained in [Ste10], Kolyvagin classes are potentially relevant to a search for a generalization of the Gross-Zagier formula that treats higher derivatives. Such a generalization is an incredibly difficult open problem and anything that might shed light on it is worth investigating. So far, finding a plausibly-provable conjecture has remained elusive.

The explicit examples in Section 8 involve rank 2 curves (instead of curves of rank ≥ 3), since the notation and computations are substantially simpler when the rank is 2. The theory and algorithms we develop apply to elliptic curves of any rank, and also to modular abelian varieties. It is thus possible to study many more general situations using our approach (see Section 9).

This paper is structured as follows. In Section 2 we give an outline of our main algorithm. Next in Section 3 we recall the BSD conjecture and give some examples, which motivate our paper. In Section 4 we recall the definition of Heegner points. In Section 5 we introduce Kolyvagin classes, make some observations, and discuss reduction of Heegner points modulo a prime over ℓ. In Section 6 we make the action of Galois on certain objects in characteristic ℓ more explicit and prove a compatibility result. In Section 7 we explain in more detail how our algorithm for computing reductions of Kolyvagin classes works. We combine our above results to obtain an algorithm to compute Kolyvagin classes, which we apply in Section 8, in which we discuss the implementation of our algorithm, tables we obtained by running it, and state some results inspired by this data. Finally, Section 9 discusses a range of related future projects.

Acknowledgement: The author would like to thank Jennifer Balakrishnan, Ralph Greenberg, Benedict Gross, Ben Howard, David Kohel, Dimitar Jetchev, Barry Mazur, Ken Ribet, Karl Rubin, Justin Walker, and Jared Weinstein for helpful discussions.

1.1 Notation and terminology

We use $\cong$ to denote a canonical isomorphism and $\approx$ to denote a noncanonical one. Unless otherwise stated, all tensor products are over $\mathbb{Z}$. We always let p, q, ℓ denote *odd* prime numbers, E an elliptic curve over $\mathbb{Q}$, and K a quadratic imaginary field such that each prime dividing the conductor N of E splits in K. Let a_n denote the nth Dirichlet series coefficient of the L-series $L(E/\mathbb{Q}, s)$ associated to E.

2 Reducing Kolyvagin Classes

As above, let E be an elliptic curve over $\mathbb{Q}$, let K be a quadratic imaginary field such that each prime dividing the conductor N of E splits in K, let p^n be an odd prime power. Let c be a squarefree product of primes that are inert in K such that for each prime $q \mid c$ we have $p^n \mid \gcd(a_q, q+1)$, where $a_q = q + 1 - \#E(\mathbb{F}_q)$. Let K_c be the ring (not ray!) class extension of K associated to c, and let σ_i be a choice of generator of $\mathrm{Gal}(K_c/K_{c/p_i})$ for each prime divisor $p_i \mid c$, and let $\sigma = (\ldots, \sigma_i, \ldots)$. As explained in Section 5 below, Kolyvagin uses Heegner points to construct a point $P_{c,\sigma} \in E(K_c)$ such that $[P_{c,\sigma}] \in (E(K_c) \otimes \mathbb{Z}/p^n\mathbb{Z})^{\mathrm{Gal}(K_c/K)}$. Under suitable hypothesis on p (e.g., the p-adic representation $\rho_{E,p}$ is surjective), Kolyvagin then uses $P_{c,\sigma}$ to define a cohomology class $\tau_{c,p^n} \in \mathrm{H}^1(K, E[p^n])$ characterized by

$$\delta([P_{c,\sigma}]) = \mathrm{res}_{K,K_c}(\tau_{c,p^n}) \in H^1(K_c, E[p^n])^{\mathrm{Gal}(K_c/K)},$$

where δ is the connecting homomorphism of Galois cohomology. (The class τ_{c,p^n} also depends on the choice of σ, but we surpress this in our notation.)

We introduce yet another prime ℓ that is also inert in K and fix a prime λ of K_c over ℓ. Reduction modulo λ induces a homomorphism $E(K_c) \otimes \mathbb{Z}/p^n\mathbb{Z} \to E(\mathbb{F}_{\ell^2}) \otimes \mathbb{Z}/p^n\mathbb{Z}$. Using Algorithm 2.1 below when $n = 1$, we compute the image z of $[P_{c,\sigma}]$ under the reduction map. When $z \neq 0$, we conclude that $\tau_{c,p}$ is also nonzero.

Algorithm 2.1.

- INPUT: *E, K, p, ℓ, c, σ, as above.*
- OUTPUT: *The (well-defined) image of $[P_{c,\sigma}]$ in $E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p\mathbb{Z})$, via reduction modulo any prime over ℓ (it does not matter which), up to some fixed nonzero scalar that is independent of c. We can compute the image of many different $P_{c,\sigma}$ with respect to a consistent choice of map.*

1. Use rational quaternion algebras and theta series of quadratic forms to directly compute a supersingular point $\overline{x}_1 \in X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}$ that is the reduction modulo λ of a choice of Heegner point $x_1 \in X_0(N)(K_1)$. (See Section 7.1.)
2. Apply a mod ℓ analogue of Kolyvagin's construction to directly obtain the reduction $\overline{Q}_{c,\sigma}$ of the "Kolyvagin derived divisor" attached to x_c as an element of $\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}})$. (See Sections 6 and 7.2.) Computing $\overline{Q}_{c,\sigma}$ closely resembles computing the image $T_c(\overline{x}_1)$ of $\overline{x}_1$ under the Hecke operator T_c using Equation (6.1), but with an appropriate choice of weighting of each summand.
3. Use linear algebra combined with refinements of results of Cornut, Ihara and Ribet (see Section 7.4) and a multiplicity one theorem (see Theorem 7.14 below) to compute a fixed nonzero scalar multiple of the image of $\overline{Q}_{c,\sigma}$, hence of $P_{c,\sigma}$, under the homomorphism of Hecke modules

$$\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}) \otimes (\mathbb{Z}/p\mathbb{Z}) \to E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p\mathbb{Z}). \qquad (2.1)$$

Remark 2.2. We emphasize that the steps of Algorithm 2.1 can all be done purely algebraically, without recourse to any numerical approximations. This contrasts with the approach of [JLS09], which provides numerical *evidence* for Kolyvagin's conjecture in one case, *without* proof. In theory the approach of [JLS09] can likely be made rigorous, but this has not been done in practice in any case, though see [Bra10] which is a step in that direction. The approach of [JLS09] can be faster for an elliptic curve with large conductor (with c *very small*); it is much worse for large c than Algorithm 2.1 (e.g., $c > 100$ would be incredibly hard).

Remark 2.3. Suppose we are only interested in verifying that the image under (2.1) of $\overline{Q}_{c,\sigma}$ is *nonzero.* Instead of the linear algebra of Step 3, we might be able to use that (2.1) is a $\mathbb{T}$-module homomorphism, where $\mathbb{T}$ is the Hecke algebra; if $\mathbb{T}\overline{Q}_{c,\sigma}$ has sufficiently large dimension, so that it cannot be contained in the nontrivial kernel, then we are done. If we take this approach and it works, we do not need to compute (2.1) at all. However, in some cases this approach cannot work, e.g., we could run into trouble if there are other elliptic curves of larger rank also of level N.

Remark 2.4. Algorithm 2.1 only computes the reduction of $P_{c,\sigma}$ up to a fixed nonzero scalar, which is enough to show that $\delta(P_{c,\sigma}) \neq 0$. The point $P_{c,\sigma}$ could in principle be normalized by finding $P_{c,\sigma}$ exactly via a numerical computation, using [JLS09] for one choice of c for which the image of $P_{c,\sigma}$ in $E(\mathbb{F}_\ell) \otimes (\mathbb{Z}/p\mathbb{Z})$ is nonzero.

To make the steps of Algorithm 2.1 explicit and machine computable, we view $\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}})$ noncanonically as the set of right ideal classes in an Eichler order R of level N in the (unique up to isomorphism) rational quaternion algebra ramified at ℓ and ∞, which we compute as explained in [Piz80, Koh01, Koh97, Ste09]. By computing representation numbers of ternary quadratic forms associated to left orders, we find the right R-ideals I whose left order admits an optimal embedding of the ring of integers $\mathcal{O}_K$ of K; this is the trick we use to compute the reduction $\overline{x}_1 \in X_0(N)(\mathbb{F}_{\ell^2})$ of x_1 modulo a prime over ℓ without ever computing x_1 itself. Then we use $\overline{x}_1$ and a parametrization of the right ideals $J \subset I$ such that $I/J \cong (\mathbb{Z}/c\mathbb{Z})^2$ to directly compute the reduction $\overline{Q}_{c,\sigma}$ (see Theorem 7.8 below). An implementation of the algorithm is included in Sage [S$^+$11].

3 The Birch and Swinnerton-Dyer Conjecture

The BSD conjecture is the main motivation for this paper, so we spend a page recalling it and emphasizing our ignorance. First we state the conjecture, then state the main theorem about it, and finish with some remarks about a curve of rank 4 and another of rank 2.

Let E be an elliptic curve over $\mathbb{Q}$. By [BCDT01, Wil95] the L-series

$$L(E, s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

attached to E extends to a holomorphic function on all of $\mathbb{C}$, hence the nonnegative integer

$$r_{\mathrm{an}}(E/\mathbb{Q}) = \mathrm{ord}_{s=1} L(E, s) \geq 0$$

is defined. The BSD conjecture was first introduced by Birch and Swinnerton-Dyer in the 1960s motivated by computer computations, and was later formulated for abelian varieties over number fields (see [Bir65, Bir71, Mil72, Tat66, Wil00]).

Conjecture 3.1 (Birch and Swinnerton-Dyer)**.** *For any elliptic curve E defined over $\mathbb{Q}$ we have*

$$\operatorname{rank} E(\mathbb{Q}) = r_{\mathrm{an}}(E/\mathbb{Q}).$$

There is also a conjectural formula of Birch and Swinnerton-Dyer for the leading coefficient of the series expansion of $L(E,s)$ about $s = 1$ (see [Lan91, III, §5] for a general formulation). This formula has now been computationally verified in many cases; see [GJP$^+$09, Mil10] where the formula is fully proved for all curves with rank ≤ 1 and conductor ≤ 5000.

Results of Kolyvagin, Gross-Zagier, and Bump-Friedberg-Hoffstein (see, e.g., [BFH90, GZ86, Kol88b]) imply the following theorem.

Theorem 3.2. *Conjecture 3.1 is true for elliptic curves E with* $\operatorname{ord}_{s=1} L(E,s) \leq 1$.

As mentioned in the introduction, Conjecture 3.1 remains completely open when $\operatorname{ord}_{s=1} L(E,s) \geq 2$. As evidence for Conjecture 3.1, we have tables of specific rank 2 and 3 curves for which the conjecture is known (see, e.g., [Crea, SW02]), and assurances that many curves have analytic rank ≤ 1 (see [BS10]). There is not a single example of a curve of rank ≥ 4 for which the conjecture has been verified. Rank 4 is difficult not because of the complexity of doing computations, but because there is, as of now, no known *algorithm* (no matter how slow) that can be used to show that $r_{\mathrm{an}}(E/\mathbb{Q}) \geq 4$.

Example 3.3. Let E be the elliptic curve $y^2 + xy = x^3 - x^2 - 79x + 289$. A 2-descent (using [Creb, S$^+$11]) and point search proves that E has algebraic rank 4, with generators $(-9,19), (-8,23), (-7,25), (4,-7)$. Applying the methods of [Cre97, Dok04] and the Gross-Zagier formula, we see that $L(E,1) = L'(E,1) = 0$, $L''(E,1)$ is *very close* to 0, and $L^{(4)}(E,1) = 214.65233\ldots$. But showing that $L''(E,1) = 0$ (which would imply Conjecture 3.1 for E) is an unsolved problem.

Assume that E is an elliptic curve with $\operatorname{ord}_{s=1} L(E,s) = 2$. Then Conjecture 3.1 asserts that $\operatorname{rank} E(\mathbb{Q}) = 2$. In the explicit examples Section 8, the Birch and Swinnerton-Dyer formula predicts that $\#\text{Ш}(E/\mathbb{Q}) = 1$, though in fact $\text{Ш}(E/\mathbb{Q})$ is not known to be finite for any of these curves (or indeed, for any curve of rank ≥ 2). The best that has been done at present for a general rank 2 curve is to verify that $\text{Ш}(E/\mathbb{Q})[p] = 0$ for (finitely) many specific p, e.g., using the algorithm of [SW11]. See the recent work of [CLS09, CLS10] on CM elliptic curves of rank 2. Also, for the rank 2 elliptic curve of conductor 389, the author used modular symbols, p-adic L-series, p-adic heights, Iwasawa theory, and results of Kato and Schneider to show that $\text{Ш}(E/\mathbb{Q})[p] = 0$ for all primes $p < 2466$, except possibly the supersingular primes $p = 107$, 599, and 1049, for which the approach of [SW11] should work, but take much longer.

4 Quadratic Imaginary Fields and Heegner Points

In this section we recall the definition of Heegner points over ring class fields, and explain how they behave under taking traces. We will use these points in the next section to construct derived Galois equivariant classes.

Let E be an elliptic curve over $\mathbb{Q}$ of conductor N, and let $\pi_E : X_0(N) \to E$ be a fixed choice of minimal modular parametrization. The main theorem of [BFH90] implies that there exists infinitely many quadratic imaginary fields $K = \mathbb{Q}(\sqrt{D})$ of discriminant $D \leq -5$ such that each prime dividing N splits in K. Fix any such K.

Fix an odd prime power p^n with $n \geq 1$. Let $c = \prod p_i$ be any product of prime numbers p_i that are each inert in K, coprime to ND, and such that

$$p^n \mid \gcd(a_{p_i},\, p_i + 1),$$

for each i. Let K_c be the ring class field associated to the conductor c. As explained in [Gro91, pg. 238], the field K_c is an abelian extension of the Hilbert class field K_1 of K, is unramified outside c, and is contained in the ray class field associated to c. Moreover, the reciprocity map of class field theory induces a canonical isomorphism

$$\mathrm{Gal}(K_c/K_1) \cong (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times, \tag{4.1}$$

where $\mathcal{O}_K$ is the ring of integer of K (see Proposition 6.2 below). Let $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$ be the order in $\mathcal{O}_K$ of conductor c. Each prime dividing N splits in K, so we can fix a *choice* $\mathfrak{n}$ of ideal in $\mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{n} \cong \mathbb{Z}/N\mathbb{Z}$.

The Heegner point associated to c is

$$x_c = \left[\Big(\mathbb{C}/\mathcal{O}_c,\;\; (\mathfrak{n} \cap \mathcal{O}_c)^{-1}/\mathcal{O}_c\Big)\right] \in X_0(N)(K_c),$$

which has image

$$y_c = \pi_E(x_c) \in E(K_c).$$

Remark 4.1. There are many possible choices of $\mathfrak{n}$ in the definition above, which are parametrized by the different choices of prime ideals of $\mathcal{O}_K$ over the prime divisors of N. These different choices are permuted by the action of the Atkin-Lehner operators. The Atkin-Lehner operators act as ± 1 on E, so y_c is well-defined up to sign, independent of the choice of $\mathfrak{n}$. See [Wat06] or [Coh07, Thm. 8.7.7] for an explicit description of the Atkin-Lehner action on Heegner points.

Motivated by the problem of constructing elements of $E(\mathbb{Q})$, it is natural to apply a trace map to y_c.

Proposition 4.2 (The Distribution Relation)**.** *We have* $\mathrm{Tr}_{K_c/K_1}(y_c) = a_c \cdot y_1 \in E(K_1)$. *More generally for each prime $q \mid c$, we have* $\mathrm{Tr}_{K_c/K_{c/q}}(y_c) = a_q \cdot y_{c/q} \in E(K_{c/q})$.

Proof. See [Gro84, §6] or [JK10, Lem. 5.2]. The key idea is that if T_c is the cth Hecke operator, then we have the following equality of divisors on $X_0(N)$:

$$T_c(x_1) = \sum_{\sigma \in \mathrm{Gal}(K_c/K_1)} \sigma(x_c).$$

To complete the proof, take the image of both sides in E and use that the Hecke operator T_c acts as a_c on E. □

Suppose E is a higher rank curve. The Gross-Zagier theorem [GZ86, §5.2] implies that the height of $\mathrm{Tr}_{K_1/K}(y_1) \in E(K)$ is a nonzero multiple of $L'(E/K, 1)$. However, $L(E/K, s) = L(E/\mathbb{Q}, s) \cdot L(E^D/\mathbb{Q}, s)$, and we assumed that $\mathrm{ord}_{s=1} L(E/\mathbb{Q}, s) \geq 2$, so $L'(E/K, 1) = 0$. Thus for all c,

$$\mathrm{Tr}_{K_c/K}(y_c) = \mathrm{Tr}_{K_1/K}(a_c y_1) \in E(K)_{\mathrm{tor}}. \tag{4.2}$$

Thus the traces of y_c are never non-torsion elements of the higher rank Mordell-group $E(\mathbb{Q})$.

5 Derived Points and Cohomology Classes, and their Reduction Modulo ℓ

In this section, we assume that p is an odd prime such that the p-adic representation $\rho_{E,p}$ is surjective.

In Section 5.1 we construct Kolyvagin's derived classes associated to Heegner points, then use these in Section 5.2 to construct Galois invariant classes. In Section 5.3 we explain how to reduce these classes modulo ℓ, and note that if the reduction is ever nonzero, then so is the class. Section 5.4 contains some consequences of nontriviality in the special case when E has analytic rank 2.

5.1 Derived points

Let p^n be a power of p, and let $c = p_1 \cdots p_t$ be a squarefree product of inert primes p_i such that $p^n \mid \gcd(a_{p_i}, p_i + 1)$. We recall the construction of Kolyvagin classes here, since it is important to emphasize the precise dependence on choice of generator of the Galois group, which impacts our algorithm. Also, we will make some remarks about this construction that appear to not be in the literature.

Let $[y_c]$ denote the image of y_c in $E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z})$. Let q be a prime divisor of c. The Galois group $\mathrm{Gal}(K_c/K_{c/q})$ is cyclic of order $q+1$. Fix a choice of generator $\sigma = \sigma_q \in \mathrm{Gal}(K_c/K_{c/q})$, let

$$P = \sum_{i=1}^{q} i\sigma^i(y_c) \in E(K_c),$$

and let $[P]$ denote the image of P in $E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z})$, so

$$[P] = \sum_{i \,\in\, \mathbb{Z}/(q+1)\mathbb{Z}} i\sigma^i([y_c]). \tag{5.1}$$

Proposition 5.1. *As above, assume that $p^n \mid \gcd(a_q, q+1)$. Then*

$$[P] \in (E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K_{c/q})}.$$

Proof. Applying our choice of generator σ of $\mathrm{Gal}(K_c/K_{c/q})$ to P, we have

$$\sigma([P]) = \sum_{i \,\in\, \mathbb{Z}/(c+1)\mathbb{Z}} \sigma i\sigma^i([y_c]) = \sum_{i \,\in\, \mathbb{Z}/(c+1)\mathbb{Z}} i\sigma^{i+1}([y_c]) \tag{5.2}$$

$$= \sum_{i \,\in\, \mathbb{Z}/(c+1)\mathbb{Z}} (i-1)\sigma^i([y_c]) = [P] - \mathrm{Tr}_{K_c/K_{c/q}}([y_c]) = [P]. \tag{5.3}$$

The first equality in (5.3) is because $p^n \mid q+1$, so we can enumerate the elements of $\mathbb{Z}/(q+1)\mathbb{Z}$ in any way we want (in fact, the notation we are using above only makes sense because $p^n \mid q+1$). The final equality in (5.3) holds since $p^n \mid a_q$ and $\mathrm{Tr}_{K_c/K_{c/q}}(y_c) = a_q y_{c/q}$, by Proposition 4.2. □

For each prime $p_i \mid c$, make a choice σ_i of generator for $\mathrm{Gal}(K_c/K_{c/p_i})$, and let $\sigma = (\sigma_1, \ldots, \sigma_t)$ be the tuple of those choices. Let

$$D_{c,\sigma} = \prod_{j=1}^{t} \sum_{i=1}^{p_j} i\sigma_j^i \in \mathbb{Z}[\mathrm{Gal}(K_c/K_1)],$$

and let

$$[P_{c,\sigma}] = \mathrm{Tr}_{K_1/K}(D_{c,\sigma}([y_c])) \in (E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K)}. \tag{5.4}$$

Remark 5.2. If we replace the hypothesis that $p^n \mid \gcd(a_q, q+1)$ with the hypothesis that $p^n \mid q+1$ and E has analytic rank ≥ 2, then we still have that $[P_{c,\sigma}] \in (E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K)}$. This is because $\mathrm{Tr}_{K_1/K}(y_1)$ is torsion and p is coprime to torsion, so the proof of Proposition 5.1 still goes through, but with an "obstruction" of $a_c y_1$, which vanishes upon taking a trace because of Equation (4.2).

Remark 5.3. The construction also generalizes if we replace the prime power p^n by the ideal I in $\mathbb{Z}$ generated by all a_q and $q+1$ for primes $q \mid c$, and we obtain

$$[P_{c,\sigma}] \in (E(K_c) \otimes (\mathbb{Z}/I))^{\mathrm{Gal}(K_c/K)}.$$

More generally, consider the modular Jacobian $J = J_0(N)$, and let I be the ideal of the Hecke algebra $\mathbb{T}$ generated by all T_q and $q+1$, for prime $q \mid c$. Then the above construction with x_c (instead of y_c) defines a class

$$[R_{c,\sigma}] = \mathrm{Tr}_{K_1/K}(D_{c,\sigma}([x_c])) \in (J(K_c) \otimes_{\mathbb{T}} (\mathbb{T}/I))^{\mathrm{Gal}(K_c/K)}$$

that maps to $[P_{c,\sigma}]$ under the natural map.

The next lemma explains how replacing σ_i by a different generator of $\mathrm{Gal}(K_c/K_{c/p_i})$ changes $[P_{c,\sigma}]$ by multiplication by an element of $(\mathbb{Z}/p^n\mathbb{Z})^\times$.

Lemma 5.4. *For every $j \in (\mathbb{Z}/(p_i+1)\mathbb{Z})^\times$, we have $[P_{c,(\ldots,\sigma_i^j,\ldots)}] = \frac{1}{j}[P_{c,\sigma}]$.*

Proof. Writing $q = p_i$ and $s = \sigma_i$, we have in $(\mathbb{Z}/(q+1)\mathbb{Z})[\mathrm{Gal}(K_c/K)]$ that

$$\sum_{i\,\in\,\mathbb{Z}/(q+1)\mathbb{Z}} is^{ji} = \sum_{i\,\in\,\mathbb{Z}/(q+1)\mathbb{Z}} \frac{i}{j}s^i = \frac{1}{j}\cdot \sum_{i\,\in\,\mathbb{Z}/(q+1)\mathbb{Z}} is^i.$$

□

Lemma 5.5. *If E has analytic rank r over $\mathbb{Q}$ and c is a product of t primes, then $\tau([P_{c,\sigma}]) = (-1)^{r+t+1}[P_{c,\sigma}]$. In particular, if $r+t$ is odd, then*

$$[P_{c,\sigma}] \in (E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/\mathbb{Q})}.$$

Proof. This is just [Gro91, Prop. 5.4(1)], which is proved by noting ([Gro91, Prop. 5.3]) that if $\tau \in \mathrm{Gal}(K_c/\mathbb{Q})$ is complex conjugation on K_c, then $\tau\sigma^i\tau = \sigma^{-i}$ for all i and we have that $\tau(y_c) = (-1)^{r+1}\sigma'(y_c) + (\text{torsion})$ for some $\sigma' \in \mathrm{Gal}(K_c/K)$. Thus $\tau([y_c]) = (-1)^{r+1}\sigma'([y_c])$, since p is coprime to any torsion. When $c = p_1 \cdots p_t$ is a product of t distinct primes, we have (using Lemma 5.4) that $\tau([P_{c,\sigma}]) = (-1)^{r+1}(-1)^t[P_{c,\sigma}]$. □

Remark 5.6. Following [How04, §1.2], we could alternatively encode the dependence on the choice of σ in a tensor product. Suppose for simplicity that c is prime. Consider the element

$$\sigma \otimes [P_{c,\sigma}] \in \mathrm{Gal}(K_c/K_1) \otimes (E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K)}.$$

This element does not depend on the choice of generator σ because for any $j \in (\mathbb{Z}/(c+1)\mathbb{Z})^\times$, if we define the element instead using the generator σ^j, by Lemma 5.4, we obtain

$$\sigma^j \otimes [P_{c,\sigma^j}] = \sigma^j \otimes \frac{1}{j}[P_{c,\sigma}] = (\sigma^j)^{1/j} \otimes [P_{c,\sigma}] = \sigma \otimes [P_{c,\sigma}],$$

where by $1/j$ we mean that element $j' \in \mathbb{Z}/p^n\mathbb{Z}$ such that $j'j = 1$. This generalizes to composite c by replacing $\mathrm{Gal}(K_c/K_1)$ by the tensor product $\bigotimes_{p_i \mid c} \mathrm{Gal}(K_c/K_{c/p_i})$.

Remark 5.7. We can define $[P_{c,\sigma}]$ without the hypothesis that each σ_i is a generator of $\mathrm{Gal}(K_c/K_{c/p_i})$. If we try to use exactly the definition given above, then the resulting $[P_{c,\sigma}]$ need not be $\mathrm{Gal}(K_c/K)$-equivariant, so we must modify the definition slightly. Let K' be the biggest subfield of K_c that is fixed by all σ_i, and let k_i (which divides p_i+1) be the order of σ_i. Let $[P] = \prod_{i=1}^{t} \sum_{j=1}^{k_i} j\sigma_i^j(y_c)$. Then the same argument as in Proposition 5.1 shows that $[P] \in (E(K_c)\otimes(\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K')}$, and we let

$$[P_{c,\sigma}] = \mathrm{Tr}_{K'/K}([P]) \in (E(K_c)\otimes(\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/K)}.$$

For example, if $c \neq 1$ and all $\sigma_i = 1$, then $[P_{c,\sigma}] = [a_c \cdot y_K] = 0$, since $p^n \mid a_c$.

For any multiple k of p^n, we have the following identity of polynomials:

$$\sum_{j=1}^{k-1} jX^j = \sum_{i=0}^{\frac{k}{p^n}-1} X^{p^n\cdot i}\cdot\left(\sum_{j=1}^{p^n-1} jX^j\right) \in (\mathbb{Z}/p^n\mathbb{Z})[X]. \tag{5.5}$$

Thus in the above construction, if we choose each σ_i to be of order exactly p^n, then we get (up to scaling by a unit) the same element $[P_{c,\sigma}]$ as if each σ_i is a generator of $\mathrm{Gal}(K_c/K_{c/p_i})$. The factorization (5.5) thus means we can alternatively view the Kolyvagin derived point construction as follows. Let K_c' be the compositum of the degree p^n subfields of each K_{p_i} for the primes $p_i \mid c$. If

$$D = \prod_{p_i\mid c}\sum_{j=1}^{p^n-1} j\sigma_i^j \in \mathbb{Z}[\mathrm{Gal}(K_c'/K_1)],$$

then

$$[P_{c,\sigma}] = \mathrm{Tr}_{K_1/K}([D(\mathrm{Tr}_{K_c/K_c'}(y_c))]).$$

5.2 Derived cohomology classes

As explained in [Gro91, §4], under our hypothesis that $\rho_{E,p}$ is surjective, the map

$$\mathrm{H}^1(K, E[p^n]) \to \mathrm{H}^1(K_c, E[p^n])^{\mathrm{Gal}(K_c/K)}$$

is an isomorphism, so $[P_{c,\sigma}]$ uniquely determines a cohomology class

$$\tau_{c,p^n} \in \mathrm{H}^1(K, E[p^n]).$$

In the rest of this short section, we make an additional observation in the special case when $r_{\mathrm{an}}(E/\mathbb{Q}) = 2$ and c is prime, since this is the situation for our data in Section 8.

Let $\mathrm{res} : \mathrm{H}^1(\mathbb{Q}, E[p^n]) \to \mathrm{H}^1(K_c, E[p^n])$ be the restriction map and δ the connecting homomorphism. Restricting res to Selmer groups, we obtain a commutative diagram:

$$\begin{tikzcd}
(E(K_c)\otimes(\mathbb{Z}/p^n\mathbb{Z}))^{\mathrm{Gal}(K_c/\mathbb{Q})} \arrow[r, hook, "\delta"] & \mathrm{Sel}^{(p^n)}(E/K_c)^{\mathrm{Gal}(K_c/\mathbb{Q})} \arrow[r] & Ш(E/K_c)[p^n]^{\mathrm{Gal}(K_c/\mathbb{Q})} \\
E(\mathbb{Q})\otimes(\mathbb{Z}/p^n\mathbb{Z}) \arrow[r, hook, "\delta"] \arrow[u, hook] & \mathrm{Sel}^{(p^n)}(E/\mathbb{Q}) \arrow[r, two heads] \arrow[u, hook, "\mathrm{res}"] & Ш(E/\mathbb{Q})[p^n] \arrow[u]
\end{tikzcd}$$

The following proposition *defines* an element τ_{c,p^n} in the Selmer group $\mathrm{Sel}^{(p^n)}(E/\mathbb{Q})$, not just in $\mathrm{H}^1(K, E[p^n])$ as above.

Proposition 5.8. *If c is prime and $r_{\mathrm{an}}(E/\mathbb{Q}) = 2$, then $\tau_{c,p^n} \in \mathrm{Sel}^{(p^n)}(E/\mathbb{Q})$.*

Proof. Since $r_{\mathrm{an}}(E/\mathbb{Q})$ is even and c is prime, Lemma 5.5 implies that $\delta([P_{c,\sigma}]) \in \mathrm{H}^1(K_c, E[p^n])^{\mathrm{Gal}(K_c/\mathbb{Q})}$. That the image of τ_{c,p^n} in $\mathrm{H}^1(\mathbb{Q}, E)[p^n]$ is locally trivially (hence in $\mathrm{Sel}^{(p^n)}(E/\mathbb{Q})$) follows from [Gro91, Prop. 6.2] with $n = c$ and $m = 1$, since $L'(E/K, 1) = 0$ hence y_K is torsion. □

5.3 Reduction modulo ℓ

The following lemma will be helpful when reducing the computation of τ_{c,p^n} to linear algebra (see Section 7.4). Below we will consider $M = E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z})$ as a module for the action of the nontrivial element $\mathrm{Frob}_\ell \in \mathrm{Gal}(\mathbb{F}_{\ell^2}/\mathbb{F}_\ell)$; we write M^- for the eigenspace of M on which Frob_ℓ acts by -1.

Lemma 5.9. *Let $p^n > 1$ be an odd prime power and let ℓ be a prime such that $p^n \mid \gcd(a_\ell, \ell+1)$. Then the groups $E(\mathbb{F}_\ell) \otimes (\mathbb{Z}/p^n\mathbb{Z})$ and $(E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^-$ are each cyclic of order p^n.*

Proof. (See [Ste10, Lem. 5.1].) We have

$$p^n \mid \gcd(a_\ell, \ell+1) \mid \ell + 1 - a_\ell = \#E(\mathbb{F}_\ell).$$

If $E(\mathbb{F}_\ell)[p]$ is noncyclic, then nondegeneracy of the Weil pairing implies that $\mu_p \subset \mathbb{F}_\ell^\times$, so $p \mid \ell - 1$, hence $p \mid \gcd(\ell - 1, \ell + 1) = 2$, which contradicts that p is odd. Thus $E(\mathbb{F}_\ell)[p]$ is cyclic, so the p-primary part of $E(\mathbb{F}_\ell)$ is cyclic of order divisible by p^n. For the second group, apply the above argument to the quadratic twist of E with trace of Frobenius $-a_\ell$, and note that p^n also divides $\gcd(-a_\ell, \ell+1)$. □

For any prime $\ell \nmid c$ that is inert in K, let λ be a prime ideal over ℓ in the ring of integers of the ring class field K_c. Define

$$z_{c,\sigma,\ell} \;=\; [P_{c,\sigma}] \pmod{\lambda} \in E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z}), \tag{5.6}$$

which is well defined, independent of the choice of λ. See [Ste10, Prop. 5.4] for the proof that $z_{c,\sigma,\ell}$ is well defined; the reason is that changing λ corresponds to acting on $[P_{c,\sigma}]$ by an automorphism, which does nothing since $[P_{c,\sigma}]$ is $\mathrm{Gal}(K_c/K)$-equivariant. Also, note that by Lemma 5.5, if $r_{\mathrm{an}}(E/\mathbb{Q}) + t$ is odd, then $z_{c,\sigma,\ell} \in E(\mathbb{F}_\ell) \otimes (\mathbb{Z}/p^n\mathbb{Z})$; if it is even, then $z_{c,\sigma,\ell} \in (E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z}))^-$, where the $-$ is for the action of the involution Frob_ℓ.

5.4 Consequences of nontriviality of the elements

We continue with the same notation and running assumptions as above. The first lemma below links verifying that $z_{c,\sigma,\ell} \neq 0$ to verifying Kolyvagin's Conjecture A [Kol91, pg. 255] (see Conjecture 1.1 above).

Lemma 5.10. *Suppose c is a squarefree product of inert primes q with $p^n \mid \gcd(a_q, q+1)$. If $z_{c,\sigma,\ell} \neq 0$, then $\tau_{c,p^n} \neq 0$.*

Proof. The nonzero element $z_{c,\sigma,\ell}$ is the image of $[P_{c,\sigma}]$ under the homomorphism

$$E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}) \longrightarrow E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z})$$

induced by reduction modulo a choice of prime ideal λ over ℓ. Thus if $z_{c,\sigma,\ell} \neq 0$, then $[P_{c,\sigma}] \neq 0$, so $\tau_{c,p^n} = \delta([P_{c,\sigma}]) \neq 0$, since δ is injective. □

Theorem 5.11. *Suppose $r_{\mathrm{an}}(E/\mathbb{Q}) = 2$ and that there exists inert primes c, ℓ (as above) such that $z_{c,\sigma,\ell} \neq 0$. Then*

$$\operatorname{rank} E(\mathbb{Q}) \leq 2$$

with equality if and only if $Ш(E/\mathbb{Q})(p)$ is finite. If $\operatorname{rank} E(\mathbb{Q}) = 2$, then $Ш(E/\mathbb{Q})[p] = 0$.

Proof. If $z_{c,\sigma,\ell} \neq 0$ then by Lemma 5.10, the Kolyvagin cohomology class $\tau_{c,p} \in \mathrm{H}^1(K, E[p])$ is nonzero, so Kolyvagin's Conjecture A [Kol91, pg. 255] is true. The desired conclusion then follows from [Ste10, Thm 4.2] (which is mainly a restatement of the main theorem of [Kol91]). □

For example, suppose E is a curve with $r_{\mathrm{an}}(E) = \operatorname{rank}(E(\mathbb{Q})) = 2$, that $Ш(E/\mathbb{Q})[2] = 0$ and that $\rho_{E,p}$ is surjective for all odd primes p. If we could somehow prove that for every prime p, there is a c with $z_{c,\sigma,\ell} \neq 0$, then Theorem 5.11 would imply that $Ш(E/\mathbb{Q}) = 0$. This would be an extremely deep result, since at present it is an open problem to prove unconditionally that the set of all pairs

$$\{(E,p) : Ш(E/\mathbb{Q})(p) \text{ is finite and } \operatorname{rank}(E) \geq 2\}$$

is infinite!

6 The Action of Galois and Reduction of Heegner Points Modulo ℓ

In this section, we prove a result (Theorem 6.6) that is crucial to giving a variant of Kolyvagin's derived points construction directly in characteristic ℓ, which is the main input to Algorithm 2.1. Note that the results in this section are on the level of the modular curve $X_0(N)$, and make no reference to a specific choice of elliptic curve over $\mathbb{Q}$ of conductor N, so they are equally useful in studying modular abelian varieties.

Theorem 6.6 below asserts that there is a compatible action of $\mathrm{Gal}(K_c/K_1)$ on two objects. Everything in the current paragraph will be made precise in Section 6.1 below. Let N be a positive integer and K a quadratic imaginary field such that each prime dividing N splits in K. Fix a choice of Heegner point $x_1 \in X_0(N)(K_1)$. For any square-free product c of primes that are inert in K, consider the support S of the divisor $T_c(x_1) \in \mathrm{Div}(X_0(N))$, where T_c is the cth Hecke operator. The Galois group $\mathrm{Gal}(K_c/K_1)$ acts transitively on S. Fix an inert prime $\ell \nmid c$ and a choice of prime λ of $\overline{\mathbb{Z}}$ over ℓ. Let $\mathbf{E}_1$ be the reduction mod λ of the enhanced elliptic curve corresponding to x_1, and consider the Eichler order $R = \mathrm{End}(\mathbf{E}_1)$. Also, as explained in Proposition 6.2, use class field theory to identify $\mathrm{Gal}(K_c/K_1)$ with $(\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$. For $x \in S$, represent $x \pmod{\lambda}$ by a right ideal class in R. Then Theorem 6.6 below asserts that *the action of $\mathrm{Gal}(K_c/K_1)$ on S is compatible with the action of $(\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ on the set of right ideals of R/cR of index c^2.* This result is somewhat complicated to state and prove, but we are amply compensated with an alternative interpretation of Kolyvagin's derived points construction.

In Section 6.1 we state our main result, then in Section 6.2 we prove it by deriving certain transformation rules for right ideals. We emphasize that in the arguments below, c is an arbitrary squarefree product of inert primes, and K is allowed to have arbitrary class number.

Remark 6.1. Reduction and the Galois action is also considered in [Cor02, §3.3], but via an adelic formulation that is less explicit and amenable to computation.

6.1 Notation and statement of theorem

In Section 6.1.1 we explain how Galois and Hecke operators act on higher Heegner points. In order to see the reduction of these points modulo ℓ, in Section 6.1.2 we introduce enhanced supersingular elliptic curves, and describe how they relate to points on modular curves. In Section 6.1.3, we explain how the Hecke operators act on divisors on enhanced curves, which will be used later in the proof of our main theorem. Finally, in Section 6.1.4 we precisely state the main theorem of this section, which is critical in reinterpreting Kolyvagin's derived classes operator in characteristic ℓ.

6.1.1 Galois and Hecke actions on Heegner points

Let N, K, c, and K_c be as above, and let $D = \mathrm{disc}(\mathcal{O}_K)$. Let $\mathcal{O}_c = \mathbb{Z}+c\mathcal{O}_K$ be the order of conductor c. Let $\mathfrak{n}$ be a choice of ideal in $\mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{n} \cong \mathbb{Z}/N\mathbb{Z}$, and let $\mathfrak{n}_c = \mathfrak{n}\cap\mathcal{O}_c$. As in [Gro84], for any order $\mathcal{O}$ (of conductor coprime to N) and any fractional $\mathcal{O}$-ideals $\mathfrak{m}$ and $\mathfrak{a}$, let $(\mathcal{O}, \mathfrak{m}, [\mathfrak{a}])$ denote the Heegner point $(\mathbb{C}/\mathfrak{a}, \mathfrak{m}^{-1}\mathfrak{a}/\mathfrak{a}) \in X_0(N)$, with endomorphism ring the order $\mathcal{O}$. In particular, let

$$x_c = (\mathcal{O}_c, \mathfrak{n}_c, [\mathcal{O}_c]) \in X_0(N)(K_c).$$

The elements of $(\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ are in bijection with the lines through the origin in the plane $\mathcal{O}_K/c\mathcal{O}_K \approx (\mathbb{Z}/c\mathbb{Z})^2$. These lines are in bijection with the sublattices of $\mathcal{O}_K$ of index c. The aforementioned sublattices are fractional $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$ ideals, and each one represents an element of the kernel of the natural map $\mathrm{Cl}(\mathcal{O}_c) \to \mathrm{Cl}(\mathcal{O}_K)$.

Proposition 6.2. *We have a commutative diagram of abelian groups:*

$$\begin{tikzcd}
1 \arrow[r] & \mathrm{Gal}(K_c/K_1) \arrow[r] \arrow[d, "\cong"] & \mathrm{Gal}(K_c/K) \arrow[r] \arrow[d, "\theta"', "\cong"] & \mathrm{Gal}(K_1/K) \arrow[r] \arrow[d, "\cong"] & 1 \\
1 \arrow[r] & (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times \arrow[r] & \mathrm{Cl}(\mathcal{O}_c) \arrow[r] & \mathrm{Cl}(\mathcal{O}_K) \arrow[r] & 1,
\end{tikzcd}$$

where the rightmost two vertical isomorphisms are induced by the Artin reciprocity map of class field theory, and the bottom row involves the bijections mentioned above.

Proof. This is standard; see, e.g., [Gro91, §3]. □

As explained in [Gro84, §4, (4.2)], for $[\mathfrak{b}] \in \mathrm{Cl}(\mathcal{O}_c)$, we have

$$(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{a})^{\theta(\mathfrak{b})} = (\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{a}\mathfrak{b}^{-1}).$$

Also [Gro84, §6], we have

$$T_c(x_1) = T_c((\mathcal{O}_K, \mathfrak{n}, \mathcal{O}_K)) = \sum_{\mathfrak{b}\subset\mathcal{O}_K} (\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b}) \in \mathrm{Div}(X_0(N)), \tag{6.1}$$

where the sum is over *sublattices* $\mathfrak{b} \subset \mathcal{O}_K$ of index c.

Remark 6.3. We emphasize: the $\mathfrak{b}$ are not *ideals* of $\mathcal{O}_K$, but merely ideals of $\mathcal{O}_c$! If they were ideals of $\mathcal{O}_K$, they would have norm $c = \#(\mathcal{O}_K/\mathfrak{b})$, but c is a product of distinct inert primes, so there are no ideals of $\mathcal{O}_K$ of norm c.

6.1.2 Enhanced supersingular elliptic curves in characteristic ℓ

We consider enhanced elliptic curves $\mathbf{E} = (E, C)$, where E is an elliptic curve and $C \subset E$ is a cyclic subgroup of order N. The terminology *enhanced elliptic curves* is used in [Rib90a, §3].

Recall that we fixed above an inert prime $\ell \nmid c$ and a prime λ of $\overline{\mathbb{Z}}$ over ℓ. The set $X_0(N)(\mathbb{F}_{\ell^2})^{\rm ss}$ of supersingular points on the mod λ reduction of $X_0(N)$ is the set of isomorphism classes of enhanced elliptic curves $\mathbf{E} = (E, C)$, where E is a supersingular elliptic curve over $\mathbb{F}_{\ell^2}$ and $C \subset E$ is a cyclic subgroup of order N.

Let $[\mathbf{E}_1] = x_1 \in X_0(N)(K_1)$, so $\mathbf{E}_1$ is a representative enhanced elliptic curve corresponding to the Heegner point x_1. Since $\mathfrak{n}$ is an $\mathcal{O}_K$-ideal, we have $\mathcal{O}_K = \mathrm{End}(\mathbf{E}_1)$, so we obtain an inclusion

$$\mathcal{O}_K = \mathrm{End}(\mathbf{E}_1) \hookrightarrow \mathrm{End}(\overline{\mathbf{E}}_1). \tag{6.2}$$

Remark 6.4. To see that Equation (6.2) is injective, note that by [ST68, Lem. 2], reduction modulo the prime λ of $\overline{\mathbb{Z}}$ induces an isomorphism $E_1[p^n] \xrightarrow{\cong} \overline{E}_1[p^n]$ for any prime power p^n with $p \neq \ell$ and p a prime of good reduction for E_1 (the lemma only asserts the map is surjective, but it is a map between finite groups of the same order, hence is an isomorphism). If $\varphi \in \mathrm{End}(\mathbf{E}_1)$ acts as 0 on $\overline{\mathbf{E}}_1$, then it acts as 0 on $\overline{E}_1[p^\infty]$, hence acts as 0 on $E_1[p^\infty]$, hence is 0 (since endomorphisms have finite degree).

The following lemma implies that

$$[\overline{\mathbf{E}}_1] \in X_0(N)(\mathbb{F}_{\ell^2})^{\rm ss}.$$

Lemma 6.5. *Suppose F is an elliptic curve defined over an extension M of K and that F has CM by an order $\mathcal{O}$ of K. Suppose that $\ell \in \mathbb{Z}$ is a prime that is inert in K such that $\ell \nmid [\mathcal{O}_K : \mathcal{O}]$. Let λ be a prime of M lying over ℓ and assume F has good reduction at λ, and let k be residue field modulo λ. Then the reduction F_k of F modulo λ is a supersingular elliptic curve.*

Proof. This is well known (see [Lan87, Ch. 10, §4, Thm. 10, Case 1] and [Sil94, Exercise 2.30]), but for the convenience of the reader we give a more conceptual proof than the ones cited above. It follows from the definition of F_k in terms of Néron models that $\mathcal{O}$ acts (functorially) on F_k. Moreover, because $\ell \nmid [\mathcal{O}_K : \mathcal{O}]$, the ℓ-torsion subgroup $F_k[\ell] = F_k(\overline{\mathbb{F}}_\ell)[\ell]$ is a vector space over the finite field $\mathcal{O}_K/(\ell) \approx \mathbb{F}_{\ell^2}$. Thus $d = \dim_{\mathbb{F}_\ell} F_k[\ell]$ is even. Since F_k is an elliptic curve over a finite field of characteristic ℓ, we have $d \leq 1$, so $d = 0$, hence F_k is supersingular. □

We view $X_0(N)(\mathbb{F}_{\ell^2})^{\rm ss}$ as explained in [Rib90a, §3], especially [Rib90a, Rmk. 3.5, pg 441], which builds on work of Deuring and Shimura. The endomorphism ring $R = \mathrm{End}(\overline{\mathbf{E}}_1)$ is an Eichler order of level N in the (unique up to isomorphism) rational quaternion algebra B ramified at ℓ and ∞. We have a bijection

$$X_0(N)(\mathbb{F}_{\ell^2})^{\rm ss} \xrightarrow{\quad\cong\quad} \{\text{ right fractional ideal classes in } R\ \}, \tag{6.3}$$

where two (nonzero) fractional right R-ideals $I, J \subset B$ are equivalent if there exists $\alpha \in B$ such that $\alpha I = J$. For any enhanced elliptic curve $\mathbf{F}$, endow $\mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F})$ with the structure of right R-module as follows: for $\varphi \in \mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F})$ and $r \in R$ we put $\varphi.r = \varphi \circ r$. This bijection sends $[\mathbf{F}]$ to the class of a right R-ideal that is isomorphic as a right R-module to the right R-module $\mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F})$. Also, we see that the right R-module $\mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F})$ is isomorphic to *some* right R-ideal I as follows. By [Mes86,

§2.4, pg. 223] or [Rib90a, Lem. 3.17], there exists an isogeny $\psi : \mathbf{F} \to \overline{\mathbf{E}}_1$. Using such an isogeny, we obtain an embedding

$$\mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F}) \hookrightarrow \mathrm{End}(\overline{\mathbf{E}}_1) = R$$

given by $\varphi \mapsto \psi \circ \varphi$, and the right ideal I is the image of $\mathrm{Hom}(\overline{\mathbf{E}}_1, \mathbf{F})$ under this embedding. Making a different choice of isogeny ψ replaces I by an equivalent right ideal.

6.1.3 Action of Hecke operators on supersingular divisors

The Hecke operators T_n act on $\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}})$, as explained in [Rib90a, pg. 443–445], and this action translates to an action on the free abelian group on the right R-ideal classes via the bijection (6.3) above, as explained in, e.g., [Koh01, §3.2]. For n any integer coprime to ℓN, we have

$$T_n([I]) = \sum_{J \subset I} [J], \tag{6.4}$$

where the sum is over right R ideals $J \subset I$ with $I/J \approx (\mathbb{Z}/n\mathbb{Z})^2$. We apply (6.4) to obtain a more explicit description of the image of the unit ideal (which corresponds to the reduction of x_1) under the Hecke operator T_c. Let

$$\overline{R} = R \otimes (\mathbb{Z}/c\mathbb{Z}) \cong R/cR.$$

Since c is coprime to N and coprime to the unique finite prime ℓ that ramifies in B, we have $R \otimes \mathbb{Z}_c \approx M_2(\mathbb{Z}_c)$, hence

$$\overline{R} \approx M_2(\mathbb{Z}/c\mathbb{Z}) \cong \bigoplus_{\text{primes } p|c} M_2(\mathbb{F}_p).$$

For any right ideal $I \subset \overline{R}$, let $\tilde{I}$ denote the inverse image of I in R under the natural surjection $R \to \overline{R}$. The right ideals of $\overline{R}$ correspond to the right ideals of R that contain cR, so the Hecke operator T_c acts on the unit ideal R via

$$T_c([R]) = \sum_{\substack{\text{right ideals } I \subset \overline{R} \\ \text{with } \overline{R}/I \approx (\mathbb{Z}/c\mathbb{Z})^2}} [\tilde{I}]. \tag{6.5}$$

More generally, for any right R-ideal J with $[R : J]$ coprime to c, we have

$$T_c([J]) = \sum_{\substack{\text{right ideals } I \subset \overline{R} \\ \text{with } \overline{R}/I \approx (\mathbb{Z}/c\mathbb{Z})^2}} [\tilde{I} \cap J].$$

6.1.4 Statement of the main theorem

As in the diagram of Proposition 6.2 above, let $[\mathfrak{a}] \in \ker(\mathrm{Cl}(\mathcal{O}_c) \to \mathrm{Cl}(\mathcal{O}_K))$ be an ideal class, and let $[\alpha] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ be the corresponding element, so $\alpha \in \mathcal{O}_K$. By replacing $\mathfrak{a}$ by an equivalent ideal, we may assume that $\mathfrak{a} = \mathbb{Z}\alpha + c\mathcal{O}_K$. Suppose $[\mathfrak{b}] \in \ker(\mathrm{Cl}(\mathcal{O}_c) \to \mathrm{Cl}(\mathcal{O}_K))$ is another ideal class, with corresponding element $[\beta]$, and let $\theta_{[\mathfrak{b}]} \in \mathrm{Gal}(K_c/K_1)$ be the corresponding automorphism. Let $I_{\mathfrak{b}} \subset \overline{R}$ be a right ideal such that

$$(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b}) \mapsto [\tilde{I}_{\mathfrak{b}}] \tag{6.6}$$

under composition of reduction modulo λ with the equivalence (6.3) above. There is such a right ideal $I_{\mathfrak{b}}$ because $(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b})$ is in the support of $T_c(x_1)$, and $[\tilde{I}_{\mathfrak{b}}]$ is in the support of $T_c(x_1 \pmod{\lambda})$ (see Equation (6.1)).

The group $\mathrm{Gal}(K_c/K_1)$ does *not* act naturally on

$$X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}} = X_0(N)(\mathcal{O}_{K_c}/\lambda)^{\mathrm{ss}},$$

since $\ell\mathcal{O}_K$ splits as a product of many primes (of which λ is one of them); of course, the "useless" decomposition subgroup of $\mathrm{Gal}(K_c/K_1)$ associated to λ (which has order 1!) does naturally act. However, as we will now see, $\mathrm{Gal}(K_c/K_1)$ acts naturally on a subset of the right ideals of $\overline{R}$. The challenge is that we need to compute what happens if we take $x_c \in X_0(N)(K_c)$, act by Galois, then map the result to $X_0(N)(\mathbb{F}_{\ell^2})$, and we can do this explicitly by instead considering the action of $\mathrm{Gal}(K_c/K_1)$ on index c^2 ideals in $\overline{R}$.

Equation (6.2) asserts that given our choice of λ there is an inclusion $\mathcal{O}_K \hookrightarrow R$, which we fix and use to define a right action of $\mathrm{Gal}(K_c/K_1)$ on certain right ideals in $\overline{R}$. For $\alpha \in \mathcal{O}_K$, let $\overline{\alpha}$ denote the image of α in $\overline{R}$. If $\sigma \in \mathrm{Gal}(K_c/K_1)$ corresponds to $[\alpha] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$, make σ act on the right on the set of right ideals I of $\overline{R}$ with $\overline{R}/I \approx (\mathbb{Z}/c\mathbb{Z})^2$ by $I^\sigma = \overline{\alpha}^{-1}I$. Finally, we state the main result of this section, which asserts that the natural right action of $\mathrm{Gal}(K_c/K_1)$ on the support of $T_c(x_1)$ in $\mathrm{Div}(X_0(N)/K_c)$ is compatible with the right action of $\mathrm{Gal}(K_c/K_1)$ that we just defined. We will prove this theorem in Section 6.2 below.

Theorem 6.6. *Let $\sigma \in \mathrm{Gal}(K_c/K_1)$, $[\mathfrak{b}] \in \ker(\mathrm{Cl}(\mathcal{O}_c) \to \mathrm{Cl}(\mathcal{O}_K))$, and let $[\tilde{I}_{\mathfrak{b}}]$ correspond to $(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b}) \pmod{\lambda}$ as in Equation 6.6 above. Then*

$$(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b})^\sigma \pmod{\lambda} \quad = \quad [\widetilde{I_{\mathfrak{b}}^\sigma}].$$

6.2 Proof of Theorem 6.6

This section is devoted to giving a proof of Theorem 6.6. When $c = 1$ the relevant objects all have cardinality 1 and the statement is trivial, so for the rest of this section we assume that $c > 1$. The strategy of the proof is to reinterpret the ideal $I_{\mathfrak{b}}$ as the right annihilator of a certain left ideal, and observe that this left ideal behaves sensibly under the action of Galois. (The proof is long because we are not sneaking any important details under the rug.)

We may assume that the representative fractional ideal $\mathfrak{b}$ is a sublattice of $\mathcal{O}_K$ of index c. Let $\mathbf{E}_1$ be the enhanced elliptic curve corresponding to the triple $(\mathcal{O}_K, \mathfrak{n}, [\mathcal{O}_K])$ and let $\mathbf{E}_{\mathfrak{b}}$ be the enhanced elliptic curve corresponding to the triple $(\mathcal{O}_c, \mathfrak{n}_c, [\mathfrak{b}])$. Let $\psi_{\mathfrak{b}} : \mathbf{E}_{\mathfrak{b}} \to \mathbf{E}_1$ be the isogeny of degree c given by the map $\mathbb{C}/\mathfrak{b} \to \mathbb{C}/\mathcal{O}_K$ that is multiplication by 1 on tangent spaces. The complementary (or dual) isogeny $\hat{\psi}_{\mathfrak{b}} : \mathbf{E}_1 \to \mathbf{E}_{\mathfrak{b}}$ is then given by the map $\mathbb{C}/\mathcal{O}_K \to \mathbb{C}/\mathfrak{b}$ induced by multiplication by c on $\mathbb{C}$. As in Section 6.1.2, we use $\psi_{\mathfrak{b}} \pmod{\lambda}$ to define a specific R-ideal $I_{\mathfrak{b}} \subset R = \mathrm{End}(\overline{\mathbf{E}}_1)$ that corresponds to $[\overline{\mathbf{E}}_{\mathfrak{b}}] \in X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}$. More precisely, the ideal $I_{\mathfrak{b}}$ is the image of $\mathrm{Hom}(\overline{\mathbf{E}}_1, \overline{\mathbf{E}}_{\mathfrak{b}})$ in R via the map $\vartheta \mapsto \overline{\psi}_{\mathfrak{b}} \circ \vartheta$, i.e.,

$$I_{\mathfrak{b}} = \{\overline{\psi}_{\mathfrak{b}} \circ \vartheta \;\; : \;\; \vartheta : \overline{\mathbf{E}}_1 \to \overline{\mathbf{E}}_{\mathfrak{b}}\} \subset R = \mathrm{End}(\overline{\mathbf{E}}_1).$$

The following lemma follows immediately from the definitions given in Section 6.1:

Lemma 6.7. *Under our fixed choices of maps and prime λ, we have*

$$[\mathbf{E}_{\mathfrak{b}}] \pmod{\lambda} \quad \longleftrightarrow \quad [I_{\mathfrak{b}}],$$

where $I_{\mathfrak{b}}$ is defined as above.

Proposition 6.10 below characterizes $I_{\mathfrak{b}}$ as an annihilator of a left R-ideal, which will be easier to work with. Let

$$J_{\mathfrak{b}} = \{\varphi \in R : \varphi(\ker(\overline{\hat{\psi}}_{\mathfrak{b}})) = 0\},$$

which is a left R-ideal. Thus $J_{\mathfrak{b}}$ is the left ideal of all endomorphisms of $\overline{\mathbf{E}}_1$ that factor through the homomorphism $\overline{\hat{\psi}}_{\mathfrak{b}} : \overline{\mathbf{E}}_1 \to \overline{\mathbf{E}}_{\mathfrak{b}}$:

$$\begin{tikzcd}
 & \overline{\mathbf{E}}_{\mathfrak{b}} \arrow[dr, dotted] & \\
\overline{\mathbf{E}}_1 \arrow[ur, "\overline{\hat{\psi}}_{\mathfrak{b}}"] \arrow[rr, "\varphi\in J_{\mathfrak{b}}"'] & & \overline{\mathbf{E}}_1
\end{tikzcd}$$

We will use the following lemma to compute the quotient abelian group $R/J_{\mathfrak{b}}$.

Lemma 6.8. *The natural map $R \to \mathrm{End}(\overline{E}_1[c])$ is surjective.*

Proof. It suffices to prove that for each prime $p \mid c$, the map

$$\varphi : R \otimes \mathbb{F}_p \to \mathrm{End}(\overline{E}_1[p]) \tag{6.7}$$

is surjective. Since R is an Eichler order of level N, N is coprime to c and $p \mid c$, we have $R \otimes \mathbb{F}_p = \mathrm{End}(\overline{E}_1) \otimes \mathbb{F}_p$. Also, since $p \neq \ell$, we have $\mathrm{End}(\overline{E}_1[p]) \approx \mathrm{End}(\mathbb{F}_p \oplus \mathbb{F}_p) \cong M_2(\mathbb{F}_p)$, and since $\overline{E}_1$ is a supersingular elliptic curve, $\dim_{\mathbb{F}_p}(R \otimes \mathbb{F}_p) = \mathrm{rank}_{\mathbb{Z}} R = 4$, so by a dimension count it suffices to prove that φ is injective. Suppose $\overline{f} = f \otimes 1 \in R \otimes \mathbb{F}_p$ is a nonzero element of $\ker(\varphi)$, with $f \in \mathrm{End}(\overline{E}_1)$. Then f acts as 0 on $\overline{E}_1[p]$, so f factors through multiplication by p, which means that there exists $g \in \mathrm{End}(\overline{E}_1)$ with $f = pg$. But then $\overline{f} = pg \otimes 1 = g \otimes p = g \otimes 0 = 0$, a contradiction. We conclude that φ is injective, hence surjective. □

Lemma 6.9. *We have $R/J_{\mathfrak{b}} \approx (\mathbb{Z}/c\mathbb{Z})^2$, where we view both sides as quotients of additive abelian groups.*

Proof. We prove this lemma by using Lemma 6.8 to reinterpret the assertion as a statement in $M_2(\mathbb{Z}/c\mathbb{Z})$, then use linear algebra modulo prime divisors of c to count dimensions. The kernel $D = \ker(\overline{\hat{\psi}}_{\mathfrak{b}}) \subset \overline{E}_1[c]$ is a cyclic group of order c. Let $\overline{J}$ be the left annihilator in $\mathrm{End}(\overline{E}_1[c]) \approx M_2(\mathbb{Z}/c\mathbb{Z})$ of D. For each prime $p \mid c$, we have $\mathrm{End}(\overline{E}_1[p]) \approx M_2(\mathbb{F}_p)$, and the factor of D in $\overline{E}_1[p]$ is of order p. The left annihilator in $M_2(\mathbb{F}_p)$ of a 1-dimensional subspace of $(\mathbb{F}_p)^2$ has $\mathbb{F}_p$-dimension 2, since it is the 2-dimensional $\mathbb{F}_p$-vector space of matrices whose rows are both a multiple of v, where v has dot product 0 with a basis for our 1-dimensional subspace. Putting these factors for each p together, we see that $\overline{J}$ is free of rank 2 over $\mathbb{Z}/c\mathbb{Z}$.

Since c kills $\ker(\overline{\hat{\psi}}_{\mathfrak{b}})$, we see that $cR \subset J_{\mathfrak{b}}$. We thus have an isomorphism of abelian groups

$$R/J_{\mathfrak{b}} \to M_2(\mathbb{Z}/c\mathbb{Z})/\overline{J}.$$

It is surjective because of Lemma 6.8. It is injective because $J_{\mathfrak{b}}$ is defined to be those endomorphisms that kill the subgroup D of $\overline{E}_1[c]$, which is a condition we can check in $\mathrm{End}(\overline{E}_1[c])$. The lemma thus follows. □

Next we use the left R-ideal $J_{\mathfrak{b}}$ to define a right R-ideal:

$$I'_{\mathfrak{b}} = \{\varphi \in R \;:\; J_{\mathfrak{b}}\varphi \subset cR\}.$$

Proposition 6.10. *We have*

$$I_{\mathfrak{b}} = I'_{\mathfrak{b}}$$

Proof. The strategy of the proof is to show that $I_{\mathfrak{b}} \subset I'_{\mathfrak{b}}$, then observe that both $I_{\mathfrak{b}}$ and $I'_{\mathfrak{b}}$ have the same index in R, so they must be equal.

To see that the inclusion $I_{\mathfrak{b}} \subset I'_{\mathfrak{b}}$ hold is a straightforward calculation using the definitions, as follows. An element $\varphi \in I_{\mathfrak{b}}$ is by definition of the form $\varphi = \overline{\psi}_{\mathfrak{b}} \circ \vartheta$, where $\vartheta : \overline{\mathbf{E}}_1 \to \overline{\mathbf{E}}_{\mathfrak{b}}$ and $\overline{\psi}_{\mathfrak{b}} : \overline{\mathbf{E}}_{\mathfrak{b}} \to \overline{\mathbf{E}}_1$, as above. Suppose $\delta \in J_{\mathfrak{b}}$, so $\delta \in \mathrm{End}(\overline{\mathbf{E}}_1)$ and $\delta(\ker(\overline{\hat{\psi}}_{\mathfrak{b}})) = 0$, hence $\delta = \delta' \circ \overline{\hat{\psi}}_{\mathfrak{b}}$ for some $\delta' : \overline{\mathbf{E}}_{\mathfrak{b}} \to \overline{\mathbf{E}}_1$. Thus

$$\delta \circ \varphi = (\delta' \circ \overline{\hat{\psi}}_{\mathfrak{b}}) \circ (\overline{\psi}_{\mathfrak{b}} \circ \vartheta) = \delta' \circ [c] \circ \vartheta \in cR,$$

which proves that $I_{\mathfrak{b}} \subset I'_{\mathfrak{b}}$.

We next prove that $[R : I'_{\mathfrak{b}}] = c^2$, as an application of Lemma 6.9. We have $c \in I'_{\mathfrak{b}}$, so $cR \subset I'_{\mathfrak{b}} \subset R$, hence $I'_{\mathfrak{b}}$ is completely determined by an ideal $\overline{I}'_{\mathfrak{b}} \subset \overline{R} = R \otimes (\mathbb{Z}/c\mathbb{Z}) \approx M_2(\mathbb{Z}/c\mathbb{Z})$. The ideal $\overline{I}'_{\mathfrak{b}}$ is the right annihilator of the left ideal $\overline{J}_{\mathfrak{b}} \subset \overline{R}$. For each prime $p \mid c$, Lemma 6.9 implies that the right annihilator mod p of $J_{\mathfrak{b}}$, i.e., the image of $I'_{\mathfrak{b}}$ in $R \otimes \mathbb{F}_p \cong M_2(\mathbb{F}_p)$, is proper and nontrivial. We conclude that $[R : I'_{\mathfrak{b}}] = c^2$.

Finally we observe that $[R : I_b] = c^2$. In light of Equation (6.5), the ideal $I_{\mathfrak{b}}$ is one of the ideals that appears in the sum in the definition of the Hecke operator T_c, so $[R : I_{\mathfrak{b}}] = c^2$. Since $[R : I'_{\mathfrak{b}}] = c^2$ and $I_{\mathfrak{b}} \subset I'_{\mathfrak{b}}$, it follows that $I_{\mathfrak{b}} = I'_{\mathfrak{b}}$, which proves the proposition. □

Suppose $[\alpha] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ with $\alpha \in \mathcal{O}_K$, and let $\mathfrak{a} \subset \mathcal{O}_K$ be the corresponding fractional $\mathcal{O}_c$-ideal (as in Section 6.1.4). Let $J_\alpha = J_{\mathfrak{a}}$. Proposition 6.12 below asserts that the natural right action of $(\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ on the left ideals in $\overline{R}$ is compatible with the natural right action of $(\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ on sublattices $\mathfrak{a} \subset \mathcal{O}_K$ of index c. Note the inverse that appears, which makes a left action into a right action (the group acting is abelian, so we are being slightly pedantic in emphasizing this). First we prove a lemma about an action on certain kernels.

Lemma 6.11. *Suppose* $[\alpha], [\beta] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ *with* $\alpha, \beta \in \mathcal{O}_K$. *Then*

$$\ker(\hat{\psi}_{\alpha\beta}) = \alpha \ker(\hat{\psi}_\beta).$$

Proof. As above, let $\mathfrak{a} \subset \mathcal{O}_K$ be the lattice of index c corresponding to $[\alpha]$. Also, recall from page 15 that the map $\hat{\psi}_\alpha : E_1 \to E_{\mathfrak{a}}$ is given over the complex numbers by the map $\mathbb{C}/\mathcal{O}_K \to \mathbb{C}/\mathfrak{a}$ induced by multiplication by the integer c on $\mathbb{C}$. We have

$$E_1[c] = \left(\frac{1}{c}\mathcal{O}_K\right)/\mathcal{O}_K \cong \mathcal{O}_K/c\mathcal{O}_K \tag{6.8}$$

and the lattice $\mathfrak{a}$ defines a rank 1 subspace of $\mathcal{O}_K/c\mathcal{O}_K$. The isomorphism (6.8) identifies $\ker(\hat{\psi}_\alpha) \subset E_1[c]$ with the image of $\mathfrak{a}$ in $\mathcal{O}_K/c\mathcal{O}_K$. If $\mathfrak{b}$ corresponds to $[\beta]$, then $\alpha\mathfrak{b} = [\alpha\beta]$, so in terms of this presentation of $E_1[c]$, the claimed equality of the lemma follows. □

Note that since $[\alpha] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$, the image $\overline{\alpha} \in \overline{R} = R \otimes (\mathbb{Z}/c\mathbb{Z})$ of α is invertible.

Proposition 6.12. *Let* α, β *be as above, let* J *be a left* R*-ideal, and let* $\overline{J}$ *denote its image in* $\overline{R}$. *Then*

$$\overline{J}_{\alpha\beta} = \overline{J}_\beta \cdot \overline{\alpha}^{-1},$$

where $\overline{\alpha}$ *is the image of* α *in* $\overline{R}$.

Proof. The reduction modulo λ map $E_1[c]$ to $\overline{E}_1[c]$ is an isomorphism since $\ell \nmid cN$ (see Remark 6.4), so reducing both sides of Lemma 6.11 modulo λ, we see that $\ker(\overline{\hat{\psi}}_{\alpha\beta}) = \alpha \ker(\overline{\hat{\psi}}_{\beta})$. Thus

$$\begin{aligned} J_{\alpha\beta} =& \{\varphi \in R : \varphi(\ker(\overline{\hat{\psi}}_{\alpha\beta})) = 0\} \\ =& \{\varphi \in R : \varphi(\alpha(\ker(\overline{\hat{\psi}}_{\beta}))) = 0\} \\ =& \{\varphi \in R : (\varphi\alpha)(\ker(\overline{\hat{\psi}}_{\beta})) = 0\} \\ =& \{\varphi \in R : \varphi\alpha \in J_\beta\} = R \cap (J_\beta \cdot \alpha^{-1}) \subset J_\beta \cdot \alpha^{-1}. \end{aligned}$$

We thus have an inclusion of (equivalent) fractional left R-ideals

$$J_{\alpha\beta} \subset J_\beta \cdot \alpha^{-1}.$$

Taking the image of both ideals in $\overline{R}$ gives an inclusion

$$\overline{J}_{\alpha\beta} \subset \overline{J}_\beta \cdot \overline{\alpha}^{-1} \subset \overline{R}.$$

Right multiplication by an invertible element in $\overline{R}$ is a bijection, so $[\overline{R} : \overline{J}_\beta \cdot \overline{\alpha}^{-1}] = [\overline{R} : \overline{J}_\beta] = c^2$, by Lemma 6.9. Since $[\overline{R} : \overline{J}_{\alpha\beta}] = c^2$, again by Lemma 6.9, it follows that $\overline{J}_{\alpha\beta} = \overline{J}_\beta \cdot \overline{\alpha}^{-1}$, as claimed.

□

Proof of Theorem 6.6. We have $\mathfrak{a}, \mathfrak{b} \subset \mathcal{O}_K$ two lattices of index c and corresponding classes

$$[\alpha], [\beta] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times.$$

Let $\sigma \in \mathrm{Gal}(K_c/K_1)$ be the automorphism corresponding to $\mathfrak{a} \in \mathrm{Cl}(\mathcal{O}_c)$. Let $\mathfrak{g} \subset \mathcal{O}_K$ be the lattice of index c corresponding to the class $[\alpha^{-1}\beta] = [\alpha]^{-1}[\beta] \in (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$, so $I_{\alpha^{-1}\beta} = I_{\mathfrak{g}}$. Then, under reduction modulo λ, we have

$$(\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{b})^\sigma = (\mathcal{O}_c, \mathfrak{n}_c, \mathfrak{a}^{-1}\mathfrak{b}) \longmapsto [I_{\alpha^{-1}\beta}].$$

For any left or right ideal I of R, let $\overline{I}$ be the image of I in $\overline{R} = R \otimes (\mathbb{Z}/c\mathbb{Z})$. By Proposition 6.10 the right ideal $\overline{I}_{\mathfrak{b}}$ is the right annihilator of the left ideal $\overline{J}_{\mathfrak{b}}$, and this is true for any $\mathfrak{b}$. By Proposition 6.12, we have that $\overline{I}_{\alpha^{-1}\beta}$ is the right annihilator of the left ideal $\overline{J}_{\alpha^{-1}\beta} = \overline{J}_\beta \cdot \overline{\alpha}$. We thus have

$$\begin{aligned} \overline{\alpha}^{-1} \cdot \overline{I}_\beta &= \overline{\alpha}^{-1} \cdot \{\varphi \in \overline{R} \ : \ \overline{J}_\beta \cdot \varphi = 0\} \\ &= \{\overline{\alpha}^{-1} \cdot \varphi \in \overline{R} \ : \ \overline{J}_\beta \cdot \varphi = 0\} \\ &= \{\varphi \in \overline{R} \ : \ \overline{J}_\beta \cdot \overline{\alpha}\varphi = 0\} \\ &= \{\varphi \in \overline{R} \ : \ \overline{J}_{\alpha^{-1}\beta} \cdot \varphi = 0\} = \overline{I}_{\alpha^{-1}\beta}, \end{aligned}$$

where in the third equality we replace φ by $\overline{\alpha}\varphi$, using that multiplication by $\overline{\alpha}$ defines a bijection $\overline{R} \to \overline{R}$. The displayed equality proves the theorem.

□

7 Reduction of Derived Classes

Let E be an elliptic curve over $\mathbb{Q}$, and let $P_{c,\sigma}$ be as in Equation (5.4) of Section 5. In this section, we apply the general results of Section 6 to give an algorithm to compute the reduction $z_{c,\sigma,\ell} \in E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p\mathbb{Z})$ (see Equation 5.6) when p is an odd prime and $E[p]$ is absolutely irreducible. We will apply this algorithm in Section 8 to verify that $[P_{c,\sigma}] \neq 0$, in specific examples. It is of interest to verify that $[P_{c,\sigma}] \neq 0$ in specific examples since, as was mentioned in Section 1, this was until now not known in even a single case for a curve E of rank ≥ 2.

We continue to assume that E and K satisfy the Heegner hypothesis. The goal of this section is to give an algorithm that we can use (in some specific examples) to verify that $[P_{c,\sigma}] \neq 0$ for some c. To do this, we consider the reduction map

$$r_\ell : E(K_c) \otimes (\mathbb{Z}/p^n\mathbb{Z}) \to E(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z}), \tag{7.1}$$

given by reducing points modulo a fixed choice of prime λ over ℓ, where $\ell \nmid c$ is a prime that is inert in K, just as at the end of Section 5. If we find one prime ℓ such that $z_{c,\sigma,\ell} = r_\ell([P_{c,\sigma}]) \neq 0$, we conclude that $[P_{c,\sigma}] \neq 0$, as desired. We will thus be concerned primarily with computing whether or not $z_{c,\sigma,\ell}$ is 0 in the case when $n = 1$.

Remark 7.1. Assume that $\text{Ш}(E/\mathbb{Q})[p] = 0$, that $r_{\text{an}}(E/\mathbb{Q}) = \text{rank}(E(\mathbb{Q})) = 2$, and that we have shown that $[P_{c,\sigma}] \neq 0$ for some prime c. Then there is an alternative approach to compute the line spanned by $P_{c',\sigma'}$ for *any* inert prime c'. Jared Weinstein and the author learned about this idea from Karl Rubin after we implemented and ran the main algorithm of this paper, and wanted to better understand the data we obtained. The algorithm builds on [How04] and the Mazur-Rubin theory of Kolyvagin systems [MR04]. This is the subject of the forthcoming paper [SW10], and we have also used this algorithm as a double check on the calculations in Section 8. Quick summary: an easy calculation shows that the line has to be in the kernel of r_c; moreover, and this is deeper, r_c fails to have maximal rank if and only if $[P_c] = 0$.

In Section 7.1 we explain how to compute the reduction map from Heegner points in characteristic 0 to supersingular points in characteristic ℓ as an application of Deuring's lifting theorem and explicit computation with ternary quadratic forms. Section 7.2 contains the promised reinterpretation of Kolyvagin's derived classes construction directly on the divisor group of supersingular points, and Section 7.3 explicitly links this construction with reduction of derived classes from characteristic 0. Section 7.4 refines a crucial surjectivity result that Cornut used in proving Mazur's conjecture, which is also extremely important to our algorithm. Finally, Section 7.5 proves a multiplicity one theorem, which ensures that we have a general algorithm, rather than just a procedure that happens to work in every case we try.

7.1 Explicit computation of the reduction map using quaternion algebras

Let ℓ be a prime that is inert in K, as above. Following [Ste09, Piz80], let $B = B_{\ell,\infty}$ be the unique (up to isomorphism) quaternion algebra ramified at ℓ and ∞, and fix an Eichler order R of level N in B.

The group of Atkin-Lehner operators of level N has order 2^ν, where ν is the number of prime divisors of N. As discussed in Remark 4.1 above, the Heegner point x_1 is only well defined up to the choice of an ideal $\mathfrak{n}$ of $\mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{n} \cong \mathbb{Z}/N\mathbb{Z}$, and there are 2^ν

choices for $\mathfrak{n}$. We temporarily write $x_{1,\mathfrak{n}}$ for the choice of Heegner point x_1 associated to the ideal $\mathfrak{n}$.

The prime ℓ is inert in K, so by Lemma 6.5, each of the points $x_{1,\mathfrak{n}}$ defines a point on $X_0(N)(K_1)$ that reduces to a supersingular point in $X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}$. Moreover, we have the bijection of Equation 6.3 between $X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}$ and a certain set of right R-ideal classes. In terms of this bijection, we compute some $\overline{x}_1 \in X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}}$ corresponding to a choice of $\mathfrak{n}$ as follows. First, we enumerate all right ideal classes $[I]$ using standard algorithms, e.g., if N is odd by applying the Hecke operator T_2 repeatedly, starting with the unit ideal, and using theta series to check equivalence (see, e.g., [Piz80, Prop. 1.18]). Then we apply Theorem 7.2 below to find an I such that $\mathcal{O}_K$ embeds in R_I.

Let I be a fractional right R-ideal, and consider the left order

$$R_I = \{x \in B : xI \subset I\}$$

associated to I. We use the Deuring lifting theorem to give an algorithm to compute $\overline{x}_1$.

Theorem 7.2 (Deuring)**.** *The bijection of Equation* (6.3) *induces a bijection*

$$\{\overline{x}_{1,\mathfrak{n}} \in X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}} : \text{ ideals } \mathfrak{n} \text{ with } \mathcal{O}_K/\mathfrak{n} \cong \mathbb{Z}/N\mathbb{Z}\} \xrightarrow{\cong} \{[I] : \mathcal{O}_K \text{ embeds in } R_I\}.$$

Proof. See [GZ85, Prop. 2.7] (see also [JK10, §2] for a generalization in which $\mathcal{O}_K$ is replaced by $\mathcal{O}_c$). □

To compute a choice of $\overline{x}_1$ thus reduces to giving an algorithm to decide whether or not $\mathcal{O}_K$ embeds in R_I. As in [Gro87, pg. 172], let $G_I \approx \mathbb{Z}^3$ be the trace zero elements in $2R_I + \mathbb{Z}$, and let $q_I : G_I \to \mathbb{Q}$ be the normalized *ternary* quadratic form got by restricting the reduced norm on B to G_I.

Lemma 7.3. *There is an embedding of $\mathcal{O}_K$ into R_I if and only if the quadratic form q_I represents the absolute value $|D_K|$ of the discriminant of $\mathcal{O}_K$.*

Proof. This follows from [Gro87, Prop. 12.9] (see also [JK10, Lem. 4.1]). □

To compute $\overline{x}_1$ we compute the quadratic form q_I for a representative I for each right ideal class in turn, and decide whether or not it represents $|D_K|$. When we find one that does, we declare that our representative element is $\overline{x}_1 = \overline{x}_{1,\mathfrak{n}}$, which is well defined up to the choice of ideal $\mathfrak{n}$. In general (e.g., when the class number of K is bigger than 1), our current formula unfortunately requires computing all $x_{1,\mathfrak{n}}$ for all $\mathfrak{n}$ (see Theorem 7.8).

7.2 Kolyvagin's derived classes construction in terms of quaternion algebras

Let I be a right ideal in our fixed choice of Eichler order R of level N such that I corresponds to $\overline{x}_{1,\mathfrak{n}}$, computed as above.

Lemma 7.4. *By replacing I by an equivalent ideal, we can arrange that $I \otimes (\mathbb{Z}/c\mathbb{Z}) = R \otimes (\mathbb{Z}/c\mathbb{Z})$.*

Proof. For any prime $r \nmid N\ell$, the graph of the Hecke operator T_r is connected (see [Mes86, §2.4, pg. 223] or [Rib90a, Lem. 3.17]). If we choose r also coprime to c, then enumerate the right ideals of R by computing the action of T_r, starting with the unit ideal, we will cover all the right ideal classes of R; in particular, there is an ideal I' equivalent to I obtained via this procedure. From the formula of Equation (6.4) for the action of Hecke operators, we see that $[R : I']$ is a power of r. Thus $I' \otimes (\mathbb{Z}/c\mathbb{Z}) = R \otimes (\mathbb{Z}/c\mathbb{Z})$, as claimed. □

Next we compute a choice of homomorphism

$$s : R \twoheadrightarrow M_2(\mathbb{Z}/c\mathbb{Z}). \tag{7.2}$$

This can be done individually for each prime divisor of c, and the maps assembled together to give s. For example, for each prime divisor $q \mid c$, one could consider the algebra $R \otimes (\mathbb{Z}/q\mathbb{Z})$ and apply [Voi, §4] to find an explicit isomorphism $R \otimes (\mathbb{Z}/q\mathbb{Z}) \to M_2(\mathbb{Z}/q\mathbb{Z})$.

Let q be any prime that is inert in K. Suppose the image of $\alpha \in \mathcal{O}_K$ generates the cyclic group

$$(\mathcal{O}_K/q\mathcal{O}_K)^\times/(\mathbb{Z}/q\mathbb{Z})^\times$$

of order $q + 1$. Using a fixed choice of embedding of $\mathcal{O}_K$ into the left order of I from above (which exists by Theorem 7.2), we view α as an element of B. Let $\overline{\alpha}$ be the canonical image of α in $M_2(\mathbb{Z}/q\mathbb{Z}) = \overline{R}/q\overline{R}$ using the splitting s of (7.2).

For each $i = 0, \ldots, q$, let

$$\overline{J}_i = \{B \in M_2(\mathbb{Z}/q\mathbb{Z}) : (1, 0)\overline{\alpha}^i B = 0\} \subset \overline{R}/q\overline{R}.$$

Suppose $[M]$ is a right ideal class of R, and (as in Lemma 7.4) choose a representative right ideal $M \subset R$ such that $q \nmid [R : M]$, so s defines a map $M \twoheadrightarrow \overline{R}$. For each i, let J_i be the inverse image of $\overline{J}_i$ in M. Define

$$D_{q,\alpha}([M]) = \sum_{i=1}^{q} i[J_i].$$

Extending linearly, we define an endomorphism

$$D_{q,\alpha} \in \mathrm{End}(\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}})).$$

Remark 7.5. We make two remarks about the above operator:

1. The map $D_{q,\alpha}$ is explicitly computable; it is closely related to computing the Hecke operator T_q, since $T_q([M]) = \sum_{i=0}^{q}[J_i]$ is almost the same as $D_{q,\alpha}([M])$, except without the coefficient in the enumeration of the J_i's.
2. The maps $D_{q,\alpha}$ typically do not commute with the Hecke operators or with each other.

Next write $c = p_1 \cdots p_t$, let $\sigma = (\sigma_1, \ldots, \sigma_t)$ with $\sigma_i \in \mathrm{Gal}(K_c/K_{c/p_i})$ be choices of generators, and let $\alpha = (\alpha_1, \ldots, \alpha_t)$ with $\alpha_i \in \mathcal{O}_K$ be the corresponding elements via the map of Equation 4.1 above. Define

$$D_{c,\alpha} = \prod_{i=1}^{t} D_{p_i,\alpha_i} \in \mathrm{End}(\mathrm{Div}(X_0(N)(\mathbb{F}_{\ell^2})^{\mathrm{ss}})).$$

7.3 Reduction of Kolyvagin's derived points

Let $f \in S_2(\Gamma_0(N))$ be a newform, let $I_f \subset \mathbb{T}$ be the annihilator of f in the Hecke algebra associated to $J_0(N)$, let $A_f = J_0(N)/I_f J_0(N)$ be the corresponding modular abelian variety with modular parametrization $\pi_f : J_0(N) \to A_f$ and let

$$\psi_f : \mathrm{Div}(X_0(N)^{\mathrm{ss}}_{\mathbb{F}_{\ell^2}}) \to A_f(\mathbb{F}_{\ell^2})$$

be the homomorphism that sends each supersingular point x to $\overline{\pi}_f(x - \infty)$, where $\overline{\pi}_f$ is the reduction modulo λ of π_f. By [BCDT01], our elliptic curve E is isogeneous to some A_f for a newform $f \in S_2(\Gamma_0(N))$ where N is the conductor of E.

Theorem 7.6. *We have the following in $A_f(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z})$:*

$$[\overline{\pi}_f(D_{c,\sigma}(y_c))] = [\psi_f\left(D_{c,\alpha}([I])\right)].$$

Proof. This follows from Theorem 6.6. □

Let

$$\mathcal{I} = \{[I] : \mathcal{O}_K \hookrightarrow R_I\}$$

be the set of all right ideal classes of R whose left order admits an embedding of $\mathcal{O}_K$. For each such $[I]$, let n_I be half the number of primitive representatives of $|D_K|$ by the ternary quadratic form q_I. Let $\mathcal{H}$ be the $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-orbit of the set of all Heegner points $x_{1,\mathfrak{n}} \in X_0(N)(K_1)$ for all ideals $\mathfrak{n} \subset \mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{n} \cong \mathbb{Z}/N\mathbb{Z}$.

Lemma 7.7. *For each $[I] \in \mathcal{I}$, the number of elements of $\mathcal{H}$ reducing to the point of $X_0(N)(\mathbb{F}_{\ell^2})$ corresponding to $[I]$ is equal to n_I.*

Proof. By [JK10, §2] there is a one-to-one correspondence between the Heegner points $x_{1,\mathfrak{n}}$ reducing to $[I]$ and $R_I^\times$ conjugacy classes of embeddings $O_K \hookrightarrow R_I$. By [JK10, Prop. 4.2] there is a $(\#R_I^\times/2)$-to-1 correspondence between embeddings $\mathcal{O}_K \hookrightarrow R_I$ and primitive representations of $|D|$ by q_I. Thus every pair of primitive representations of $|D|$ by q_I corresponds to $\#R_I^\times$ embeddings, so half the number of primitive representatives is the number of $R_I^\times$ conjugacy classes of embeddings. □

Theorem 7.8. *Let ν be the number of distinct prime divisors of N. We have the following in $A_f(\mathbb{F}_{\ell^2}) \otimes (\mathbb{Z}/p^n\mathbb{Z})$:*

$$\overline{\pi}_f([P_{c,\sigma}]) = 2^{-\nu} \cdot \sum_{[I]\in\mathcal{I}} n_I \cdot [\psi_f\left(D_{c,\alpha}([I])\right)] \tag{7.3}$$

Proof. This follows by combining Lemma 7.7 and Theorem 7.6, and noting that $\mathcal{H}$ is a disjoint union of $[K_1 : K]$ Atkin-Lehner orbits, each of size 2^ν. Thus in computing the sum on the right of (7.3) we are computing $\mathrm{Tr}_{K_1/K}(D_{c,\sigma}(y_c))$ separately 2^ν times, hence we divide out this extra factor of 2^ν, which is harmless since p is odd. □

We still have not explained how to explicitly compute the map ψ_f, so Theorem 7.8 does not yet yield an algorithm. In Section 7.4 we will establish that ψ_f is surjective after tensoring with $\mathbb{Z}/p\mathbb{Z}$, and in Section 7.5 we give conditions under which ψ_f is uniquely determined up to scalars by being Hecke equivariant ("multiplicity one"), which means we can compute ψ_f up to a scalar. Alternatively, as mentioned in Remark 2.3, we can sometimes instead avoid computing ψ_f at all if we know ψ_f is surjective by instead verifying that the $\mathbb{T}$-span of $\sum n_I D_{c,\alpha}([I])$ is all of $X \otimes \mathbb{F}_p$.

7.4 Map from the supersingular module to an optimal abelian variety quotient

Let ℓ be an inert prime that does not divide the level N, and let $k = \mathbb{F}_{\ell^2} \approx \mathcal{O}_K/\ell\mathcal{O}_K$, which is a finite field of order ℓ^2. The Hecke algebra $\mathbb{T}$ acts via correspondences on many objects attached to the modular curves $X_0(N)$ and $X_1(N)$, e.g., via endomorphisms on the Jacobian $J_0(N)$ and also on

$$X = \mathrm{Div}(X_0(N)(k)^{\mathrm{ss}}) \qquad \text{and} \qquad X^0 = \mathrm{Div}^0(X_0(N)(k)^{\mathrm{ss}}). \tag{7.4}$$

Also, $\mathbb{T}$ acts on the *Shimura subgroup* $\Sigma = \ker(J_0(N) \to J_1(N))$. We say that a $\mathbb{T}$-module M is *Eisenstein* (in the sense of [Maz77]) if for any prime $p \nmid N$, the operator $T_p - (1+p)$ annihilates M. For example, [Rib88, Thm. 1] asserts that Σ is Eisenstein.

Let $J = J_0(N)_k$, and consider the natural $\mathbb{T}$-module homomorphism

$$X \to J(k) \tag{7.5}$$

that sends a divisor $D \in X$ to the equivalence class of the degree zero divisor $D - \deg(D)\infty$ in the Jacobian.

Proposition 7.9 (Ribet). *The cokernel S of the induced map*

$$X^0 \to J(k) \tag{7.6}$$

is the Cartier dual $\Sigma^\vee$ of Σ, and the $\mathbb{T}$-module $\Sigma^\vee$ is Eisenstein.

Proof. The following argument is due to Ribet (see [Rib10]). Let F be the ℓth power Frobenius endomorphism of J and let V be its dual. We have $J(k) = J[1-F^2]$. This kernel is Cartier dual to $J[1-V^2]$, since it is obtained by dualizing the following exact sequence (see [Mum70, §15, pg. 143] and [Mil86, §11]):

$$0 \to J[1-F^2] \to J \xrightarrow{1-F^2} J \to 0.$$

Ribet proved in 1983 (see [Pra95, Prop. 3.6]) that the subgroup $J[1-V^2]$ contains the reduction modulo ℓ of the Shimura subgroup Σ of J, and S is the annihilator of Σ in the natural perfect pairing between $J[1-F^2]$ and $J[1-V^2]$. The content of [Pra95, Prop. 3.6] is that the supersingular group is "as large as possible" in the sense that it is the full annihilator.

In the pairing between $J[1-F^2]$ and $J[1-V^2]$, there is the standard formula $\langle Tx, y\rangle = \langle x, T^\dagger y\rangle$, where the dagger refers to the Rosati involution of $\mathrm{End}(J)$ and T is a Hecke operator. The Hecke operators T_n with n coprime to N are self dual with respect to the Rosati involution.

To see that the group $J[1-F^2]/S$ is Eisenstein in the sense that $T_p = 1+p$ on this quotient for p prime to N, let η be the difference $T_p - (1+p)$, which is self dual with respect to the Rosati involution, since T_p is self dual and multiplication by the integer $(1+p)$ is also self dual. For $x \in J[1-F^2]$, we want to show that $\eta(x)$ is in the supersingular divisor class group; by [Pra95, Prop. 3.6], as mentioned above, this is the same as showing that $\langle \eta(x), y\rangle = 0$ for all $y \in \Sigma$. However, η annihilates Σ (see [Rib88, Thm. 1]), so

$$\langle \eta(x), y\rangle = \langle x, \eta(y)\rangle = 0.$$

□

The following proposition is a refinement of [Cor02, Prop. 4.4]. An *optimal quotient* A of $J_0(N)$ is any quotient of $J_0(N)$ by an abelian subvarieties (see [CS01, §3] for the basic properties of optimal quotients). For example, the abelian varieties A_f of Section 7.3 above are, up to isogeny, the simple optimal quotients of $J_0(N)$ that satisfy the hypothesis of Proposition 7.10 below.

Proposition 7.10. *Let A be any abelian variety optimal quotient of $J_0(N)$ such that $\ker(J_0(N) \to A)$ is Hecke stable, let $\mathfrak{m}$ be a non-Eisenstein maximal ideal of $\mathbb{T}$, and let X^0 be as in Equation (7.6). Then the natural map*

$$X^0 \to A(k) \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) \tag{7.7}$$

is surjective. In particular, if $A[\mathfrak{m}]$ is irreducible, then (7.7) is surjective.

Proof. As above, let S be the image of X^0 in $J(k)$, and let S_A be the image of S in $A(k)$. Also, let $Q = A(k)/S_A$. In light of Proposition 7.9, we have a commutative diagram of $\mathbb{T}$-modules with exact rows

$$\begin{tikzcd}
0 \arrow[r] & S \arrow[r] \arrow[d] & J(k) \arrow[r] \arrow[d] & \Sigma^\vee \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & S_A \arrow[r] & A(k) \arrow[r] & Q \arrow[r] & 0.
\end{tikzcd}$$

Since A is an optimal quotient of $J_0(N)$, there is an abelian variety B such that we have an exact sequence $0 \to B \to J_0(N) \to A \to 0$ of abelian varieties over $\mathbb{Q}$ with good reduction at ℓ (since $\ell \nmid N$). This sequences reduces to an exact sequence $0 \to B_k \to J \to A_k \to 0$ over k by [BLR90, §7.5, Thm. 4] (we have "$e < p-1$", since $p = \ell$ is odd and $e = 1$). Lang's theorem (see [Lan56] or [Ser88, §VI.4]) implies that $\mathrm{H}^1(k, B_k) = 0$, so $J(k) \to A(k)$ is surjective. The snake lemma then implies that the vertical map $\Sigma^\vee \to Q$ is surjective.

If $\Sigma^\vee \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) \cong \Sigma^\vee/\mathfrak{m}\Sigma^\vee$ is nonzero then $I = \mathrm{Ann}_{\mathbb{T}}(\Sigma^\vee/\mathfrak{m}\Sigma^\vee)$ equals $\mathfrak{m}$ since $\mathfrak{m}$ is maximal. Every $\eta_q = T_q - (q+1)$ for $q \nmid N$ is in I, since $\Sigma^\vee$ is Eisenstein by Proposition 7.9. But some $\eta_q \notin \mathfrak{m}$, since $\mathfrak{m}$ is non-Eisenstein, a contradiction. Thus $\Sigma^\vee \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) = 0$, so upon tensoring the rightmost vertical surjection of the above diagram with $\mathbb{T}/\mathfrak{m}$, we conclude that $Q \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) = 0$. Tensoring the bottom row over $\mathbb{T}$ with $\mathbb{T}/I$ and using that tensor product is right exact again then implies that (7.7) is surjective.

Since $\mathfrak{m}$ is a maximal ideal such that $A[\mathfrak{m}]$ is irreducible (which implies by definition that $A[\mathfrak{m}] \neq 0$), there is a prime $q \nmid N$ such that $\eta_q = T_q - (1+q)$ does not act as 0 on $A[\mathfrak{m}]$, since otherwise $A[\mathfrak{m}]$ would have semisimplification the reducible representation $1 \oplus \chi$, where χ is the cyclotomic character. Thus $\mathfrak{m}$ is non-Eisenstein, and the first part of the proposition proves the second claim. □

7.5 Multiplicity one theorem

The results of this section may be viewed as a partial generalization of [Rib99, Theorem. 2.3] and [Eme02, Thm. 4.2, Thm. 4.6] to more general levels. In particular, we prove under mild hypothesis that the multiplicity of a certain submodule of the $\mathbb{T}$-module $\mathrm{Div}(X_0(N)^{\mathrm{ss}}_{\mathbb{F}_{\ell^2}}) \otimes \mathbb{F}_p$ is 1. Our proof proceeds by finding a natural injective map from this submodule into $J_0(N\ell)[p]$, and observing that the image lies in a 1-dimensional subspace, as a consequence of a general multiplicity one result for $J_1(N\ell)$. For any positive integer N, let $\mathbb{T}(N)$ denote the ring of Hecke operators acting on $S_2(\Gamma_0(N))$.

Let N be a positive integer and ℓ a prime that does divide N, and let $X = \mathrm{Div}(X_0(N)^{\mathrm{ss}}_{\mathbb{F}_{\ell^2}})$, as in Equation (7.4). Let $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ be a newform of level N and let $\mathfrak{m}_0$ be a maximal ideal of $\mathbb{T}(N)$ such that the following three conditions simultaneously hold:

1. $\mathfrak{m}_0$ has odd residue characteristic p,
2. $a_\ell, \ell + 1 \in \mathfrak{m}_0$, and
3. the 2-dimensional mod p Galois representation ρ attached to $\mathfrak{m}_0$ is absolutely irreducible.

By Ribet's level raising theorem (see [Rib90b]), for each choice of ± 1, there is a maximal ideal $\mathfrak{m}$ in the Hecke algebra $\mathbb{T} = \mathbb{T}(N\ell)$ such that $\rho_{\mathfrak{m}} \approx \rho$ and $T_\ell \pm 1 \in \mathfrak{m}$. Letting $J = J_0(N\ell)$, as explained in [RS01, §3.3], we have

$$J[\mathfrak{m}] \cong \bigoplus_{i=1}^{t} \rho, \tag{7.8}$$

for some integer $t \geq 1$ called the *multiplicity of* $\mathfrak{m}$. That $t \geq 1$ follows from an argument of Mazur, as explained in [RS01, §3.3].

Proposition 7.11. *We have* $\dim_{\mathbb{T}/\mathfrak{m}} \mathrm{Hom}(X, \mu_p)[\mathfrak{m}] \leq t$.

Proof. The proof is inspired by [Rib94, Prop. 7.7], though that argument takes place in the midst of a proof by contradiction.

Let $G_\ell \approx \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ be the decomposition subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ associated to our fixed choice of prime λ of $\overline{\mathbb{Z}}$ over ℓ, and let $I_\ell \subset G_\ell$ be the inertia subgroup. Let $k = \mathbb{F}_\ell$, and let J_k be the special fiber of the Néron model of J at k. By [ST68, Lem. 2], we have $J_k[\mathfrak{m}] \cong J[\mathfrak{m}]^{I_\ell}$, and because ρ is unramified at ℓ, we have $J[\mathfrak{m}]^{I_\ell} = J[\mathfrak{m}]$, so $J_k[\mathfrak{m}] \cong J[\mathfrak{m}]$.

Let Φ be the component group of J_k. As explained in [CS01, §4], we have a diagram with an exact row and exact column, where T is the toric part of J_k^0 and B is an abelian variety:

$$\begin{tikzcd}
 & 0 \arrow[d] & & & \\
 & T \arrow[d] & & & \\
0 \arrow[r] & J_k^0 \arrow[r] \arrow[d] & J_k \arrow[r] & \Phi \arrow[r] & 0 \\
 & B \arrow[d] & & & \\
 & 0 & & &
\end{tikzcd}$$

Moreover, $X^0 \cong \mathrm{Hom}(T, \mathbb{G}_m)$, so $T \cong \mathrm{Hom}(X^0, \mathbb{G}_m)$ and $T[p] = \mathrm{Hom}(X^0, \mu_p)$. Hence

$$\mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}] = T[\mathfrak{m}] \hookrightarrow J_k^0[\mathfrak{m}] \subset J_k[\mathfrak{m}] \cong J[\mathfrak{m}]. \tag{7.9}$$

The representation ρ arises from level N, so is unramified at ℓ. The characteristic polynomial of $\rho(\mathrm{Frob}_\ell)$ is $x^2 - a_\ell x + \ell$. Our hypothesis 2 on a_ℓ and $\ell + 1$ imply that

$$x^2 - a_\ell x + \ell = x^2 - 1 \in \mathbb{F}_p[x].$$

Since $x^2 - 1 = (x-1)(x+1)$ and p is odd, we have a decomposition of $\mathbb{T}[D]$-modules $J[\mathfrak{m}] \cong J[\mathfrak{m}]^+ \oplus J[\mathfrak{m}]^-$ and (7.8) implies that the two summands have dimension t. Here we are using that $J[\mathfrak{m}] = \oplus\rho$; if V is the space underlying ρ, then V has dimension 2 and the characteristic polynomial of Frob_ℓ on V is $(x-1)(x+1)$, so V^+ and V^- each have dimension 1.

By [Rib90a, Prop. 3.7–3.8], the action of Frob_ℓ on X^0 is via $-T_\ell$. Since $T_\ell \pm 1 \in \mathfrak{m}$ (for some choice of sign), the action of Frob_ℓ on the $\mathbb{T}[D]$-module $\mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}]$ is by $\pm\ell$ (because Frob_ℓ acts on μ_p by ℓth powering). Since $\ell + 1 \in \mathfrak{m}_0$, we have $\ell \equiv \pm 1 \pmod{p}$, so we conclude that Frob_ℓ acts on $\mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}]$ as either $+1$ or -1. Thus the sequence of inclusions of Equation (7.9) sends $\mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}]$ to a submodule of $J[\mathfrak{m}]^\pm$ for one choice of sign, from which we conclude that

$$\dim_{\mathbb{T}/\mathfrak{m}} \mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}] \leq \dim_{\mathbb{T}/\mathfrak{m}} J[\mathfrak{m}]^\pm = t.$$

□

Lemma 7.12. *We have $X/\mathfrak{m}X \cong X^0/\mathfrak{m}X^0$. (In fact, this lemma is true for any non-Eisenstein maximal ideal $\mathfrak{m}$.)*

Proof. It follows from the explicit description of Hecke operators (see Section 6.1.3) that we have an exact sequence $0 \to X^0 \to X \xrightarrow{\deg} \mathbb{Z} \to 0$, where $\mathbb{T}$ acts on $\mathbb{Z}$ by $T_r = r + 1$ for r a prime coprime to $N\ell$. Tensoring this exact sequence over $\mathbb{T}$ with $\mathbb{T}/\mathfrak{m}$ yields an exact sequence

$$\mathrm{Tor}_1^{\mathbb{T}}(\mathbb{Z}, \mathbb{T}/\mathfrak{m}) \to X/\mathfrak{m}X \to X^0/\mathfrak{m}X^0 \to \mathbb{Z} \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) \to 0.$$

Since $\mathfrak{m}$ is non-Eisenstein, $\mathbb{Z} \otimes_{\mathbb{T}} (\mathbb{T}/\mathfrak{m}) = 0$ and

$$\mathrm{Tor}_1^{\mathbb{T}}(\mathbb{Z}, \mathbb{T}/\mathfrak{m}) = \mathrm{Tor}_1^{\mathbb{T}}(\mathbb{T}/\mathfrak{m}, \mathbb{Z}) = \mathbb{Z}[\mathfrak{m}] = 0.$$

□

Recall that $\mathfrak{m}$ is any maximal ideal of level $N\ell$ arising from level raising, as explained above (7.8) at the beginning of this section.

Proposition 7.13. *We have $\dim_{\mathbb{T}/\mathfrak{m}} \mathrm{Hom}(X, \mu_p)[\mathfrak{m}] \geq 1$.*

Proof. Let $A = A_f$ be the optimal quotient of $J_0(N)$ attached to f, let $k = \mathbb{F}_{\ell^2}$, and let $\mathbb{T} = \mathbb{T}(N)$. Consider the $\mathbb{T}[\mathrm{Frob}_\ell]$-module $M = A(k) \otimes \mathbb{T}/\mathfrak{m}_0$. Proposition 7.10 implies that the $\mathbb{T}$-module homomorphism

$$X^0 \to M \cong M^+ \oplus M^-$$

is surjective. Projection onto a one-dimensional $\mathbb{T}/\mathfrak{m}_0$-subspace of each of M^+ and M^- defines a nonzero element of $\mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}]$ for each of the two possible choices of $\mathfrak{m}$. Note that $\mathrm{Frob}_\ell^2 = 1$ on $A[\mathfrak{m}]$ by hypothesis, so $A[\mathfrak{m}](\overline{k}) = A[\mathfrak{m}](k)$. Here we also use that $\dim_{\mathbb{T}/\mathfrak{m}} A[\mathfrak{m}] \geq 1$ (see [RS01, §3.3]).

It is elementary that every element of $\mathrm{Hom}(X, \mu_p)[\mathfrak{m}]$ factors through $X/\mathfrak{m}X$ and likewise for X^0, so by Lemma 7.12 we have

$$\mathrm{Hom}(X, \mu_p)[\mathfrak{m}] \cong \mathrm{Hom}(X/\mathfrak{m}X, \mu_p)[\mathfrak{m}] \cong \mathrm{Hom}(X^0/\mathfrak{m}X^0, \mu_p)[\mathfrak{m}] \cong \mathrm{Hom}(X^0, \mu_p)[\mathfrak{m}].$$

□

Theorem 7.14. *If $p \nmid N$, then* $\dim_{\mathbb{T}/\mathfrak{m}} \mathrm{Hom}(X, \mu_p)[\mathfrak{m}] = 1$.

Proof. In light of the above two propositions, it suffices to show that $t = 1$, where t is the multiplicity in Equation (7.8). Let f be a cuspidal eigenform in $S_2(\Gamma_0(N\ell))$ such that $\mathrm{Ann}_{\mathbb{T}}(f) \subset \mathfrak{m}$, and view f as an element of $S_2(\Gamma_1(N\ell))$. Let $\mathfrak{m}_1$ be the inverse image of $\mathfrak{m}$ in $\mathbb{T}_1 = \mathbb{T}_1(N\ell)$ under the natural map $\mathbb{T}_1 \to \mathbb{T}$. Since $p > 2$ and $p \nmid N\ell$, [Edi92, Th. 9.2, part 1] implies that $\dim_{\mathbb{T}_1/\mathfrak{m}_1} J_1(N\ell)[\mathfrak{m}_1] = 2$. The inclusion $J_0(N\ell) \to J_1(N\ell)$ has kernel the Shimura subgroup, which is Eisenstein (by [Rib88, Thm. 1]), so $J_0(N\ell)[\mathfrak{m}] \hookrightarrow J_1(N\ell)[\mathfrak{m}_1]$. Since $t \geq 1$, this inclusion implies that $t = 1$. □

8 Implementation and Data

We implemented in Sage[1] algorithms based on the above results, and used them to compute $z_{c,\sigma,\ell}$ for 10 different rank 2 curves, and various primes ℓ, primes $q = 3, 5, 7$, discriminants D of class number 1, and primes c, as in Table 8.1. Let r_ℓ be the reduction map from Equation (7.1). We choose the pairs (E, ℓ) so that r_ℓ is surjective and if ℓ_1 and ℓ_2 are the first two primes for a given elliptic curve E, then $\ker(r_{\ell_1}) \cap \ker(r_{\ell_2}) = 0$. For each pair (E, ℓ) in the table, we considered all fundamental discriminants $D \leq -5$ such that $K = \mathbb{Q}(\sqrt{D})$ has class number 1, satisfies the Heegner hypothesis for E, has $\mathrm{ord}_{s=1} L(E^D, s) \leq 1$, and for which ℓ is inert. The restriction to class number 1 is not essential.

8.1 Tables

Table 8.1: Rank 2 curves, discriminants, and primes for which we computed $z_{c,\sigma,\ell}$.

E	D	p	ℓ
389a1	-7	3	5
389a1	-7	3	17
389a1	-7	3	41
389a1	-7	5	19
389a1	-11	3	17
389a1	-11	3	41
389a1	-11	5	19
389a1	-19	3	41
389a1	-67	3	5
389a1	-67	3	41
433a1	-8	5	79
433a1	-8	5	199
433a1	-11	3	17
433a1	-11	3	41
433a1	-11	5	79
563a1	-8	3	23
563a1	-163	3	17
563a1	-163	3	23
571b1	-7	3	47
571b1	-7	7	97
571b1	-7	7	167
571b1	-8	3	47
571b1	-8	5	29
571b1	-8	5	149
571b1	-8	7	167
571b1	-19	5	29
571b1	-19	7	97
571b1	-19	7	167
571b1	-67	3	11
571b1	-67	7	97
643a1	-8	3	29
643a1	-11	3	29
643a1	-19	3	29
643a1	-43	3	29
643a1	-67	3	11
655a1	-19	3	29
681c1	-8	3	23
709a1	-7	3	5
709a1	-7	3	47
709a1	-43	3	5
709a1	-67	3	5
709a1	-163	3	5
718b1	-7	3	5
997c1	-19	3	41
997c1	-67	3	41

[1] All computations in this section can be done in Version 4.6.1 using the free open source software Sage [S$^+$11]. Our implementation was peer reviewed by John Cremona for inclusion in Sage. Some relevant output files from running the computation can be found at `http://wstein.org/home/wstein/db/kolyconj/`. All computations were done under Linux (Ubuntu and Redhat) on several NSF-funded Sun Fire X4450 servers with 24 2.6Ghz cores and 128GB RAM each, at University of Washington and University of Georgia, and the computations took a few weeks CPU time.

We refer to elliptic curves using Cremona's notation (see [Crea]). Table 8.1 has columns E, D, p, ℓ. Each row has the property that E has rank 2, ℓ is inert in the field $K = \mathbb{Q}(\sqrt{D})$, and K satisfies the Heegner hypothesis for E. Also, we have $p \mid \gcd(\ell+1, a_\ell(E))$. We selected these examples because the $\mathbb{Z}$-rank of $\mathrm{Div}(X_0(N)^{\mathrm{ss}}_{\mathbb{F}_{\ell^2}})$ is relatively small (the dimensions are in Table 8.2).

The Tamagawa numbers of all of our curves are 1 or 2, and in all cases $\rho_{E,p}$ is surjective (see Proposition 8.1).

Table 8.2 contains data about the points $z_{c,\sigma,\ell}$. The columns labeled E, D, p, and ℓ correspond exactly to the entries in Table 8.2. The column labeled dim gives the dimension of $\mathrm{Div}(X_0(N)^{\mathrm{ss}}_{\mathbb{F}_{\ell^2}})$; this dimension directly impacts the runtime of our implementation. The column labeled $\max c$ contains the largest c such that we managed to compute $z_{c,\sigma,\ell}$. The columns labeled "$= 0$" and "$\neq 0$" are a count of how many $z_{c,\sigma,\ell}$ are 0 and not 0 among those we computed; note that for each c, ℓ we compute $z_{c,\sigma,\ell}$ for only one choice of generator σ (see below for how we chose σ), since other choices of σ would yield a nonzero scalar multiple, hence we often just write $z_{c,\ell}$. The columns labeled $z_{c,\ell} = 0$ and $z_{c,\ell} \neq 0$ give the first few c such that $z_{c,\ell}$ is zero or nonzero, respectively.

A consistency check on Table 8.2 comes from the rows labeled (**389a1**, $-7, 3, 17$) and (**389a1**, $-7, 3, 41$), since the reduction maps

$$E(\mathbb{Q}) \to E(\mathbb{F}_\ell) \otimes (\mathbb{Z}/3\mathbb{Z})$$

have the same kernel for $\ell = 17$ and 41. Hence the $z_{c,17} \neq$ if and only if $z_{c,41} \neq 0$, which was indeed the case in the range of our computations.

In every single case in Table 8.2 we find at least one c such that $z_{c,\ell} \neq 0$, so Conjecture 1.1 is true in these cases.

One initially surprising numerical observation we made is that the classes $\tau_{c,p}$ appear to *not* be equidistributed in the most naive possible sense. For example, in our computations with $p = 3$, the 0 subspace occurs about twice as much as any other subspace. Once we know that one class is nonzero, the exact asymptotic distribution of *all* classes can then be determined as an application of work of Mazur-Rubin, B. Howard [How04], and the Chebotarev density theorem. See the forthcoming paper [SW10]. As mentioned in Remark 7.1 above, this also leads to an alternate way to compute $\tau_{c,p}$ up to scaling. This provided an convincing double check on the correctness of our tables.

Tables 8.3–8.4 provide further details about the distribution of elements of

$$\mathrm{Sel}^{(p)}(E/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^2$$

coming from this construction. The first 5 columns labeled E, D, p, ℓ_1 and ℓ_2 specify an elliptic curve, fundamental discriminant D, a prime p and two primes ℓ_1 and ℓ_2, chosen from the data summarized in Table 8.2. As mentioned above, the primes ℓ_1 and ℓ_2 are chosen so that the intersection of the two reduction maps to $E(\mathbb{F}_{\ell_i}) \otimes (\mathbb{Z}/p\mathbb{Z})$ is 0. Since the Selmer group has dimension 2 and in our implementation we chose the generator $\sigma \in \mathrm{Gal}(K_c/K_1) \cong (\mathcal{O}_K/c\mathcal{O}_K)^\times/(\mathbb{Z}/c\mathbb{Z})^\times$ to be $\sqrt{D} + n$ with $n \geq 1$ minimal, where $D = \mathrm{disc}(K)$. This allows us to deduce the subspace spanned by $\tau_{c,p}$ in $\mathrm{Sel}^{(p)}(E/\mathbb{Q})$ with respect to some unknown basis for $\mathrm{Sel}^{(p)}(E/\mathbb{Q})$. The column labeled $\tau_{c,p}$ gives the normalized generator for this subspace. The next column, labeled # gives the number of c such that $\tau_{c,p}$ spans the given subspace, and the last column gives the first few such primes c.

Table 8.2: Data about $z_{c,\sigma,\ell}$.

E	D	q	ℓ	dim	$\max c$	$= 0$	$\neq 0$	c with $z_{c,\ell} = 0$	c with $z_{c,\ell} \neq 0$
389a1	-7	3	5	130	19031	152	121	17, 173, 227, 269	41, 59, 83, 587
389a1	-7	3	17	520	14657	122	92	41, 83, 173, 227	5, 59, 503, 587
389a1	-7	3	41	1300	11681	102	74	17, 83, 173, 227	5, 59, 503, 587
389a1	-7	5	19	586	28229	32	67	349, 509, 769, 2539	419, 929, 1049, 1399
389a1	-11	3	17	520	14717	116	101	29, 41, 83, 107	233, 263, 347, 479
389a1	-11	3	41	1300	14879	117	104	17, 29, 83, 107	233, 263, 347, 479
389a1	-11	5	19	586	22189	24	60	239, 569, 1759, 1999	149, 349, 359, 769
389a1	-19	3	41	1300	14699	132	98	29, 53, 107, 227	59, 113, 173, 449
389a1	-67	3	5	130	23663	170	147	41, 113, 281, 347	53, 233, 599, 653
389a1	-67	3	41	1300	15473	129	82	53, 113, 281, 587	5, 233, 347, 503
433a1	-8	5	79	2822	15199	19	30	1319, 2269, 2549, 3079	199, 389, 1039, 1669
433a1	-8	5	199	7162	11149	14	26	1319, 1879, 2269, 2549	79, 389, 1039, 1669
433a1	-11	3	17	580	12473	91	88	131, 239, 293, 359	41, 83, 107, 197
433a1	-11	3	41	1448	11579	82	84	239, 281, 293, 359	17, 83, 107, 131
433a1	-11	5	79	2822	15329	12	37	1889, 2309, 3079, 4759	409, 1289, 1319, 1669
563a1	-8	3	23	1034	14813	113	109	197, 263, 311, 383	47, 173, 191, 269
563a1	-163	3	17	752	15887	123	93	137, 293, 311, 887	23, 59, 191, 269
563a1	-163	3	23	1034	15149	114	92	137, 311, 521, 569	17, 59, 191, 269
571b1	-7	7	97	4576	12011	15	32	167, 503, 937, 1511	349, 839, 881, 1063
571b1	-7	7	167	7914	9547	16	16	97, 503, 937, 1063	349, 839, 881, 1483
571b1	-8	5	149	7056	11159	5	43	29, 1319, 2239, 7639	79, 229, 349, 359
571b1	-8	7	167	7914	12109	8	13	1063, 1861, 2141, 2309	349, 503, 839, 1511
571b1	-19	5	29	1336	15259	16	33	79, 1709, 2179, 2339	439, 829, 1229, 1319
571b1	-19	7	97	4576	13789	9	23	2309, 2953, 4157, 7349	167, 839, 1063, 1511
571b1	-19	7	167	7914	10639	9	13	97, 1063, 1861, 2141	839, 1511, 1931, 3989
571b1	-67	3	11	478	16889	129	108	239, 281, 353, 521	191, 233, 251, 311
571b1	-67	7	97	4576	12641	9	14	503, 2239, 4157, 4507	937, 1063, 1861, 2309
643a1	-8	3	29	1504	12527	104	82	47, 71, 149, 173	167, 263, 359, 431
643a1	-11	3	29	1504	12953	91	93	83, 131, 149, 197	167, 173, 263, 359
643a1	-19	3	29	1504	12143	107	86	89, 293, 509, 641	71, 113, 167, 173
643a1	-43	3	29	1504	12647	102	83	89, 131, 137, 149	71, 113, 503, 521
643a1	-67	3	11	538	14753	115	104	113, 137, 191, 251	197, 311, 353, 443
655a1	-19	3	29	1848	12149	96	77	59, 89, 113, 167	53, 179, 227, 257
681c1	-8	3	23	1672	11909	101	81	29, 47, 167, 263	191, 317, 479, 557
709a1	-7	3	5	238	16061	131	107	47, 257, 269, 419	59, 83, 227, 353
709a1	-7	3	47	2724	9833	92	56	257, 269, 419, 503	5, 59, 83, 227
709a1	-43	3	5	238	16319	131	118	149, 233, 389, 503	137, 179, 227, 257
709a1	-67	3	5	238	16301	133	109	179, 197, 233, 353	137, 239, 281, 503
709a1	-163	3	5	238	16883	138	107	233, 239, 353, 479	59, 137, 149, 257
718b1	-7	3	5	360	15137	122	100	41, 47, 131, 167	101, 251, 353, 839
997c1	-19	3	41	3328	8297	66	63	179, 227, 269, 449	113, 173, 383, 677
997c1	-67	3	41	3328	8231	76	61	179, 191, 311, 347	113, 197, 383, 647

Table 8.3: Data about normalized elements $\tau_{c,p} \in \mathrm{Sel}^{(q)}(E/\mathbb{Q})$ (part 1 of 2)

E	D	q	ℓ_1	ℓ_2	$\tau_{c,p}$	#	at most first 10 primes c
389a1	-7	3	5	17	$(0,0)$	87	173, 227, 269, 479, 509, 761, 797, 929, 1013, 1181
					$(0,1)$	30	503, 773, 1049, 1193, 1487, 2897, 3359, 4157, 5333, 5843
					$(1,0)$	35	41, 83, 857, 1151, 1553, 1637, 1907, 2141, 2393, 2441
					$(1,1)$	34	59, 587, 941, 1307, 1571, 1721, 2273, 2399, 3407, 3797
					$(1,2)$	27	1091, 1217, 1931, 2579, 3191, 3779, 4493, 5477, 6011, 6173
389a1	-7	3	5	41	$(0,0)$	75	17, 173, 227, 269, 479, 509, 761, 797, 929, 1013
					$(0,1)$	25	503, 773, 1049, 1193, 1487, 2897, 3359, 4157, 5333, 5843
					$(1,0)$	27	83, 857, 1151, 1553, 1637, 1907, 2141, 2393, 2441, 2477
					$(1,1)$	29	59, 587, 941, 1307, 1571, 1721, 2273, 2399, 3407, 3797
					$(1,2)$	19	1091, 1217, 1931, 2579, 3191, 3779, 4493, 5477, 6011, 6173
389a1	-67	3	5	41	$(0,0)$	95	113, 281, 587, 857, 1013, 1049, 1187, 1481, 1571, 1583
					$(0,1)$	25	347, 503, 683, 929, 1319, 1487, 2129, 2687, 3947, 4157
					$(1,0)$	34	53, 653, 1151, 1553, 1907, 2207, 2393, 2417, 2423, 3167
					$(1,1)$	26	233, 599, 1181, 1217, 1409, 2657, 3779, 4019, 5387, 5477
					$(1,2)$	30	941, 1307, 1709, 1721, 2339, 2549, 2909, 3467, 3797, 3821
433a1	-8	5	79	199	$(0,0)$	11	1319, 2269, 2549, 3079, 3319, 4349, 4759, 4799, 6949, 7879
					$(0,1)$	3	6719, 8389, 8669
					$(1,0)$	3	1879, 4549, 6679
					$(1,1)$	4	1669, 2879, 5119, 5399
					$(1,2)$	3	5839, 6029, 9949
					$(1,3)$	6	2239, 3389, 4079, 5639, 7589, 11149
					$(1,4)$	9	389, 1039, 2309, 2749, 4789, 6599, 7669, 9349, 9679
433a1	-11	3	17	41	$(0,0)$	63	239, 293, 359, 503, 563, 659, 761, 821, 1097, 1217
					$(0,1)$	21	131, 677, 1031, 1427, 1601, 1979, 2129, 2213, 3797, 4451
					$(1,0)$	19	281, 479, 857, 1019, 1949, 2207, 2309, 2609, 4421, 5147
					$(1,1)$	36	83, 107, 701, 941, 953, 1091, 1223, 1667, 1913, 2087
					$(1,2)$	26	197, 263, 431, 887, 2741, 2837, 3137, 3209, 3659, 3803
563a1	-163	3	17	23	$(0,0)$	88	137, 311, 887, 929, 953, 1217, 1223, 1367, 1583, 1733
					$(0,1)$	28	293, 983, 1433, 1553, 2213, 2843, 3923, 4397, 4691, 5927
					$(1,0)$	26	521, 569, 587, 863, 1289, 1427, 1637, 3167, 3863, 4481
					$(1,1)$	31	59, 269, 353, 509, 977, 1709, 1979, 2399, 2801, 3413
					$(1,2)$	32	191, 317, 761, 827, 1283, 1409, 1871, 3779, 3911, 4049

Table 8.4: Data about normalized elements $\tau_{c,p} \in \mathrm{Sel}^{(q)}(E/\mathbb{Q})$ (part 2 of 2)

E	D	q	ℓ_1	ℓ_2	$\tau_{c,p}$	#	at most first 10 primes c
571b1	-7	7	97	167	$(0,0)$	9	503, 937, 1511, 3989, 4157, 4507, 6691, 7349, 9421
					$(0,1)$	2	2239, 7489
					$(1,0)$	6	1063, 1861, 2141, 2309, 5039, 8581
					$(1,1)$	2	349, 9547
					$(1,2)$	2	5417, 6131
					$(1,3)$	4	881, 1931, 2099, 5683
					$(1,4)$	2	839, 1483
					$(1,5)$	2	3163, 6229
					$(1,6)$	2	2953, 6719
571b1	-19	7	97	167	$(0,0)$	4	2309, 2953, 4157, 7349
					$(0,1)$	1	7489
					$(1,0)$	4	1063, 1861, 2141, 8581
					$(1,1)$	2	3989, 10639
					$(1,2)$	3	5417, 6131, 9883
					$(1,3)$	2	1931, 5683
					$(1,4)$	2	839, 1511
					$(1,5)$	1	6691
					$(1,6)$	2	6719, 10331
709a1	-7	3	5	47	$(0,0)$	62	257, 269, 419, 593, 839, 857, 881, 929, 971, 1433
					$(0,1)$	17	479, 1091, 1319, 1553, 2243, 4049, 4259, 4289, 4973, 5519
					$(1,0)$	30	503, 647, 677, 1049, 1151, 1181, 1301, 1613, 1697, 2267
					$(1,1)$	16	353, 521, 563, 1097, 1427, 1637, 1949, 2579, 2621, 2687
					$(1,2)$	22	59, 83, 227, 773, 983, 1259, 2897, 2939, 3779, 4721

Table 8.5: Data about **non-scaled** elements $\tau_{c,p} \in \mathrm{Sel}^{(q)}(E/\mathbb{Q})$ (part 1 of 2)

E	D	q	ℓ_1	ℓ_2	$\tau_{c,p}$	#	at most first 13 primes c
389a1	-7	3	5	17	$(0,0)$	87	173, 227, 269, 479, 509, 761, 797, 929, 1013, 1181, 1319, 1511, 1601
					$(0,1)$	15	1487, 2897, 3359, 4157, 5843, 6317, 6653, 6803, 7229, 7901, 8237, 9551, 10559
					$(0,2)$	15	503, 773, 1049, 1193, 5333, 6971, 8069, 9371, 9623, 10457, 11483, 11681, 13151
					$(1,0)$	21	41, 83, 857, 1553, 1637, 2393, 2441, 2477, 3167, 4217, 6053, 6221, 7103
					$(1,1)$	16	1307, 1571, 1721, 2399, 3407, 4091, 4721, 5171, 6389, 6977, 7451, 8501, 8627
					$(1,2)$	17	1217, 3191, 3779, 5477, 6011, 6173, 6947, 8363, 8951, 9173, 9929, 11087, 11927
					$(2,0)$	14	1151, 1907, 2141, 3461, 3617, 6257, 7019, 7727, 10463, 10589, 11171, 12101, 12983
					$(2,1)$	10	1091, 1931, 2579, 4493, 8039, 10163, 10433, 13313, 13331, 14621
					$(2,2)$	18	59, 587, 941, 2273, 3797, 4457, 4751, 4973, 5309, 6569, 7817, 8111, 8123
389a1	-7	3	5	41	$(0,0)$	75	17, 173, 227, 269, 479, 509, 761, 797, 929, 1013, 1181, 1319, 1511
					$(0,1)$	13	1487, 2897, 3359, 4157, 5843, 6317, 6653, 6803, 7229, 7901, 8237, 9551, 10559
					$(0,2)$	12	503, 773, 1049, 1193, 5333, 6971, 8069, 9371, 9623, 10457, 11483, 11681
					$(1,0)$	16	83, 857, 1553, 1637, 2393, 2441, 2477, 3167, 4217, 6053, 6221, 7103, 8573
					$(1,1)$	14	1307, 1571, 1721, 2399, 3407, 4091, 4721, 5171, 6389, 6977, 7451, 8501, 8627
					$(1,2)$	12	1217, 3191, 3779, 5477, 6011, 6173, 6947, 8363, 8951, 9173, 9929, 11087
					$(2,0)$	11	1151, 1907, 2141, 3461, 3617, 6257, 7019, 7727, 10463, 10589, 11171
					$(2,1)$	7	1091, 1931, 2579, 4493, 8039, 10163, 10433
					$(2,2)$	15	59, 587, 941, 2273, 3797, 4457, 4751, 4973, 5309, 6569, 7817, 8111, 8123
389a1	-67	3	5	41	$(0,0)$	95	113, 281, 587, 857, 1013, 1049, 1187, 1481, 1571, 1583, 1811, 1889, 2531
					$(0,1)$	10	347, 503, 683, 929, 1487, 4157, 5639, 13649, 14051, 14969
					$(0,2)$	15	1319, 2129, 2687, 3947, 4583, 4673, 5867, 6551, 6653, 7109, 8807, 9371, 10259
					$(1,0)$	16	53, 1151, 1553, 2417, 2423, 3167, 3461, 5279, 5741, 7583, 8741, 8819, 9521
					$(1,1)$	13	233, 1217, 2657, 3779, 5387, 7649, 7757, 8039, 9041, 10973, 12659, 14879, 15053
					$(1,2)$	12	1721, 3467, 3821, 5171, 5231, 6143, 10331, 13613, 14033, 14321, 14669, 14717
					$(2,0)$	18	653, 1907, 2207, 2393, 3617, 4229, 4253, 4937, 5471, 6221, 7019, 7547, 7643
					$(2,1)$	18	941, 1307, 1709, 2339, 2549, 2909, 3797, 4463, 5237, 6779, 7481, 8627, 8849
					$(2,2)$	13	599, 1181, 1409, 4019, 5477, 7331, 8093, 8243, 11087, 11489, 12263, 12671, 15083
433a1	-8	5	79	199	$(0,0)$	11	1319, 2269, 2549, 3079, 3319, 4349, 4759, 4799, 6949, 7879, 11069
					$(0,1)$	1	8669
					$(0,2)$	0	
					$(0,3)$	0	
					$(0,4)$	2	6719, 8389
					$(1,0)$	2	1879, 6679
					$(1,1)$	2	1669, 5119
					$(1,2)$	1	6029
					$(1,3)$	0	
					$(1,4)$	2	389, 2749
					$(2,0)$	1	4549
					$(2,1)$	2	3389, 11149
					$(2,2)$	0	
					$(2,3)$	1	6599
					$(2,4)$	1	9949
					$(3,0)$	0	
					$(3,1)$	1	5839
					$(3,2)$	6	1039, 2309, 4789, 7669, 9349, 9679
					$(3,3)$	1	2879
					$(3,4)$	1	5639
					$(4,0)$	0	
					$(4,1)$	0	
					$(4,2)$	3	2239, 4079, 7589
					$(4,3)$	0	
					$(4,4)$	1	5399

Table 8.6: Data about **non-scaled** elements $\tau_{c,p} \in \mathrm{Sel}^{(q)}(E/\mathbb{Q})$ (part 2 of 2)

E	D	q	ℓ_1	ℓ_2	$\tau_{c,p}$	#	at most first 13 primes c
433a1	-11	3	17	41	$(0,0)$	63	239, 293, 359, 503, 563, 659, 761, 821, 1097, 1217, 1319, 1487, 1613
					$(0,1)$	11	131, 677, 1031, 1979, 2213, 3797, 4451, 5939, 9437, 9473, 11483
					$(0,2)$	10	1427, 1601, 2129, 4517, 5189, 5507, 5711, 5741, 9257, 10247
					$(1,0)$	13	281, 479, 857, 1949, 2207, 2309, 2609, 4421, 5147, 5297, 5519, 10067, 10691
					$(1,1)$	19	107, 701, 941, 1091, 2087, 2969, 3119, 3527, 4133, 4583, 5279, 5309, 7127
					$(1,2)$	17	197, 431, 887, 2741, 2837, 3209, 3659, 3803, 4241, 4253, 4523, 6701, 7229
					$(2,0)$	6	1019, 5231, 5639, 7211, 9467, 10457
					$(2,1)$	9	263, 3137, 6269, 6299, 7829, 8147, 8861, 9941, 10589
					$(2,2)$	17	83, 953, 1223, 1667, 1913, 2459, 2591, 3533, 4157, 6113, 6221, 6761, 7487
563a1	-163	3	17	23	$(0,0)$	88	137, 311, 887, 929, 953, 1217, 1223, 1367, 1583, 1733, 1811, 1907, 2243
					$(0,1)$	15	983, 2843, 4397, 5927, 6389, 6869, 7949, 8093, 8363, 8669, 8753, 11159, 11489
					$(0,2)$	13	293, 1433, 1553, 2213, 3923, 4691, 7673, 8273, 11069, 11243, 12569, 14699, 15149
					$(1,0)$	12	521, 587, 1637, 4583, 5507, 6449, 8429, 11969, 12161, 12959, 13649, 13907
					$(1,1)$	12	59, 353, 977, 1979, 2399, 2801, 3413, 4217, 4241, 6701, 10289, 10709
					$(1,2)$	14	191, 761, 827, 3911, 4391, 6863, 8111, 9419, 9491, 9521, 10133, 12491, 13751
					$(2,0)$	14	569, 863, 1289, 1427, 3167, 3863, 4481, 4793, 4799, 6323, 6983, 7703, 10067
					$(2,1)$	18	317, 1283, 1409, 1871, 3779, 4049, 4673, 5783, 6143, 6317, 6971, 9341, 9803
					$(2,2)$	19	269, 509, 1709, 3617, 4283, 4721, 6551, 7727, 9371, 9887, 10301, 10391, 12497
709a1	-7	3	5	47	$(0,0)$	62	257, 269, 419, 593, 839, 857, 881, 929, 971, 1433, 1487, 1511, 1571
					$(0,1)$	7	479, 1091, 4259, 5519, 6299, 6359, 7481
					$(0,2)$	10	1319, 1553, 2243, 4049, 4289, 4973, 5843, 5927, 6053, 6803
					$(1,0)$	16	647, 1049, 1151, 1181, 1697, 2957, 3449, 4283, 4637, 5879, 6047, 7187, 7229
					$(1,1)$	10	353, 563, 1097, 1427, 1637, 2621, 2687, 3191, 5897, 6221
					$(1,2)$	7	59, 227, 1259, 4721, 4919, 7829, 7937
					$(2,0)$	14	503, 677, 1301, 1613, 2267, 2693, 2903, 3491, 3671, 4217, 5393, 8627, 9467
					$(2,1)$	15	83, 773, 983, 2897, 2939, 3779, 4751, 5381, 6173, 6317, 6737, 6977, 8123
					$(2,2)$	6	521, 1949, 2579, 3659, 6011, 7649

8.2 Appendix: remarks about surjectivity of Galois representations

In order to pass from $[P_{c,\sigma}]$ to $\tau_{c,p^n} \in \mathrm{H}^1(K, E[p^n])$ in Section 5.2, we assumed that p is an odd prime such that

$$\rho_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Z}_p)$$

is surjective. If we assume that E does not have CM (as will be the case for our examples), the p-adic representation $\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}_p)$ is surjective for all but finitely many p. Moreover, we can compute all primes p such that $\rho_{E,p}$ is not surjective, as explained in [GJP$^+$09, §2.1] and implemented in Sage (see also forthcoming work of A. Sutherland [Sut09]). For example, we have the following proposition:

Proposition 8.1. *If E is a rank 2 elliptic curve with conductor < 1058, then $\rho_{E,p}$ is surjective for all odd primes p.*

Proof. Using the algorithm of [GJP$^+$09, §2.1] as implemented in [S$^+$11] shows that the mod-p representations $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ are surjective for all rank 2 curves E of conductor < 1058 and all primes p. As explained in [GJP$^+$09, §2.1], this implies that the p-adic representation $\rho_{E,p}$ is surjective for $p \geq 5$.

It remains to deal with $p = 3$. For $p = 3$ we use the method of [Elk06], namely that it is enough to check that $j(E) - f(x)$ has no rational zero, where $f(x)$ is the function

$$f(x) = \frac{3^7 \cdot (x^2-1)^3 \cdot (x^6+3x^5+6x^4+x^3-3x^2+12x+16)^3 \cdot (2x^3+3x^2-3x-5)}{(x^3-3x-1)^9}$$

of degree 27 from [Elk06, pg. 5]. Elkies remarks (see [Elk10]) that there is a minus sign in the formula in [Elk06, pg. 5] that does not belong, as we verify by trying the integral specializations tabulated on [Elk06, pg. 7], and also by factoring $f - 1728$. Doing this computation for our curves yields the claimed result.

□

Remark 8.2. Andrew Sutherland used the techniques of [Sut09] to show [Sut10] that "the rank 2 elliptic curves with conductor less than 1058 all have surjective Galois images in $\mathrm{GL}_2(\mathbb{Z}/16\mathbb{Z})$." We thus also expect that $\rho_{E,2}$ is surjective for all rank 2 curves with conductor less than 1058.

Remark 8.3. The rank 2 curve 1058c1 has a rational 3 isogeny.

9 Related Projects

There are several future projects that are suggested by this paper, and we briefly sketch some of the most promising ones here.

We can do the same computations as we do here, but for modular abelian varieties A_f attached to newforms with $\mathrm{ord}_{s=1} L(f,s) \geq 2$. There is a table of such abelian varieties in [AS05]. For example, we carried out this computation for the modular abelian variety **1061b** of dimension 2 and indeed verified the natural higher dimensional analogue of Kolyvagin's conjecture for this abelian variety (for $p = 3$). Note that Kolyvagin appears to have never explicitly made such a conjecture, though of course he considers modular abelian varieties in [KL89]. We could also use our method to show that $Ш(A_f/\mathbb{Q})[p] = 0$ for a particular A_f, even when $\operatorname{ord} L(f,s) \leq 1$. This may require extending Kolyvagin's structure theorem to abelian varieties, or otherwise making results of [KL89] more explicit.

We could verify Conjecture 1.1 for the rank 3 elliptic curve of conductor 5077, and possibly some other rank 3 curves. Indeed, Jennifer Balakrishnan and the author have verified Conjecture 1.1 at least for **5077a** for $p = 3$.

It would be of interest to generalize Algorithm 2.1 to treat the case p^n with $n > 1$ or the case when $\rho_{E,p}$ is reducible. We could also consider an example such as the rank 2 curve **916c1** and $p = 3$ in which p divides a Tamagawa number.

Since we are doing explicit computation, it would also be interesting to closely investigate the case $p = 2$; this is particularly exciting when $r_{\mathrm{an}}(E/\mathbb{Q}) = 2$, since, after a harmless trace (as in Remark 5.7), we find that the points y_c, for c prime, are defined over **real quadratic** extensions of $\mathbb{Q}$, and define explicit elements of $\mathrm{Sel}^{(2)}(E/\mathbb{Q})$ that define globally trivial [2]-coverings $X \to E$. For example, if we take E to be **389a**, $K = \mathbb{Q}(\sqrt{-7})$ and $c = 3$, then y_3 is defined over a cyclic degree 4 extension K_3 of K; the trace z_3 of y_3 to the quadratic subfield of K_3 is defined over the real quadratic field $\mathbb{Q}(\sqrt{21})$; it is the point

$$z_3 = \left(-\frac{131}{84}, \frac{1091}{3528}\sqrt{21} - \frac{1}{2}\right).$$

Also, we find that $0 \neq \tau_{3,2} = \delta((0,0)) \in \mathrm{Sel}^{(2)}(E/\mathbb{Q})$. Is there any connection between these Heegner points over real quadratic fields and Stark-Heegner points?

Much of the work of Kolyvagin and Gross-Zagier has been generalized to totally real fields by Zhang and his students. Likewise, it would be of interest to see to what extent the results of this paper generalize to totally real fields.

It would also be of interest to investigate rank 2 curves E for which E^D exhibits some unusual behavior, e.g., nontrivial odd Ш or rank ≥ 3. For example, for E the curve **389a** of rank 2, and $K = \mathbb{Q}(\sqrt{-264})$, which has class number 8, the twist E^D has rank 3, so Kolyvagin's structure theorem implies that $[P_{c,\sigma}] = 0$ for all prime c, and it would be interesting to (a) computationally observe this, and (b) find a c that is a product of primes for which $[P_{c,\sigma}] \neq 0$. Similarly, if we take $K = \mathbb{Q}(\sqrt{-667})$, then K has class number 4 and $5 \mid \#Ш(E^D/\mathbb{Q})$; thus we expect that $[P_{c,5}] = 0$ for all prime c. Again it would beinteresting to observe this computationally, and find a prime c such that $[P_{c,5^2}] \neq 0$.

As a challenge, we could attempt to verify Conjecture 1.1 for the rank 4 elliptic curve of conductor 234446 given by the equation $y^2 + xy = x^3 - x^2 - 79x + 289$. This computation is at the edge of feasible, so it will require very sophisticated linear algebra or some other new idea.

References

[AS05] Agashe Agashe and William Stein, *Visible evidence for the Birch and Swinnerton-Dyer conjecture for modular abelian varieties of analytic rank zero*, Math. Comp. **74** (2005), no. 249, 455–484 (electronic), With an appendix by J. Cremona and B. Mazur, `http://wstein.org/papers/shacomp/`. MR 2085902

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic), http://math.stanford.edu/~conrad/papers/tswfinal.pdf. MR 2002d:11058

[BFH90] Daniel Bump, Solomon Friedberg, and Jeffrey Hoffstein, *Nonvanishing theorems for L-functions of modular forms and their derivatives*, Invent. Math. **102** (1990), no. 3, 543–618, http://wstein.org/papers/bib/bump-friedberg-hoffstein-nonvanishing.pdf. MR 1074487 (92a:11058)

[Bir65] B. J. Birch, *Conjectures concerning elliptic curves*, Proceedings of Symposia in Pure Mathematics, VIII, Amer. Math. Soc., Providence, R.I., 1965, http://wstein.org/papers/bib/Birch-Conjectures_Concerning_Elliptic_Curves.pdf, pp. 106–112. MR 30 #4759

[Bir71] B. J. Birch, *Elliptic curves over* **Q***: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, http://wstein.org/papers/bib/Birch-Elliptic_curves_over_Q-A_Progress_Report.pdf, pp. 396–400.

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR 91i:14034

[Bra10] Robert Bradshaw, *Provable Computation of Motivic L-functions*, University of Washington Ph.D. Thesis under William Stein (2010), http://www.sagemath.org/files/thesis/bradshaw-thesis-2010.pdf.

[BS10] M. Bhargava and A. Shankar, *Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves*, Preprint (2010), http://arxiv.org/abs/1006.1002.

[BS11] R. Bradshaw and W. A. Stein, *Heegner Points and the Arithmetic of Elliptic Curves over Ring Class Extensions*, Submitted (2011), http://wstein.org/papers/bs-heegner/.

[BY09] Jan Hendrik Bruinier and Tonghai Yang, *Faltings heights of CM cycles and derivatives of L-functions*, Invent. Math. **177** (2009), no. 3, 631–681, http://arxiv.org/abs/0807.0502. MR 2534103

[CFO$^+$08] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, *Explicit n-descent on elliptic curves. I. Algebra*, J. Reine Angew. Math. **615** (2008), 121–155. MR 2384334 (2009g:11067)

[CLS09] J. Coates, Z. Liang, and R. Sujatha, *The Tate-Shafarevich group for elliptic curves with complex multiplication*, J. Algebra **322** (2009), no. 3, 657–674. MR 2531216 (2010e:11052)

[CLS10] ———, *The Tate-Shafarevich group for elliptic curves with complex multiplication II*, Preprint (2010), http://arxiv.org/abs/1005.4206.

[Coh07] Henri Cohen, *Number theory. Vol. I. Tools and Diophantine equations*, Graduate Texts in Mathematics, vol. 239, Springer, New York, 2007. MR 2312337 (2008e:11001)

[Cor02] Christophe Cornut, *Mazur's conjecture on higher Heegner points*, Invent. Math. **148** (2002), no. 3, 495–523, http://www.math.jussieu.fr/~cornut/papers/mcinv_published.pdf.

[Crea] J. E. Cremona, *Elliptic Curves Data*, http://www.warwick.ac.uk/~masgaj/ftp/data/.

[Creb] ______, mwrank *(computer software)*, http://www.warwick.ac.uk/~masgaj/mwrank/.

[Cre97] ______, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, http://www.warwick.ac.uk/~masgaj/book/fulltext/.

[CS01] B. Conrad and W. A. Stein, *Component groups of purely toric quotients*, Math. Res. Lett. **8** (2001), no. 5-6, 745–766, http://wstein.org/papers/compgrp/. MR 2003f:11087

[ÇW08] Mirela Çiperiani and Andrew Wiles, *Solvable points on genus one curves*, Duke Math. J. **142** (2008), no. 3, 381–464, http://www.ma.utexas.edu/users/mirela/solvable.pdf. MR 2412044 (2009m:11092)

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149, http://arxiv.org/abs/math/0207280. MR 2068888 (2005f:11128)

[Edi92] B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594, http://www.math.leidenuniv.nl/~edix/public_html_rennes/publications/weight.html.

[Elk06] Noam Elkies, *Elliptic curves with 3-adic Galois representation surjective mod 3 but not mod 9*, Preprint (2006), http://arxiv.org/abs/math/0612734.

[Elk10] ______, *Email: surjective p-adic repn for some rank 2 curves*, Personal Communication (2010), http://wstein.org/papers/bib/2010-elkies-3.txt.

[Eme02] Matthew Emerton, *Supersingular elliptic curves, theta series and weight two modular forms*, J. Amer. Math. Soc. **15** (2002), no. 3, 671–714 (electronic), http://www.math.northwestern.edu/~emerton/pdffiles/two.pdf. MR 1896237 (2003b:11038)

[GJP$^+$09] G. Grigorov, A. Jorza, S. Patrikis, C. Tarnita, and W. Stein, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. **78** (2009), 2397–2425, http://wstein.org/papers/bsdalg/.

[Gro84] Benedict H. Gross, *Heegner points on $X_0(N)$*, Modular forms (Durham, 1983), Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., Horwood, Chichester, 1984, http://wstein.org/papers/bib/gross-heegner_points_on_X0N.pdf, pp. 87–105. MR 803364 (87f:11036b)

[Gro87] ______, *Heights and the special values of L-series*, Number theory (Montreal, Que., 1985), CMS Conf. Proc., vol. 7, Amer. Math. Soc., Providence, RI, 1987, http://wstein.org/papers/bib/Gross-Heights_and_the_Special_Values_of_L-series.pdf, pp. 115–187. MR 894322 (89c:11082)

[Gro91] B. H. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, http://wstein.org/papers/bib/gross-kolyvagins_work_on_modular_elliptic_curves.pdf, pp. 235–256.

[GZ85] Benedict H. Gross and Don B. Zagier, *On singular moduli*, J. Reine Angew. Math. **355** (1985), 191–220, http://wstein.org/papers/bib/gross-zagier-on_singular_moduli.pdf. MR 772491 (86j:11041)

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320, http://wstein.org/papers/bib/Gross-Zagier_Heegner_points_and_derivatives_of_Lseries.pdf. MR 87j:11057

[How04] Benjamin Howard, *The Heegner point Kolyvagin system*, Compos. Math. **140** (2004), no. 6, 1439–1472, https://www2.bc.edu/~howardbe/Research/heegner.pdf. MR 2098397 (2006a:11070)

[JK10] Dimitar Jetchev and Ben Kane, *Equidistribution of Heegner Points and Ternary Quadratic Forms*, Preprint (2010), http://arxiv.org/abs/0908.3905.

[JLS09] Dimitar Jetchev, Kristin Lauter, and William Stein, *Explicit Heegner points: Kolyvagin's conjecture and non-trivial elements in the Shafarevich-Tate group*, J. Number Theory **129** (2009), no. 2, 284–302, http://wstein.org/papers/kolyconj/. MR 2473878 (2009m:11080)

[KL89] V. A. Kolyvagin and D. Y. Logachev, *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Algebra i Analiz **1** (1989), no. 5, 171–196, http://wstein.org/papers/bib/kolyvagin-logachev-finiteness_of_the_shafarevich-tate_group_and_the_group_of_rational_points_for_some_modular_abelian_varieties.pdf.

[Koh97] D. R. Kohel, *Computing modular curves via quaternions*, Fourth CANT Conference (1997), http://wstein.org/papers/bib/kohel-sydney.pdf.

[Koh01] David R. Kohel, *Hecke module structure of quaternions*, Class field theory—its centenary and prospect (Tokyo, 1998), Adv. Stud. Pure Math., vol. 30, Math. Soc. Japan, Tokyo, 2001, http://wstein.org/papers/bib/kohel-hecke.pdf, pp. 177–195. MR 1846458 (2002i:11059)

[Kol88a] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E, \mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671, http://wstein.org/papers/bib/kolyvagin-finitess_of_EQ_and_sha_for_a_subclass.pdf. MR 89m:11056

[Kol88b] ———, *The Mordell-Weil and Shafarevich-Tate groups for Weil elliptic curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 6, 1154–1180, 1327, http://wstein.org/papers/bib/kolyvagin-on_the_mw_and_sha_groups.pdf. MR 90f:11035

[Kol91] ———, *On the structure of Selmer groups*, Math. Ann. **291** (1991), no. 2, 253–259, http://wstein.org/papers/stein-ggz/references/kolyvagin-structure_of_selmer_groups/. MR 93e:11073

[Lan56] S. Lang, *Algebraic groups over finite fields*, Amer. J. Math. **78** (1956), 555–563, http://wstein.org/papers/bib/Lang-Algebraic_Groups_Over_Finite_Fields.pdf. MR 19,174a

[Lan87] ———, *Elliptic functions*, second ed., Graduate Texts in Mathematics, vol. 112, Springer-Verlag, New York, 1987, With an appendix by J. Tate. MR 890960 (88c:11028)

[Lan91] ———, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR 93a:11048

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978), http://archive.numdam.org/article/PMIHES_1977__47__33_0.pdf.

[McC91] W. G. McCallum, *Kolyvagin's work on Shafarevich-Tate groups*, L-functions and arithmetic (Durham, 1989), Cambridge Univ. Press, Cambridge, 1991, http://wstein.org/papers/bib/mccallum-kolyvagin.pdf, pp. 295–316. MR 92m:11062

[Mes86] J.-F. Mestre, *La méthode des graphes. Exemples et applications*, Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata) (1986), 217–242, See http://wstein.org/papers/rank4/mestre-fr.pdf and http://wstein.org/papers/rank4/mestre-en.pdf (English translation).

[Mil72] J. S. Milne, *On the arithmetic of abelian varieties*, Invent. Math. **17** (1972), 177–190, http://wstein.org/papers/bib/Milne-On_the_Arithmetic_of_Abelian_Varieties.pdf. MR 48 #8512

[Mil86] ———, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[Mil10] Robert L. Miller, *Proving the Birch and Swinnerton-Dyer conjecture for specific elliptic curves of analytic rank zero and one*, http://arxiv.org/abs/1010.2431, 2010.

[MR04] Barry Mazur and Karl Rubin, *Kolyvagin systems*, Mem. Amer. Math. Soc. **168** (2004), no. 799, viii+96. MR 2031496 (2005b:11179)

[Mum70] D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.

[Piz80] A. Pizer, *An algorithm for computing modular forms on* $\Gamma_0(N)$, J. Algebra **64** (1980), no. 2, 340–390, http://wstein.org/papers/bib/pizer-algorithm_for_computing_modular_forms_on_gamma0.pdf.

[Pra95] Dipendra Prasad, *Ribet's theorem: Shimura-Taniyama-Weil implies Fermat*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993–1994), CMS Conf. Proc., vol. 17, Amer. Math. Soc., Providence, RI, 1995, http://wstein.org/papers/bib/prasad-ribets_theorem-shimura_taniyama_weil_implies_fermat.pdf, pp. 155–177. MR 1357211 (96j:11072)

[Rib88] Kenneth A. Ribet, *On the component groups and the Shimura subgroup of* $J_0(N)$, Séminaire de Théorie des Nombres, 1987–1988 (Talence, 1987–1988), Univ. Bordeaux I, Talence, 1988, `http://math.berkeley.edu/~ribet/Articles/bx_87.pdf`, pp. Exp. No. 6, 10. MR 993107 (91b:11070)

[Rib90a] K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), no. 2, 431–476, `http://math.berkeley.edu/~ribet/Articles/invent_100.pdf`.

[Rib90b] ———, *Raising the levels of modular representations*, Séminaire de Théorie des Nombres, Paris 1987–88, Birkhäuser Boston, Boston, MA, 1990, `http://math.berkeley.edu/~ribet/Articles/dpp.pdf`, pp. 259–271.

[Rib94] ———, *Report on mod ℓ representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Motives (Seattle, WA, 1991), Amer. Math. Soc., Providence, RI, 1994, `http://math.berkeley.edu/~ribet/Articles/motives.pdf`, pp. 639–676.

[Rib99] ———, *Torsion points on $J_0(N)$ and Galois representations*, Arithmetic theory of elliptic curves (Cetraro, 1997), Springer, Berlin, 1999, `http://math.berkeley.edu/~ribet/Articles/cime.pdf`, pp. 145–166. MR 2001b:11054

[Rib10] Kenneth A. Ribet, *Email: supersingular points on elliptic curves modulo ℓ*, Personal Communication (2010), `http://wstein.org/papers/bib/2010-ribet-eis.txt`.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, `http://wstein.org/papers/serre/`, pp. 143–232. MR 2002h:11047

[S$^+$11] W. A. Stein et al., *Sage Mathematics Software (Version 4.6.2)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[Ser88] J-P. Serre, *Algebraic groups and class fields*, Springer-Verlag, New York, 1988, Translated from the French.

[Sil94] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Springer-Verlag, New York, 1994.

[ST68] J-P. Serre and J. T. Tate, *Good reduction of abelian varieties*, Ann. of Math. (2) **88** (1968), 492–517, `http://wstein.org/papers/bib/Serre-Tate-Good_Reduction_of_Abelian_Varieties.pdf`.

[Ste09] William Stein, *Computational Number Theory Course Notes*, Unpublished (2009), `http://wiki.wstein.org/09/583e`.

[Ste10] William Stein, *Toward a Generalization of the Gross-Zagier Conjecture*, Internat. Math. Res. Notices (2010), `http://wstein.org/papers/stein-ggz/`.

[Sut09] Andrew Sutherland, *Computing the image of Galois representations attached to an elliptic curve*, Talk Slides (2009), `http://math.mit.edu/~drew/ImageOfGalois.pdf`.

[Sut10] ———, *Email: surjective p-adic repn for some rank* 2 *curves*, Personal Communication (2010), http://wstein.org/papers/bib/2010-sutherland-16.txt.

[SW02] William Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, http://wstein.org/ecdb, pp. 267–275. MR 2041090 (2005h:11113)

[SW10] William Stein and Jared Weinstein, *Kolyvagin classes on elliptic curves: structure, distribution, and algorithms*, In preparation (2010).

[SW11] William Stein and Christian Wuthrich, *Computations About Tate-Shafarevich Groups Using Iwasawa Theory*, in preparation (2011), http://wstein.org/papers/shark/.

[Tat66] J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1965/66, http://wstein.org/papers/bib/tate-bsd.pdf, pp. Exp. No. 306, 415–440.

[Vat02] V. Vatsal, *Uniform distribution of Heegner points*, Invent. Math. **148** (2002), no. 1, 1–46, http://www.math.ubc.ca/~vatsal/research/uniform3.pdf. MR 1892842 (2003j:11070)

[Voi] John Voight, *Identifying the matrix ring: algorithms for quaternion algebras and quadratic forms*, Submitted, http://arxiv.org/abs/1004.0994 or http://www.cems.uvm.edu/~voight/articles/quatalgs-040110.pdf.

[Wat06] Mark Watkins, *Some remarks on Heegner point computations*, Preprint (2006), http://arxiv.org/abs/math/0506325.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551, http://users.tpg.com.au/nanahcub/flt.pdf.

[Wil00] ———, *The Birch and Swinnerton-Dyer Conjecture*, http://www.claymath.org/prize_problems/birchsd.htm.

[YZZ11] X. Yuan, S. Zhang, and W. Zhang, *Triple product L-series and Gross-Schoen cycles I: split case*, Preprint (2011), http://www.math.columbia.edu/~yxy/preprints/triple.pdf.

[Zha01] Shou-Wu Zhang, *Gross-Zagier formula for* GL_2, Asian J. Math. **5** (2001), no. 2, 183–290, http://intlpress.com/AJM/p/2001/5_2/AJM-5-2-183-290.pdf. MR 1868935 (2003k:11101)

[Zha04] ———, *Gross-Zagier formula for* GL(2)*. II*, Heegner points and Rankin *L*-series, Math. Sci. Res. Inst. Publ., vol. 49, Cambridge Univ. Press, Cambridge, 2004, http://www.math.columbia.edu/~szhang/papers/gzes.pdf, pp. 191–214. MR 2083213

42 Algorithms for the Arithmetic of Elliptic Curves using Iwasawa Theory, with C. Wuthrich

MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

ALGORITHMS FOR THE ARITHMETIC OF ELLIPTIC CURVES USING IWASAWA THEORY

WILLIAM STEIN AND CHRISTIAN WUTHRICH

Abstract. We explain how to use results from Iwasawa theory to obtain information about p-parts of Tate-Shafarevich groups of specific elliptic curves over $\mathbb{Q}$. Our method provides a practical way to compute $\#Ш(E/\mathbb{Q})(p)$ in many cases when traditional p-descent methods are completely impractical and also in situations where results of Kolyvagin do not apply, e.g., when the rank of the Mordell-Weil group is greater than 1. We apply our results along with a computer calculation to show that $Ш(E/\mathbb{Q})[p] = 0$ for the 1,534,422 pairs (E, p) consisting of a non-CM elliptic curve E over $\mathbb{Q}$ with conductor $\leq$ 30,000, rank ≥ 2, and good ordinary primes p with $5 \leq p < 1000$ and surjective mod-p representation.

1. Introduction

The papers [GJP$^+$09, Mil10] describe verification of the Birch and Swinnerton-Dyer conjecture for elliptic curves of conductor ≤ 5000 with rank ≤ 1 by a computational application of Euler system results of Kato and Kolyvagin combined with explicit descent. The main motivation for the present paper is to develop algorithms using Iwasawa theory, in order to enable verification of the conjecture in new directions, e.g., large-scale verification of assertions about $Ш(E/\mathbb{Q})$, when E has rank at least 2. The present paper naturally complements related projects by Perrin-Riou [PR03] and Coates [CLS09, Coa11]. Moreover, we fill small gaps in the literature (e.g., precision bounds in Section 3) and take the opportunity to correct errors in the literature (e.g., Lemma 4.2) that we found in the course of implementing algorithms.

In Sections 2–7 we recall the main objects and theorems involved in the classical and p-adic Birch and Swinnerton-Dyer conjectures (BSD conjectures), correct some minor errors in the literature, and state a tight error bound that is essential for rigorous computation with p-adic L-series. These sections gather together disparate results and provide unified notation and fill minor gaps. In Section 3, we define p-adic L-functions and explain how to compute them. Next we define the p-adic regulator, treating separately the cases of split multiplicative and supersingular reduction, and recall p-adic analogues of the BSD conjecture. In Section 6, we recall the basic definitions and results for the algebraic p-adic L-functions defined using Iwasawa theory. This leads to the statement of the main conjecture and Kato's theorem.

2010 *Mathematics Subject Classification.* Primary .

The first author was supported by NSF grants DMS-0555776 and DMS-0821725.

©2011 William Stein and Christian Wuthrich

In Section 8 we discuss using p-adic results to bound $Ш(E)(p)$ when E has analytic rank 0, and Section 9 covers the case when the analytic rank is 1. In Section 10 we describe a conditional algorithm for computing the rank of an elliptic curve that uses p-adic methods and hence differs in key ways from the standard n-descent approach. Similarly, Section 11 contains an algorithm that applies to curves of any rank, and either computes $Ш(E/\mathbb{Q})(p)$ or explicitly disproves some standard conjecture. In Section 12 we give examples that illustrate the algorithms described above in numerous cases, including verifying for a rank 2 curve E that $Ш(E/\mathbb{Q})(p) = 0$ for a large number of p, as predicted by the BSD conjecture. In particular, we prove the following theorem via a computation of p-adic regulators and p-adic L-functions, which provides evidence for the BSD conjecture for curves of rank at least 2:

Theorem 1.1. *Let X be the set of 1,534,422 pairs (E,p), where E is a non-CM elliptic curve over $\mathbb{Q}$ with rank at least 2 and conductor $\leq$30,000, and $p \geq 5$ is a good ordinary prime for E with $p < 1000$ such that the mod p representation is surjective. Then $Ш(E/\mathbb{Q})[p] = 0$ for each of the pairs in X.*

1.1. **Background.** Let E be an elliptic curve defined over $\mathbb{Q}$ and let

$$y^2 + a_1 x y + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6 \tag{1.1}$$

be the unique global minimal Weierstrass equation for E with $a_1, a_3 \in \{0,1\}$ and $a_2 \in \{-1,0,1\}$. Mordell proved that the set of rational points $E(\mathbb{Q})$ is an abelian group of finite rank $r = \operatorname{rank}(E(\mathbb{Q}))$. Birch and Swinnerton-Dyer conjectured that $r = \operatorname{ord}_{s=1} L(E,s)$, where $L(E,s)$ is the Hasse-Weil L-function of E (see Conjecture 2.1 below). We call $r_{\mathrm{an}} = \operatorname{ord}_{s=1} L(E,s)$ the analytic rank of E, which is defined since $L(E,s)$ can be analytically continuted to all $\mathbb{C}$ (see [BCDT01]).

There is no known algorithm (procedure that has been proved to terminate) that computes r in all cases. We can computationally obtain upper and lower bounds in any particular case. One way to give a lower bound on r is to search for linearly independent points of small height via the method of descent. We can also use constructions of complex and p-adic Heegner points in some cases to bound the rank from below. To compute an upper bound on the rank r, in the case of analytic ranks 0 and 1, we can use Kolyvagin's work on Euler systems of Heegner points; for general rank, the only known method is to do an n-descent for some integer $n > 1$. The 2-descents implemented by Cremona [Cre97], by Simon [Sim02] in Pari [PAR11] (and SAGE [S$^+$11b]), and the $2,3,4$, etc., descents in Magma [BCP97] (see also [CFO$^+$08, CFO$^+$09, CFO$^+$11]), are particularly powerful. But they may fail in practice to compute the exact rank due to the presence of 2 or 3-torsion elements in the Tate-Shafarevich group.

The Tate-Shafarevich group $Ш(E/\mathbb{Q})$ is a torsion abelian group associated to $E/\mathbb{Q}$. It is the kernel of the localization map loc in the exact sequence

$$0 \longrightarrow Ш(E/\mathbb{Q}) \longrightarrow \mathrm{H}^1(\mathbb{Q},E) \xrightarrow{\mathrm{loc}} \bigoplus_v \mathrm{H}^1(\mathbb{Q}_v,E),$$

where the sum runs over all places v in $\mathbb{Q}$. The arithmetic importance of this group lies in its geometric interpretation. There is a bijection from $Ш(E/\mathbb{Q})$ to the $\mathbb{Q}$-isomorphism classes of principal homogeneous spaces $C/\mathbb{Q}$ of E which have points everywhere locally. In particular, such a C is a curve of genus 1 defined over $\mathbb{Q}$ whose Jacobian is isomorphic to E. Nontrivial elements in $Ш(E/\mathbb{Q})$ correspond to curves C that defy the Hasse principle, i.e., have a point over every completion of $\mathbb{Q}$, but have no points over $\mathbb{Q}$.

Conjecture 1.2. *Shafarevich and Tate The group* $\text{Ш}(E/\mathbb{Q})$ *is finite.*

The rank r and the Tate-Shafarevich group $\text{Ш}(E/\mathbb{Q})$ are encoded in the Selmer groups of E. Fix a prime p, and let $E(p)$ denote the $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-module of all torsion points of E whose orders are powers of p. The Selmer group $\mathcal{S}\rceil\updownarrow_p(E/\mathbb{Q})$ is defined by the following exact sequence:

$$0 \longrightarrow \mathcal{S}\rceil\updownarrow_p(E/\mathbb{Q}) \longrightarrow \mathrm{H}^1(\mathbb{Q}, E(p)) \longrightarrow \bigoplus_v \mathrm{H}^1(\mathbb{Q}_v, E)\,.$$

Likewise, for any positive integer m, the m-Selmer group is defined by the exact sequence

$$0 \to \mathcal{S}\rceil\updownarrow^{(m)}(E/\mathbb{Q}) \to \mathrm{H}^1(\mathbb{Q}, E[m]) \longrightarrow \bigoplus_v \mathrm{H}^1(\mathbb{Q}_v, E)$$

where $E[m]$ is the subgroup of elements of order dividing m in E.

It follows from the Kummer sequence that there are short exact sequences

$$0 \longrightarrow E(\mathbb{Q})/mE(\mathbb{Q}) \longrightarrow \mathcal{S}\rceil\updownarrow^{(m)}(E/\mathbb{Q}) \longrightarrow \text{Ш}(E/\mathbb{Q})[m] \longrightarrow 0$$

and

$$0 \longrightarrow E(\mathbb{Q}) \otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow \mathcal{S}\rceil\updownarrow_p(E/\mathbb{Q}) \longrightarrow \text{Ш}(E/\mathbb{Q})(p) \longrightarrow 0\,.$$

If the Tate-Shafarevich group is finite, then the $\mathbb{Z}_p$-corank of $\mathcal{S}\rceil\updownarrow_p(E/\mathbb{Q})$ is equal to the rank r of $E(\mathbb{Q})$.

The finiteness of $\text{Ш}(E/\mathbb{Q})$ is only known for curves of analytic rank 0 and 1, in which case computation of Heegner points and Kolyvagin's work on Euler systems gives an explicit computable multiple of its order [GJP$^+$09]. The group $\text{Ш}(E/\mathbb{Q})$ is not known to be finite for even a single elliptic curve with $r_{\mathrm{an}} \geq 2$. In such cases, the best we can do using current techniques is hope to bound the p-part $\text{Ш}(E/\mathbb{Q})(p)$ of $\text{Ш}(E/\mathbb{Q})$, for specific primes p. Even this might not a priori be possible, since it is not known that $\text{Ш}(E/\mathbb{Q})(p)$ is finite. However, if it were the case that $\text{Ш}(E/\mathbb{Q})(p)$ is finite (as Conjecture 1.2 asserts), then this could be verified by computing Selmer groups $\mathcal{S}\rceil\updownarrow^{(p^n)}(E/\mathbb{Q})$ for sufficiently many n (see, e.g., [SS04]). Note that practical unconditional computation of $\mathcal{S}\rceil\updownarrow^{(p^n)}(E/\mathbb{Q})$ via the method of descent is prohibitively difficult for all but a few very small p^n.

We present in this paper two algorithms using p-adic L-functions $\mathcal{L}_p(E,T)$, which are p-adic analogs of the complex function $L(E,s)$ (see Section 3 for the definition). Both algorithms rely heavily on the work of Kato [Kat04], which is a major breakthrough in the direction of a proof of the p-adic version of the BSD conjecture (see Section 5). The possibility of using these results to compute information about the Tate-Shafarevich group is well known to specialists and was for instance mentioned in [Col04] which gives a nice overview of the p-adic BSD conjecture. For supersingular primes such methods were used by Perrin-Riou in [PR03] to calculate $\text{Ш}(E/\mathbb{Q})(p)$ in many interesting cases when p is a prime of supersingular reduction.

Our first algorithm, which we describe in Section 10, finds a provable upper bound for the rank r of $E(\mathbb{Q})$ by computing approximations to the p-adic L-series for various small primes p. Any upper bound on the vanishing of $\mathcal{L}_p(E,T)$ at $T=0$ is also an upper bound on the rank r.

The second algorithm, which we discuss in Section 11, gives a new method for computing bounds on the order of $\text{Ш}(E/\mathbb{Q})(p)$, for specific primes p. We will exclude $p = 2$, since traditional descent methods work well at $p = 2$, and Iwasawa theory is not as well developed for $p = 2$. We also exclude some primes p, e.g., those for which E has additive reduction, since much of the theory we rely on has not yet been developed in this case.

Our second algorithm uses again the p-adic L-functions $\mathcal{L}_p(E, T)$, but also requires that the full Mordell-Weil group $E(\mathbb{Q})$ is known. Its output, if it yields any information, is a proven upper bound on the order of $\text{Ш}(E/\mathbb{Q})(p)$; in particular, we expect it to often prove the finiteness of the p-primary part of the Tate-Shafarevich group. But it will not in general be able to give any information about the structure of $\text{Ш}(E/\mathbb{Q})(p)$ as an abelian group or any information on its elements. For such finer results on the Tate-Shafarevich group, one general method is to use p^n-descents as described above. In some cases, we can also use visibility [AS02] to relate $\text{Ш}(E/\mathbb{Q})(p)$ to Mordell-Weil groups of other elliptic curves or abelian varieties. Assuming Kolyvagin's conjecture, it may also be possible to compute the structure of $\text{Ш}(E/\mathbb{Q})(p)$, for E of any rank, by making Kolyvagin's Euler system explicit in some cases (see forthcoming work of the first author and Jared Weinstein that builds on [Kol91b], and the remarks at the end of [Kol91a]). The computability of our upper bound on $\#\text{Ш}(E/\mathbb{Q})(p)$ relies on several conjectures, such as the finiteness of $\text{Ш}(E/\mathbb{Q})(p)$ and Conjectures 4.1 and 4.4 on the nondegeneracy of the p-adic height on E.

Under the assumption of the main conjecture (see Section 7), the number output by our algorithm equals the order of $\text{Ш}(E/\mathbb{Q})(p)$. There are several cases when this conjecture is known to hold by Greenberg and Vatsal in [GV00], by Grigorov in [Gri05], and in a forthcoming paper by Skinner and Urban [SU10]. In particular, under appropriate hypotheses, [SU10] prove the main conjecture for elliptic curves with good ordinary reduction (see Theorem 7.5 below). Thus in some cases, the upper bound on $\text{Ш}(E/\mathbb{Q})(p)$ that we obtain is actually a lower bound too, if all the computations go through, e.g. the p-adic height is nondegenerate and we find enough points to verify that the rank is equal to the order of vanishing.

Note that our algorithms can in principle be extended to give bounds in some cases on the rank of $E(K)$ and $\#\text{Ш}(E/K)(p)$ for number fields K which are abelian extensions of $\mathbb{Q}$ (here we still assume E is defined over $\mathbb{Q}$).

Acknowledgments. It is a pleasure to thank John Coates, Henri Darmon, Jerôme Grand'maison, Ralph Greenberg and Dimitar Jetchev for helpful discussions and comments. We are also greatly indebted to Robert Pollack who shared his code for computing p-adic L-functions and helped with the error estimates in Section 3. We also thank Mark Watkins, who independently implemented in Magma some of the algorithms of this paper, and in so doing found bugs in our implementation and discovered mistakes in an early draft of this manuscript.

2. The Birch and Swinnerton-Dyer conjecture

Let E be an elliptic curve defined over $\mathbb{Q}$. If the BSD conjecture (Conjecture 2.1 below) were true, it would yield an algorithm to compute both the rank r and the order of $\text{Ш}(E/\mathbb{Q})$.

Let E be an elliptic curve over $\mathbb{Q}$, and let $L(E, s)$ be the Hasse-Weil L-function associated to the $\mathbb{Q}$-isogeny class of E. According to [BCDT01] (which completes work initiated in [Wil95]), the function $L(E, s)$ is holomorphic on the whole complex plane. Let ω_E be the invariant differential $dx/(2y + a_1x + a_3)$ of the minimal

Weierstrass equation (1.1) of E. We write $\Omega_E = \int_{E(\mathbb{R})} \omega_E \in \mathbb{R}_{>0}$ for the Néron period of E.

Conjecture 2.1. *Birch and Swinnerton-Dyer*

(1) *The order of vanishing of the Hasse-Weil function $L(E,s)$ at $s=1$ is equal to the rank $r = \operatorname{rank}(E(\mathbb{Q}))$.*

(2) *The leading coefficient $L^*(E,1)$ of the Taylor expansion of $L(E,s)$ at $s=1$ satisfies*

$$\frac{L^*(E,1)}{\Omega_E} = \frac{\prod_v c_v \cdot \#\text{Ш}(E/\mathbb{Q})}{(\#E(\mathbb{Q})_{\mathrm{tor}})^2} \cdot \operatorname{Reg}(E/\mathbb{Q}) \tag{2.1}$$

where the Tamagawa numbers are denoted by c_v and $\operatorname{Reg}(E/\mathbb{Q})$ is the regulator of E, i.e., the discriminant of the Néron-Tate canonical height pairing on $E(\mathbb{Q})$.

Below we write $\#\text{Ш}(E/\mathbb{Q})_{\mathrm{an}}$ for the order of $\text{Ш}(E/\mathbb{Q})$ that is predicted by Conjecture 2.1.

Cassels proved in [Cas65] that if Conjecture 2.1 is true for an elliptic curve E over $\mathbb{Q}$, then it is true for all curves that are $\mathbb{Q}$-isogenous to E.

Proposition 2.2 (Manin)**.** *If Conjecture 2.1 is true, then there is an algorithm to compute r and $\#\text{Ш}(E/\mathbb{Q})$.*

Proof. Manin proved this result in [Man71, §11], but we recall the essential ideas here. By searching for points in $E(\mathbb{Q})$ we obtain a lower bound on r, which gets closer to the true rank r the longer we run the search. At some point this lower bound will equal r, but without using further information we have no way to know if that has occurred. As explained, e.g., in [Cre97, Coh07, Dok04], we can for any k compute $L^{(k)}(E,1)$ to any precision. Such computations yield upper bounds on r_{an}. In particular, if we compute $L^{(k)}(E,1)$ and it is nonzero (to the precision of our computation), then $r_{\mathrm{an}} \leq k$. Eventually this method will also converge to give the correct value of r_{an}, though again without further information we do not know when this will occur. However, if we know Conjecture 2.1, we know that $r = r_{\mathrm{an}}$, hence at some point the lower bound on r computed using point searches will equal the upper bound on r_{an} computed using the L-series. At this point, by Conjecture 2.1 we know the true value of both r and r_{an}.

Once r is known, we can compute $E(\mathbb{Q})$ via a point search (as explained in [Cre97, §3.5] or [Ste07a, §1.2]), hence we can approximate $\operatorname{Reg}(E/\mathbb{Q})$ to any desired precision. All quantities in (2.1) except $\#\text{Ш}(E/\mathbb{Q})$ can then be approximated to any desired precision. Solving for $\#\text{Ш}(E/\mathbb{Q})$ in (2.1) and computing all other quantities to large enough precision to determine the integer $\#\text{Ш}(E/\mathbb{Q})_{\mathrm{an}}$ then determines $\#\text{Ш}(E/\mathbb{Q})$, as claimed. □

The above algorithm would only produce the order of $\text{Ш}(E/\mathbb{Q})$ but no information about its structure as an abelian group. We could compute the structure of $\text{Ш}(E/\mathbb{Q})$ by computing the group $\mathcal{S}\rceil\updownarrow^{(n)}(E/\mathbb{Q})$ where $n^2 = \#\text{Ш}(E/\mathbb{Q})$, which is possible since $\mathcal{S}\rceil\updownarrow^{(m)}(E/\mathbb{Q})$ is computable for all m. The algorithms in Section 10 and 11 mimic the ideas of the proof of Proposition 2.2, but they replace the complex L-function by a p-adic L-series and use that much is known unconditionally about p-adic analogues of the BSD conjecture.

3. The p-adic L-function

We will assume for the rest of this article that E does not admit complex multiplication, though curves with complex multiplication are an area of active research for these methods (see e.g., [Rub99, PR04, CLS09, CLS10]).

Formulating a p-adic analogue of the BSD conjecture requires a p-adic analogue of the analytic function $L(E,s)$, as introduced by Mazur and Swinnerton-Dyer [MSD74, MTT86]. In this section, we recall the definition of this p-adic L-function, and fill a gap in the literature by giving a *complete recipe* for how to compute it in all cases, including proven error bounds on each coefficient.

Let $\pi\colon X_0(N) \longrightarrow E$ be the modular parametrization and let c_π be the Manin constant, i.e., the positive integer satisfying $c_\pi \cdot \pi^*\omega_E = 2\pi i f(\tau)d\tau$ with f the newform associated to E. When E is an optimal quotient (so the dual map $E \to \mathrm{Jac}(X_0(N))$ is injective), Manin conjectured that $c_\pi = 1$, and much work has been done toward this conjecture (see [Edi91, ARS06]).

Given a rational number r, define

$$\lambda^+(r) = -\pi i \cdot \left(\int_r^{i\infty} f(\tau)\, d\tau + \int_{-r}^{i\infty} f(\tau)\, d\tau \right) \in \mathbb{R}.$$

There is a basis $\{\gamma_+, \gamma_-\}$ of $H_1(E,\mathbb{Z})$ such that $\int_{\gamma_+} \omega_E$ is equal to Ω_E if $E(\mathbb{R})$ is connected and to $\frac{1}{2}\,\Omega_E$ otherwise. By a theorem of Manin [Man72], we know that $\lambda^+(r)$ belongs to $\mathbb{Q}\cdot\Omega_E$. For all $r \in \mathbb{Q}$, the *modular symbol* $[r]^+ \in \mathbb{Q}$ is

$$[r]^+ = \frac{\lambda^+(r)}{\Omega_E}.$$

In particular, we have $[0]^+ = L(E,1)\cdot\Omega_E^{-1}$. The quantity $[r]^+$ can be computed algebraically using modular symbols and linear algebra (see [Cre97] and [Ste07b]).

Let p be a prime of semistable reduction. We write[1] a_p for the trace of Frobenius. Suppose first that E has good reduction at p, and let $\tilde{E}$ denote the reduction of a minimal model of E modulo p. Then $N_p = p + 1 - a_p$ is the number of points on $\tilde{E}(\mathbb{F}_p)$. Let $X^2 - a_p \cdot X + p$ be the characteristic polynomial of Frobenius and let $\alpha \in \bar{\mathbb{Q}}_p$ be a root of this polynomial such that $\mathrm{ord}_p(\alpha) < 1$. There are two choices of α if E has supersingular reduction at p and there is a single possibility for α when E has good ordinary reduction at p. Next suppose E has bad multiplicative reduction at p. Then a_p is 1 if the reduction is split multiplicative and a_p is -1 if it is nonsplit multiplicative reduction. In either multiplicative case, we define $\alpha = a_p$.

As in [MTT86, §I.10], define a measure on $\mathbb{Z}_p^\times$ with values in $\mathbb{Q}(\alpha)$ by

$$\mu_\alpha(a + p^k\mathbb{Z}_p) = \begin{cases} \frac{1}{\alpha^k}\cdot\left[\frac{a}{p^k}\right]^+ - \frac{1}{\alpha^{k+1}}\cdot\left[\frac{a}{p^{k-1}}\right]^+ & \text{if } E \text{ has good reduction,} \\ \frac{1}{\alpha^k}\cdot\left[\frac{a}{p^k}\right]^+ & \text{otherwise.} \end{cases}$$

for any $k \geq 1$ and $a \in \mathbb{Z}_p^\times$ (by $\left[\frac{a}{p^k}\right]^+$ we mean $\left[\frac{a'}{p^k}\right]^+$ where $a' \in \mathbb{Z}$ is equivalent to a modulo p^k, which is well defined because of the modular symbols relations). Given a continuous character χ on $\mathbb{Z}_p^\times$ with values in the completion $\mathbb{C}_p$ of the algebraic closure of $\mathbb{Q}_p$, we may integrate χ against μ_α.

[1] The context should make it clear if we mean traces a_p of Frobenius, coefficients a_i as in (1.1), or series coefficients as in Proposition 3.1.

We assume henceforth that p is odd.[2] As in [MTT86, §I.13], any invertible element x of $\mathbb{Z}_p^\times$ can be written as $\omega(x) \cdot \langle x \rangle$ where $\omega(x)$ is a $(p-1)$-st root of unity and $\langle x \rangle$ belongs to $1 + p\mathbb{Z}_p$. We call ω the Teichmüller character. We define the analytic p-adic L-function by

$$L_\alpha(E, s) = \int_{\mathbb{Z}_p^\times} \langle x \rangle^{s-1} \, d\mu_\alpha(x) \quad \text{ for all } s \in \mathbb{Z}_p.$$

where by $\langle x \rangle^{s-1}$ we mean $\exp_p((s-1) \cdot \log_p(\langle x \rangle))$, and $\exp_p$ and $\log_p$ are the p-adic exponential and logarithm. The function $L_\alpha(E, s)$ extends to a locally analytic function in s on the disc defined by $|s-1| < 1$ (see the first proposition of [MTT86, §I.13]).

Let ${}_\infty G$ be the Galois group of the cyclotomic extension $\mathbb{Q}(\mu_{p^\infty})$ obtained by adjoining to $\mathbb{Q}$ all p-power roots of unity. By κ we denote the cyclotomic character ${}_\infty G \longrightarrow \mathbb{Z}_p^\times$. Because the cyclotomic character is an isomorphism, choosing a topological generator γ in $\Gamma = {}_\infty G^{(p-1)}$ amounts to picking a generator $\kappa(\gamma)$ of $1 + p\mathbb{Z}_p^\times$. With this choice, we may convert the function $L_\alpha(E, s)$ into a p-adic power series in $T = \kappa(\gamma)^{s-1} - 1$. We write $\mathcal{L}_\alpha(E, T)$ for this series in $\mathbb{Q}_p(\alpha)[\![T]\!]$. We have

$$\mathcal{L}_\alpha(E, T) = \int_{\mathbb{Z}_p^\times} (1+T)^{\frac{\log_p(\langle x \rangle)}{\log_p(\kappa(\gamma))}} d\mu_\alpha(x)\,. \tag{3.1}$$

For each integer $n \geq 1$, define a polynomial

$$P_n(T) = \sum_{a=1}^{p-1} \left(\sum_{j=0}^{p^{n-1}-1} \mu_\alpha\left(\omega(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot (1+T)^j \right) \in \mathbb{Q}_p(\alpha)[T].$$

Note that $P_n(T)$ depends on the choice of α, but for simplicity we do not include α in the notation.

Proposition 3.1. *We have*

$$\lim_{n \to \infty} P_n(T) = \mathcal{L}_\alpha(E, T),$$

where the convergence is coefficient-by-coefficient, in the sense that if $P_n(T) = \sum_j a_{n,j} T^j$ *and* $\mathcal{L}_\alpha(E, T) = \sum_j a_j T^j$*, then* $\lim_{n \to \infty} a_{n,j} = a_j$*.*

We now give a proof of this convergence and in doing so obtain an explicit upper bound for $|a_j - a_{n,j}|$, which is critical to making the computation of $\mathcal{L}_\alpha(E, T)$ algorithmic, and which appears to not be explicitly stated in the literature.

For any choice ζ_r of p^r-th root of unity in $\mathbb{C}_p$, let χ_r be the $\mathbb{C}_p$-valued character of $\mathbb{Z}_p^\times$ of order p^r obtained by composing the map $\langle\ \rangle : \mathbb{Z}_p^\times \to 1 + p\mathbb{Z}_p$ defined above with the map $1 + p\mathbb{Z}_p \to \mathbb{C}_p^*$ that sends $1 + p$ to ζ_r. Note that the conductor of χ_r is p^{r+1}.

Lemma 3.2. *Let* ζ_r *be a* p^r*-th root of unity with* $1 \leq r \leq n-1$*, and let* χ_r *be the corresponding character of order* p^r*, as above. Then*

$$P_n(\zeta_r - 1) = \int_{\mathbb{Z}_p^\times} \chi_r \ d\mu_\alpha.$$

In particular, note that the right hand side does not depend on n*.*

[2] Everything in this section can be done for $p = 2$ with $1 + p$ replaced by an integer that is congruent to 5 modulo 8, and various other slight modifications.

Proof. Writing $\chi = \chi_r$, we have

$$\begin{aligned} P_n(\zeta_r - 1) &= \sum_{a=1}^{p-1} \sum_{j=0}^{p^{n-1}-1} \mu_\alpha \left(\omega(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot \zeta_r^j \\ &= \sum_{a=1}^{p-1} \sum_{j=0}^{p^{n-1}-1} \mu_\alpha \left(\omega(a)(1+p)^j + p^n\mathbb{Z}_p\right) \cdot \chi\left((1+p)^j\right) \\ &= \sum_{b\in(\mathbb{Z}/p^n\mathbb{Z})^\times} \mu_\alpha\left(b + p^n\mathbb{Z}_p\right) \cdot \chi(b) = \int_{\mathbb{Z}_p^\times} \chi \; d\mu_\alpha. \end{aligned}$$

In the second to the last equality, we use that

$$(\mathbb{Z}/p^n\mathbb{Z})^\times \cong (\mathbb{Z}/p\mathbb{Z})^\times \times (1 + p(\mathbb{Z}/p^n\mathbb{Z}))$$

to sum over lifts of $b \in (\mathbb{Z}/p^n\mathbb{Z})^\times$ of the form $\omega(a)(1+p)^j$, i.e., a Teichmüller lift times a power of $(1+p)^j$. In the last equality, we use that χ has conductor dividing p^n, so is constant on the residue classes modulo p^n, and use the Riemann sums definition of the given integral. □

For each positive integer n, let $w_n(T) = (1+T)^{p^n} - 1$.

Corollary 3.3. *We have in $\mathbb{Q}_p(\alpha)[T]$ that*

$$w_{n-1}(T) \textit{ divides } P_{n+1}(T) - P_n(T).$$

Proof. By Lemma 3.2, $P_{n+1}(T)$ and $P_n(T)$ agree on $\zeta_j - 1$ for $0 \le j \le n-1$ and any choice ζ_j of p^j-th root of unity, so their difference vanishes on every root of the polynomial $w_{n-1}(T) = (1+T)^{p^{n-1}} - 1$. The claimed divisibility follows, since $w_{n-1}(T)$ has distinct roots. □

Lemma 3.4. *Let $f(T) = \sum_j b_j T^j$ and $g(T) = \sum_j c_j T^j$ be in $\mathcal{O}[T]$ with $\mathcal{O}$ the ring of integers of a finite field extension of $\mathbb{Q}_p$. If $f(T)$ divides $g(T)$, then*

$$\mathrm{ord}_p(c_j) \ge \min_{0\le i\le j} \mathrm{ord}_p(b_i).$$

Proof. We have $f(T)k(T) = g(T)$ with $k(T) \in \mathcal{O}[T]$. The lemma follows by using the definition of polynomial multiplication and the nonarchimedean property of ord_p. □

As above, let $a_{n,j}$ be the j-th coefficient of the polynomial $P_n(T)$. Let

$$c_n = \max(0, -\min_j \mathrm{ord}_p(a_{n,j}))$$

so that $p^{c_n} P_n(T) \in (\mathbb{Z}_p[\alpha])[T]$. For any $j > 0$, let

$$e_{n,j} = \min_{1\le i\le j} \mathrm{ord}_p \binom{p^n}{i}.$$

Proposition 3.5. *For all $n \ge 0$, we have $a_{n+1,0} = a_{n,0}$, and for $j > 0$,*

$$\mathrm{ord}_p(a_{n+1,j} - a_{n,j}) \ge e_{n-1,j} - \max(c_n, c_{n+1}).$$

Proof. Corollary 3.3 implies that there is a polynomial $h(T) \in \mathbb{Q}_p(\alpha)[T]$ with $w_{n-1}(T) \cdot h(T) = P_{n+1}(T) - P_n(T)$. Let $c \leq \max(c_n, c_{n+1})$ be the integer such that $p^c \cdot (P_{n+1}(T) - P_n(T)) \in \mathbb{Z}_p[\alpha][T]$ is primitive. Multiply both sides of the above equation by p^c, to get

$$w_{n-1}(T) \cdot p^c h(T) = p^c P_{n+1}(T) - p^c P_n(T) \in \mathbb{Z}_p[\alpha][T].$$

The right hand side is primitive and integral, so it is reducible in $\mathbb{Z}_p[\alpha][T]$. Since $w_{n-1}(T)$ is integral, we must have $p^c h(T) \in \mathbb{Z}_p[\alpha][T]$. Applying Lemma 3.4 and renormalizing by p^c gives $c + \mathrm{ord}_p(a_{n+1,j} - a_{n,j}) \geq e_{n-1,j}$, so

$$\mathrm{ord}_p(a_{n+1,j} - a_{n,j}) \geq e_{n-1,j} - c \geq e_{n-1,j} - \max(c_n, c_{n+1}).$$

□

Lemma 3.6. *The c_k are uniformly bounded above.*

Proof. Tracing through the definitions and using that $\mathrm{ord}_p(1/\alpha) > 1$, we see that the lemma is equivalent to showing that the modular symbol $[x]^+$ appearing in the definition of μ_α has bounded denominator. By the Abel-Jacobi theorem, the quotient of the image of the modular symbol map $[x]$ modulo $\mathbb{Z}^2 \approx H_1(E, \mathbb{Z})$ is equal to the image of the cuspidal subgroup C of $J_0(N)$. In particular, a bound on the denominator of $[x]^+$ is the largest power of p that divides the exponent of the image of C in $E(\bar{\mathbb{Q}})$. The claim follows since C is finite, since it is generated by finitely many "Manin symbols" as explained in [Man72, Thm. 2.7] or [Cre97, Ch. 2], and C is torsion as noted on the footnote of [Man72, pg. 35]. □

For j fixed, $e_{n-1,j} - \max(c_{n+1}, c_n)$ goes to infinity as n grows since the c_k are uniformly bounded above, by Lemma 3.6. Thus, $\{a_{n,j}\}$ is a Cauchy sequence and Proposition 3.5 implies that

$$\mathrm{ord}_p(a_j - a_{n,j}) \geq e_{n-1,j} - \max(c_n, c_{n+1}).$$

3.1. **The p-adic multiplier.** In this section we specialize the definition of p-adic multiplier from [MTT86, §I.14] to the case of an elliptic curve. For a prime p of good reduction, we define the p-adic multiplier by

$$\epsilon_p = \left(1 - \tfrac{1}{\alpha}\right)^2 . \tag{3.2}$$

Note that $\mathrm{ord}_p(\epsilon_p)$ is equal to 2 $\mathrm{ord}_p(N_p)$ where $N_p = p + 1 - a_p$ is the number of points in $\tilde{E}(\mathbb{F}_p)$.

For a prime of bad multiplicative reduction, we put

$$\epsilon_p = 1 - \tfrac{1}{\alpha} = \begin{cases} 0 & \text{if } p \text{ is split multiplicative,} \\ 2 & \text{if } p \text{ is nonsplit.} \end{cases}$$

3.2. **Interpolation property.** The p-adic L-function constructed above satisfies an interpolation property with respect to the complex L-function (see [MTT86, §I.14]). For instance, we have that

$$\mathcal{L}_\alpha(E, 0) = L_\alpha(E, 1) = \int_{\mathbb{Z}_p^\times} d\mu_\alpha = \epsilon_p \cdot \frac{L(E, 1)}{\Omega_E} .$$

A similar formula holds when integrating nontrivial characters of $\mathbb{Z}_p^\times$ against $d\mu_\alpha$. If χ is the character on ${}_\infty G$ sending γ to a root of unity ζ of exact order p^n, then

$$\mathcal{L}_\alpha(E,\zeta-1)=\frac{1}{\alpha^{n+1}}\cdot\frac{p^{n+1}}{G(\chi^{-1})}\cdot\frac{L(E,\chi^{-1},1)}{\Omega_E}\,.$$

Here $G(\chi^{-1})$ is the Gauss sum and $L(E,\chi^{-1},1)$ is the Hasse-Weil L-function of E twisted by χ^{-1}.

3.3. **The good ordinary case.** Suppose that the reduction of the elliptic curve at the prime p is good and ordinary, so a_p is not divisible by p. As mentioned before, in this case there is a unique choice of root α of the characteristic polynomial x^2-a_px+p that satisfies $\mathrm{ord}_p(\alpha)<1$. Since α is an algebraic integer, this implies that $\mathrm{ord}_p(\alpha)=0$, so α is a unit in $\mathbb{Z}_p$. We get therefore a unique p-adic L-function that we will denote simply by $\mathcal{L}_p(E,T)=\mathcal{L}_\alpha(E,T)$.

Proposition 3.7. *Let E be an elliptic curve with good ordinary reduction at a prime $p>2$ such that $E[p]$ is irreducible. Then the series $\mathcal{L}_p(E,T)$ belongs to $\mathbb{Z}_p[\![T]\!]$.*

Proof. See [GV00, Prop. 3.7] with $\chi=1$. □

We next illustrate the above material with a few numerical examples, one for each type of reduction. Let $E_0/\mathbb{Q}$ be the curve

$$E_0\colon\quad y^2+xy=x^3-x^2-4x+4 \tag{3.3}$$

which is labeled 446d1 in Cremona's tables [Cre]. The Mordell-Weil group $E_0(\mathbb{Q})$ is isomorphic to $\mathbb{Z}^2$ generated by the points $(2,0)$ and $(1,-1)$. We consider the prime $p=5$ where E_0 has good and ordinary reduction. As the number of points $N_p=10$ is divisible by p, this is an anomalous prime in the terminology of [Maz72]. Using [S$^+$11b], we compute an approximation to the p-adic L-series as explained above with $n=5$ to find

$$\begin{aligned}\mathcal{L}_5(E_0,T)\ &=\mathbf{O}(5^4)\cdot T+(5+5^2+3\cdot5^3+\mathbf{O}(5^4))\cdot T^2\\&\quad+(2\cdot5+3\cdot5^2+3\cdot5^3+\mathbf{O}(5^4))\cdot T^3+(4\cdot5^2+4\cdot5^3+\mathbf{O}(5^4))\cdot T^4\\&\quad+(4\cdot5+4\cdot5^2+\mathbf{O}(5^3))\cdot T^5\\&\quad+(1+2\cdot5+5^2+4\cdot5^3+\mathbf{O}(5^4))\cdot T^6+\mathbf{O}(T^7)\,.\end{aligned}$$

We see that the order of vanishing is at least 1 as follows. The interpolation formula implies that $\mathcal{L}_5(E_0,0)=0$ since $[0]^+=0$. We will give an explanation for the vanishing of the coefficient of T^1 later in the comments right after Theorem 6.1. We remark that the coefficient of T^2 has valuation 1, but the coefficient of T^6 is a unit.

3.4. **Multiplicative case.** We separate the cases of split and nonsplit multiplicative reduction. In fact, if the reduction is nonsplit, then the description of the good ordinary case applies just the same. But if the reduction is split multiplicative (the "exceptional case" in [MTT86]), then the p-adic L-series must have a trivial zero, i.e., $\mathcal{L}_p(E,0)=0$ because $\epsilon_p=0$. By a result of Greenberg and Stevens [GS93] (see also [Kob06] for a proof using Kato's Euler system), we know that

$$\left.\frac{d\,\mathcal{L}_p(E,T)}{d\,T}\right|_{T=0}=\frac{1}{\log_p\kappa(\gamma)}\cdot\frac{\log_p(q_E)}{\mathrm{ord}_p(q_E)}\cdot\frac{L(E,1)}{\Omega_E}$$

where q_E denotes the Tate period of E over $\mathbb{Q}_p$. It is now known thanks to [BDGP96] that $\log_p(q_E)$ is nonzero. Hence we define the p-adic $\mathscr{L}$-invariant as

$$\mathscr{L}_p = \frac{\log_p(q_E)}{\operatorname{ord}_p(q_E)} \neq 0\,. \tag{3.4}$$

We refer to [Col10] for a detailed discussion of the different $\mathscr{L}$-invariants and their connections.

3.5. **The supersingular case.** In the supersingular case, that is when $a_p \equiv 0 \pmod{p}$, we have two roots α and β both of valuation $\frac{1}{2}$. An analysis of the functions $\mathcal{L}_\alpha$ and $\mathcal{L}_\beta$ is in [Pol03]. The series $\mathcal{L}_\alpha(E,T)$ might not have integral coefficients in $\mathbb{Q}_p(\alpha)$. Nevertheless we can still extract two integral series $\mathcal{L}_p^{\pm}(E,T)$. We will not need this description.

There is a way of rewriting the p-adic L-series which relates more easily to the p-adic height defined in the next section. We follow Perrin-Riou's description in [PR03].

As before, ω_E denotes the chosen invariant differential on E. Let $\eta_E = x \cdot \omega_E$. The pair $\{\omega_E, \eta_E\}$ forms a basis of the Dieudonné module

$$D_p(E) = \mathbb{Q}_p \otimes \mathrm{H}^1_{\mathrm{dR}}(E/\mathbb{Q}).$$

This $\mathbb{Q}_p$-vector space comes equipped with a canonical Frobenius endomorphism φ that acts on it linearly. We normalize it in the following way, which makes it equal to $\frac{1}{p}\cdot F$ with F being the Frobenius as used in [MST06] and [Ked01, Ked03, Ked04]. Let t be any uniformizer at the point O_E at infinity on E, e.g., take $t = -\frac{x}{y}$. Let ν be a class in $D_p(E)$ represented by the differential $\sum c_n \cdot t^{n-1}\, dt$ with $c_n \in \mathbb{Q}_p$. Then $\varphi(\nu)$ can be represented by the differential $\sum c_n \cdot t^{pn-1}\, dt$. In particular $\varphi(dt) = t^{p-1}\, dt$. The characteristic polynomial of φ is equal to $X^2 - p^{-1}\, a_p\, X + p^{-1}$.

Write $\mathcal{L}_\alpha(E,T)$ as $G(T) + \alpha \cdot H(T)$ with $G(T)$ and $H(T)$ in $\mathbb{Q}_p[\![T]\!]$. Then we define

$$\mathcal{L}_p(E,T) = G(T)\cdot\omega_E \;+\; a_p \cdot H(T)\cdot\omega_E \;-\; p\cdot H(T)\cdot\varphi(\omega_E)\,,$$

which we view as a formal power series with coefficients in $D_p(E)\otimes\mathbb{Q}_p[\![T]\!]$, which contains exactly the same information as $\mathcal{L}_\alpha(E,T)$. See [PR03, §1] for a direct definition. Since the invariant differential ω_E depends on the choice of the Weierstrass equation (1.1), the expression $\mathcal{L}_p(E,T)$ is also dependent on this choice. However, if we write the series in the basis $\{\omega_E, \varphi(\omega_E)\}$ rather than in $\{\omega_E, \eta_E\}$, then the coordinates as above are independent. The D_p-valued L-series satisfies again certain interpolation properties,[3] e.g.,

$$(1-\varphi)^{-2}\,\mathcal{L}_p(E,0) = \frac{L(E,1)}{\Omega_E}\cdot\omega_E \,\in\, D_p(E)\,.$$

See Section 12.2 for an example.

3.6. **Additive case.** The case of additive reduction is much harder to treat, though we are optimistic that such a treatment is possible. We have not tried to include the possibility of additive reduction in our algorithm, especially because the existence of the p-adic L-function is not yet guaranteed in general. Note that there are two interesting papers [Del98] and [Del02] of Delbourgo on this subject.

[3] Perrin-Riou writes in [PR03] the multiplier as $(1-\varphi)^{-1}\cdot(1-p^{-1}\varphi^{-1})$ and she multiplies the right hand side with $L(E/\mathbb{Q}_p,1)^{-1} = N_p\cdot p^{-1}$. It is easy to see that $(1-\varphi)\cdot(1-p^{-1}\varphi^{-1}) = 1-\varphi+(\varphi - a_p\cdot p^{-1}) + p^{-1} = N_p\cdot p^{-1}$.

3.7. **Quadratic twists.** When the curve E is not semistable, we can try to use the modular symbols of a quadratic twist $E^\dagger$ of E in the computation of the p-adic L-function for E. This leads to dramatic speedups when the quadratic twist has lower conductor than E.

Suppose that there exists a fundamental discriminant D of a quadratic field satisfying the following conditions:

- p does not divide D,
- D^2 divides N,
- $M = N/D^2$ is coprime to D, and
- the conductor $N^\dagger$ of the quadratic twist $E^\dagger$ of E by D is of the form $M \cdot Q$ with Q dividing D.

Then $\psi = (\frac{D}{\cdot})$ is the Dirichlet character associated to the quadratic field $\mathbb{Q}(\sqrt{D})$ over which E and $E^\dagger$ become isomorphic. Let $f_{\scriptscriptstyle E}^\dagger$ be the newform of level $N^\dagger$ associated to the isogeny class of $E^\dagger$. As explained in [MTT86, §II.11], the twist of $f_{\scriptscriptstyle E}^\dagger$ by ψ is equal to $f_{\scriptscriptstyle E}$ and we can use their formula (I.8.3)

$$f_{\scriptscriptstyle E}(\tau) = \frac{1}{G(\psi)} \sum_{u \bmod |D|} \psi(u) \cdot f_{\scriptscriptstyle E}^\dagger\Bigl(\tau + \frac{u}{|D|}\Bigr). \tag{3.5}$$

Here $G(\psi)$ is as before the Gauss sum of ψ, whose value we know to be the square root $\sqrt{D}$ of D in $\mathbb{R}_{>0}$ or in $i\cdot\mathbb{R}_{>0}$. Let $c_{\mathbb{R}}$ be the number of connected components of $E(\mathbb{R})$, which is also the number of connected components of $E^\dagger(\mathbb{R})$. We write $\Omega_{E^\dagger}^-$ for $c_{\mathbb{R}} \cdot \int_{\gamma^-} \omega_{E^\dagger}$, similar to $\Omega_{E^\dagger} = \Omega_{E^\dagger}^+ = c_{\mathbb{R}} \cdot \int_{\gamma^+} \omega_{E^\dagger}$ with the notations from (3.1). We also put

$$\lambda^-(r) = \pi i \cdot \left(\int_r^{i\infty} - \int_{-r}^{i\infty} \right) f(\tau)\, d\tau$$

and $[r]^- = \lambda^-(r)/\Omega_{\scriptscriptstyle E}^-$. As for the modular symbol $[r]^+$, we have $[r]^- \in \mathbb{Q}$. Following [MTT86], we define the quantity η such that

$$\sqrt{D} \cdot \Omega_E^+ = \eta \cdot \Omega_{E^\dagger}^{\operatorname{sign}(D)}\,.$$

It is known that η is either 1 or 2.

Now we can compute the modular symbol $[r]^+$ for the curve E in terms of modular symbols for $E^\dagger$. Suppose first that $D > 0$.

$$\begin{aligned}
\lambda_E^+(r) &= \pi i \cdot \left(\int_r^{i\infty} + \int_{-r}^{i\infty} \right) \frac{1}{\sqrt{D}} \sum_{u=1}^{D-1} \psi(u) f_{\scriptscriptstyle E}^\dagger\Bigl(\tau + \frac{u}{D}\Bigr) d\tau \\
&= \frac{\pi i}{\sqrt{D}} \sum_{u=1}^{D-1} \psi(u) \int_{r+u/D}^{i\infty} f_{\scriptscriptstyle E}^\dagger(\tau) d\tau \\
&\qquad + \frac{\pi i}{\sqrt{D}} \sum_{v=1}^{D-1} \psi(D-v) \int_{-r}^{i\infty} f_{\scriptscriptstyle E}^\dagger\Bigl(\tau + 1 - \frac{v}{D}\Bigr) d\tau \\
&= \frac{\pi i}{\sqrt{D}} \sum_{u=1}^{D-1} \psi(u) \left(\int_{r+u/D}^{i\infty} + \int_{-r-u/D}^{i\infty} \right) f_{\scriptscriptstyle E}^\dagger(\tau) d\tau \\
&= \frac{1}{\sqrt{D}} \sum_{u=1}^{D-1} \psi(u)\, \lambda_{E^\dagger}^+\Bigl(r + \frac{u}{D}\Bigr).
\end{aligned}$$

We used that $\psi(u) = \operatorname{sign}(D)\,\psi(D-u)$, that $f_E^\dagger(\tau+1) = f_E^\dagger(\tau)$ and Equation (3.5). Similarly for $D<0$, we find

$$\lambda_E^+(r) = \frac{-1}{\sqrt{D}} \sum_{u=1}^{|D|-1} \psi(u)\,\lambda_{E^\dagger}^-\Bigl(r + \frac{u}{D}\Bigr).$$

Therefore, we have for any fundamental discriminant D

$$[r]_E^+ = \frac{\operatorname{sign}(D)}{\eta} \sum_{u=1}^{|D|-1} \Bigl(\frac{D}{u}\Bigr) \cdot \Bigl[r + \frac{u}{D}\Bigr]_{E^\dagger}^{\operatorname{sign}(D)}.$$

We can also express the unit eigenvalue α of Frobenius in terms of the corresponding $\alpha^\dagger$ unit eigenvalue for $E^\dagger$ as

$$\alpha = \psi(p) \cdot \alpha^\dagger.$$

In summary, we can evaluate the approximations to the p-adic L-function of E using only modular symbols of the curve $E^\dagger$ with smaller conductor. The estimations for the error of these approximations remain exactly the same.

We recalled that the computation of the modular symbols $[r]^\pm$ can be done purely algebraically. Unfortunately, the algebraic computation determines them only up to one single fixed choice of sign. If $[0]^+$ is nonzero, we can simply compare the value of the modular symbol at 0 with $L(E,1)/\Omega_E$ and adjust the sign when needed. If $L(E,1)=0$, we can use the above formula to compute $[0]_{E^\dagger}^+$ for some quadratic twist $E^\dagger$ with nonvanishing L-value. So we can easily adjust the unknown sign. Also, if we only know the modular symbols up to a rational multiple, we can use these formulae to scale them.

We should also add here that we can not possibly do a similar thing with quartic or sextic twists when they exist. This is due to the fact that the extension over which the twists become isomorphic is no longer an abelian extension. So we would have to twist the modular symbols with a Galois representation of dimension at least 2. Nevertheless there is a way of using these twists for computing the p-adic L-function as explained in [CLS09], using the fact that these curves have complex multiplication.

4. p-adic heights

The second term that we will generalize in the BSD formula is the real-valued regulator. In p-adic analogues of the conjecture we replace it by a p-adic regulator, which we define using a p-adic analogue of the height pairing. We follow here the generalized version [BPR93] and [PR03].

Let ν be an element of the Dieudonné module $D_p(E)$ (see Section 3.5). We will define a p-adic height function $h_\nu\colon E(\mathbb{Q}) \longrightarrow \mathbb{Q}_p$ which depends linearly on the vector ν. Hence it is sufficient to define it on the basis $\omega = \omega_E$ and $\eta = \eta_E$.

If $\nu = \omega$, then we define

$$h_\omega(P) = \log_E(P)^2$$

where $\log_E$ is the linear extension of the p-adic elliptic logarithm

$$\log_{\hat{E}}\colon \hat{E}(p\mathbb{Z}_p) \longrightarrow p\mathbb{Z}_p$$

defined on the formal group $\hat{E}$, by integrating our fixed differential ω_E.

For $\nu = \eta$, we define the p-adic sigma function of Bernardi as in [Ber81] to be the solution σ of the differential equation

$$-x = \frac{d}{\omega_{E}}\left(\frac{1}{\sigma}\cdot\frac{d\sigma}{\omega_{E}}\right)$$

such that $\sigma(O_{E}) = 0$, $\frac{d\sigma}{\omega_E}(O_{E}) = 1$, and $\sigma(-P) = -\sigma(P)$. If we denote by $t = -\frac{x}{y}$ the uniformizer at O_{E}, we may develop the sigma function as a series in t:

$$\sigma(t) = t + \frac{a_1}{2}\,t^2 + \frac{a_1^2 + a_2}{3}\,t^3 + \frac{a_1^3 + 2a_1a_2 + 3a_3}{4}\,t^4 + \cdots \in \mathbb{Q}(\!(t)\!),$$

where the a_i are the coefficients of the Weierstrass equation (1.1). As a function on the formal group $\hat{E}(p\mathbb{Z}_p)$, it converges for all t with $\operatorname{ord}_p(t) > \frac{1}{p-1}$.

We say that a point P in $E(\mathbb{Q})$ has good reduction at a prime p if P reduces to the identity component of the special fiber of the Néron model of E at p. Given a point P in $E(\mathbb{Q})$ there exists a multiple $m\cdot P$ such that $\sigma(m\cdot P)$ converges and such that $m\cdot P$ has good reduction at all primes. Denote by $e(m\cdot P)\in\mathbb{Z}$ the square root of the denominator of the x-coordinate of $m\cdot P$. Define

$$h_\eta(P) = \frac{2}{m^2}\cdot\log_p\left(\frac{e(m\cdot P)}{\sigma(m\cdot P)}\right).$$

Bernardi [Ber81] proves that this function is quadratic and satisfies the parallelogram law.

Finally, if $\nu = a\,\omega + b\,\eta$ then put

$$h_\nu(P) = a\,h_\omega(P) + b\,h_\eta(P)\,.$$

Since this function is quadratic and satisfies the parallelogram law, it induces a bilinear symmetric pairing $\langle\cdot,\cdot\rangle_\nu$ with values in $\mathbb{Q}_p$ defined by

$$\langle P, Q\rangle_\nu = \frac{1}{2}\cdot\big(h_\nu(P+Q) - h_\nu(P) - h_\nu(Q)\big)\,.$$

Note that all these definitions are dependent on the choice of the Weierstrass equation. It is easy to verify that the pairing is zero if one of the points is a torsion point.

4.1. **The good ordinary case.** Since we have only a single p-adic L-function in the case that the reduction is good ordinary, we have also to pin down a canonical choice of a p-adic height function. This was first done by Schneider [Sch82] and Perrin-Riou [PR82]. We refer to [MT91] and [MST06] for more details.

Let $\nu_\alpha = a\,\omega + b\,\eta$ be an eigenvector of φ on $D_p(E)$ associated to the eigenvalue $\frac{1}{\alpha}$. The value $e_2 = \mathbf{E}_2(E,\omega_{E}) = -12\cdot\frac{a}{b}$ is the value of the Katz p-adic Eisenstein series of weight 2 at (E,ω_{E}). If a point P has good reduction at all primes and lies in the range of convergence of $\sigma(t)$, we define the canonical p-adic height of P to be

$$\begin{aligned}\hat{h}_p(P) &= \frac{1}{b}\cdot h_{\nu_\alpha}(P)\\ &= -\frac{a}{b}\cdot\log_{E}(P)^2 + 2\,\log\left(\frac{e(P)}{\sigma(P)}\right)\\ &= 2\,\log_p\left(\frac{e(P)}{\exp(\frac{e_2}{24}\log_{E}(P)^2)\cdot\sigma(P)}\right) = 2\,\log_p\left(\frac{e(P)}{\sigma_p(P)}\right). \qquad (4.1)\end{aligned}$$

The function σ_p, defined by the last line, is called the canonical sigma-function, see [MT91]; it is known to lie in $\mathbb{Z}_p[\![t]\!]$. The p-adic height defined here is up to a factor of 2 the same as in [MST06].[4] It is also important to note that the function $\hat{h}_p$ is independent of the Weierstrass equation.

We write $\langle\cdot,\cdot\rangle_p$ for the canonical p-adic height pairing on $E(\mathbb{Q})$ associated to $\hat{h}_p$, and $\mathrm{Reg}_p(E/\mathbb{Q})$ for the discriminant of the height pairing on $E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tor}}$.

Conjecture 4.1. *Schneider* [Sch82] *The canonical p-adic height is nondegenerate on $E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tor}}$. In other words, the canonical p-adic regulator* $\mathrm{Reg}_p(E/\mathbb{Q})$ *is nonzero.*

Apart from the special case treated in [Ber82] of curves with complex multiplication of rank 1, there are hardly any results on this conjecture. See also [Wut04].

We return to our running example curve E_0 from Section 3.3. The methods of [MST06, Har08] permit us to quickly compute to relatively high precision the p-adic regulator of E_0. We have

$$\mathbf{E}_2(E_0,\omega_{E}) = 3\cdot 5 + 4\cdot 5^2 + 5^3 + 5^4 + 5^5 + 2\cdot 5^6 + 4\cdot 5^7 + 3\cdot 5^9 + \mathbf{O}(5^{10}),$$

and the regulator associated to the canonical p-adic height is

$$\mathrm{Reg}_p(E_0/\mathbb{Q}) = 2\cdot 5 + 2\cdot 5^2 + 5^4 + 4\cdot 5^5 + 2\cdot 5^7 + 4\cdot 5^8 + 2\cdot 5^9 + \mathbf{O}(5^{10}). \tag{4.2}$$

4.2. **The multiplicative case.** When E has multiplicative reduction at p, if we want to have the same closed formula in the p-adic version of the BSD conjecture for multiplicative primes as for other ordinary primes, the p-adic height has to be changed slightly. We use the description of the p-adic regulator given in [MTT86, §II.6]. Alas, their formula is not correct, as explained in [Wer98], so we use the corrected version.

If the reduction is nonsplit multiplicative, we use the same formula (4.1) to define the p-adic height as for the good ordinary case.

We assume for the rest of this section that the reduction is split multiplicative. We use Tate's p-adic uniformization (see for instance in [Sil94, Ch. V]). We have an explicit description of the height pairing in [Sch82]. Let q_E be the Tate parameter of the elliptic curve E over $\mathbb{Q}_p$, so we have an analytic homomorphism $\psi\colon \bar{\mathbb{Q}}_p^\times \longrightarrow E(\bar{\mathbb{Q}}_p)$ whose kernel is precisely $q_E^{\mathbb{Z}}$. The image of $\mathbb{Z}_p^\times$ under ψ is equal to the subgroup of points of $E(\mathbb{Q}_p)$ lying on the connected component of the reduction modulo p of the Néron model of E. Let C be the constant such that $\psi^*(\omega_E) = C\cdot\frac{du}{u}$ where u is a uniformizer of $\mathbb{Q}_p^\times$ at 1. The value of the p-adic Eisenstein series of weight 2 is

$$e_2 = \mathbf{E}_2(E,\omega_E) = C^2\cdot\left(1 - 24\cdot\sum_{n\geq 1}\sum_{d\mid n} d\cdot q_E^n\right).$$

Then we use the formula as in the good ordinary case to define the canonical sigma function $\sigma_p(t(P)) = \exp(\frac{e_2}{24}\log_E(P)^2)\cdot\sigma(t(P))$. We could also have used directly the formula

$$\sigma_p(u) = \frac{u-1}{u^{1/2}}\cdot\prod_{n\geq 1}\frac{(1-q_E^n\cdot u)(1-q_E^n/u)}{(1-q_E^n)^2}$$

[4] This factor is needed if we do not want to modify the p-adic version of the BSD conjecture (Conjecture 5.1).

where $u \in 1+p\mathbb{Z}_p$ is the unique preimage of $P \in \widehat{E}(p\mathbb{Z}_p)$ under the Tate parametrization ψ, where $\widehat{E}$ is the formal group of E at p.

Let P be a point in $E(\mathbb{Q})$ having good reduction at all finite places and with trivial reduction at p. Then we define

$$\hat{h}_p(P) = 2\log_p\left(\frac{e(P)}{\sigma_p(t(P))}\right) - \frac{\log_p(u)^2}{\log_p(q_{E})}$$

with u as above. The p-adic regulator is formed as before but with this modified p-adic height $\hat{h}_p$.

4.3. **The supersingular case.** In the supersingular case, we do not find a canonical p-adic height with values in $\mathbb{Q}_p$. Instead, the height has values in the Dieudonné module $D_p(E)$, as explained in [BPR93] and [PR03].

First, if the rank of the curve is 0, we define the p-adic regulator of $E/\mathbb{Q}$ to be $\omega = \omega_{E} \in D_p(E)$. Thus assume for the rest of this section that the rank r of $E(\mathbb{Q})$ is positive. Let $\nu = a\,\omega + b\,\eta$ be any element of $D_p(E)$ not lying in $\mathbb{Q}_p\,\omega$, (so $b \neq 0$). It can be easily checked that the value of

$$H_p(P) = \frac{1}{b}\cdot\left(h_\nu(P)\cdot\omega - h_\omega(P)\cdot\nu\right) \in D_p(E)$$

is independent of the choice of ν. We will call this the D_p-valued height on $E(\mathbb{Q})$. But note that it depends on the choice of the Weierstrass equation of E: if we change coordinates by putting

$$x' = u^2\cdot x + r \qquad \text{and} \qquad y' = u^3\cdot y + s\cdot x + t, \tag{4.3}$$

then the D_p-valued height $H_p'(P)$ computed in the new coordinates x', y' will satisfy $H_p'(P) = \frac{1}{u}\cdot H_p(P)$ for all points $P \in E(\mathbb{Q})$.

On $D_p(E)$ there is a canonical alternating bilinear form $[\cdot,\cdot]$ characterized by the property that $[\omega_{E}, \eta_{E}] = 1$. Write $\mathrm{Reg}_\nu \in \mathbb{Q}_p$ for the regulator of h_ν on $E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tor}}$. Then we have the following lemma which is a corrected version[5] of [PR03, Lem. 2.6].

Lemma 4.2. *Suppose that the rank r of $E(\mathbb{Q})$ is positive. There exists a unique element $\mathrm{Reg}_p(E/\mathbb{Q})$ in $D_p(E)$ such that for all $\nu \in D_p(E)$ not in $\mathbb{Q}_p\omega$, we have*

$$[\mathrm{Reg}_p(E/\mathbb{Q}), \nu] = \frac{\mathrm{Reg}_\nu}{[\omega,\nu]^{r-1}}. \tag{4.4}$$

Furthermore, if the rank r is 1, then $\mathrm{Reg}_p(E/\mathbb{Q}) = H_p(P)$ for a generator P. If the Weierstrass equation is changed as in (4.3), the regulator $\mathrm{Reg}_p'(E/\mathbb{Q})$ computed in the new equation satisfies $\mathrm{Reg}_p'(E/\mathbb{Q}) = \frac{1}{u}\cdot\mathrm{Reg}_p(E/\mathbb{Q})$.

We call $\mathrm{Reg}_p(E/\mathbb{Q}) \in D_p(E)$ the *D_p-valued regulator* of $E/\mathbb{Q}$, or better, of the chosen Weierstrass equation.

Proof. Since h_ω is made out of the square of the linear function $\log_{E}$, the matrix of the associated pairing on a basis $\{P_i\}$ of $E(\mathbb{Q})$ modulo torsion has entries of

[5] The wrong normalization in [PR03] only influences the computations for curves of rank greater than 1. It seems that, by chance, the computations in [PR03] were done with a ν in $D_p(E)$ such that $[\omega,\nu] = 1$, so that the normalization did not enter into the end results.

the form $\log_{E}(P_i)\cdot\log_{E}(P_j)$ and hence has rank 1. Therefore the regulator of the pairing associated to $\nu = a\cdot\omega + b\cdot\eta$ is equal to

$$\operatorname{Reg}_{a\omega+b\eta} = a\cdot b^{r-1}\cdot X + b^{r}\cdot Y$$

for some constants X and Y. In fact, we must have $X = \operatorname{Reg}_{\omega+\eta} - \operatorname{Reg}_{\eta}$ and $Y = \operatorname{Reg}_{\eta}$. Therefore the expression on the right hand side of (4.4) is linear in ν. More explicitly, we may define

$$\operatorname{Reg}_p(E/\mathbb{Q}) = Y\cdot\omega - X\cdot\eta.$$

The formula for the case of rank 1 is then also immediate. The variance of the regulator with the change of the equation can be checked just as for H_p. □

We continue to assume that the rank r of $E/\mathbb{Q}$ is positive, as in Lemma 4.2. Define the *fine Mordell-Weil group* as in [Wut07] to be the kernel

$$\mathfrak{M}(E/\mathbb{Q}) = \ker\left(E(\mathbb{Q})\otimes\mathbb{Z}_p \longrightarrow E(\mathbb{Q}_p)^{p\text{-adic completion}}\right),$$

which is a free $\mathbb{Z}_p$-module of rank $r-1$. The bilinear form associated to the normalized p-adic height

$$\frac{h_\nu(P)}{[\omega,\nu]},$$

can be restricted to obtain a pairing

$$\langle\cdot,\cdot\rangle_0\colon \mathfrak{M}(E/\mathbb{Q})\times(E(\mathbb{Q})\otimes\mathbb{Z}_p)\longrightarrow\mathbb{Q}_p\,.$$

It is then independent of the choice of $\nu\notin\mathbb{Q}_p\omega$. We call the regulator of this bilinear form $\langle\cdot,\cdot\rangle_0$ on a basis of $\mathfrak{M}(E/\mathbb{Q})$ the *fine regulator* $\operatorname{Reg}_0(E/\mathbb{Q})\in\mathbb{Q}_p$, which is an element of $\mathbb{Q}_p$ defined up to multiplication by a unit in $\mathbb{Z}_p$.

Lemma 4.3. *Suppose there exists a point Q in $E(\mathbb{Q})\otimes\mathbb{Z}_p$ such that $\mathfrak{M}(E/\mathbb{Q}) + \mathbb{Z}_pQ = E(\mathbb{Q})\otimes\mathbb{Z}_p$. Then*

$$[\operatorname{Reg}_p(E/\mathbb{Q}_p),\omega] \equiv \log_{E}(Q)^2\cdot\operatorname{Reg}_0(E/\mathbb{Q}) \pmod{\mathbb{Z}_p^\times}.$$

Proof. From the proof of the Lemma 4.2, we only have to show that

$$X = \operatorname{Reg}_{\omega+\eta} - \operatorname{Reg}_{\eta} \equiv h_\omega(Q)\operatorname{Reg}_0(E/\mathbb{Q}).$$

By hypothesis, there is a basis of $\mathfrak{M}(E/\mathbb{Q})$ that we can complete to a basis of $E(\mathbb{Q})\otimes\mathbb{Z}_p$ by adding Q to it. If M is the matrix of the pairing for η in this basis, then the matrix for $\omega+\eta$ is obtained by changing the entry for $\langle Q,Q\rangle$ by adding $h_\omega(Q)$ to it. Since X is the difference of the two determinants, it is $h_\omega(Q)$ times the determinant of $\langle\cdot,\cdot\rangle_\eta$ on the basis of $\mathfrak{M}(E/\mathbb{Q})$, which equals $\operatorname{Reg}_0(E/\mathbb{Q})$ by definition. □

This lemma proves the last equality in [PR03, §2]. We should mention that the formula just above it, linking $\operatorname{Reg}_p(E/\mathbb{Q})$ to $H_p(Q)\cdot\operatorname{Reg}_0(E/\mathbb{Q})$, is not known to hold as it can not be assumed in general that we can find a point Q as in the lemma above which is orthogonal to $\mathfrak{M}(E/\mathbb{Q})$. In particular, the D_p-valued regulator $\operatorname{Reg}_p(E/\mathbb{Q})$ is nonzero provided the fine regulator does not vanish, because $\log_{E}(Q)\neq 0$.

Conjecture 4.4. *Perrin-Riou* [PR93, Conjecture 3.3.7.i] *The fine regulator of $E/\mathbb{Q}$ is nonzero for all primes p. In particular, $\operatorname{Reg}_p(E/\mathbb{Q})\neq 0$ for all primes where E has supersingular reduction.*

Conjecture 3.3.7.ii' in [PR93], which asserts that Reg_ν is nonzero for at least one ν, is implied by the above conjecture. This is explained in remark iii) following the conjecture there, if we use the fact that the weak Leopoldt conjecture is now known for E and p.

We have presented here how to compute the p-adic regulator in the basis $\{\omega, \eta\}$, but in order to compare it later to the leading term of the p-adic L-function, it is better to write it in terms of the basis $\{\omega, \varphi(\omega)\}$. In particular, we would then have a vector whose coordinates are independent of the chosen Weierstrass equation.

In [BPR93, pg. 232], there is an algorithm for computing the action of φ by successive approximation using the expansion of ω in terms of a uniformizer t. It is dramatically more efficient to replace this by the computation of φ using Monsky-Washnitzer cohomology as explained in [Ked01, Ked03, Ked04, Har08].

4.4. **Normalization.** In view of Iwasawa theory, it is natural to normalize the heights and the regulators depending on the choice of the generator γ. In this way the heights depend on the choice of an isomorphism $\Gamma \longrightarrow \mathbb{Z}_p$ rather than on the $\mathbb{Z}_p$-extension only. This normalization can be achieved by simply dividing $\hat{h}_p(P)$ and $h_\nu(P)$ by $\kappa(\gamma)$. The regulators will be divided by $\log_p \kappa(\gamma)^r$ where r is the rank of $E(\mathbb{Q})$. Hence we write

$$\mathrm{Reg}_\gamma(E/\mathbb{Q}) = \frac{\mathrm{Reg}_p(E/\mathbb{Q})}{\log(\kappa(\gamma))^r}.$$

5. The p-adic Birch and Swinnerton-Dyer conjecture

5.1. **The ordinary case.** The following conjecture is due to Mazur, Tate and Teitelbaum [MTT86]. Rather than formulating it for the function $L_\alpha(E, s)$, we state it directly for the series $\mathcal{L}_p(E, T)$. It is then a statement about the expansion of this function at $T = 0$ rather than at $s = 1$.

Conjecture 5.1. *Mazur, Tate and Teitelbaum* [MTT86] *Let E be an elliptic curve with good ordinary reduction or with multiplicative reduction at a prime p.*

- *The order of vanishing of the p-adic L-function $\mathcal{L}_p(E, T)$ at $T = 0$ is equal to the rank $r = \mathrm{rank}(E(\mathbb{Q}))$, unless E has split multiplicative reduction at p in which case the order of vanishing is equal to $r + 1$.*
- *The leading term $\mathcal{L}_p^*(E, 0)$ satisfies*

$$\mathcal{L}_p^*(E,0) = \epsilon_p \cdot \frac{\prod_v c_v \cdot \#\text{Ш}(E/\mathbb{Q})}{(\#E(\mathbb{Q})_{\mathrm{tor}})^2} \cdot \mathrm{Reg}_\gamma(E/\mathbb{Q}) \tag{5.1}$$

unless the reduction is split multiplicative in which case the leading term is

$$\mathcal{L}_p^*(E,0) = \frac{\mathscr{L}_p}{\log(\kappa(\gamma))} \cdot \frac{\prod_v c_v \cdot \#\text{Ш}(E/\mathbb{Q})}{(\#E(\mathbb{Q})_{\mathrm{tor}})^2} \cdot \mathrm{Reg}_\gamma(E/\mathbb{Q}), \tag{5.2}$$

where $\mathscr{L}_p$ is as in Equation (3.4).

The conjecture asserts exact equality, not just equality up to a p-adic unit. However, the current approaches to the conjecture, which involve the main conjecture of Iwasawa theory, prove results up to a p-adic unit, since the characteristic power series is only defined up to a unit, as we will see in Section 7.

Again, we consider the curve E_0 (see Equation (3.3)) for an example in the good ordinary case. For this curve, we have $\prod c_v = 2$ and $E_0(\mathbb{Q})_{\mathrm{tor}} = 0$, so all the terms in the expression above can be computed except for the unknown size of $Ш(E_0/\mathbb{Q})$. The p-adic BSD conjecture predicts that

$$\#Ш(E_0/\mathbb{Q}) = 1 + \mathbf{O}(5^3)$$

which is consistent with the complex BSD conjecture, which predicts that $Ш(E_0/\mathbb{Q})$ is trivial.

5.2. **The supersingular case.** The conjecture in the case of supersingular reduction is given in [BPR93] and [PR03]. The conjecture relates an algebraic and an analytic value in the $\mathbb{Q}_p$-vector space $D_p(E)$ of dimension 2. (The fact that we have two coordinates was used by Kurihara and Pollack in [KP07] to construct global points via a p-adic analytic computation.)

Conjecture 5.2. *Bernardi and Perrin-Riou* [BPR93] *Let E be an elliptic curve with supersingular reduction at a prime p.*

- *The order of vanishing of the D_p-valued L-series $\mathcal{L}_p(E,T)$ at $T=0$ is equal to the rank r of $E(\mathbb{Q})$.*
- *The leading term $\mathcal{L}_p^*(E,0)$ satisfies*

$$(1-\varphi)^{-2}\cdot\mathcal{L}_p^*(E,0) = \frac{\prod_v c_v\cdot\#Ш(E/\mathbb{Q})}{(\#E(\mathbb{Q})_{\mathrm{tor}})^2}\cdot\mathrm{Reg}_\gamma(E/\mathbb{Q}) \quad \in D_p(E)\,. \tag{5.3}$$

We emphasize that both sides of (5.3) are dependent on the Weierstrass equation. But under a change of the form $x' = u^2\cdot x + r$, they both get multiplied by $\frac{1}{u}$ and hence the conjecture is independent of this choice.

6. Iwasawa theory of elliptic curves

We suppose from now on that $p>2$. Let ${}_\infty\mathbb{Q}$ be the cyclotomic $\mathbb{Z}_p$-extension of $\mathbb{Q}$, which is a Galois extension of $\mathbb{Q}$ whose Galois group is Γ. Let Λ be the completed group algebra $\mathbb{Z}_p[\![\Gamma]\!]$. We use a fixed topological generator γ of Γ to identify Λ with $\mathbb{Z}_p[\![T]\!]$ by sending γ to $1+T$. Any finitely generated Λ-module admits a decomposition up to quasi-isomorphism as a direct sum of elementary Λ-modules. Denote by ${}_n\mathbb{Q}$ the n-th layer of the $\mathbb{Z}_p$-extension, so ${}_n\mathbb{Q}$ is a subfield of ${}_\infty\mathbb{Q}$ and $\mathrm{Gal}({}_n\mathbb{Q}/\mathbb{Q})\approx\mathbb{Z}/p^n\mathbb{Z}$. As in Section 1.1, we define the p-Selmer group of E over ${}_n\mathbb{Q}$ by the exact sequence

$$0\longrightarrow\mathcal{S}\rceil\updownarrow_p(E/{}_n\mathbb{Q})\longrightarrow\mathrm{H}^1({}_n\mathbb{Q},E(p))\longrightarrow\bigoplus_v\mathrm{H}^1({}_n\mathbb{Q}_v,E)$$

with the product running over all places v of ${}_n\mathbb{Q}$. Over the full $\mathbb{Z}_p$-extension, we define $\mathcal{S}\rceil\updownarrow_p(E/{}_\infty\mathbb{Q})$ to be the direct limit $\varinjlim\mathcal{S}\rceil\updownarrow_p(E/{}_n\mathbb{Q})$ with respect to the maps induced by the restriction maps $\mathrm{H}^1({}_n\mathbb{Q},E(p))\longrightarrow\mathrm{H}^1({}_{n+1}\mathbb{Q},E(p))$. The group $\mathcal{S}\rceil\updownarrow_p(E/{}_\infty\mathbb{Q})$ encodes information about the growth of the rank of $E({}_n\mathbb{Q})$ and of the size of $Ш(E/{}_n\mathbb{Q})(p)$ as n tends to infinity. We will consider the Pontryagin dual

$$X(E/{}_\infty\mathbb{Q}) = \mathrm{Hom}\Big(\mathcal{S}\rceil\updownarrow_p(E/{}_\infty\mathbb{Q}),\mathbb{Q}_p/\mathbb{Z}_p\Big)\,,$$

which is a finitely generated Λ-module (see [CS00]). For further introduction to these objects, see [Gre01].

6.1. **The ordinary case.** Assume that the reduction at p is either good ordinary or of multiplicative type. Kato's theorem (see [Kat04, Thm. 17.4]), which uses the work of Rohrlich [Roh84], states that $X(E/_{\infty}\mathbb{Q})$ is a torsion Λ-module, so we may associate to it a characteristic series

$$f_{E}(T) \in \mathbb{Z}_p[\![T]\!] \tag{6.1}$$

that is well-defined up to multiplication by a unit in $\mathbb{Z}_p[\![T]\!]^{\times}$.

The following result is due to Schneider [Sch85] and Perrin-Riou [PR82], and the multiplicative case is due to Jones [Jon89]. Note that it uses the analytic and algebraic p-adic height defined by Schneider in [Sch82]; taking into account the mentioned correction by Werner, these heights agree with the height in Section 4.2.

Theorem 6.1 (Schneider, Perrin-Riou, Jones)**.** *The order of vanishing of $f_{E}(T)$ at $T = 0$ is at least equal to the rank r. It is equal to r if and only if the p-adic height pairing is nondegenerate (Conjecture 4.1) and the p-primary part of the Tate-Shafarevich group $Ш(E/\mathbb{Q})(p)$ is finite (Conjecture 1.2). In this case the leading term of the series $f_{E}(T)$ has the same valuation as*

$$\epsilon_p \cdot \frac{\prod_v c_v \cdot \#Ш(E/\mathbb{Q})(p)}{(\#E(\mathbb{Q})(p))^2} \cdot \mathrm{Reg}_{\gamma}(E/\mathbb{Q}),$$

unless the reduction is split multiplicative in which case the same formula holds with ϵ_p replaced by $\mathscr{L}_p/\log(\kappa(\gamma))$.

Let us consider again our running example curve E_0. We have computed the 5-adic regulator and found that it is nonzero. The above theorem shows that the order of vanishing of $f_{E_0}(T)$ is at least equal to the rank. The finiteness of $Ш(E_0/\mathbb{Q})(5)$ is now equivalent to the statement that the order of vanishing of $f_{E_0}(T)$ is equal to the rank 2 of E_0. If this is the case, then the leading coefficient has valuation equal to

$$\mathrm{ord}_5(f^*_{E_0}(0)) = \mathrm{ord}_5(\#Ш(E_0/\mathbb{Q})(5)) + 1,$$

since $\mathrm{ord}_5(\mathrm{Reg}_5(E_0/\mathbb{Q})) = 1$ by Equation (4.2) and c_v, ϵ_5 and torsion are coprime to 5.

For general E, if the valuation of the leading term of $f_E(T)$ is positive we call p an *irregular*[6] prime for E. For irregular primes either the Mordell-Weil rank of E over $_{\infty}\mathbb{Q}$ is larger than the rank of $E(\mathbb{Q})$ or the Tate-Shafarevich group $Ш(E/_{\infty}\mathbb{Q})$ is no longer finite or both. We will determine exactly what happens for E_0 with $p = 5$ in Section 7.1 below.

6.2. **The supersingular case.** The supersingular case is more complicated, since the Λ-module $X(E/_{\infty}\mathbb{Q})$ is not torsion. A beautiful approach to the supersingular case has been found by Pollack [Pol03] and Kobayashi [Kob03]. As mentioned above (in Section 3.5), there are two p-adic series $\mathcal{L}_p^{\pm}(E,T)$ to which will correspond two new Selmer groups $X^{\pm}(E/_{\infty}\mathbb{Q})$, which are Λ-torsion. Despite the advantages of this ±-theory, we use the approach of Perrin-Riou here (see [PR03, §3]).

Let T_pE be the Tate module and define $\mathbb{H}^1_{\mathrm{loc}}$ to be the projective limit of the cohomology groups $\mathrm{H}^1({}_n\mathbb{Q}_{\mathfrak{p}}, T_pE)$ with respect to the corestriction maps. Here ${}_n\mathbb{Q}_{\mathfrak{p}}$ is the localization of ${}_n\mathbb{Q}$ at the unique prime $\mathfrak{p}$ above p. Perrin-Riou [PR94]

[6] For a good introduction to such terminology and the basics of Iwasawa theory of elliptic curves, we refer the reader to [Gre99].

constructed a Λ-linear Coleman map Col from $\mathbb{H}^1_{\text{loc}}$ to a submodule of $\mathbb{Q}_p[\![T]\!] \otimes D_p(E)$.

Define the fine Selmer group to be the kernel

$$\mathcal{R}(E/_n\mathbb{Q}) = \ker\left(\mathcal{S}\rceil\updownarrow(E/_n\mathbb{Q}) \longrightarrow E(_n\mathbb{Q}_\mathfrak{p}) \otimes \mathbb{Q}_p/\mathbb{Z}_p\right) .$$

It is again a consequence of the work of Kato, namely [Kat04, Thm. 12.4], that the Pontryagin dual $Y(E/_\infty\mathbb{Q})$ of $\mathcal{R}(E/_\infty\mathbb{Q})$ is a Λ-torsion module. Denote by $g_{E}(T)$ its characteristic series.

Let Σ be any finite set of places in $\mathbb{Q}$ containing the places of bad reduction for E and the places ∞ and p. Let $G_\Sigma(_n\mathbb{Q})$ denote the Galois group of the maximal extension of $_n\mathbb{Q}$ unramified at all places which do not lie above Σ. Next we define $\mathbb{H}^1_{\text{glob}}$ as the projective limit of $\mathrm{H}^1(G_\Sigma(_n\mathbb{Q}), T_pE)$. It is a Λ-module of rank 1 and it is independent of the choice of Σ.

By Kato again, the Λ-module $\mathbb{H}^1_{\text{glob}}$ is torsion-free and $\mathbb{H}^1_{\text{glob}} \otimes \mathbb{Q}_p$ has $\Lambda \otimes \mathbb{Q}_p$-rank 1. Choose now any element $_\infty c$ in $\mathbb{H}^1_{\text{glob}}$ such that $Z_c = \mathbb{H}^1_{\text{glob}}/(\Lambda \cdot {_\infty c})$ is Λ-torsion. Typically such a choice could be the "zeta element" of Kato, i.e., the image of his Euler system in $\mathbb{H}^1_{\text{glob}}$. Write $h_c(T)$ for the characteristic series of Z_c. Then we define an algebraic equivalent of the $D_p(E)$-valued L-series by

$$f_{E}(T) = \mathrm{Col}(_\infty c) \cdot g_{E}(T) \cdot h_c(T)^{-1} \in \mathbb{Q}_p[\![T]\!] \otimes D_p(E)$$

where by $\mathrm{Col}(_\infty c)$ we mean the image under the Coleman map Col of the localization of $_\infty c$ to $\mathbb{H}^1_{\text{loc}}$. The resulting series $f_{E}(T)$ is independent of the choice of $_\infty c$. Of course, $f_{E}(T)$ is again only defined up to multiplication by a unit in $\Lambda^\times$.

Again we have a result due to Perrin-Riou [PR93]:

Theorem 6.2 (Perrin-Riou). *The order of vanishing of $f_{E}(T)$ at $T = 0$ is at least equal to the rank r. It is equal to r if and only if the $D_p(E)$-valued regulator $\mathrm{Reg}_p(E/\mathbb{Q})$ is nonzero (Conjecture 4.4) and the p-primary part of the Tate-Shafarevich group $\text{Ш}(E/\mathbb{Q})(p)$ is finite (Conjecture 1.2). In this case the leading term of the series $(1-\varphi)^{-2}\, f_{E}(T)$ has the same valuation as*

$$\prod_{\upsilon} c_\upsilon \cdot \#\text{Ш}(E/\mathbb{Q})(p) \cdot \mathrm{Reg}_p(E/\mathbb{Q}) \, .$$

Note that the proof of this theorem in the appendix of [PR03] for the supersingular case uses the formula in Lemma 4.3 rather than the wrong definition of the regulator. Also we simplified the right hand term in comparison to (5.3), because the reduction at p is supersingular, so $N_p \equiv 1 \pmod{p}$, hence $\#E(\mathbb{Q})_{\text{tor}}$ must be a p-adic unit.

7. The Main Conjecture

The main conjecture links the two p-adic power series (3.1) and (6.1) of the previous sections. We formulate everything simultaneously for the ordinary and the supersingular case, even though they are of a quite different nature. We still assume that $p \neq 2$.

Conjecture 7.1. *Main conjecture of Iwasawa theory for elliptic curves If E has good or nonsplit multiplicative reduction at p, then there exists an element $u(T)$ in $\Lambda^\times$ such that $\mathcal{L}_p(E,T) = f_{E}(T) \cdot u(T)$. If the reduction of E at p is split multiplicative, then there exists such a $u(T)$ in $T \cdot \Lambda^\times$.*

Our statement above of the main conjecture for supersingular primes is equivalent to Kato's formulation in [Kat04, Conj. 12.10] and to Kobayashi's version in [Kob03]. In the notation of Section 6.2, it asserts that $g_{E}(T) = h_c(T)$, where c is Kato's zeta element.

Much is now known about this conjecture. To the elliptic curve E we attach the p-adic representation

$$\rho_p : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(T_p(E)) \approx \mathrm{GL}_2(\mathbb{Z}_p)$$

and its reduction

$$\bar{\rho}_p : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[p]) \approx \mathrm{GL}_2(\mathbb{F}_p).$$

Serre [Ser72] proved that $\bar{\rho}_p$ is almost always surjective (note our running hypothesis that E does not have complex multiplication) and that for curves with multiplicative reduction at p, surjectivity can only fail when there is an isogeny of degree p defined over $\mathbb{Q}$ (see [Ser96] and [RS01, Prop. 1.1] for the case $p = 2$ of this statement, though the theorem below has the hypothesis that p is odd).

Proposition 7.2. *If $p \geq 5$ then $\bar{\rho}_p$ is surjective if and only if ρ_p is surjective.*

Proof. See [GJP$^+$09, §2.1] for references for this and related results. □

Kato's Theorem 7.3. *Suppose that E has semistable reduction at p and that ρ_p is surjective. Then there exists a series $d(T)$ in Λ such that $\mathcal{L}_p(E,T) = f_{E}(T) \cdot d(T)$. If the reduction is split multiplicative then T divides $d(T)$.*

The main ingredient for this theorem is in [Kat04, Thm. 17.4], which addresses the good ordinary case when $\bar{\rho}_p$ is surjective. For the exceptional case we refer to [Kob06], which treats the case of split multiplicative reduction (i.e., where exceptional zeroes appear).

For the remaining cases, we obtain only a weaker statement:

Kato's Theorem 7.4. *Suppose that $\bar{\rho}_p$ is not surjective. Then there is an integer $m \geq 0$ such that $f_{E}(T)$ divides $p^m \cdot \mathcal{L}_p(E,T)$.*

Greenberg and Vatsal [GV00] have shown that in certain cases the main conjecture holds when $E[p]$ is reducible. Recently, Skinner-Urban have proved the main conjecture in many more cases. The following is a slightly weaker form of [SU10, Thm. 1]:

Theorem 7.5 (Skinner-Urban)**.** *Suppose that E has good ordinary reduction at p, that ρ_p is surjective and that there exists a prime q of multiplicative reduction such that $\bar{\rho}_p$ is ramified at q. Then the main conjecture holds, i.e., $\mathcal{L}_p(E,T)$ is equal to $f_{E}(T)$, up to a unit in Λ.*

The condition on the extra prime q is satisfied if E has split multiplicative reduction at q and p does not divide the Tamagawa number c_q. If E has non-split multiplicative reduction, one has to check that p does not divide the Tamagawa number over the unramified quadratic extension of $\mathbb{Q}_q$. Equivalently, in both cases of multiplicative reduction, the representation $\bar{\rho}_p$ is ramified at q if $p \nmid \mathrm{ord}_q(\Delta_E)$, as explained in [RS01, §2.4].

7.1. **The Example.** Consider again the curve E_0 (see Equation (3.3)) and the good ordinary prime $p = 5$. Kato's theorem implies that $f_{E_0}(T)$ divides $\mathcal{L}_p(E_0, T)$. Since we have found two linearly independent points of infinite order in $E_0(\mathbb{Q})$, we know that the rank of $E_0(\mathbb{Q})$ is at least 2. Hence the order of vanishing of $f_{E_0}(T)$ at $T = 0$ is at least 2 and, by Theorem 7.3, so is the order of vanishing of $\mathcal{L}_p(E_0, T)$. By explicitly computing an approximation to $\mathcal{L}_p(E_0, T)$ we see that the order of vanishing cannot be larger than 2. Therefore the rank of $E_0(\mathbb{Q})$ is equal to the order of vanishing of the p-adic L-series.

But we know more now. The fact that the order of vanishing of $f_{E_0}(T)$ is equal to 2 shows that the 5-primary part of $Ш(E_0/\mathbb{Q})$ cannot be infinite. We compute the p-adic valuation of the leading term of $f_{E_0}(T)$ by approximating $\mathrm{Reg}_p(E)$ and using Theorem 6.1. Comparing the leading term of $\mathcal{L}_p(E_0, T)$, which has valuation 1, and the leading term of $f_{E_0}(T)$, which has valuation $1 + \mathrm{ord}_5(\#Ш(E_0/\mathbb{Q})(5))$, shows that *the* 5*-primary part of* $Ш(E_0/\mathbb{Q})$ *is trivial*.

Moreover, the series $f_{E_0}(T)$ and $\mathcal{L}_p(E_0, T)$ have the same leading term, which implies that the main conjecture holds, i.e., $f_{E_0}(T) \in \mathcal{L}_p(E_0, T) \cdot \Lambda^\times$. By analyzing the series $\mathcal{L}_p(E_0, T)$, one can show that

$$f_{E_0}(T) = T \cdot ((T + 1)^5 - 1) \cdot u(T)$$

for a unit $u(T) \in \Lambda^\times$. Let ${}_1\mathbb{Q}$ be the first layer of the $\mathbb{Z}_5$-extension of $\mathbb{Q}$. Unless the Tate-Shafarevich group $Ш(E/{}_1\mathbb{Q})(5)$ is infinite, Iwasawa theory predicts that the rank of the Mordell-Weil group $E_0({}_1\mathbb{Q})$ is 6. Doing a quick search it is easy to find points of infinite order in $E_0({}_1\mathbb{Q})$ which are not defined over $\mathbb{Q}$. Therefore, we know that the rank of $E_0({}_1\mathbb{Q})$ and of $E_0({}_\infty\mathbb{Q})$ is 6 and that $Ш(E_0/{}_1\mathbb{Q})(5)$ and $Ш(E_0/{}_\infty\mathbb{Q})(5)$ are finite. For more examples of such factorizations of p-adic L-series we refer to [Pol].

8. If the L-series does not vanish

Suppose the Hasse-Weil L-function $L(E, s)$ does not vanish at $s = 1$. In this case, Kolyvagin proved that $E(\mathbb{Q})$ and $Ш(E/\mathbb{Q})$ are finite. In particular, Conjecture 1.2 is valid; also, Conjectures 4.1 and 4.4 are trivially true in this case.

Let $p > 2$ be a prime of semistable reduction such that the representation $\bar{\rho}_p$ is surjective. By the interpolation property, we know that $\mathcal{L}_p(E, 0)$ is nonzero, unless E has split multiplicative reduction.

8.1. **The good ordinary case.** In the ordinary case we have

$$\epsilon_p^{-1} \cdot \mathcal{L}_p(E, 0) = \frac{L(E, 1)}{\Omega_E} = [0]^+,$$

which is a nonzero rational number by [Man72]. In the following inequality, we use Theorem[7] 6.1 of Perrin-Riou and Schneider for the first equality and Kato's

[7] In the case of analytic rank 0, the theorem is actually relatively easy and well explained in [CS00, Ch. 3].

Theorem 7.3 on the main conjecture for the inequality in the second line.

$$\begin{aligned}\operatorname{ord}_p\left(\epsilon_p\cdot\frac{\prod_v c_v\cdot\#\text{Ш}(E/\mathbb{Q})(p)}{(\#E(\mathbb{Q})(p))^2}\right) &= \operatorname{ord}_p(f_E(0))\\ &\le \operatorname{ord}_p(\mathcal{L}_p(E,0))\\ &= \operatorname{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right)+\operatorname{ord}_p(\epsilon_p)\end{aligned}$$

Hence, we have the following upper bound on the p-primary part of the Tate-Shafarevich group

$$\begin{aligned}\operatorname{ord}_p\left(\#\text{Ш}(E/\mathbb{Q})(p)\right) &\le \operatorname{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right)-\operatorname{ord}_p\left(\frac{\prod c_v}{(\#E(\mathbb{Q})_{\text{tor}})^2}\right)\\ &= \operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})_{\text{an}}).\end{aligned}\tag{8.1}$$

Under the assumption of the main conjecture, this is sharp. In particular, if the conditions of Theorem 7.5 are satified for p, then we have the equality

$$\operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)) = \operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})_{\text{an}}).$$

This is Theorem 2.a in [SU10].

8.2. **The multiplicative case.** If the reduction is nonsplit, then the above holds just the same, because in all the theorems involved the nonsplit case never differs from the good ordinary case (only the split multiplicative case is exceptional). If instead the reduction is split multiplicative, we have that $\mathcal{L}_p(E,0)=0$ and

$$\mathcal{L}_p'(E,0)=\frac{\mathscr{L}_p}{\log\kappa(\gamma)}\cdot\frac{L(E,1)}{\Omega_E}=\frac{\mathscr{L}_p}{\log\kappa(\gamma)}\cdot[0]^+\neq 0\,.$$

Since the p-adic multiplier is the same on the algebraic as on the analytic side, we can once again compute as above to obtain the same bound (8.1).

8.3. **The supersingular case.** For the supersingular $D_p(E)$-valued series, we have

$$(1-\varphi)^{-2}\cdot\mathcal{L}_p(E,0)=\frac{L(E,1)}{\Omega_E}\cdot\omega_E=[0]^+\cdot\omega_E,$$

which is a nonzero element of $D_p(E)$. The $D_p(E)$-valued regulator $\operatorname{Reg}_p(E/\mathbb{Q})$ is equal to ω_E. We may therefore concentrate solely on the coordinate in ω_E. Write $\operatorname{ord}_p(f_E(0))$ for the p-adic valuation of the leading coefficient of the ω_E-coordinate of $f_E(T)$. Again we obtain an inequality by using Theorem 6.2:

$$\begin{aligned}\operatorname{ord}_p\left(\prod_v c_v\cdot\#\text{Ш}(E/\mathbb{Q})(p)\right) &= \operatorname{ord}_p((1-\varphi)^{-2}\,f_E(0))\\ &\le \operatorname{ord}_p((1-\varphi)^{-2}\,\mathcal{L}_p(E,0))\\ &= \operatorname{ord}_p\left(\frac{L(E,1)}{\Omega_E}\right).\end{aligned}$$

So we have once again that $\#\text{Ш}(E/\mathbb{Q})(p)$ is bounded from above by the highest power of p dividing $\#\text{Ш}(E/\mathbb{Q})_{\text{an}}$.

8.4. **Conclusion.** Summarizing the above computations, we have

Theorem 8.1 (Kato, Perrin-Riou, Schneider)**.** *Let E be an elliptic curve such that $L(E,1) \neq 0$. Then $\text{Ш}(E/\mathbb{Q})$ is finite and*

$$\#\text{Ш}(E/\mathbb{Q}) \;\Big|\; C \cdot \frac{L(E,1)}{\Omega_E} \cdot \frac{(\#E(\mathbb{Q})_{\text{tor}})^2}{\prod c_v}$$

where C is a product of a power of 2 and of powers of primes of additive reduction and of powers of primes for which the representation $\bar{\rho}_p$ is not surjective.

This improves [Rub00, Cor. 3.5.19].

9. If the L-series vanishes to the first order

We suppose for this section that E has good ordinary reduction at p and that the complex L-series $L(E,s)$ has a zero of order 1 at $s = 1$. Kolyvagin's theorem implies that $\text{Ш}(E/\mathbb{Q})$ is finite and that the rank of $E(\mathbb{Q})$ is equal to 1. Let P be a choice of generator of the Mordell-Weil group modulo torsion. Suppose that the p-adic height $\hat{h}_p(P)$ is nonzero. A theorem of Perrin-Riou in [PR87] asserts the following equality of rational numbers:

$$\frac{1}{\text{Reg}(E/\mathbb{Q})} \cdot \frac{L'(E,1)}{\Omega_E} = \frac{1}{\text{Reg}_p(E/\mathbb{Q})} \cdot \frac{\mathcal{L}_p'(E,0)}{(1-\frac{1}{\alpha})^2 \cdot \log(\kappa(\gamma))},$$

where, on the left hand side, the canonical real-valued regulator $\text{Reg}(E/\mathbb{Q}) = \hat{h}(P)$ appears along with the leading coefficient of $L(E,s)$, while, on the right hand side, we have the p-adic regulator $\text{Reg}_p(E/\mathbb{Q}) = \hat{h}_p(P)$ and the leading term of the p-adic L-series. By the BSD conjecture (or its p-adic analogue), this rational number should be equal to $\prod c_v \cdot \#\text{Ш}(E/\mathbb{Q}) \cdot (\#E(\mathbb{Q})_{\text{tor}})^{-2}$. By Kato's theorem, we know that the characteristic series $f_E(T)$ of the Selmer group divides $\mathcal{L}_p(E,T)$, at least up to a power of p. Hence the series $f_E(T)$ has a zero of order 1 at $T = 0$ and its leading term divides the above rational number in $\mathbb{Q}_p$ (here we use that $E(\mathbb{Q})$ has rank 1 so $T \mid f_E(T)$). Imposing the additional hypothesis that ρ_p is surjective, Theorem 7.3 implies the above divisibility over $\mathbb{Z}_p$ (rather than just up to a power of p), and we thus arrive at the following theorem.

Theorem 9.1 (Kato, Perrin-Riou)**.** *Let $E/\mathbb{Q}$ be an elliptic curve with good ordinary reduction at the odd prime p. Assume that the p-adic regulator of E is nonzero. Suppose that the representation ρ_p is surjective. If $L(E,s)$ has a simple zero at $s = 1$, then*

$$\begin{aligned}\text{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)) &\leq \text{ord}_p\left(\frac{(\#E(\mathbb{Q})_{\text{tor}})^2}{\prod c_v} \cdot \frac{1}{\text{Reg}(E/\mathbb{Q})} \cdot \frac{L'(E,1)}{\Omega_E}\right) \\ &= \text{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)_{\text{an}}).\end{aligned}$$

In other words the upper bound asserted by the BSD conjecture is true up to a factor involving only bad and supersingular primes, and primes p for which $\bar{\rho}_p$ is not surjective or the p-adic regulator is 0.

The above theorem has as a hypothesis that the reduction is good ordinary, because this is the only case when we know a proof of the p-adic Gross-Zagier formula. It would be interesting to obtain a generalization of the p-adic Gross-Zagier formula to the supersingular case.

10. Algorithm for an upper bound on the rank

Let $E/\mathbb{Q}$ be an elliptic curve. In this section we explain how to compute upper bounds on the rank r of the Mordell-Weil group $E(\mathbb{Q})$. For this purpose, we choose a prime p satisfying the following conditions:

- $p > 2$,
- E has good reduction at p.

By computing the analytic p-adic L-function $\mathcal{L}_p(E,T)$ to a certain precision, we find an upper bound, say b, on the order of vanishing of $\mathcal{L}_p(E,T)$ at $T=0$. Note that a theorem of Rohrlich [Roh84] guarantees that $\mathcal{L}_p(E,T)$ is not zero. Then

$$b \;\geq\; \operatorname{ord}_{T=0} \mathcal{L}_p(E,T) \;\geq\; \operatorname{ord}_{T=0} f_E(T) \;\geq\; r$$

by Kato's Theorems 7.3 and 7.4 and by Theorems 6.1 and 6.2. Hence we have an upper bound on the rank r.

Proposition 10.1. *The computation of an approximation of the p-adic L-series of E for an odd prime p of good reduction produces an upper bound on the rank r of the Mordell-Weil group $E(\mathbb{Q})$.*

By searching for points of small height on E, we also obtain a lower bound on the rank r. Simultaneously, we can increase the precision of the computation of the p-adic L-function in order to try to lower the bound b. Eventually, the lower bound is equal to the upper bound, unless the p-adic BSD Conjecture 5.1 or 5.2 is false. This is similar to the conditional algorithm described in Proposition 2.2, except that we do know here that our upper bounds are unconditional. We do not know unconditionally that this procedure terminates after finitely many steps. Summarizing we can claim the following.

Proposition 10.2. *Let E be an elliptic curve, and assume that there is a prime p of good reduction such that the p-adic BSD conjecture is true. Then there is an algorithm that computes the rank r of E using p-adic L-functions.*

Of course, the procedure for computing bounds on the rank r using m-descents has the same properties: it tries to determine the rank by searching for points and by bounding r from above by the rank of the various m-Selmer groups. Unless all the p-primary parts of the Tate-Shafarevich group are infinite, this procedure returns the rank r after a finite number of steps.

But the two algorithms are fundamentally different, since the m-descent algorithm is fast and there are optimized implementations for small m, but it would be prohibitively time-consuming for larger m (e.g., $m \geq 13$). In contrast, computing the p-adic L-series even for p around 1000 is reasonably efficient, assuming one can compute the relevant modular symbols spaces.

10.1. **Technical remarks.**

The second condition above (good reduction) on the prime p is too strict. We may actually allow primes of multiplicative reduction, too. Of course in the exceptional case, when E has split multiplicative reduction, the upper bound b on the order of vanishing of the p-adic L-function $\mathcal{L}_p(E,T)$ at $T=0$ satisfies $b \geq r+1$.

Note that, assuming that the p-adic BSD conjecture holds, it is easy to predict the needed precision in the computation of the p-adic L-series. So we can compute immediately with the precision that should be sufficient and concentrate on the search for points of small heights.

For practical purposes, we take p as small as possible. The computation of the leading term of $\mathcal{L}_p(E,T)$ using the algorithm of Section 3 for curves of higher rank r is time-consuming for large p. Also we should avoid primes p with supersingular or split multiplicative reduction as there the needed precision is much higher and the computation of b is much slower.

Also the speed of the computation of $\mathcal{L}_p(E,T)$ using modular symbols depends on the size of the conductor. As the conductor grows, the determination of the rank, when it is larger than 1, using the descent method becomes much more efficient than the use of p-adic L-series computed using modular symbols following the linear algebra algorithm of [Cre97]. However, using p-adic L-series may provide an advantage when considering families of quadratic twists.

An advantage to the descent method is that the determination of the m-Selmer group for some $m>1$ can be used for the search of points of infinite order. If the elements of the Selmer group can be expressed as coverings, it is more efficient to search for rational points on the coverings rather than on the elliptic curve itself.

11. The algorithm for the Tate-Shafarevich group

The second algorithm takes as input an elliptic curve E and a prime p and tries to compute an upper bound on the p-primary part of $Ш(E/\mathbb{Q})$. To apply the results above, we impose the following conditions on (E,p):

- $p>2$,
- E does not have additive reduction at p,
- the image of $\bar{\rho}_p$ is the full group $\mathrm{GL}_2(\mathbb{F}_p)$.

As mentioned above, these conditions apply to all but finitely many primes p.

Algorithm 11.1. Given an elliptic curve $E/\mathbb{Q}$ and a prime p satisfying the above conditions, this procedure either gives an upper bound for $\#Ш(E/\mathbb{Q})(p)$ or terminates with an error.

(1) Attempt to determine the rank r and the full Mordell-Weil group $E(\mathbb{Q})$. Exit with an error if we fail to do this.

(2) Compute higher and higher approximations to the p-adic regulator of E over $\mathbb{Q}$ using the algorithm in [MST06, Har08]. Exit with an error if after a predetermined number of steps, the p-adic height pairing is not shown to be nondegenerate.

(3) Using modular symbols, compute an approximation of the coefficient $\mathcal{L}_p^*(E,0)$ of the leading term of the p-adic L-series $\mathcal{L}_p(E,T)$. If the order of vanishing
$$\mathrm{ord}_{T=0}\,\mathcal{L}_p(E,T)$$
is equal to r (or $r+1$ if E has split multiplicative reduction at p), then we print that $Ш(E/\mathbb{Q})(p)$ is finite, otherwise we increase the precision of the computation of $\mathcal{L}_p(E,T)$. If, after some prespecified cutoff, this fails to prove that $\mathrm{ord}_{T=0}\,\mathcal{L}_p(E,T)=r$ (or $r+1$), then exit with an error.

(4) Compute the remaining information, including the Tamagawa numbers c_v and the p-adic multiplier ϵ_p. If p is a good ordinary prime or a prime at which E has nonsplit multiplicative reduction, let
$$\begin{aligned} b_p = &\,\mathrm{ord}_p(\mathcal{L}_p^*(E,0)) - \mathrm{ord}_p(\epsilon_p) \\ &- \sum_v \mathrm{ord}_p(c_v) - \mathrm{ord}_p(\mathrm{Reg}_\gamma(E/\mathbb{Q})). \end{aligned}$$

If p is supersingular, let

$$b_p = \operatorname{ord}_p((1-\varphi)^{-2}\,\mathcal{L}_p^*(E,0)) - \operatorname{ord}_p(\operatorname{Reg}_p(E/\mathbb{Q})) - \sum_v \operatorname{ord}_p(c_v).$$

Finally, if E has split multiplicative reduction at p, let

$$\begin{aligned} b_p &= \operatorname{ord}_p(\mathcal{L}_p^*(E,0)) - \operatorname{ord}_p(\mathscr{L}_p) \\ &\quad - \sum_v \operatorname{ord}_p(c_v) - \operatorname{ord}_p(\operatorname{Reg}_\gamma(E/\mathbb{Q}))\,. \end{aligned}$$

(5) Output that $\#\text{Ш}(E/\mathbb{Q})(p)$ is bounded by p^{b_p}.

Proof. At Step 4, we have shown that Conjecture 4.1 (or Conjecture 4.4 in the supersingular case) on the nondegeneracy of the p-adic regulator holds and that $\text{Ш}(E/\mathbb{Q})(p)$ is indeed finite by Theorem 6.1 (or Theorem 6.2 in the supersingular case). Moreover this theorem shows that

$$\operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)) = \operatorname{ord}_p(f_E^*(0)) + \operatorname{ord}_p\left(\frac{(\#E(\mathbb{Q})(p))^2}{\epsilon_p\cdot\prod_v c_v}\cdot\frac{1}{\operatorname{Reg}_\gamma(E/\mathbb{Q})}\right)$$

in the ordinary case (or the same formula where ϵ_p is replaced by $\mathscr{L}_p$ in the split multiplicative case) and

$$\operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p)) = \operatorname{ord}_p((1-\varphi)^{-2}\,f_E^*(0)) - \operatorname{ord}_p(\operatorname{Reg}_p(E/\mathbb{Q})) - \sum_v \operatorname{ord}_p(c_v)$$

in the supersingular case. Note that $\#E(\mathbb{Q})(p) = 1$ since we assumed that $\bar\rho_p$ is surjective. Finally, we use Kato's Theorem 7.3 that

$$\operatorname{ord}_p(f_E^*(0)) \le \operatorname{ord}_p(\mathcal{L}_p^*(E,0))$$

to prove that b_p is indeed an upper bound on $\operatorname{ord}_p(\#\text{Ш}(E/\mathbb{Q})(p))$. □

In the next proposition we summarize the discussion of this section.

Proposition 11.2. *Let E be an elliptic curve and $p>2$ a prime for which E has semistable reduction. If Conjectures 4.1 and 4.4 hold and if we are able to determine the Mordell-Weil group of E, then there is a algorithm to verify that the p-primary part of $\text{Ш}(E/\mathbb{Q})$ is finite. If moreover the representation $\bar\rho_p$ is surjective, then the algorithm produces an upper bound on $\#\text{Ш}(E/\mathbb{Q})(p)$. If Conjecture 7.1 holds then the result of the algorithm is equal to the order of $\text{Ш}(E/\mathbb{Q})(p)$.*

11.1. **Technical remarks.** In Step 1 of Algorithm 11.1 we may use several ways to determine the rank and the Mordell-Weil group. E.g., first compute the modular symbol $[0]^+$. If it is not zero, we have that $L(E,1)\neq 0$ and the rank has to be 0. If the order of vanishing of $L(E,s)$ at $s=1$ is 1, we may use Heegner points to find the full Mordell-Weil group, which then is of rank 1. Otherwise we use descent methods or the algorithm in the previous section to bound the rank from above and search for points to find a lower bound. When enough points are found to generate a group of finite index, we saturate the group using infinite descent in order to find the full group $E(\mathbb{Q})$. In practice this step does not create any problems as Step 3 is usually computationally more difficult.

In Step 3, it is easy to determine the precision that will be needed to compute the p-adic valuation of the leading term $\mathcal{L}_p^*(E,0)$ if we assume the complex and the p-adic version of the BSD conjecture. Hence it is easy to decide when to exit at this step.

The algorithm exits with an error only if the Mordell-Weil group could not be determined (in Step 1), if Conjecture 4.1 or 4.4 is wrong (in Step 2), if the p-primary part of $\text{Ш}(E/\mathbb{Q})$ is infinite or if the main conjecture is false (both in Step 3). Hence only weaker variants of the p-adic Birch and Swinnerton-Dyer conjecture are needed.

Another application of the algorithm is the following remark. If, for a given elliptic curve E and a prime p, the algorithm yields as output that the p-primary part of $\text{Ш}(E/\mathbb{Q})$ is trivial, then the algorithm has actually also proved the main conjecture for E and p. Because we know by then that $\mathcal{L}_p(E,T)$ and the characteristic series $f_{E}(T)$ of the Selmer group have the same order of vanishing at $T=0$ and the leading terms have the same valuation. Since, by Kato's theorem $f_{E}(T)$ divides $\mathcal{L}_p(E,T)$, we know then that the quotient is a unit in $\mathbb{Z}_p[\![T]\!]$. Such calculations and especially this remark on how to verify the main conjecture in special cases are already contained in [PR03] for supersingular primes p.

12. Numerical results

The algorithms described above were implemented by the authors in Sage (see [S$^+$11b]) and all of the calculations given below can be carried out using Sage and PSage [S$^+$11a].

12.1. **A split multiplicative example.**

To give an example of a curve with split multiplicative reduction, we use the same curve as before (see Equation (3.3))

$$E_0\colon \quad y^2 + x\,y = x^3 - x^2 - 4\,x + 4$$

but with the prime $p=223$. Of course, there is no hope in practice that an explicit 223-descent could be used to compute the order of $\text{Ш}(E_0/\mathbb{Q})(223)$. However, we can compute the p-adic regulator and the $\mathscr{L}$-invariant to high precision quickly using Tate's parametrization of E_0:

$$\begin{aligned}\operatorname{Reg}_p(E_0/\mathbb{Q}) &= 153\cdot 223^2 + 125\cdot 223^3 + 124\cdot 223^4 + \mathbf{O}(223^5),\\ \mathscr{L} &= 179\cdot 223 + 85\cdot 223^2 + 30\cdot 223^3 + \mathbf{O}(223^4).\end{aligned}$$

The computation of the p-adic L-series is more time consuming[8]. But as we only need the first p-adic digit to prove the triviality of $\text{Ш}(E_0/\mathbb{Q})(223)$, we only need to sum over $222\cdot 223$ modular symbols. This yields

$$\mathcal{L}_p(E_0,T) = \mathbf{O}(223^4) + \mathbf{O}(223^1)\cdot T + \mathbf{O}(223^1)\cdot T^2 + (139 + \mathbf{O}(223))\cdot T^3 + \mathbf{O}(T^4).$$

In fact, we know that the first three coefficients vanish as we are in the exceptional case, so the leading term has valuation 0. From these computations, we see that the p-adic BSD conjecture predicts that

$$\#\text{Ш}(E_0/\mathbb{Q}) \equiv 1 \pmod{223};$$

in particular, we may conclude that $\text{Ш}(E_0/\mathbb{Q})(223) = 0$.

[8] The optimized implementation mentioned in Section 12.4 does this entire computation in less than one second total time, including the modular symbols space computation.

12.2. **A supersingular example.** Let E be the elliptic curve

$$E\colon\quad y^2 + x = x^3 + x^2 + 2x + 2$$

listed as curve 1483a1 in Cremona's tables. The curve has rank 2 generated by $(-1,0)$ and $(0,1)$. The reduction of E at $p=5$ is supersingular. The p-adic L-series is

$$\begin{aligned}\mathcal{L}_p(E,T) =& \big((1+\mathbf{O}(5))\cdot T^2 + (1+\mathbf{O}(5))\cdot T^3 + \mathbf{O}(T^4)\big)\cdot \omega_{E}\\ &+ \big((4\cdot 5+\mathbf{O}(5^2))\cdot T^2 + (4\cdot 5+\mathbf{O}(5^2))\cdot T^3 + \mathbf{O}(T^4)\big)\cdot \varphi(\omega_E)\end{aligned}$$

where we have already taken in account that the first two terms vanish. We compute the normalized D_p-valued regulator

$$\begin{aligned}\mathrm{Reg}_\gamma(E/\mathbb{Q}) =& \big(1+2\cdot 5+3\cdot 5^2+5^3+\mathbf{O}(5^5)\big)\cdot \omega_{E}\\ &+ \big(4\cdot 5+4\cdot 5^2+4\cdot 5^3+5^4+2\cdot 5^5+\mathbf{O}(5^6)\big)\cdot \varphi(\omega_{E})\,.\end{aligned}$$

Hence the p-adic BSD conjecture predicts that

$$\begin{aligned}&\big(1+\mathbf{O}(5)\big)\,\omega_{E} + \big(4\cdot 5+\mathbf{O}(5^2)\big)\,\varphi(\omega_{E}) =\\ &\qquad \#Ш(E/\mathbb{Q})\cdot\Big(\big(1+\mathbf{O}(5)\big)\,\omega_{E} + \big(4\cdot 5+\mathbf{O}(5^2)\big)\,\varphi(\omega_{E})\Big).\end{aligned}$$

In particular, we have shown that $Ш(E/\mathbb{Q})(5)$ is trivial. It follows from Iwasawa-theoretic consideration as in [PR03] that, if $\#Ш(E/_n\mathbb{Q})(5) = 5^{e_n}$ then

$$e_n = \frac{p}{p^2-1}\cdot p^n + \mathbf{O}(1)\,.$$

12.3. **An example whose Tate-Shafarevich group is nontrivial.** Let E be the elliptic curve given by

$$E\colon\quad y^2 + x\,y = x^3 + 16353089\,x - 335543012233$$

which is labeled 858k2 in [Cre]. The curve has rank 0 and is semistable, and the full BSD conjecture predicts that the Tate-Shafarevich group $Ш(E/\mathbb{Q})$ consists of two copies of $\mathbb{Z}/7\mathbb{Z}$.

We may compute the 7-adic L-series, which yields

$$\begin{aligned}\mathcal{L}_7(E,T) =& 7^2\cdot(2\cdot 7^2+7^3+7^4+3\cdot 7^5+\mathbf{O}(7^6) + (5\cdot 7^2+\mathbf{O}(7^3))\cdot T\\ &+ (3+4\cdot 7+5\cdot 7^2+\mathbf{O}(7^3))\cdot T^2 + \mathbf{O}(T^3))\end{aligned}$$

On the algebraic side, we find that the constant term of the characteristic series of E has valuation $2+\mathrm{ord}_7(\#Ш(E/\mathbb{Q}))$. So our algorithm yields the correct upper bound, that $\#Ш(E/\mathbb{Q})(7) \le 7^2$. We can change to the curve 858k1 with a 7-isogeny and prove there directly that the upper bound on the 7-primary part of the Tate-Shafarevich group is 1, so by isogeny invariance of the Birch and Swinnerton-Dyer conjecture it follows that $\#Ш(E/\mathbb{Q})(7) = 7^2$. (Of course, this can be shown with other methods for this curve of rank 0, e.g., by using Heegner points.) Since we know the exact order of $Ш(E/\mathbb{Q})$, we deduce that the main conjecture holds. (Also, this can be deduced from Theorem 7.5 taking $q = 11$.)

Once again we learn even more from the computation of the p-adic L-series. Iwasawa theory tells us that the order of the Tate-Shafarevich group grows quickly (for an ordinary prime) in the $\mathbb{Z}_7$-extension. Namely if $\#Ш(E/_n\mathbb{Q}) = 7^{e_n}$ then $e_n = 2\cdot 7^n + 2\cdot n + \mathbf{O}(1)$.

12.4. **Tate-Shafarevich Groups of Elliptic Curves of Rank at Least 2.** According to [Cre], for every elliptic curve with rank ≥ 2 and conductor up to 130,000, the BSD conjecture predicts that $Ш(E/\mathbb{Q}) = 0$. In this section, we describe the computation we did to verify Theorem 1.1, which gives evidence for this observation, at least up to conductor 30,000.

Consider a pair (E, p) consisting of

(1) an optimal elliptic curve E defined over $\mathbb{Q}$ with rank $r \geq 2$ and conductor $\leq$ 30,000, and
(2) a good ordinary prime p with $5 \leq p <$ 1,000 such that $\bar{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p])$ is surjective.

There are 9,679 such curves E and 1,534,422 such pairs (E, p). For each pair, we do the following:

(1) Show that $r = \mathrm{ord}_T\, \mathcal{L}_p(E, T)$.
(2) Compute the conjectural order of $Ш(E/\mathbb{Q})$ according to Conjecture 5.1 mod p, and check that it is $1 + O(p)$.

As explained in the proof of Algorithm 11.1, our hypotheses on p then imply that $Ш(E/\mathbb{Q})[p] = 0$. As evidence for Conjecture 5.1 and as a double check on our implementation, we also verify the conjecture to precision $O(p)$ for each pair (E, p).

(1) We compute[9] approximations to $\mathcal{L}_p(E, T)$ that are sufficient to show that $\mathrm{ord}_T(\mathcal{L}_p(E, T)) = r$. For 1,523,413 of our 1,534,422 pairs (E, p), we did this by computing $P_2 \equiv \mathcal{L}_p(E, T) \pmod{(p, T^5)}$; for the remaining 11,009 pairs, we computed to higher precision.
(2) For all of our pairs (E, p), we computed the p-adic regulator $\mathrm{Reg}_p(E) \in \mathbb{Q}_p$ to precision at least $O(p^{12})$. In all cases this computation confirmed that $\mathrm{Reg}_p(E) \neq 0$.
(3) With the above data for our pairs (E, p), it was then straightforward to compute the conjectural order of $Ш(E/\mathbb{Q})$ according to Conjecture 5.1, and in all cases we got $1 + O(p)$, so $Ш(E/\mathbb{Q})[p] = 0$.

Remark 12.1. In fact, we carried out the regulator calculation mentioned above for all pairs (E, p) with $5 \leq p < 1000$ good ordinary for which the conductor of E is $\leq$ 130,000 and the rank is ≥ 2. A selection of large $\mathrm{ord}_p(\mathrm{Reg}_p(E))$ is given in Table 1. For example, for the first curve 53770a1 with $p = 7$, the conductor factors as $53770 = 2 \cdot 5 \cdot 19 \cdot 283$, the Tamagawa numbers are $12, 2, 6, 1$, which are all coprime to 7, we have $Ш(E/\mathbb{Q})_{\mathrm{an}} = 1$, and $N_7 = 9$, which is coprime to 7, but

$$\mathrm{Reg}_7(E) = 7^7 \cdot 419257219506 + O(7^{21})$$

is divisible by a rather large power of 7. The leading coefficient of the 7-adic L-series vanishes to order $7 - \mathrm{rank}(E)$, as expected, so $Ш(E/\mathbb{Q})(7) = 0$:

$$\begin{aligned}\mathcal{L}_7(E, T) = O(7^9) + O(7^6)T + \left(6 \cdot 7^5 + O(7^6)\right) T^2 + \left(3 \cdot 7^5 + O(7^6)\right) T^3 \\ + \left(5 + 5 \cdot 7 + 2 \cdot 7^4 + 7^5 + O(7^6)\right) T^4 + O(T^5)\end{aligned}$$

Remark 12.2. A very hard case is $(E, p) = (17856j1, 757)$, in which E has rank 2 and

$$\mathrm{Reg}_p(E) = 261 \cdot 757^4 + 531 \cdot 757^5 + 293 \cdot 757^6 + 309 \cdot 757^7 + \cdots$$

[9] The computation of the approximate p-adic L-series for all of our pairs (E, p) took *several months of CPU time* using an optimized implementation of the algorithm of Section 3.

The leading coefficient of the 757-adic L-series must be divisible by 757^2, so we must compute $\mathcal{L}_7(E,T) \pmod{757^3}$, which is enormously time consuming, even with our highly optimized implementation, since each power of p increases the complexity by a factor of p (and, in addition, we use slower arbitrary precision arithmetic to avoid overflow). The computation took over two months of CPU time, and yielded

$$\mathcal{L}_{757}(T) = O(757^3) + O(757^3)T + \left(399 \cdot 757^2 + O(757^3)\right) T^2 + \cdots$$

Thus the p-adic BSD conjecture predicts that $\#\text{Ш}(E/\mathbb{Q})(757) \equiv 1 \pmod{757}$, hence $\text{Ш}(E/\mathbb{Q})[757] = 0$.

TABLE 1. Various examples in which $\mathrm{ord}_p(\mathrm{Reg}_p(E))$ is large

Curve	Rank	p	$\mathrm{Reg}_p(E)$
53770a1	2	7	$7^7 \cdot 419257219506 + O(7^{21})$
60237b1	2	7	$7^7 \cdot 195984223121 + O(7^{21})$
65088bm1	2	5	$5^7 \cdot 3628814228 + O(5^{21})$
71236b1	2	5	$5^7 \cdot 2905505203 + O(5^{21})$
74220b1	2	7	$7^7 \cdot 411568240919 + O(7^{21})$
82096e1	2	11	$11^7 \cdot 163096174634581 + O(11^{21})$
91143f1	2	17	$17^7 \cdot 32722747582988964 + O(17^{21})$
101552a1	2	5	$5^7 \cdot 1575344534 + O(5^{21})$
116634k1	2	5	$5^7 \cdot 1877361868 + O(5^{21})$
121212q1	2	5	$5^7 \cdot 5806958402 + O(5^{21})$
123888bm1	2	7	$7^7 \cdot 537125029809 + O(7^{21})$
127368d1	2	13	$13^7 \cdot 485242111874635 + O(13^{21})$
27448d1	3	5	$5^6 \cdot 115188708423 + O(5^{22})$
53122a1	3	5	$5^6 \cdot 31988633 + O(5^{22})$
90953a1	3	7	$7^6 \cdot 28674298268349 + O(7^{22})$

Let E be the elliptic curve 389a of rank 2. We verified for a large number of primes p that $\text{Ш}(E/\mathbb{Q})[p] = 0$.

Theorem 12.3. *Let E be the rank 2 elliptic curve of conductor 389. Then for* 2 *and all 5,005 good ordinary primes $p <$ 48,859 except $p =$ 16,231 we have $\text{Ш}(E/\mathbb{Q})[p] = 0$. For each such p, the p-adic BSD conjectural order of* Ш *is congruent to* 1 *modulo p. This only excludes the following bad or supersingular primes and the good ordinary prime 16,231:*

$$\begin{aligned} p =& 107, 389, 599, 1049, 2957, 6661, 8263, 9397, 9551, 14633, 15101, 28591, \\ & 30671, 30869, 31799, 34781, 36263, 45161. \end{aligned}$$

Proof. This is a computation similar to the one described above that takes several weeks CPU time. □

Remark 12.4. For the prime $p =$ 16,231, we have $\mathrm{ord}_p(\mathrm{Reg}_p) = 3$ instead of $2 = \mathrm{rank}(E)$. Thus the computation is roughly 16,231 times as difficult as it is for nearby primes using our algorithm, so we estimate it would take several CPU years to finish. It should be possible to instead deal with this exceptional case efficiently using the overconvergent modular symbols approach of Pollack-Stevens [PS11], when a suitable implementation is available.

Remark 12.5. We have excluded supersingular primes from this section not because our algorithms do not apply (they do apply), but because our implementations are significantly slower in this case. We hope to address this shortcoming in future work.

References

[ARS06] Amod Agashe, Kenneth Ribet, and William A. Stein, *The Manin constant*, Pure Appl. Math. Q. **2** (2006), no. 2, 617–636.

[AS02] Amod Agashe and William Stein, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory **97** (2002), no. 1, 171–185.

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

[BCP97] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[Ber81] Dominique Bernardi, *Hauteur p-adique sur les courbes elliptiques*, Seminar on Number Theory, Paris 1979–80, Progr. Math., vol. 12, Birkhäuser Boston, 1981, pp. 1–14.

[Ber82] Daniel Bertrand, *Valuers de fonctions thêta et hauteur p-adiques*, Seminar on Number Theory, Paris 1980-81, Progr. Math., vol. 22, Birkhäuser Boston, 1982, pp. 1–11.

[BPR93] Dominique Bernardi and Bernadette Perrin-Riou, *Variante p-adique de la conjecture de Birch et Swinnerton-Dyer (le cas supersingulier)*, C. R. Acad. Sci. Paris Sér. I Math. **317** (1993), no. 3, 227–232.

[BDGP96] Katia Barré-Sirieix, Guy Diaz, François Gramain, and Georges Philibert, *Une preuve de la conjecture de Mahler-Manin*, Invent. Math. **124** (1996), no. 1-3, 1–9.

[Cas65] J. W. S. Cassels, *Arithmetic on curves of genus 1. VIII. On conjectures of Birch and Swinnerton-Dyer*, J. Reine Angew. Math. **217** (1965), 180–199.

[CFO$^+$08] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, *Explicit n-descent on elliptic curves. I. Algebra*, J. Reine Angew. Math. **615** (2008), 121–155.

[CFO$^+$09] ———, *Explicit n-descent on elliptic curves. II. Geometry*, J. Reine Angew. Math. **632** (2009), 63–84.

[CFO$^+$11] ———, *Explicit n-descent on elliptic curves. III. Algorithms*, Preprint. `http://arxiv.org/abs/1107.3516`, 2011.

[CLS09] J. Coates, Z. Liang, and R. Sujatha, *The Tate-Shafarevich group for elliptic curves with complex multiplication*, J. Algebra **322** (2009), no. 3, 657–674.

[CLS10] ———, *The Tate-Shafarevich group for elliptic curves with complex multiplication II*, Milan J. Math. **78** (2010), no. 2, 395–416. MR 2781846

[Coa11] John Coates, *The enigmatic Tate-Shafarevich group*, 2010 Proceedings of International Congress of Chinese Mathematicians (2011).

[Coh07] Henri Cohen, *Number theory. Vol. II. Analytic and modern tools*, Graduate Texts in Mathematics, vol. 240, Springer, New York, 2007. MR MR2312338

[Col04] Pierre Colmez, *La conjecture de Birch et Swinnerton-Dyer p-adique*, Astérisque (2004), no. 294, ix, 251–319.

[Col10] ———, *Invariants $\mathscr{L}$ et dérivées de valeurs propres de Frobenius*, Astérisque (2010), no. 331, 13–28.

[Cre] J.E. Cremona, *Elliptic Curves Data*, `http://www.warwick.ac.uk/~masgaj/ftp/data/INDEX.html`.

[Cre97] John E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, 1997.

[CS00] John Coates and Ramdorai Sujatha, *Galois cohomology of elliptic curves*, Tata Institute of Fundamental Research Lectures on Mathematics, vol. 88, Narosa Publishing House, 2000.

[Del98] Daniel Delbourgo, *Iwasawa theory for elliptic curves at unstable primes*, Compositio Math. **113** (1998), no. 2, 123–153.

[Del02] ———, *On the p-adic Birch, Swinnerton-Dyer conjecture for non-semistable reduction*, J. Number Theory **95** (2002), no. 1, 38–71.

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149. MR MR2068888 (2005f:11128)

[Edi91] Bas Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Progr. Math., vol. 89, Birkhäuser Boston, Boston, MA, 1991, pp. 25–39.

[GJP$^+$09] G. Grigorov, A. Jorza, S. Patrikis, C. Tarnita, and W. Stein, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. **78** (2009), 2397–2425, `http://wstein.org/papers/bsdalg/`.

[Gre99] Ralph Greenberg, *Iwasawa theory for elliptic curves*, Arithmetic theory of elliptic curves (Cetraro, 1997), Lecture Notes in Math., vol. 1716, Springer, Berlin, 1999, pp. 51–144.

[Gre01] ———, *Introduction to Iwasawa theory for elliptic curves*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 407–464. MR 1860044 (2003a:11067)

[Gri05] Grigor Tsankov Grigorov, *Kato's Euler System and the Main Conjecture*, Ph.D. thesis, Harvard University, 2005.

[GS93] Ralph Greenberg and Glenn Stevens, *p-adic L-functions and p-adic periods of modular forms*, Invent. Math. **111** (1993), no. 2, 407–447.

[GV00] Ralph Greenberg and Vinayak Vatsal, *On the Iwasawa invariants of elliptic curves*, Invent. Math. **142** (2000), no. 1, 17–63.

[Har08] David Harvey, *Efficient computation of p-adic heights*, LMS J. Comput. Math. **11** (2008), 40–59.

[Jon89] John W. Jones, *Iwasawa L-functions for multiplicative abelian varieties*, Duke Math. J. **59** (1989), no. 2, 399–420.

[Kat04] Kazuya Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, Cohomologies p-adiques et application arithmétiques. III, Astérisque, vol. 295, Société Mathématique de France, Paris, 2004.

[Ked01] Kiran S. Kedlaya, *Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology*, J. Ramanujan Math. Soc. **16** (2001), no. 4, 323–338.

[Ked03] K. S. Kedlaya, *Errata for: "Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology" [J. Ramanujan Math. Soc.* **16** *(2001), no. 4, 323–338*, J. Ramanujan Math. Soc. **18** (2003), no. 4, 417–418, Dedicated to Professor K. S. Padmanabhan. MR 2 043 934

[Ked04] K. Kedlaya, *Computing zeta functions via p-adic cohomology*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 3076, Springer, Berlin, 2004, pp. 1–17.

[Kob03] Shin-ichi Kobayashi, *Iwasawa theory for elliptic curves at supersingular primes*, Invent. Math. **152** (2003), no. 1, 1–36.

[Kob06] Shinichi Kobayashi, *An elementary proof of the Mazur-Tate-Teitelbaum conjecture for elliptic curves*, Doc. Math. (2006), no. Extra Vol., 567–575 (electronic).

[Kol91a] V. A. Kolyvagin, *On the structure of Shafarevich-Tate groups*, Algebraic geometry (Chicago, IL, 1989), Lecture Notes in Math., vol. 1479, Springer, Berlin, 1991, pp. 94–121.

[Kol91b] V. A. Kolyvagin, *On the structure of Selmer groups*, Math. Ann. **291** (1991), no. 2, 253–259.

[KP07] Masato Kurihara and Robert Pollack, *Two p-adic L-functions and rational points on elliptic curves with supersingular reduction*, L-functions and Galois representations, London Math. Soc. Lecture Note Ser., vol. 320, Cambridge Univ. Press, Cambridge, 2007, pp. 300–332.

[Man71] J. I. Manin, *Cyclotomic fields and modular curves*, Russian Math. Surveys **26** (1971), no. 6, 7–78.

[Man72] Ju. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66.

[Maz72] B. Mazur, *Rational points of abelian varieties with values in towers of number fields*, Invent. Math. **18** (1972), 183–266.

[Mil10] Robert L. Miller, *Proving the Birch and Swinnerton-Dyer conjecture for specific elliptic curves of analytic rank zero and one*, `http://arxiv.org/abs/1010.2431`, 2010.

[MSD74] Barry Mazur and Peter Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61.

[MST06] Barry Mazur, William Stein, and John Tate, *Computation of p-adic heights and log convergence*, Doc. Math. (2006), no. Extra Vol., 577–614 (electronic).

[MT91] Barry Mazur and John Tate, *The p-adic sigma function*, Duke Math. J. **62** (1991), no. 3, 663–688.

[MTT86] Barry Mazur, John Tate, and J. Teitelbaum, *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer*, Invent. Math. **84** (1986), no. 1, 1–48.

[PAR11] The PARI Group, Bordeaux, *PARI/GP, version* `2.5`, 2011, available from `http://pari.math.u-bordeaux.fr/`.

[Pol] Robert Pollack, *Tables of Iwasawa invariants of elliptic curves*, `http://math.bu.edu/people/rpollack/Data/data.html`.

[Pol03] ———, *On the p-adic L-function of a modular form at a supersingular prime*, Duke Math. J. **118** (2003), no. 3, 523–558.

[PR82] Bernadette Perrin-Riou, *Descente infinie et hauteur p-adique sur les courbes elliptiques à multiplication complexe*, Invent. Math. **70** (1982), no. 3, 369–398.

[PR87] ———, *Fonctions L p-adiques, théorie d'Iwasawa et points de Heegner*, Bull. Soc. Math. France **115** (1987), no. 4, 399–456.

[PR93] ———, *Fonctions L p-adiques d'une courbe elliptique et points rationnels*, Ann. Inst. Fourier (Grenoble) **43** (1993), no. 4, 945–995.

[PR94] ———, *Théorie d'Iwasawa des représentations p-adiques sur un corps local*, Invent. Math. **115** (1994), no. 1, 81–161, With an appendix by Jean-Marc Fontaine.

[PR03] ———, *Arithmétique des courbes elliptiques à réduction supersingulière en p*, Experiment. Math. **12** (2003), no. 2, 155–186.

[PR04] Robert Pollack and Karl Rubin, *The main conjecture for CM elliptic curves at supersingular primes*, Ann. of Math. (2) **159** (2004), no. 1, 447–464.

[PS11] Robert Pollack and Glenn Stevens, *Overconvergent modular symbols and p-adic L-functions*, Ann. Sci. Éc. Norm. Supér. (4) **44** (2011), no. 1, 1–42.

[Roh84] David E. Rohrlich, *On L-functions of elliptic curves and cyclotomic towers*, Invent. Math. **75** (1984), no. 3, 409–423.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, `http://wstein.org/papers/serre/`, pp. 143–232. MR 2002h:11047

[Rub99] Karl Rubin, *Elliptic curves with complex multiplication and the conjecture of Birch and Swinnerton-Dyer*, Arithmetic theory of elliptic curves (Cetraro, 1997), Lecture Notes in Math., vol. 1716, Springer, Berlin, 1999, pp. 167–234.

[Rub00] ———, *Euler systems*, Annals of Mathematics Studies, vol. 147, Princeton University Press, Princeton, NJ, 2000, Hermann Weyl Lectures. The Institute for Advanced Study.

[S^+11a] W. A. Stein et al., *Psage Library*, 2011, `http://code.google.com/p/purplesage/`.

[S^+11b] ———, *Sage Mathematics Software (Version 4.6.2)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[Sch82] Peter Schneider, *p-adic height pairings. I*, Invent. Math. **69** (1982), no. 3, 401–409.

[Sch85] ———, *p-adic height pairings. II*, Invent. Math. **79** (1985), no. 2, 329–374.

[Ser72] Jean-Pierre Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331.

[Ser96] ———, *Travaux de Wiles (et Taylor, …). I*, Astérisque (1996), no. 237, Exp. No. 803, 5, 319–332, Séminaire Bourbaki, Vol. 1994/95.

[Sil94] Joseph H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 151, Springer-Verlag, New York, 1994.

[Sim02] Denis Simon, *Computing the rank of elliptic curves over number fields*, LMS J. Comput. Math. **5** (2002), 7–17 (electronic).

[SS04] Edward F. Schaefer and Michael Stoll, *How to do a p-descent on an elliptic curve*, Trans. Amer. Math. Soc. **356** (2004), no. 3, 1209–1231.

[Ste07a] William Stein, *The Birch and Swinnerton-Dyer Conjecture, a Computational Approach*, 2007, `http://wstein.org/books/bsd/`.

[Ste07b] ———, *Modular forms, a computational approach*, Graduate Studies in Mathematics, vol. 79, American Mathematical Society, Providence, RI, 2007, With an appendix by Paul E. Gunnells. MR MR2289048

[SU10] C. Skinner and D. Urban, *The Iwasawa Main Conjecture for* GL_2, `http://www.math.columbia.edu/\%7Eurban/eurp/MC.pdf`.

[Wer98] Annette Werner, *Local heights on abelian varieties and rigid analytic uniformization*, Doc. Math. **3** (1998), 301–319.

[Wil95] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Wut04] Christian Wuthrich, *On p-adic heights in families of elliptic curves*, J. London Math. Soc. (2) **70** (2004), no. 1, 23–40.

[Wut07] ———, *Iwasawa theory of the fine Selmer group*, J. Algebraic Geom. **16** (2007), 83–108.

Department of Mathematics, University of Washington
E-mail address: `wstein@uw.edu`

School of Mathematical Sciences, University of Nottingham
E-mail address: `christian.wuthrich@nottingham.ac.uk`

43 Sage: Creating a Viable Free Open Source Alternative to Magma, Maple, Mathematica, and MATLAB

1

Sage: Creating a Viable Free Open Source Alternative to Magma, Maple, Mathematica, and MATLAB

William Stein[1]

[1] The work was supported by NSF grant DUE-1022574.

1.1 Introduction

The goal of the Sage project (`http://www.sagemath.org`) is to create a viable free open source alternative to Magma, Maple$^{\text{TM}}$, Mathematica®, and MATLAB®, which are the most popular non-free closed source mathematical software systems.[1] Magma is (by far) the most advanced non-free system for structured abstract algebraic computation, Mathematica and Maple are popular and highly developed systems that shine at symbolic manipulation, and MATLAB is the most popular system for applied numerical mathematics. Together there are over 3,000 employees working at the companies that produce the four Ma's listed above, which take in over a hundred million dollars of revenue annually.

By a viable free alternative to the Ma's, we mean a system that will have the important mathematical features of each Ma, with comparable speed. It will have 2d and 3d graphics, an interactive graphical user interface, and documentation, including books, papers, school and college curriculum materials, etc. A single alternative to all of the Ma's is not necessarily a drop-in replacement for any of the Ma's; in particular, it need not run programs written in the custom languages of those systems. Thus an alternative may be philosophically different than the open source system Octave, which understands the MATLAB source language and attempts to implement the entire MATLAB library. Development could instead focus on implementing functions that users demand, rather than systematically trying to implement every single function of the Ma's. The culture, architecture, and general look and feel of such a system would be very different than that of the Ma's.

In Section 1.2 we explain some of the motivation for starting the Sage project, in Section 1.3 we describe the basic architecture of Sage, and in Section 1.4 we sketch aspects of the history of the project.

1.2 Motivation for Starting Sage

Each of the Ma's cost substantial money, and is hence expensive for me, my collaborators, and students. The Ma's are not *owned by the community* like Sage is, or Wikipedia is, for that matter.

The Ma's are closed, which means that the implementation of some

[1] Maple is a trademark of Waterloo Maple Inc. Mathematica is a registered trademark of Wolfram Research Incorporated. MATLAB is a registered trademark of MathWorks. I will refer to the four systems together as "the Ma's" in the rest of this article.

algorithms are secret, in which case you are not allowed to modify or extend them.

> "You should realize at the outset that while knowing about the internals of Mathematica may be of intellectual interest, it is usually much less important in practice than you might at first suppose. Indeed, in almost all practical uses of Mathematica, issues about how Mathematica works inside turn out to be largely irrelevant. Particularly in more advanced applications of Mathematica, it may sometimes seem worthwhile to try to analyze internal algorithms in order to predict which way of doing a given computation will be the most efficient. [...] But most often the analyses will not be worthwhile. For the internals of Mathematica are quite complicated.."
> – The Mathematica Documentation

The philosophy espoused in Sage, and indeed by the vast open source software community, is exactly the opposite. We want you to know about the internals, and when they are quite complicated, we want you to help make them more understandable. Indeed, Sage's growth depends on *you* analyzing how Sage works, improving it, and contributing your improvements back.

```
sage: crt(2, 1, 3, 5)  # Chinese Remainder Theorem
11
sage: crt?         # ? = documentation and examples
Returns a solution to a Chinese Remainder Theorem...
...
sage: crt??        # ?? = source code
def crt(...):
...
    g, alpha, beta = XGCD(m, n)
    q, r = (b - a).quo_rem(g)
    if r != 0:
        raise ValueError("No solution ...")
    return (a + q*alpha*m) % lcm(m, n)
```

Moreover, by browsing `http://hg.sagemath.org/sage-main/`, you can see exactly who wrote or modified any particular line of code in the Sage library, when they did it, and why. Everything included in Sage is free and open source, and it will forever remain that way.

> "I see open source as Science. If you don't spread your ideas in the open, if you don't allow other people to look at how your ideas work and verify that they work, you are not doing Science, you are doing Witchcraft. Traditional software development models, where you keep things inside a company and hide what you are doing, are basically Witchcraft. Open source is all about the fact that it is open; people can actually look at what you are doing, and they can improve it, and they can build on top of it. [...] One of my favorite quotes from history is Newton: 'If I had seen further, it has been by standing on the shoulders of giants.'"

– Linus Torvalds.
Listen at `http://www.youtube.com/watch?v=bt_Y4pSdsHw`

The design decisions of the Ma's are not made openly by the community. In contrast, important decisions about Sage development are made via open public discussions and voting that is archived on public mailing lists with thousands of subscribers.

Every one of the Ma's uses a special mathematics-oriented interpreted programming language, which locks you into their product, makes writing some code outside mathematics unnecessarily difficult, and impacts the number of software engineers that are experts at programming in that language. In contrast, the user language of Sage is primarily the mainstream free open source language Python `http://python.org`, which is one of the world's most popular interpreted programming languages. The Sage project neither invented nor maintains the underlying Python language, but gains immediate access to the IPython shell, Python scientific libraries (such as NumPy, SciPy, CVXopt and MatPlotLib), and a large Python community with major support from big companies such as Google. In comparison to Python, the Ma's are small players in terms of language development. Thus for Sage most of the problems of language development are handled by someone else.

The bug tracking done for three of four of the Ma's is currently secret[2], which means that there is no published accounting of all known bugs, the status of work on them, and how bugs are resolved. But the Ma's do have many bugs; see the release notes of each new version, which lists bugs that were fixed[3]. Sage also has bugs, which are all publicly tracked at `http://trac.sagemath.org`, and there are numerous "Bug Days" workshops devoted entirely to fixing bugs in Sage. Moreover, all discussion about resolving a given bug, including peer review of solutions, is publicly archived. We note that sadly even some prize winning[4] free open source systems, such as GAP `http://www.gap-system.org/`, do not have an open bug tracking system, resulting in people reporting the same bugs over and over again.

Each of the Ma's is a combination of secret unchangeable compiled code and less secret interpreted code. Users with experience programming in compiled languages such as Fortran or C++ may find the loss of a compiler to be frustrating. None of the Ma's has an optimizing compiler that converts programs written in their custom interpreted language to a

[2] MATLAB has an open bug tracker, though it requires free registration to view.

[3] See also `http://cybertester.com/` and `http://maple.bug-list.org/`.

[4] Jenks Prize, 2008

fast executable binary format that is not interpreted at runtime.[5] In contrast, Sage is tightly integrated with Cython[6] `http://www.cython.org`, which is a ython-to-C/C++ compiler that speeds up code execution and has support for statically declaring data types (for potentially enormous speedups) and natively calling existing C/C++/Fortran code. For example, enter the following in a cell of the Sage notebook (e.g., `http://sagenb.org`):

```
def python_sum2(n):
    s = int(0)
    for i in xrange(1, n+1):
        s += i*i
    return s
```

Then enter the following in another cell:

```
%cython
def cython_sum2(long n):
    cdef long i, s = 0
    for i in range(1, n+1):
        s += i*i
    return s
```

The second implementation, despite looking nearly identical, is nearly a hundred times faster than the first one (your timings may vary).

```
sage: timeit('python_sum2(2*10^6)')
5 loops, best of 3: 154 ms per loop
sage: timeit('cython_sum2(2*10^6)')
125 loops, best of 3: 1.76 ms per loop
sage: 154/1.76
87.5
```

Of course, it is better to choose a different algorithm. In case you don't remember a closed form expression for the sum of the first n squares, Sage can deduce it:

```
sage: var('k, n')
sage: factor(sum(k^2, k, 1, n))
```

[5] MATLAB has a compiler, but "the source code is still interpreted at run-time, and performance of code should be the same whether run in standalone mode or in MATLAB." Mathematica also has a `Compile` function, but simply compiles expressions to a different internal format that is interpreted, much like Sage's `fast_callable` function.

[6] The Cython project has received extensive contributions from Sage developers, and is very popular in the world of Python-based scientific computing.

```
1/6*(n + 1)*(2*n + 1)*n
```

And now our simpler fast implementation is:

```
def sum2(n):
    return n*(2*n+1)*(n+1)/6
```

Just as above, we can also use the Cython compiler:

```
%cython
def c_sum2(long n):
    return n*(2*n+1)*(n+1)/6
```

Comparing times, we see that Cython is 10 times faster:

```
sage: n = 2*10^6
sage: timeit('sum2(n)')
625 loops, best of 3: 1.41 microseconds per loop
sage: timeit('c_sum2(n)')
625 loops, best of 3: 0.145 microseconds per loop
sage: 1.41/.145
9.72413793103448
```

In this case, the enhanced speed comes at a cost, in that the answer is *wrong* when the input is large enough to cause an overflow:

```
sage: c_sum2(2*10^6)   # WARNING: overflow
-407788678951258603
```

Cython is very powerful, but to fully benefit from it, one must understand machine level arithmetic data types, such as long, int, float, etc. With Sage you have that option.

1.3 What is Sage?

The goal of Sage is to compete with the Ma's, and the intellectual property at our disposal is the complete range of GPL-compatibly licensed open source software.

Sage is a self-contained free open source *distribution* of about 100 open source software packages and libraries[7] that aims to address all

[7] See the list of packages in Sage at `http://sagemath.org/packages/standard/`. The list includes R, Pari, Singular, GAP, Maxima, GSL, Numpy, Scipy, ATLAS, Matplotlib, and many other popular programs.

computational areas of pure and applied mathematics. The download of Sage contains all dependencies required for the normal functioning of Sage, including Python itself. Sage includes a substantial amount of code that provides a unified Python-based *interface* to these other packages. Sage also includes a library of new code written in Python, Cython and C/C++, which implements a huge range of algorithms.

1.4 History

I made the first release of Sage in February 2005, and at the time called it "**S**oftware for **A**rithmetic **G**eometry **E**xperimentation." I was a serious user of, and contributor to, Magma at the time, and was motivated to start Sage for many of the reasons discussed above. In particular, I was personally frustrated with the top-down closed development model of Magma, the fact that *several million lines* of the source code of Magma are closed source, and the fees that my colleagues had to pay in order to use the substantial amount of code that I contributed to Magma. Despite my early naive hope that Magma would be open sourced, it never was. So I started Sage motivated by the dream that someday the single most important item of software I use on a daily basis would be free and open. David Joyner, David Kohel, Joe Wetherell, and Martin Albrecht were also involved in the development of Sage during the first year.

In February 2006, the National Science Foundation funded a 2-day workshop called "Sage Days 2006" at UC San Diego, which had about 40 participants and speakers from several open and closed source mathematical software projects. After doing a year of fulltime mostly solitary work on Sage, I was surprised by the positive reception of Sage by members of the mathematical research community. What Sage promised was something many mathematicians wanted. Whether or not Sage would someday deliver on that promise was (and for many still is) an open question.

I had decided when I started Sage that I would make it powerful enough for my research, with or without the help of anybody else, and was pleasantly surprised at this workshop to find that many other people were interested in helping, and understood the shortcomings of existing open source software, such as GAP and PARI, and the longterm need to move beyond Magma. Six months later, I ran another Sage Days workshop, which resulted in numerous talented young graduate students, including David Harvey, David Roe, Robert Bradshaw, and Robert Miller, getting involved in Sage development. I used startup money from University of Washington to hire Alex Clemesha as a fulltime employee to implement 2d graphics and help create a notebook interface to Sage. I also learned that there was much broader interest in such a system, and stopped referring to Sage as being exclusively for "arithmetic geometry"; instead, Sage became "**S**oftware for **A**lgebra and **G**eometry **E**xperimentation." Today the acronym is deprecated.

The year 2007 was a major turning point for Sage. Far more people got involved with development, we had four Sage Days workshops, and prompted by Craig Citro, we instituted a requirement that all new code must have tests for 100% of the functions touched by that code, and every modification to Sage must be peer reviewed. Our peer review process is much more open than in mathematical research journals; everything that happens is publicly archived at `http://trac.sagemath.org`. During 2007, I also secured some funding for Sage development from Microsoft Research, Google, and NSF. Also, a German graduate student studying cryptography, Martin Albrecht presented Sage at the Trophées du Libre competition in France, and Sage won first place in "Scientific Software", which led to a huge amount of good publicity, including articles in many languages around the world and appearances[8] on the front page of `http://slashdot.org`.

In 2008, I organized 7 Sage Days workshops at places such as IPAM (at UCLA) and the Clay Mathematics Institute, and for the first time, several people besides me made releases of Sage. In 2009, we had 8 more Sage Days workshops, and the underlying foundations of Sage improved, including development of a powerful coercion architecture. This *coercion model* systematically determines what happens when performing operations such as `a + b`, when `a` and `b` are elements of potentially different rings (or groups, or modules, etc.).

```
sage: R.<x> = PolynomialRing(ZZ)
sage: f = x + 1/2; f
x + 1/2
sage: parent(f)
Univariate Polynomial Ring in x over Rational Field
```

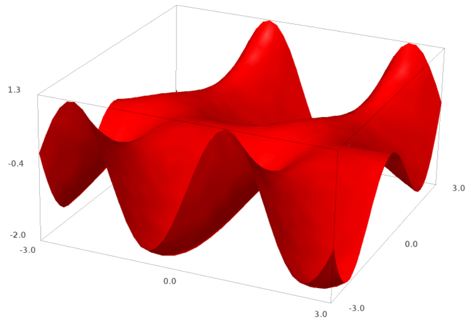
We compare this with Magma (V2.17-4), which has a more ad hoc coercion system:

[8] For example, `http://science.slashdot.org/story/07/12/08/1350258/Open-Source-Sage-Takes-Aim-at-High-End-Math-Software`

```
> R<x> := PolynomialRing(IntegerRing());
> x + 1/2
    ^
Runtime error in '+': Bad argument types
Argument types given: RngUPolElt[RngInt], FldRatElt
```

Robert Bradshaw and I also added support for beautiful browser-based 3D graphics to Sage, which involved writing a 3D graphics library, and adapting the free open source JMOL Java library (see `http://jmol.sourceforge.net/`) for rendering molecules to instead plot mathematical objects.

```
sage: f(x,y) = sin(x - y) * y * cos(x)
sage: plot3d(f, (x,-3,3), (y,-3,3), color='red')
```



In 2009, following a huge amount of porting work by Mike Hansen, development of algebraic combinatorics in Sage picked up substantial momentum, with the switch of the entire MuPAD-combinat group to Sage (forming sage-combinat `http://wiki.sagemath.org/combinat`), only months before the formerly free system MuPAD®[9] was bought out by Mathworks (makers of MATLAB). In addition to work on Lie theory by Dan Bump, this also led to a massive amount of work on a category theoretic framework for Sage by Nicolas Thiery.

In 2010, there were 13 Sage Days workshops in many parts of the world, and grant funding for Sage significantly improved, including new NSF funding for undergraduate curriculum development. I also spent much of my programming time during 2010–2011 developing a number theory library called psage `http://code.google.com/p/purplesage/`, which is currently not included in Sage, but can be easily installed.

[9] MuPAD is a registered trademark of SciFace Software GmbH & Co.

Many aspects of Sage make it an ideal tool for teaching mathematics, so there's a steadily growing group of teachers using it: for example, there have been MAA PREP workshops on Sage for the last two years, and a third is likely to run next summer, there are regular posts on the Sage lists about setting up classroom servers, and there is an NSF-funded project called UTMOST (see `http://utmost.aimath.org/`) devoted to creating undergraduate curriculum materials for Sage.

The page `http://sagemath.org/library-publications.html` lists 101 accepted publications that use Sage, 47 preprints, 22 theses, and 16 books, and there are surely many more "in the wild" that we are not aware of. According to Google Analytics, the main Sage website gets about 2,500 absolute unique visitors per day, and the website `http://sagenb.org`, which allows anybody to easily use Sage through their web browser, has around 700 absolute unique visitors per day.

For many mathematicians and students, Sage is today the mature, open source, and free foundation on which they can build their research program.

44 Numerical computation of Chow-Heegner points associated to pairs of elliptic curves

Numerical computation of Chow-Heegner points associated to pairs of elliptic curves[*][†]

William Stein

February 5, 2012

Abstract

In this paper, we consider a special case of Chow-Heegner points that has a simple concrete description due to Shouwu Zhang. Given a pair E, F of nonisogenous elliptic curves, and surjective morphisms $\varphi_E : X_0(N) \to E$ and $\varphi_F : X_0(N) \to F$ of curves over $\mathbb{Q}$, we associate a rational point $P \in E(\mathbb{Q})$. We describe a numerical approach to computing P, state some motivating results of Zhang et al. about the height of P, and present a table of data.

1 Introduction: Zhang's Construction

Consider a pair E, F of nonisogenous elliptic curves over $\mathbb{Q}$ and fix surjective morphisms from $X_0(N)$ to each curve. We do *not* assume that N is the conductor of either E or F, though N is necessarily a multiple of the conductor.

$$\begin{tikzcd} & X_0(N) \arrow[dl, "\varphi_E"'] \arrow[dr, "\varphi_F"] & \\ E & & F \end{tikzcd}$$

Let $(\varphi_E)_* : \mathrm{Div}(X_0(N)) \to \mathrm{Div}(E)$ and $\varphi_F^* : \mathrm{Div}(F) \to \mathrm{Div}(X_0(N))$ be the pushforward and pullback maps on divisors on algebraic curves. Let $Q \in F(\mathbb{C})$ be any point, and let

$$P_{\varphi_E,\varphi_F,Q} = \sum (\varphi_E)_* \varphi_F^*(Q) \in E(\mathbb{C}),$$

where $\sum$ means the sum of the points in the divisor using the group law on E, i.e., given a divisor $D = \sum n_i P_i \in \mathrm{Div}(E)$, we have $(\sum D) - \infty \sim D - \deg(D)\infty$, which uniquely determines $\sum D$.

[*] Work on this paper was partly supported by National Science Foundation Grant No. DMS-0821725.

[†] A modified version of this paper will be published as an appendix to [DDLR11].

Proposition 1.1. *The point $P_{\varphi_E,\varphi_F,Q}$ does not depend on the choice of Q.*

Proof. The composition $(\varphi_E)_* \circ \varphi_F^*$ induces a homomorphism of elliptic curves

$$\psi : \mathrm{Pic}^0(F) = \mathrm{Jac}(F) \to \mathrm{Jac}(E) = \mathrm{Pic}^0(E).$$

Our hypothesis that E and F are nonisogenous implies that $\psi = 0$. We denote by $[D]$ the linear equivalence class of a divisor in the Picard group. If $Q' \in F(\mathbb{C})$ is another point, then under the above composition of maps,

$$[Q - Q'] \mapsto [(\varphi_E)_*\varphi_F^*(Q) - (\varphi_E)_*\varphi_F^*(Q')] = [P_Q - P_{Q'}].$$

Thus the divisor $P_Q - P_{Q'}$ is linearly equivalent to 0. But F has genus 1, so there is no rational function on F of degree 1, hence $P_Q = P_{Q'}$, as claimed. □

We let $P_{\varphi_E,\varphi_F} = P_{\varphi_E,\varphi_F,Q} \in E(\mathbb{C})$, for any choice of Q.

Corollary 1.2. *We have $P_{\varphi_E,\varphi_F} \in E(\mathbb{Q})$.*

Proof. Taking $Q = \mathcal{O} \in F(\mathbb{Q})$, we see that the divisor $(\varphi_E)_* \circ \varphi_F^*(Q)$ is rational, so its sum is also rational. □

In the rest of this paper, we write $P_{E,F} = P_{\varphi_E,\varphi_F}$ when E and F are both optimal curves of the same conductor N, and φ_E and φ_F are as in Section 5.

1.1 Outline

In Section 2 we discuss an example in which E and F both have conductor 37. Section 3 is about a formula of Yuan-Zhang-Zhang for the height of $P_{E,F}$ in terms of the derivative of an L-function, in some cases. In Section 4, we discuss the connection between this paper and the paper [DDLR11] about computing Chow-Heegner points using iterated integrals. The heart of the paper is Section 5, which describes our numerical approach to approximating $P_{E,F}$. Finally, Section 5.2 presents a table of points $P_{E,F}$.

Acknowledgements: We would like to thank Ralph Greenberg, Ken Ribet, Barry Mazur, Karl Rubin, Shouwu Zhang, and Jon Bober for helpful conversations related to this paper, and especially thank Xinyi Yuan for introducing us to this topic, Henri Darmon for encouraging us to flesh out the details and write it up for publication, and Victor Rotger for clarifying some issues in Section 3. We would also like to thank John Cremona for refereeing the corresponding Sage [S$^+$11] code and Cremona, Bas Edixhoven and Samit Dasgupta for contributions to Section 5.1.

2 Example: $N = 37$

The smallest conductor curve of rank 1 is the curve E with Cremona label 37a (see [Cre]). The paper [MSD74] discusses the modular curve $X_0(37)$ in detail.

It gives the affine equation $y^2 = -x^6 - 9x^4 - 11x^2 + 37$ for $X_0(37)$, and describes how $X_0(37)$ is equipped with three independent involutions w, T and S. The quotient of $X_0(37)$ by w is E, the quotient by T is an elliptic curve F with $F(\mathbb{Q}) \approx \mathbb{Z}/3\mathbb{Z}$ and Cremona label 37b, and the quotient by S is the projective line $\mathbb{P}^1$.

$$\begin{tikzcd}
 & X_0(37) \arrow[dl, "\varphi_E"'] \arrow[d, "\varphi_F"] \arrow[dr] & \\
E = X/w & F = X/T & \mathbb{P}^1 = X/S
\end{tikzcd}$$

The maps φ_E and φ_F have degree 2, by virtue of being induced by an involution. As explained in [MSD74], the fiber over $Q = 0 \in F(\mathbb{Q})$ contains 2 points:

1. the cusp $[\infty] \in X_0(37)(\mathbb{Q})$, and
2. the noncuspidal affine rational point $(-1, -4) = T(\infty) \in X_0(37)(\mathbb{Q})$.

We have $\varphi_E([\infty]) = 0 \in E(\mathbb{Q})$, and [MSD74, Prop. 3, pg. 30] implies that

$$\varphi_F((-1, -4)) = (6, 14) = -6(0, -1),$$

where $(0, -1)$ generates $E(\mathbb{Q})$. We conclude that

$$P_{E,F} = (6, 14) \qquad \text{and} \qquad [E(\mathbb{Q}) : \mathbb{Z}P_{E,F}] = 6.$$

On [MSD74, pg. 31], they remark: "It would be of the utmost interest to link this index to something else in the theory."

This remark motivates our desire to compute more examples. Unfortunately, it is very difficult to generalize the above approach directly, since it involves computations with $X_0(37)$ and its quotients that rely on explicit defining equations. Just as there are multiple approaches to computing Heegner points, there are several approaches to computing $P_{E,F}$:

- a Gross-Zagier style formula for the height of $P_{E,F}$ (see Section 3),
- explicit evaluation of iterated integrals (see Section 4), and
- numerical approximation of the fiber in the upper half plane over a point on F using a polynomial approximation to φ_F (see Section 5).

This paper is mainly about the last approach listed above.

3 The Formula of Yuan-Zhang-Zhang

Consider a special case of the triple product L-function of [GK92]

$$L(E, F, F, s) = L(E, s) \cdot L(E, \mathrm{Sym}^2(F), s), \tag{1}$$

where E and F are elliptic curves of the same conductor N, and all L-functions are normalized so that $1/2$ is the center of the critical strip. The following theorem is proved in [YZZ11]:

Theorem 3.1 (Yuan-Zhang-Zhang). *Assume that the local root number of $L(E, F, F, s)$ at every prime of bad reduction is $+1$ and that the root number at infinity is -1. Then $\hat{h}(P_{E,F}) = (*) \cdot L'(E, F, F, \frac{1}{2})$, where $(*)$ is nonzero.*

Remark 3.2. The above formula resembles the Gross-Zagier formula

$$\hat{h}(P_K) = (*) \cdot (L(E/\mathbb{Q}, s) \cdot L(E^K/\mathbb{Q}, s))'|_{s=\frac{1}{2}},$$

where K is a quadratic imaginary field satisfying certain hypotheses.

If one could evaluate $L'(E, F, F, \frac{1}{2})$, e.g., by applying the algorithm of [Dok04], along with the factor $(*)$ in the theorem, this would yield an algorithm to compute $\pm P_{E,F} \pmod{E(\mathbb{Q})_{\mathrm{tor}}}$ when the root number hypothesis is satisfied. Unfortunately, it appears that nobody has numerically evaluated the formula of Theorem 3.1 in any interesting cases.

When E and F have the same squarefree conductor N, [GK92, §1] implies that the local root number of $L(E, F, F, s)$ at p is the same as the local root number of E at p; computing the local root number when the level is not square free is more complicated.

Proposition 3.3. *Assume that E and F have the same squarefree conductor N, that the local root numbers of E at primes $p \mid N$ are all $+1$ (equivalently, that we have $a_p(E) = -1$) and that $r_{\mathrm{an}}(E/\mathbb{Q}) = 1$. Then $L(E, \mathrm{Sym}^2 F, \frac{1}{2}) \neq 0$ if and only if $\hat{h}(P_{E,F}) \neq 0$.*

Proof. By hypothesis, we have $L(E, \frac{1}{2}) = 0$ and $L'(E, \frac{1}{2}) \neq 0$. Theorem 3.1 and the factorization (1) imply that

$$\hat{h}(P_{E,F}) = (*) \cdot L'(E, \frac{1}{2}) \cdot L(E, \mathrm{Sym}^2 F, \frac{1}{2}),$$

from which the result follows. □

Section 5.2 contains numerous examples in which E has rank 1, F has rank 0, and yet $P_{E,F}$ is a torsion point. The first example is when E is 91b and F is 91a. Then $P_{E,F} = (1, 0)$ is a torsion point (of order 3). In this case, we cannot apply Proposition 3.3 since $\varepsilon_7 = \varepsilon_{13} = -1$ for E. Another example is when E is 99a and F is 99c, where we have $P_{E,F} = 0$, and $\varepsilon_3 = \varepsilon_{11} = +1$, but Proposition 3.3 does not apply since the level is not square free. Fortunately, we found an example with squarefree level $158 = 2 \cdot 79$: here E is 158b, F is 158d, we have $P_{E,F} = 0$ and $\varepsilon_2 = \varepsilon_{79} = +1$, so Proposition 3.3 implies that $L(E, \mathrm{Sym}^2 F, \frac{1}{2}) = 0$.

4 Iterated Complex Path Integrals

The paper [DDLR11] contains a general approach using iterated path integrals to compute certain Chow-Heegner points, of which $P_{E,F}$ is a specific instance. Comparing our data (Section 5.2) with theirs, we find that if E and F are optimal elliptic curves over $\mathbb{Q}$ of the same conductor $N \leq 100$, if $e, f \in S_2(\Gamma_0(N))$

are the corresponding newforms, and if $P_{f,e,1} \in E(\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}$ the associated Chow-Heegner point in the sense of [DDLR11], then $2P_{E,F} = P_{f,e,1}$. This is (presumably) a consequence of [DRS11].

5 A Numerical Approach to Computing $P_{E,F}$

The numerical approach to computing P that we describe in this section uses relatively little abstract theory. It is inspired by work of Delaunay (see [Del02]) on computing the fiber of the map $X_0(389) \to E$ over rational points on the rank 2 curve E of conductor 389. We make no guarantee about how many digits of our approximation to $P_{E,F}$ are correct, instead viewing this as an algorithm to produce something that is useful for experimental mathematics only.

Let $\mathfrak{h}$ be the upper half plane, and let $Y_0(N) = \Gamma_0(N)\backslash\mathfrak{h} \subset X_0(N)$ be the affine modular curve. Let E and F be nonisogenous optimal elliptic curve quotients of $X_0(N)$, with modular parametrization maps φ_E and φ_F, and assume both Manin constants are 1. Let Λ_E and Λ_F be the period lattices of E and F, so $E \cong \mathbb{C}/\Lambda_E$ and $F \cong \mathbb{C}/\Lambda_F$. Viewed as a map $[\tau] \mapsto \mathbb{C}/\Lambda_E$, we have, using square brackets to denote equivalence classes, that

$$\varphi_E([\tau]) = \left[\sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n\tau}\right],$$

and similarly for φ_F, where $a_n = a_n(E)$ are the L-series coefficients of E (see [Cre97, §2.10], which uses the oppositive sign convention). For any positive integer B, define the polynomial

$$\varphi_{E,B} = \sum_{n=1}^{B} \frac{a_n}{n} T^n \in \mathbb{Q}[T],$$

and similarly for $\varphi_{F,B}$.

To approximate $P_{E,F}$, we proceed as follows. First we make some choices, and after making these choices we run the algorithm, which will either find a "probable" numerical approximation to $P_{E,F}$ or fail.

- $y \in \mathbb{R}_{>0}$ – minimum imaginary part of points in fiber,
- $d \in \mathbb{Z}_{>0}$ – degree of the first approximation to φ_F in Step 1 below,
- $r \in \mathbb{R}_{\neq 0}$ – real number specified to b bits of precision that defines $Q \in \mathbb{C}/\Lambda$,
- b' – bits of precision when dividing points into $\Gamma_0(N)$ orbits, and
- n – number of trials before we give up and output FAIL.

We compute $P_{E,F,Q}$ using an approach that will always fail if Q is a ramification point. Our algorithm will also fail if any points in the fiber over Q are cusps. This is why we do not allow $r = 0$. One can modify the algorithm to work when Q is an unramified torsion point by using modular symbols and keeping track of images of cusps.

To increase our confidence that we have computed the right point $P_{E,F}$, we often carry out the complete computation with more than one choice of r.

1. **Low precision roots:** Compute all complex double precision roots of the polynomial $\varphi_{F,d} - r$. One way to do this is to use "balanced QR reduction of the companion matrix", as implemented in GSL.[1] Record the roots that correspond to $\tau \in \mathfrak{h}$ with $\mathrm{Im}(\tau) \geq y$.

2. **High precision roots:** Compute B such that if $\mathrm{Im}(\tau) \geq y$, then

$$\left| \sum_{n=B+1}^{\infty} \frac{a_n(F)}{n} \tau^n \right| < 2^{-b},$$

   where b is the number of bits of precision of r. Summing the tail end of the series and using that $|a_n| \leq n$ (see [GJP+09, Lem. 2.9]), we find that

$$B = \left\lceil \frac{\log(2^{-(b+1)} \cdot (1 - e^{-2\pi y_1}))}{-2\pi y} \right\rceil$$

   works. Next, compute the polynomial $\varphi_{F,B} \in \mathbb{Q}[T]$, and use Newton iteration to refine all roots saved in Step 1 to roots α of $f = \varphi_{F,B} - r \in \mathbb{R}[T]$ such that $|f(\alpha)| < 2^{-b}$. Save those roots that correspond to $\tau \in \mathfrak{h}$ with $\mathrm{Im}(\tau) \geq y$.

3. $\Gamma_0(N)$**-orbits:** Divide the τ's from Step 2 into $\Gamma_0(N)$-equivalence classes, testing equivalence to the chosen bit precision $b' \leq b$, as explained in Section 5.1. It is easy to efficiently compute the modular degree $m_F = \deg(\varphi_F)$ (see [Wat02]). If we find m_F distinct $\Gamma_0(N)$ classes of points, we suspect that we have found the fiber over $[r]$, so we map each element of the fiber to E using φ_E and sum, then apply the elliptic exponential to obtain $P_{E,F}$ to some precision, then output this approximation and terminate. If we find more than m_F distinct classes, there was an error in the choices of precision in our computation, so we output FAIL (and suggest either increasing b or decreasing b').

4. **Try again:** We did not find enough points in the fiber. Systematically replace r by $r + m\Omega_F$, for $m = 1, -1, 2, -2, \ldots$, where Ω_F is the least real period of F, then try again going to Step 1 and including the new points found. If upon trying n choices $r + m\Omega_F$ in a row we find no new points, we output FAIL and terminate the algorithm.

5.1 Determining $\Gamma_0(N)$ equivalency

The field of meromorphic functions invariant under $\Gamma_0(N)$ is generated by $j(z)$ and $j(Nz)$, so if two points z_1 and z_2 in the upper half plane are equivalent under $\Gamma_0(N)$, then z_1 and z_2 are equivalent under $\mathrm{SL}_2(\mathbb{Z})$ and Nz_1 and Nz_2 are also equivalent under $\mathrm{SL}_2(\mathbb{Z})$. Because of singularities in the affine curve defined

[1] GSL is the the GNU scientific library, which is part of Sage [S+11]. Rough timings of GSL for this computation: it takes less than a half second for degree 500, about 5 seconds for degree 1000, about 45 seconds for degree 2000, and several minutes for degree 3000.

by $j(z)$ and $j(Nz)$, the converse is *not* true: for example, $z_1 = (-2+i)/5$ and $z_2 = (2+i)/5$ are equivalent under $\mathrm{SL}_2(\mathbb{Z})$ as are $5z_1$ and $5z_2$, but z_1 and z_2 are not equivalent under $\Gamma_0(5)$. This is why the algorithm we give below must take into account singularities.

Suppose we are given arbitrary z_1 and z_2 in the upper half plane. We first find $g_1, g_2 \in \mathrm{SL}_2(\mathbb{Z})$ such that $w_i = g_i(z_i)$ is the canonical representative for z_i in the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, as explained in [Cre97, §2.14] but using interval arithmetic to avoid rounding errors. If $w_1 \neq w_2$, then z_1 and z_2 are not equivalent under $\mathrm{SL}_2(\mathbb{Z})$, so they cannot be equivalent under $\Gamma_0(N)$. Thus let $w = g_1(z_1) = g_2(z_2)$. The elements of $\mathrm{PSL}_2(\mathbb{Z})$ that send z_1 to z_2 are the finitely many elements $g_2^{-1}Ag_1$, for $A \in \mathrm{Stab}(w)$, so we check whether any $g_2^{-1}Ag_1$ is in $\Gamma_0(N)$. The only elements of the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ with nontrivial stabilizers are $w = i$, with stabilizer generated by $S \in \mathrm{PSL}_2(\mathbb{Z})$ of order 2, and $w = e^{2\pi i/3}$ with stabilizer generated by ST, where T corresponds to $z \mapsto z+1$.

5.2 Data

We implemented the above algorithm in Sage [S$^+$11][2]. The columns of the tables below are as follows. The columns labeled E and F contain Cremona labels for elliptic curves, and those labeled r_E and r_F contain the corresponding ranks. The column labeled $E(\mathbb{Q})$ gives a choice of generators $P_1, P_2, \ldots$ for the Mordell-Weil group, with r_E points of infinite order listed first, then 0, 1 or 2 torsion points listed with a subscript of their order. The column labeled $P_{E,F}$ contains a rational point close to the numerically computed Chow-Heegner point, represented in terms of the generators P_i from the column labeled $E(\mathbb{Q})$, where P_1 is the first generator, P_2 the second, and so on. The columns labeled m_E and m_F give the modular degrees of E and F. The column labeled ε's contains the local root numbers of $L(E,s)$ at each bad prime. The notes column refers to the notes after the table, which give information about the input parameters needed to compute $P_{E,F}$.

We believe that the values of $P_{E,F}$ are "likely" to be correct, but we emphasize again that *they are not proven correct.* In the table we give an exact point, but the algorithm computes a numerical approximation $\tilde{P}_{E,F}$ to $P_{E,F} \in E(\mathbb{Q})$. We find what we call $P_{E,F}$ in the table by running through several hundred low height points in $E(\mathbb{Q})$ and find the one closest to $\tilde{P}_{E,F}$; in all cases, the coordinates of the point we list are within 10^{-5} of the coordinates of $\tilde{P}_{E,F}$.

The table contains *every* pair E, F of nonisogenous optimal elliptic curves of the same conductor $N \leq 184$ with $r_E = 1$, and *most* (but not all) with $N \leq 250$. It also contains a few additional miscellaneous examples, e.g., with $r_E = 0$ and some of larger conductor with $r_F = 2$. Most rows took only a few seconds to compute, though ones with m_F large in some cases took much longer; the total CPU time to compute the entire table was about 8 hours. Unless otherwise noted, we used $y = 10^{-4}$, $d = 500$, $b' = 20$, and $r = 0.1$ with 53 bits of precision,

[2] See `http://trac.sagemath.org/sage_trac/ticket/11975`.

as in Section 5. We also repeated all computations with at least one additional value of $r \neq 0.1$, and in every case got the same result (usually we used $r = 0.2$).

5.3 Discussion

In the table we always have $2 \mid [E(\mathbb{Q})_{/\,\mathrm{tor}} : \mathbb{Z}P_{E,F}]$. In may be possible to prove this in some cases by using that when $r_{\mathrm{an}}(E) = 1$ then the sign in the functional equation for $L(E,s)$ is -1, so at least one nontrivial Atkin-Lehner involution w_q acts as $+1$ on E, which means that the map $X_0(N) \to E$ factors through $X_0(N) \to X_0(N)/w_q$. Also, there are four cases in which the index $[E(\mathbb{Q})/_{\mathrm{tor}} : \mathbb{Z}P_{E,F}]$ is divisible by a prime $\ell \geq 5$. They are (106b, 106c, $\ell = 11$), (118a, 118d, $\ell = 7$), (121b, 121d, $\ell = 7$), and (158b, 158c, $\ell = 7$). These prime divisors do not appear to have anything to do with the invariants of E and F, individually.

E	ε_p's	r_E	$E(\mathbb{Q})$	m_E	F	r_F	m_F	$P_{E,F}$	Notes
37a	$+$	1	$(0,-1)$	2	37b	0	2	$-6P_1$	
37b	$-$	0	$(8,18)_3$	2	37a	1	2	P_1	
57a	$++$	1	$(2,1)$	4	57c	0	12	$8P_1$	
57a	$++$	1	$(2,1)$	4	57b	0	3	$-8P_1$	
57b	$-+$	0	$(7/4,-11/8)_2,(1,-1)_2$	3	57a	1	4	0	
57b	$-+$	0	$(7/4,-11/8)_2,(1,-1)_2$	3	57c	0	12	0	
57c	$-+$	0	$(2,4)_5$	12	57a	1	4	$3P_1$	
57c	$-+$	0	$(2,4)_5$	12	57b	0	3	P_1	
58a	$++$	1	$(0,-1)$	4	58b	0	4	$8P_1$	
58b	$-+$	0	$(-1,2)_5$	4	58a	1	4	$3P_1$	
77a	$++$	1	$(2,3)$	4	77b	0	20	$24P_1$	(1)
77a	$++$	1	$(2,3)$	4	77c	0	6	$-4P_1$	
89a	$+$	1	$(0,-1)$	2	89b	0	5	$4P_1$	
91a	$++$	1	$(0,0)$	4	91b	1	4	$4P_1$	
91b	$--$	1	$(-1,3),(1,0)_3$	4	91a	1	4	P_2	
92b	$--$	1	$(1,1)$	6	92a	0	2	0	
99a	$++$	1	$(2,0),(-1,0)_2$	4	99b	0	12	$-4P_1$	
99a	$++$	1	$(2,0),(-1,0)_2$	4	99c	0	12	0	
99a	$++$	1	$(2,0),(-1,0)_2$	4	99d	0	6	$2P_1$	
102a	$+++$	1	$(2,-4),(0,0)_2$	8	102b	0	16	$-8P_1$	(1)
102a	$+++$	1	$(2,-4),(0,0)_2$	8	102c	0	24	$32P_1$	
106b	$++$	1	$(2,1)$	8	106a	0	6	$-4P_1$	
106b	$++$	1	$(2,1)$	8	106c	0	48	$-88P_1$	
106b	$++$	1	$(2,1)$	8	106d	0	10	$12P_1$	
112a	$++$	1	$(0,-2),(-2,0)_2$	8	112b	0	4	0	
112a	$++$	1	$(0,-2),(-2,0)_2$	8	112c	0	8	0	
118a	$++$	1	$(0,-1)$	4	118b	0	12	$-8P_1$	(1)
118a	$++$	1	$(0,-1)$	4	118c	0	6	$4P_1$	
118a	$++$	1	$(0,-1)$	4	118d	0	38	$-28P_1$	
121b	$+$	1	$(4,5)$	4	121a	0	6	$4P_1$	
121b	$+$	1	$(4,5)$	4	121c	0	6	$4P_1$	
121b	$+$	1	$(4,5)$	4	121d	0	24	$-28P_1$	(2)
123a	$--$	1	$(-4,1),(-1,4)_5$	20	123b	1	4	0	
123b	$++$	1	$(1,0)$	4	123a	1	20	$4P_1$	
124a	$--$	1	$(-2,1),(0,1)_3$	6	124b	0	6	0	
128a	$+$	1	$(0,1),(-1,0)_2$	4	128b	0	8	0	
128a	$+$	1	$(0,1),(-1,0)_2$	4	128c	0	4	0	
128a	$+$	1	$(0,1),(-1,0)_2$	4	128d	0	8	0	
129a	$++$	1	$(1,-5)$	8	129b	0	15	$-8P_1$	
130a	$+--$	1	$(-6,10),(-1,10)_6$	24	130b	0	8	0	
130a	$+--$	1	$(-6,10),(-1,10)_6$	24	130c	0	80	0	
135a	$++$	1	$(4,-8)$	12	135b	0	36	0	(1)
136a	$--$	1	$(-2,2),(0,0)_2$	8	136b	0	8	0	
138a	$+++$	1	$(1,-2),(-2,1)_2$	8	138b	0	16	$-16P_1$	(1)
138a	$+++$	1	$(1,-2),(-2,1)_2$	8	138c	0	8	$-8P_1$	
141a	$--$	1	$(-3,-5)$	28	141b	0	12	0	
141a	$--$	1	$(-3,-5)$	28	141c	0	6	0	
141a	$--$	1	$(-3,-5)$	28	141d	1	4	0	

E	ε_p's	r_E	$E(\mathbb{Q})$	m_E	F	r_F	m_F	$P_{E,F}$	Notes
141a	- -	1	$(-3,-5)$	28	141e	0	12	0	
141d	$++$	1	$(0,-1)$	4	141a	1	28	$-12P_1$	
141d	$++$	1	$(0,-1)$	4	141b	0	12	$4P_1$	
141d	$++$	1	$(0,-1)$	4	141c	0	6	$4P_1$	
141d	$++$	1	$(0,-1)$	4	141e	0	12	$4P_1$	
142a	- -	1	$(1,1)$	36	142b	1	4	0	
142a	- -	1	$(1,1)$	36	142c	0	9	0	
142a	$--$	1	$(1,1)$	36	142d	0	4	0	
142a	$--$	1	$(1,1)$	36	142e	0	324	0	(2)
142b	$++$	1	$(-1,0)$	4	142a	1	36	$4P_1$	(1)
142b	$++$	1	$(-1,0)$	4	142c	0	9	$-4P_1$	
142b	$++$	1	$(-1,0)$	4	142d	0	4	$4P_1$	
142b	$++$	1	$(-1,0)$	4	142e	0	324	$8P_1$	(2)
152a	$++$	1	$(-1,-2)$	8	152b	0	8	0	
153a	$++$	1	$(0,1)$	8	153b	1	16	$8P_1$	
153a	$++$	1	$(0,1)$	8	153c	0	8	$8P_1$	
153a	$++$	1	$(0,1)$	8	153d	0	24	0	
153b	$--$	1	$(5,-14)$	16	153a	1	8	0	
153b	$--$	1	$(5,-14)$	16	153d	0	24	0	
154a	$+++$	1	$(5,3),(-6,3)_2$	24	154b	0	24	$-24P_1$	
154a	$+++$	1	$(5,3),(-6,3)_2$	24	154c	0	16	$16P_1$	
155a	$--$	1	$(5/4,31/8),(0,2)_5$	20	155b	0	8	0	
155a	$--$	1	$(5/4,31/8),(0,2)_5$	20	155c	1	4	0	
155c	$++$	1	$(1,-1)$	4	155a	1	20	$-12P_1$	
155c	$++$	1	$(1,-1)$	4	155b	0	8	$4P_1$	
156a	$-+-$	1	$(1,1),(2,0)_2$	12	156b	0	12	0	(1)
158a	$--$	1	$(-1,-4)$	32	158b	1	8	0	
158a	$--$	1	$(-1,-4)$	32	158c	0	48	0	(1)
158a	$--$	1	$(-1,-4)$	32	158d	0	40	0	
158a	$--$	1	$(-1,-4)$	32	158e	0	6	0	
158b	$++$	1	$(0,-1)$	8	158a	1	32	$-8P_1$	
158b	$++$	1	$(0,-1)$	8	158c	0	48	$-56P_1$	(1)
158b	$++$	1	$(0,-1)$	8	158d	0	40	0	
158b	$++$	1	$(0,-1)$	8	158e	0	6	$-8P_1$	
160a	$++$	1	$(2,-2),(1,0)_2$	8	160b	0	8	0	
162a	$++$	1	$(-2,4),(1,1)_3$	12	162b	0	6	0	
162a	$++$	1	$(-2,4),(1,1)_3$	12	162c	0	6	0	
162a	$++$	1	$(-2,4),(1,1)_3$	12	162d	0	12	0	
170a	$+--$	1	$(0,2),(1,-1)_2$	16	170d	0	12	0	
170a	$+--$	1	$(0,2),(1,-1)_2$	16	170e	0	20	0	
171b	$--$	1	$(2,-5)$	8	171a	0	12	0	
171b	$--$	1	$(2,-5)$	8	171c	0	96	0	(1)
171b	$--$	1	$(2,-5)$	8	171d	0	32	0	
175a	$--$	1	$(2,-3)$	8	175b	1	16	0	(1)
175a	$--$	1	$(2,-3)$	8	175c	0	40	0	(1)
175b	$++$	1	$(-3,12)$	16	175a	1	8	$16P_1$	
175b	$++$	1	$(-3,12)$	16	175c	0	40	$16P_1$	(1)
176c	$--$	1	$(1,-2)$	8	176b	0	8	0	(1)

E	ε_p's	r_E	$E(\mathbb{Q})$	m_E	F	r_F	m_F	$P_{E,F}$	Notes
176c	$--$	1	$(1,-2)$	8	176a	0	16	0	
176c	$--$	1	$(1,-2)$	8	176b	0	8	0	(1)
184a	$--$	1	$(0,1)$	8	184c	0	12	0	
184a	$--$	1	$(0,1)$	8	184d	0	24	0	
184b	$++$	1	$(2,-1)$	8	184a	1	8	0	
184b	$++$	1	$(2,-1)$	8	184c	0	12	0	
184b	$++$	1	$(2,-1)$	8	184d	0	24	0	
185a	$++$	1	$(4,-13)$	48	185b	1	8	$8P_1$	
185a	$++$	1	$(4,-13)$	48	185c	1	6	$24P_1$	
185b	$--$	1	$(0,2)$	8	185c	1	6	0	
185c	$++$	1	$(-5/4,3/8),(-1,0)_2$	6	185b	1	8	$2P_1$	
189a	$++$	1	$(-1,-2)$	12	189b	1	12	$-12P_1$	
189a	$++$	1	$(-1,-2)$	12	189c	0	12	$12P_1$	
189b	$--$	1	$(-3,9),(3,0)_3$	12	189a	1	12	0	
189b	$--$	1	$(-3,9),(3,0)_3$	12	189c	0	12	0	
190a	$-+-$	1	$(13,-47)$	88	190b	1	8	0	
190a	$-+-$	1	$(13,-47)$	88	190c	0	24	0	(1)
190b	$+++$	1	$(1,2)$	8	190c	0	24	$16P_1$	(1)
192a	$++$	1	$(3,2),(-1,0)_2$	8	192b	0	8	0	
192a	$++$	1	$(3,2),(-1,0)_2$	8	192c	0	8	0	
192a	$++$	1	$(3,2),(-1,0)_2$	8	192d	0	8	0	
196a	$--$	1	$(0,-1)$	6	196b	0	42	0	(1)
198a	$+--$	1	$(-1,-4),(-4,2)_2$	32	198b	0	32	0	(1)
198a	$+--$	1	$(-1,-4),(-4,2)_2$	32	198c	0	32	0	
198a	$+--$	1	$(-1,-4),(-4,2)_2$	32	198d	0	32	0	(1)
198a	$+--$	1	$(-1,-4),(-4,2)_2$	32	198e	0	160	0	(1)
200b	$--$	1	$(-1,1),(-2,0)_2$	8	200c	0	24	0	
200b	$--$	1	$(-1,1),(-2,0)_2$	8	200d	0	40	0	(1)
200b	$--$	1	$(-1,1),(-2,0)_2$	8	200e	0	24	0	
201a	$++$	1	$(1,-2)$	12	201b	1	12	$4P_1$	
201b	$--$	1	$(-1,2)$	12	201a	1	12	0	
201c	$++$	1	$(16,-7)$	60	201a	1	12	$-24P_1$	
201c	$++$	1	$(16,-7)$	60	201b	1	12	$8P_1$	
203b	$--$	1	$(2,-5)$	8	203a	0	48	0	
203b	$--$	1	$(2,-5)$	8	203c	0	12	0	
205a	$--$	1	$(-1,8),(2,1)_4$	12	205b	0	16	0	
205a	$--$	1	$(-1,8),(2,1)_4$	12	205c	0	8	0	
208a	$--$	1	$(4,-8)$	16	208c	0	12	0	
208a	$--$	1	$(4,-8)$	16	208d	0	48	0	(1)
208b	$++$	1	$(4,4)$	16	208a	1	16	0	(1)
208b	$++$	1	$(4,4)$	16	208c	0	12	0	
208b	$++$	1	$(4,4)$	16	208d	0	48	0	(1)
212a	$--$	1	$(2,2)$	12	212b	0	21	0	
214a	$--$	1	$(0,-4)$	28	214b	1	12	0	(1)
214a	$--$	1	$(0,-4)$	28	214d	0	12	0	
214b	$++$	1	$(0,0)$	12	214a	1	28	$-8P_1$	(1)
214b	$++$	1	$(0,0)$	12	214d	0	12	$-4P_1$	

E	ε_p's	r_E	$E(\mathbb{Q})$	m_E	F	r_F	m_F	$P_{E,F}$	Notes
214c	$++$	1	$(11,10)$	60	214a	1	28	$-4P_1$	(1)
214c	$++$	1	$(11,10)$	60	214d	0	12	$16P_1$	
214c	$++$	1	$(11,10)$	60	214b	1	12	$12P_1$	(1)
216a	$++$	1	$(-2,-6)$	24	216b	0	24	0	
219a	$++$	1	$(2,-1)$	12	219c	1	60	$-12P_1$	(1)
219a	$++$	1	$(2,-1)$	12	219b	1	12	$-4P_1$	
216a	$++$	1	$(-2,-6)$	24	216d	0	72	0	
219b	$--$	1	$(-3/4,-1/8),(0,1)_3$	12	219a	1	12	0	
219b	$--$	1	$(-3/4,-1/8),(0,1)_3$	12	219c	1	60	0	(1)
219c	$++$	1	$(-6,7),(10,-5)_2$	60	219a	1	12	$-12P_1$	
219c	$++$	1	$(-6,7),(10,-5)_2$	60	219b	1	12	$4P_1$	
220a	$--+$	1	$(-7,11),(15,55)_6$	36	220b	0	12	0	
224a	$++$	1	$(1,2),(0,0)_2$	8	224b	0	8	0	
225a	$++$	1	$(1,1)$	8	225b	0	40	0	(1)
225e	$--$	1	$(-5,22)$	48	225a	1	8	0	(1)
225e	$--$	1	$(-5,22)$	48	225b	0	40	0	(1)
228b	$-+-$	1	$(3,6)$	24	228a	0	18	0	
232a	$++$	1	$(2,-4)$	16	232b	0	16	0	
234c	$+++$	1	$(1,-2),(-2,1)_2$	16	234b	0	48	0	(1)
234c	$+++$	1	$(1,-2),(-2,1)_2$	16	234e	0	20	0	(1)
235a	$--$	1	$(-2,3)$	12	235c	0	18	0	(1)
236a	$--$	1	$(1,-1)$	6	236b	0	14	0	
238a	$--+$	1	$(24,100),(-8,4)_2$	112	238b	1	8	0	(1)
238a	$--+$	1	$(24,100),(-8,4)_2$	112	238c	0	16	0	(1)
238a	$--+$	1	$(24,100),(-8,4)_2$	112	238d	0	16	0	(1)
238b	$+++$	1	$(1,1),(0,0)_2$	8	238a	1	112	$12P_1$	(1)
238b	$+++$	1	$(1,1),(0,0)_2$	8	238c	0	16	$-4P_1$	(1)
238b	$+++$	1	$(1,1),(0,0)_2$	8	238d	0	16	$4P_1$	(1)
240c	$+++$	1	$(1,2),(0,0)_2$	16	240a	0	16	0	
240c	$+++$	1	$(1,2),(0,0)_2$	16	240d	0	16	0	(1)
243a	$+$	1	$(1,0)$	6	243b	0	9	0	(1)
245a	$--$	1	$(7,17)$	48	245c	1	32	0	
246d	$+++$	1	$(3,-6),(4,-2)_2$	48	246a	0	84	$24P_1$	(1)
446a	$++$	1	$(4,-6)$	24	446d	2	88	0	(2)
446b	$--$	1	$(5,-10)$	56	446d	2	88	0	(2)
446d	$+-$	2	-	88	446a	1	12	0	(1)
446d	$+-$	2	-	88	446b	1	56	0	(1)
681a	$++$	1	$(4,4)$	32	681c	2	96	$-24P_1$	(2)

Notes:

(1) We used $y = 10^{-5}$ and $d = 1500$, which takes a few minutes.

(2) We used $y = \frac{1}{2} \cdot 10^{-5}$ and $d = 3000$, which takes over an hour.

References

[Cre] J. E. Cremona, *Elliptic Curves Data*, `http://www.warwick.ac.uk/~masgaj/ftp/data/`.

[Cre97] ———, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, `http://www.warwick.ac.uk/~masgaj/book/fulltext/`.

[DDLR11] Henri Darmon, Michael Daub, Sam Lichtenstein, and Victor Rotger, *The Effective Computation of Iterated Integrals and Chow-Heegner Points on Triple Products*, In Preparation (2011).

[Del02] Christophe Delaunay, *Formes modulaires et invariants de courbes elliptiques définies sur* **Q**, Thèse de Doctorat, Université Bordeaux I, available at `http://math.univ-lyon1.fr/~delaunay/`.

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149, `http://arxiv.org/abs/math/0207280`. MR 2068888 (2005f:11128)

[DRS11] Henri Darmon, Victor Rotger, and Ignacio Sols, *Iterated integrals, diagonal cycles and rational points on elliptic curves*, Preprint (2011), `http://www-ma2.upc.edu/vrotger/docs/DRS1.pdf`.

[GJP$^+$09] G. Grigorov, A. Jorza, S. Patrikis, C. Tarnita, and W. Stein, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. **78** (2009), 2397–2425, `http://wstein.org/papers/bsdalg/`.

[GK92] Benedict H. Gross and Stephen S. Kudla, *Heights and the central critical values of triple product L-functions*, Compositio Math. **81** (1992), no. 2, 143–209, `http://www.numdam.org.offcampus.lib.washington.edu/item?id=CM_1992__81_2_143_0`. MR 1145805 (93g:11047)

[MSD74] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61. MR 50 #7152

[S$^+$11] W. A. Stein et al., *Sage Mathematics Software (Version 4.8)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[Wat02] M. Watkins, *Computing the modular degree of an elliptic curve*, Experiment. Math. **11** (2002), no. 4, 487–502 (2003). MR 1 969 641

[YZZ11] X. Yuan, S. Zhang, and W. Zhang, *Triple product L-series and Gross-Schoen cycles I: split case*, Preprint (2011), `http://www.math.columbia.edu/~yxy/preprints/triple.pdf`.

45 A Database Of Elliptic Curves Over $\mathbf{Q}(\sqrt{5})$ – First Report, with J. Bober, A. Deines, A. Klages-Mundt, B. LeVeque, R. A. Ohana, A. Rabindranath, P. Sharaba

A DATABASE OF ELLIPTIC CURVES OVER $\mathbb{Q}(\sqrt{5})$—FIRST REPORT

JONATHAN BOBER, ALYSON DEINES, ARIAH KLAGES-MUNDT, BENJAMIN LEVEQUE, R. ANDREW OHANA, ASHWATH RABINDRANATH, PAUL SHARABA, WILLIAM STEIN

Abstract. We describe a tabulation of (conjecturally) modular elliptic curves over the field $\mathbb{Q}(\sqrt{5})$ up to the first curve of rank 2. Using an efficient implementation of an algorithm of Lassina Dembélé [Dem05], we computed tables of Hilbert modular forms of weight $(2, 2)$ over $\mathbb{Q}(\sqrt{5})$, and via a variety of methods we constructed corresponding elliptic curves, including (again, conjecturally) all elliptic curves over $\mathbb{Q}(\sqrt{5})$ that have conductor with norm less than or equal to 1831.

1. Introduction

1.1. **Elliptic Curves over $\mathbb{Q}$.** Tables of elliptic curves over $\mathbb{Q}$ have been of great value in mathematical research. Some of the first such tables were those in Antwerp IV [BK75], which included all elliptic curves over $\mathbb{Q}$ of conductor up to 200, and also a table of all elliptic curves with bad reduction only at 2 and 3.

Cremona's book [Cre97] gives a detailed description of algorithms that together output a list of all elliptic curves over $\mathbb{Q}$ of any given conductor, along with extensive data about each curve. The proof that his algorithm outputs *all* curves of given conductor had to wait for the proof of the full modularity theorem in [BCDT01]. Cremona has subsequently computed tables [Cre] of all elliptic curves over $\mathbb{Q}$ of conductor up to 220,000, including Mordell-Weil groups and other extensive data about each curve; he expects to soon reach his current target, conductor 234,446, which is the smallest known conductor of a rank 4 curve.

In a different direction, Stein-Watkins (see [SW02, BMSW07]) created a table of 136,832,795 elliptic curves over $\mathbb{Q}$ of conductor $\leq 10^8$, and a table of 11,378,911 elliptic curves over $\mathbb{Q}$ of prime conductor $\leq 10^{10}$. There are many curves of large discriminant missing from the Stein-Watkins tables, since these tables are made by enumerating curves with relatively small defining equations, and discarding those of large conductor, rather than systematically finding all curves of given conductor no matter how large the defining equation.

1.2. **Why $\mathbb{Q}(\sqrt{5})$?** Like $\mathbb{Q}$, the field $F = \mathbb{Q}(\sqrt{5})$ is a totally real field, and many of the theorems and ideas about elliptic curves over $\mathbb{Q}$ have been generalized to totally real fields. As is the case over $\mathbb{Q}$, there is a notion of modularity of elliptic curves over F, and work of Zhang [Zha01] has extended many results of Gross-Zagier [GZ86] and Kolyvagin [Kol91] to the context of elliptic curves over totally real fields.

This work is supported by NSF grant DMS-0757627, administered by the American Institute of Mathematics.

If we order totally real number fields K by the absolute value of their discriminant, then $F = \mathbb{Q}(\sqrt{5})$ comes next after $\mathbb{Q}$ (the Minkowski bound implies that $|D_K| \geq (n^n/n!)^2$, where $n = [K : \mathbb{Q}]$, so if $n \geq 3$ then $|D_K| > 20$). That 5 divides $\mathrm{disc}(F) = 5$ thwarts attempts to easily generalize the method of Taylor-Wiles to elliptic curves over F, which makes $\mathbb{Q}(\sqrt{5})$ even more interesting. The field F also has 31 CM j-invariants, which is far more than any other quadratic field (see Section 5). Letting $\varphi = \frac{1+\sqrt{5}}{2}$, we have that the group of units $\{\pm 1\} \times \langle \varphi \rangle$ of the ring $R = \mathcal{O}_F = \mathbb{Z}[\varphi]$ of integers of F is infinite, leading to additional complications. Finally, F has even degree, which makes certain computations more difficult, as the cohomological techniques of [GV11] are not available.

1.3. **Modularity conjecture.**

The following conjecture is open:

Conjecture 1.1 (Modularity)**.** *The set of L-functions of elliptic curves over F equals the set of L-functions associated to cuspidal Hilbert modular newforms over F of weight $(2,2)$ with rational Hecke eigenvalues.*

Given the progress on modularity theorems initiated by [Wil95], we are optimistic that Conjecture 1.1 will be proved. *We officially assume Conjecture 1.1 for the rest of this paper.*

In Section 2 we sketch how to compute Hilbert modular forms using arithmetic in quaternion algebras. Section 3 gives numerous methods for finding an elliptic curve corresponding to a Hilbert modular form. Section 4 addresses how to find all curves that are isogenous to a given curve. In Section 5 we enumerate the CM j-invariants in F. We discuss some projects for future work in Section 6. Finally, Section 7 contains tables that summarize various information about our dataset [BDKM$^+$12].

Acknowledgements. We would like to thank John Cremona, Tom Fisher, Noam Elkies, Richard Taylor, and John Voight for helpful conversations. We would especially like to thank Joanna Gaski for providing (via the method of Section 3.1) the explicit table of elliptic curves that kickstarted this project. We used `Sage` [S$^+$12] extensively throughout this project.

2. Computing Hilbert modular forms over F

In Section 2.1 we sketch Dembélé's approach to computing Hilbert modular forms over F, then in Section 2.2 we make some remarks about our fast implementation.

2.1. **Hilbert modular forms and quaternion algebras.**

Dembélé [Dem05] introduced an algebraic approach via the Jacquet-Langlands correspondence to computing Hilbert modular forms of weight $(2,2)$ over F. The Hamiltonian quaternion algebra $F[i,j,k]$ over F is ramified exactly at the two infinite places, and contains the maximal order

$$S = R\left[\frac{1}{2}(1 - \overline{\varphi}i + \varphi j),\ \frac{1}{2}(-\overline{\varphi}i + j + \varphi k),\ \frac{1}{2}(\varphi i - \overline{\varphi}j + k),\ \frac{1}{2}(i + \varphi j - \overline{\varphi}k)\right].$$

For any nonzero ideal $\mathfrak{n}$ in $R = \mathcal{O}_F$, let $\mathbb{P}^1(R/\mathfrak{n})$ be the set of equivalence classes of column vectors with two coprime entries $a, b \in R/\mathfrak{n}$ modulo the action of $(R/\mathfrak{n})^*$. We use the notation $[a : b]$ to denote the equivalence class of $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. For each prime $\mathfrak{p} \mid \mathfrak{n}$, we fix a choice of isomorphism $F[i,j,k] \otimes F_{\mathfrak{p}} \approx M_2(F_{\mathfrak{p}})$, which induces a left action of S^* on $\mathbb{P}^1(R/\mathfrak{n})$. The Jacquet-Langlands correspondence implies that the space of Hilbert modular forms of level $\mathfrak{n}$ and weight $(2,2)$ is noncanonically isomorphic as a module over the Hecke algebra

$$\mathbb{T} = \mathbb{Z}[T_{\mathfrak{p}} : \mathfrak{p} \text{ nonzero prime ideal of } R]$$

to the finite dimensional complex vector space $V = \mathbb{C}[S^* \backslash \mathbb{P}^1(R/\mathfrak{n})]$. The action of $T_{\mathfrak{p}}$, for $p \nmid \mathfrak{n}$, is $T_{\mathfrak{p}}([x]) = \sum[\alpha x]$, where the sum is over the classes $[\alpha] \in S/S^*$ with $N_{\mathrm{red}}(\alpha) = \pi_{\mathfrak{p}}$ (reduced quaternion norm), where $\pi_{\mathfrak{p}}$ is a fixed choice of totally positive generator of $\mathfrak{p}$.

2.2. **Remarks on Computing with $\mathbb{P}^1(R/\mathfrak{n})$.** In order to implement the algorithm sketched in Section 2.1, it is critical that we can compute with $\mathbb{P}^1(R/\mathfrak{n})$ very, very quickly. For example, to apply the method of Section 3.7 below, in some cases we have to compute tens of thousands of Hecke operators. Thus in this section we make some additional remarks about this fast implementation.

When $\mathfrak{n} = \mathfrak{p}^e$ is a prime power, it is straightforward to efficiently enumerate representative elements of $\mathbb{P}^1(R/\mathfrak{p}^e)$, since each element $[x : y]$ of $\mathbb{P}^1(R/\mathfrak{p}^e)$ has a unique representative of the form $[1 : b]$ or $[a : 1]$ with a divisible by $\mathfrak{p}$, and these are all distinct. It is easy to put any $[x : y]$ in this canonical form and enumerate the elements of $\mathbb{P}^1(R/\mathfrak{p}^e)$, after choosing a way to enumerate the elements of $R/\mathfrak{p}^e$. An enumeration of $R/\mathfrak{p}^e$ is easy to give once we decide on how to represent $R/\mathfrak{p}^e$.

In general, factor $\mathfrak{n} = \prod_{i=1}^m \mathfrak{p}_i^{e_i}$. We have a bijection $\mathbb{P}^1(R/\mathfrak{n}) \cong \prod_{i=1}^m \mathbb{P}^1(R/\mathfrak{p}_i^{e_i})$, which allows us to reduce to the prime power case, at the expense of having to compute the bijection $R/\mathfrak{n} \cong \prod R/\mathfrak{p}_i^{e_i}$. To this end, we *represent elements* of $R/\mathfrak{n}$ as m-tuples in $\prod R/\mathfrak{p}_i^{e_i}$, thus making computation of the bijection trivial.

To minimize dynamic memory allocation, thus speeding up the code by an order of magnitude, in the implementation we make some arbitrary bounds; this is not a serious constraint, since the linear algebra needed to isolate eigenforms for levels beyond this bound is prohibitive. We assume $m \leq 16$ and each individual $p_i^{e_i} \leq 2^{31}$, where p_i is the residue characteristic of $\mathfrak{p}_i$. In all cases, we represent an element of $R/\mathfrak{p}_i^{e_i}$ as a pair of 64-bit integers, and represent an element of $R/\mathfrak{n}$ as an array of 16 pairs of 64-bit integers. We use this representation in all cases, even if $\mathfrak{n}$ is divisible by less than 16 primes; the gain in speed coming from avoiding dynamic memory allocation more than compensates for the wasted memory.

Let $\mathfrak{p}^e$ be one of the prime power factors of $\mathfrak{n}$, and let p be the residue characteristic of $\mathfrak{p}$. We have one of the following cases:

- $\mathfrak{p}$ splits in R; then $R/\mathfrak{p} \cong \mathbb{Z}/p\mathbb{Z}$ and we represent elements of $R/\mathfrak{p}^e$ as pairs $(a, 0)$ mod p^e with the usual addition and multiplication in the first factor.
- $\mathfrak{p}$ is inert in R; then $R/\mathfrak{p}^e \cong (\mathbb{Z}/p^e\mathbb{Z})[x]/(x^2 - x - 1)$, and we represent elements by pairs $(a, b) \in \mathbb{Z}/p^e\mathbb{Z}$ with multiplication

$$(a, b)(c, d) = (ac + bd, ad + bd + bc) \mod p^e.$$

- $\mathfrak{p}$ is ramified and $e = 2f$ is even; this is exactly the same as the case when $\mathfrak{p}$ is inert but with e replaced by f, since $R/\mathfrak{p}^e R \cong (\mathbb{Z}/p^f\mathbb{Z})[x]/(x^2 - x - 1)$.
- $\mathfrak{p}$ is ramified (so $p = 5$) and $e = 2f - 1$ is odd; the ring $A = R/\mathfrak{p}^e$ is trickier than the rest, because it is *not* of the form $\mathbb{Z}[x]/(m, g)$ where $m \in \mathbb{Z}$ and $g \in \mathbb{Z}[x]$. We have $A \approx (\mathbb{Z}/5^f\mathbb{Z})[x]/(x^2 - 5, 5^{f-1}x)$, and represent elements

of A as pairs $(a,b) \in (\mathbb{Z}/5^f) \times (\mathbb{Z}/5^{f-1}\mathbb{Z})$, with arithmetic given by

$$(a,b) + (c,d) = (a + c \mod 5^f,\ \ b + d \mod 5^{f-1})$$
$$(a,b) \cdot (c,d) = (ac + 5bd \mod 5^f,\ \ ad + bc \mod 5^{f-1}).$$

We find that $\varphi \in R \mapsto (1/2, 1/2)$.

3. Strategies for finding an elliptic curve attached to a Hilbert modular form

In this section we describe various strategies to find an elliptic curve associated to each of the Hilbert modular forms computed in Section 2. Let f be a rational cuspidal Hilbert newform of weight $(2,2)$ as in Section 2. According to Conjecture 1.1, there is some elliptic curve E_f over F such that $L(f,s) = L(E_f,s)$. (Note that E_f is only well defined up to isogeny.) Unlike the case for elliptic curves over $\mathbb{Q}$ (see [Cre97]), there seems to be no known *efficient* direct algorithm to find E_f. Nonetheless, there are several approaches coming from various directions, which are each efficient in some cases.

Everywhere below, we continue to assume that Conjecture 1.1 is true and assume that we have computed (as in Section 2) the Hecke eigenvalues $a_\mathfrak{p} \in \mathbb{Z}$ of all rational Hilbert newforms of some level $\mathfrak{n}$, for $\mathrm{Norm}(\mathfrak{p}) \leq B$ a good prime, where B is large enough to distinguish newforms. In some cases we will need far more $a_\mathfrak{p}$ in order to compute with the L-function attached to a newform. We will also need the $a_\mathfrak{p}$ for bad $\mathfrak{p}$ in a few cases, which we obtain using the functional equation for the L-function (as an application of Dokchitser's algorithm [Dok04]).

We define the *norm conductor* of an elliptic curve over F to be the absolute norm of the conductor ideal of the curve.

In Section 3.1 we give a very simple enumeration method for finding curves, then in Section 3.2 we refine it by taking into account point counts modulo primes; together, these two methods found a substantial fraction of our curves. Sections 3.3 and 3.4 describe methods for searching in certain families of curves, e.g., curves with a torsion point of given order or curves with a given irreducible mod ℓ Galois representation. Section 3.5 is about how to find all twists of a curve with bounded norm conductor. In Section 3.6 we mention the Cremona-Lingham algorithm, which relies on computing all S-integral points on many auxiliary curves. Finally, Section 3.7 explains in detail an algorithm of Dembélé that uses explicit computations with special values of L-functions to find curves.

3.1. **Extremely naive enumeration.**

The most naive strategy is to systematically enumerate elliptic curves $E : y^2 = x^3 + ax + b$, with $a, b \in R$, and for each E, to compute $a_\mathfrak{p}(E)$ for $\mathfrak{p}$ not dividing $\mathrm{Disc}(E)$ by counting points on E reduced modulo $\mathfrak{p}$. If all the $a_\mathfrak{p}(E)$ match with those of the input newform f up to the bound B, we then compute the conductor $\mathfrak{n}_E$, and if it equals $\mathfrak{n}$, we conclude from the sufficient largeness of B that E is in the isogeny class of E_f.

Under our hypotheses, this approach provides a deterministic and terminating algorithm to find all E_f. However, it can be extremely slow when $\mathfrak{n}$ is small but the simplest curve in the isogeny class of E_f has large coefficients. For example, using this search method it would be infeasible to find the curve (3.1) computed by Fisher using the visibility of Ш[7].

3.2. **Sieved enumeration.** A refinement to the approach discussed above uses the $a_{\mathfrak{p}}$ values to impose congruence conditions modulo $\mathfrak{p}$ on $\#\tilde{E}(R/\mathfrak{p})$. If f is a newform with Hecke eigenvalues $a_{\mathfrak{p}}$, then $\#\tilde{E}_f(R/\mathfrak{p}) = \mathbf{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$. Given $\mathfrak{p}$ not dividing the level $\mathfrak{n}$, we can find all elliptic curves modulo $\mathfrak{p}$ with the specified number of points, especially when $\mathbf{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$ has few prime factors. We impose these congruence conditions at multiple primes $\mathfrak{p}_i$, use the Chinese Remainder Theorem, and lift the resulting curves modulo $R/(\prod \mathfrak{p}_i)$ to non-singular curves over R.

While this method, like the previous one, will eventually terminate, it too is very ineffective if every E in the class of isogenous curves corresponding to f has large coefficients. However in practice, by optimally choosing the number of primes $\mathfrak{p}_i$, a reasonably efficient implementation of this method can be obtained.

3.3. **Torsion families.** We find elliptic curves of small conductor by specializing explicit parametrizations of families of elliptic curves over F having specified torsion subgroups. We use the parametrizations of [Kub76].

Theorem 3.1 (Kamienny-Najman, [KN12]). *The following is a complete list of torsion structures for elliptic curves over F:*

$$\begin{aligned} &\mathbb{Z}/m\mathbb{Z}, && 1 \le m \le 10, \quad m = 12,\\ &\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2m\mathbb{Z}, && 1 \le m \le 4,\\ &\mathbb{Z}/15\mathbb{Z}. \end{aligned}$$

Moreover, there is a unique curve with 15*-torsion.*

We use the following proposition to determine in which family to search.

Proposition 3.2. *Let ℓ be a prime and E a curve over F. Then $\ell \mid \#E'(F)_{\mathrm{tor}}$ for some curve E' in the isogeny class of E if and only if $\ell \mid \mathbf{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$ for all odd primes $\mathfrak{p}$ at which E has good reduction.*

Proof. If $\ell \mid \#E'(F)_{\mathrm{tor}}$, from the injectivity of the reduction map at good primes [Kat81, Appendix], we have that $\ell \mid \#\tilde{E}'(\mathbb{F}_{\mathfrak{p}}) = \mathbf{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$. The converse statement is one of the main results of [Kat81]. □

By applying Proposition 3.2 for all $a_{\mathfrak{p}}$ with $\mathfrak{p}$ up to some bound, we can decide whether or not it is *likely* that some curve in the isogeny class of E contains an F-rational ℓ-torsion point. If this is the case, then we search over those families of curves with rational ℓ-torsion. With a relatively small search space, we thus find many curves with large coefficients more quickly than with the algorithm of Section 3.1. For example, we first found the curve E given by

$$y^2 + \varphi y = x^3 + (27\varphi - 43)\,x + (-80\varphi + 128)$$

with norm conductor 145 by searching for curves with torsion subgroup $\mathbb{Z}/7\mathbb{Z}$.

3.4. **Congruence families.** Suppose that we are searching for a curve E and we already know another curve E' with $E[\ell] \approx E'[\ell]$, where ℓ is some prime and $E[\ell]$ is irreducible. If $\ell = 7, 11$ then we can use techniques of Fisher [Fis12] to attempt to search through the finitely many curves with mod ℓ Galois representation isomorphic to $E[\ell]$. We used this approach to find the curve E given by

$$\begin{aligned} y^2 + \varphi xy = x^3 + (\varphi - 1)\,x^2 + &\\ (-257364\varphi - 159063)\,x &+ (-75257037\varphi - 46511406) \end{aligned} \tag{3.1}$$

with conductor $-38\varphi+10$, which has norm 1476. Given just the $a_{\mathfrak{p}}$, we noticed that $E[7] \approx E'[7]$, where E' has norm conductor 369, then Fisher used a Magma [BCP97] program to find rational points on a certain quartic surface that parametrize curves with the same $E'[7]$. Fortunately, our curve E was amongst those curves. We had already found E' via a naive search, since it is given by the equation $y^2+(\varphi+1)\,y = x^3+(\varphi-1)\,x^2+(-2\varphi)\,x$.

3.5. **Twisting.** Let E be an elliptic curve F. A *twist* E' of E is a curve over F that is isomorphic to E over some extension of F. A *quadratic twist* is a twist in which the extension has degree 2. We can use twisting to find curves that may otherwise be difficult to find as follows: starting with a known elliptic curve E of some (small) conductor, we compute its twists of conductor up to some bound, and add them to our table.

More explicitly, if E is given by $y^2 = x^3+ax+b$ and $d \in F^*$, then the twist E^d of E by d is given by $dy^2 = x^3+ax+b$; in particular, we may assume that d is square free. The following is well known:

Proposition 3.3. *If $\mathfrak{n}$ is the conductor of E and $d \in F^*$ is coprime to $\mathfrak{n}$, then the conductor of E^d is divisible by $d^2\mathfrak{n}$.*

Proof. There are choices of Weierstrass equations such that $\Delta(E^d) = 2^{12}d^6\Delta(E)$, where Δ is the discriminant. Thus the curve E^d has bad reduction at each prime that divides d, because twisting introduces a 6th power of the squarefree d into the discriminant, and d is coprime to $\Delta(E)$, so no change of Weierstrass equation can remove this 6th power. Moreover, E^d is isomorphic to E over an extension of the base field, so E^d has potentially good reduction at each prime dividing d. Thus the reduction at each prime dividing d is additive. The conductor is unchanged at the primes dividing $\mathfrak{n}$ because of the formula relating the conductor, discriminant and reduction type (see [Sil92, App. C,§15]), that formation of Néron models commutes with unramified base change, and the fact that at the primes that divide $\mathfrak{n}$ the minimal discriminant of E^d is the same as that of E. □

To find all twists E^d with norm conductor at most B, we twist E by all d of the form $\pm\varphi^\delta d_0 d_1$, where $\delta \in \{0,1\}$, d_0 is a product of a fixed choice of generators for the prime divisors of $\mathfrak{n}$, d_1 is a squarefree product of a fixed choice of generators of primes not dividing $\mathfrak{n}$, and $|\mathbf{N}(d_1)| \le \sqrt{B/C}$, where C is the norm of the product of the primes that exactly divide $\mathfrak{n}$. We know from 3.3 that this search is exhaustive.

For example, let E be given by $y^2+xy+\varphi y = x^3+(-\varphi-1)\,x^2$ of conductor $5\varphi-3$ having norm 31. Following the above strategy to find twists of norm conductor $\le B := 1831$, we have $C = 31$ and square-free d_1 such that $|\mathbf{N}(d_1)| \le \sqrt{B/C} \approx 7.6\ldots$. Thus $d_1 \in \{1,2,\varphi,2\varphi\}$ and checking all possibilities for $\varphi^\delta d_0 d_1$, we find the curve $E^{-\varphi-2}$ having norm conductor 775 and the curve $E^{5\varphi-3}$ having norm conductor 961. Other twists have larger norm conductors, e.g., E^2 has norm conductor $126976 = 2^{12}\cdot 31$.

3.6. **Curves with good reduction outside S.** We use the algorithm of Cremona and Lingham from [CL07] to find all elliptic curves E having good reduction at primes outside of a finite set $\mathcal{S}$ of primes in F. This algorithm has limitations over a general number field K due to the difficulty of finding a generating set for $E(K)$ and points on E defined over $\mathcal{O}_K$. Using Cremona's Magma implementation of the algorithm, we found several curves not found by other methods, e.g., y^2+

$(\varphi+1)\,xy+y = x^3-x^2+(-19\varphi-39)\,x+(-143\varphi-4)$, which has norm conductor 1331.

3.7. **Special values of twisted L-series.** In [Dem08], Lassina Dembélé outlines some methods for finding modular elliptic curves from Hilbert modular forms over real quadratic fields. Formally, these methods are not proven to be any better than a direct search procedure, as they involve making a large number of guesses, and a priori we do not know just how many guesses we will need to make. And unlike other methods described in this paper, this method requires many Hecke eigenvalues, and computing these takes a lot of time. However, this method certainly works extremely well in many cases, and after tuning it by using large tables of curves that we had already computed, we are able to use it to find more curves that we would have had no hope of finding otherwise; we will give an example of one of these curves later.

When the level $\mathfrak{n}$ is not square, Dembélé's method relies on computing or guessing periods of the curve by using special values of L-functions of twists of the curve. In particular, the only inputs required are the level of the Hilbert modular form and its L-series. So we suppose that we know the level $\mathfrak{n}=(N)$ of the form, where N is totally positive, and that we have sufficiently many coefficients of its L-series $a_{\mathfrak{p}_1}, a_{\mathfrak{p}_2}, a_{\mathfrak{p}_3}, \ldots$.

Let σ_1 and σ_2 denote the embeddings of F into the real numbers, with $\sigma_1(\varphi)\approx 1.61803\ldots$. For an elliptic curve E over F we get two associated embeddings into the complex numbers, and hence a pair of period lattices. We let Ω_E^+ and Ω_E^-, which we refer to as the periods of E, be the least real and imaginary periods of the lattice which come from the embedding σ_1, and as the period lattices are interchanged when E is replaced with its conjugate curve, we let $\Omega_{\overline{E}}^+$ and $\Omega_{\overline{E}}^-$ denote the least real and imaginary periods of the lattice under the embedding σ_2.

For ease, we write

$$\begin{aligned} \Omega_E^{++} &= \Omega_E^+\Omega_{\overline{E}}^+ & \qquad \Omega_E^{+-} &= \Omega_E^+\Omega_{\overline{E}}^- \\ \Omega_E^{-+} &= \Omega_E^-\Omega_{\overline{E}}^+ & \qquad \Omega_E^{--} &= \Omega_E^-\Omega_{\overline{E}}^-. \end{aligned}$$

We refer to these numbers as the *mixed periods* of E.

3.7.1. *Recovering the curve from its mixed periods.* If we know these mixed periods to sufficient precision, it is not hard to recover the curve E. Without the knowledge of the discriminant of the curve, we do not know the lattice type of the curve and its conjugate, but there are only a few possibilities for what they might be. This gives us a few possibilities for the j-invariant of E. Observe that $\sigma_1(j(E))$ is either $j(\tau_1(E))$ or $j(\tau_2(E))$ and $\sigma_2(j(E))$ is either $j(\tau_1(\overline{E}))$ or $j(\tau_2(\overline{E}))$, where

$$\begin{aligned} \tau_1(E) &= \frac{\Omega_E^{-+}}{\Omega_E^{++}} = \frac{\Omega_E^-}{\Omega_E^+} & \qquad \tau_2(E) &= \frac{1}{2}\left(1+\frac{\Omega_E^{-+}}{\Omega_E^{++}}\right) = \frac{1}{2}\left(1+\frac{\Omega_E^-}{\Omega_E^+}\right) \\ \tau_1(\overline{E}) &= \frac{\Omega_E^{+-}}{\Omega_E^{++}} = \frac{\Omega_{\overline{E}}^-}{\Omega_E^+} & \qquad \tau_2(\overline{E}) &= \frac{1}{2}\left(1+\frac{\Omega_E^{+-}}{\Omega_E^{++}}\right) = \frac{1}{2}\left(1+\frac{\Omega_{\overline{E}}^-}{\Omega_{\overline{E}}^+}\right) \end{aligned}$$

and $j(\tau)$ is the familiar

$$j(\tau) = e^{-2\pi i\tau} + 744 + 196884e^{2\pi i\tau} + 21493760e^{4\pi i\tau} + \cdots .$$

We try each pair of possible embeddings for $j(E)$ in turn, and recognize possibilities for $j(E)$ as an algebraic number. We then construct curves E' corresponding to each possibility for $j(E)$. By computing a few $a_{\mathfrak{p}}(E)$, we should be able to determine whether we have chosen the correct j-invariant, in which case E' will be a twist of E. We can then recognize which twist it is in order to recover E.

In practice, of course, as we have limited precision, and as $j(E)$ will not be an algebraic integer, it may not be feasible to directly determine its exact value, especially if its denominator is large.

To get around the problem of limited precision, we suppose that we have some extra information; namely, the discriminant Δ_E of the curve we are looking for. With Δ_E in hand we can directly determine which τ to choose: if $\sigma_1(\Delta_E) > 0$ then $\sigma_1(j(E)) = j(\tau_1(E))$, and if $\sigma_1(\Delta_E) < 0$ then $\sigma_1(j(E)) = j(\tau_2(E))$, and similarly for σ_2. We then compute $\sigma_1(c_4(E)) = (j(\tau)\sigma_1(\Delta_E))^{1/3}$ and $\sigma_2(c_4(E)) = (j(\tau')\sigma_2(\Delta_E))^{1/3}$.

Using the approximations of the two embeddings of c_4, we can recognize c_4 approximately as an algebraic integer. Specifically, we compute

$$\alpha = \frac{\sigma_1(c_4) + \sigma_2(c_4)}{2} \quad \text{and} \quad \beta = \frac{\sigma_1(c_4) - \sigma_2(c_4)}{2\sqrt{5}}.$$

Then $c_4 = \alpha + \beta\sqrt{5}$, and we can find c_6.

In practice, there are two important difficulties we must overcome: we do not know Δ_E and it may be quite difficult to get high precision approximations to the mixed periods, and thus we may not be able to easily compute c_4. Thus, we actually proceed by choosing a Δ_{guess} from which we compute half-integers α and β and an integer $a + b\varphi \approx \alpha + \beta\sqrt{5}$, arbitrarily rounding either a or b if necessary. We then make some choice of search range M, and for each pair of integers m and n, bounded in absolute value by M, we try each $c_{4,\text{guess}} = (a + m) + (b + n)\varphi$.

Given $c_{4,\text{guess}}$, we attempt to solve

$$c_{6,\text{guess}} = \pm\sqrt{c_{4,\text{guess}}^3 - 1728\Delta_{\text{guess}}},$$

and, if we can, we use these to construct a curve E_{guess}. If E_{guess} has the correct conductor and the correct Hecke eigenvalues, we declare that we have found the correct curve; otherwise, we proceed to the next guess.

For a choice of Δ_{guess}, we will generally start with the conductor N_E, and then continue by trying unit multiples and by adding in powers of factors of N_E.

3.7.2. *Guessing the mixed periods.* We have thus far ignored the issue of actually finding the mixed periods of the curve that we are looking for. Finding them presents an extra difficulty as our procedure involves even more guesswork. Dembélé's idea is to use special values of twists of the L-function $L(f, s)$. Specifically, we twist by primitive quadratic Dirichlet characters over $\mathcal{O}_F$, which are homomorphisms $\chi : (\mathcal{O}_F/\mathfrak{c})^* \to \pm 1$, pulled back to $\mathcal{O}_F$.

In the case of odd prime conductor, which we will stick to here, there is just a single primitive quadratic character, which is the quadratic residue symbol. A simple way to compute it is by making a table of squares, or by choosing a primitive root of $g \in (\mathcal{O}_F/\mathfrak{c})^*$, assigning $\chi(g) = -1$, and again making a table by extending multiplicatively. Alternatively, one could use a reciprocity formula as described in [BS10]. For general conductor, one can compute with products of characters having prime conductor.

For a given f and a primitive χ, we can construct the twisted L-function

$$L(f,\chi,s) = \sum_{\mathfrak{m}\subseteq\mathcal{O}_F} \frac{\chi(m)a_{\mathfrak{m}}}{N(\mathfrak{m})^s},$$

where m is a totally positive generator of $\mathfrak{m}$. (Note that χ is not well defined on ideals, but *is* well defined on totally positive generators of ideals.) $L(f,\chi,s)$ will satisfy a functional equation similar to that of $L(f,s)$, but the conductor is multiplied by $\mathrm{Norm}(\mathfrak{c})^2$ and the sign is multiplied by $\chi(-N)$. The key to finding the mixed periods of E is contained in the following conjecture that Dembélé has distilled from [BDG04], and we have stated specifically for $\mathbb{Q}(\sqrt{5})$.

Conjecture 3.4. *If χ is a primitive quadratic character with conductor $\mathfrak{c}$ relatively prime to the conductor of E, with $\chi(\varphi) = s'$ and $\chi(1-\varphi) = s$, (where $s, s' \in \{+,-\} = \{\pm 1\}$), then*

$$\Omega_E^{s,s'} = c_\chi \tau(\overline{\chi}) L(E,\chi,1)\sqrt{5},$$

for some integer c_χ, where $\tau(\chi)$ is the Gauss sum

$$\tau(\chi) = \sum_{\alpha \bmod \mathfrak{c}} \chi(\alpha)\exp\left(2\pi i \operatorname{Tr}\left(\alpha/m\sqrt{5}\right)\right),$$

with m a totally positive generator of $\mathfrak{c}$.

Remark 3.5. The Gauss sum is more innocuous than it seems. For odd conductor $\mathfrak{c}$ it is of size $\sqrt{\mathrm{Norm}(\mathfrak{c})}$, while for an even conductor it is of size $\sqrt{2\,\mathrm{Norm}(\mathfrak{c})}$. Its sign is a 4-th root of unity, and whether it is real or imaginary can be deduced directly from the conjecture, as it matches with the sign of $\Omega_E^{s,s'}$. In particular, $\tau(\chi)$ is real when $\chi(-1) = 1$ and imaginary when $\chi(-1) = -1$, which is a condition on $\mathrm{Norm}(\mathfrak{c}) \bmod 4$, as $\chi(-1) \equiv \mathrm{Norm}(\mathfrak{c}) \pmod 4$. This can all be deduced, for example, from [BS10].

Also, note that Dembélé writes this conjecture with an additional factor of $4\pi^2$; this factor does not occur with the definition of $L(f,s)$ that we have given.

Remark 3.6. Contained in this conjecture is the obstruction to carrying out the method described here when $\mathfrak{n}$ is a square. In this case, the sign of $L(f,\chi,s)$ will be completely determined by whether or not $\chi(\varphi) = \chi(1-\varphi)$, so we can only obtain information about either Ω^{--} and Ω^{++} or Ω^{-+} and Ω^{+-}, and we need three of these values to find E.

With this conjecture in place, we can describe a method for guessing the mixed periods of E. Now, to proceed, we construct four lists of characters up to some conductor bound M (we are restricting to odd prime modulus here for simplicity, as primitivity is ensured, but this is not necessary):

$$S^{s,s'} = \{\chi \bmod \mathfrak{p} : \chi(\varphi) = s', \chi(1-\varphi) = s, (\mathfrak{p},\mathfrak{n}) = 1, \mathrm{Norm}(\mathfrak{p}) < M, \chi(-N) = \epsilon_E\}.$$

Here $s, s' \in \{+,-\} = \pm 1$ again. We will consider these lists to be ordered by the norms of the conductors of the characters in increasing order, and index their elements as $\chi_0^{s,s'}, \chi_1^{s,s'}, \chi_2^{s,s'}, \ldots$. For each character we compute the central value of the twisted L-function to get four new lists

$$\mathcal{L}^{s,s'} = \{i^{ss'}\sqrt{5\,\mathrm{Norm}(\mathfrak{p})}L(E,\chi,1), \chi \in S^{s,s'}\} = \{\mathcal{L}_0^{s,s'}, \mathcal{L}_1^{s,s'}, \ldots\}.$$

These numbers should now all be integer multiples of the mixed periods, so to get an idea of which integer multiples they might be, we compute each of the ratios

$$\frac{\mathcal{L}_0^{s,s'}}{\mathcal{L}_k^{s,s'}} = \frac{c_{\chi_0^{s,s'}}}{c_{\chi_k^{s,s'}}} \in \mathbb{Q}, \quad k = 1, 2, \ldots,$$

attempt to recognize these as rational numbers, and choose as an initial guess

$$\Omega_{E,\text{guess}}^{ss'} = \mathcal{L}_0^{s,s'} \left(\operatorname{lcm} \left\{ \text{numerator} \left(\frac{\mathcal{L}_0^{s,s'}}{\mathcal{L}_k^{s,s'}} \right), k = 1, 2, \ldots \right\} \right)^{-1}.$$

3.7.3. *An example.* We give an example of an elliptic curve that we were only able to find by using this method. At level $\mathfrak{n} = (-38\varphi + 26)$ we found a newform f, computed

$$a_{(2)}(f) = -1,\ a_{(-2\varphi+1)}(f) = 1,\ a_{(3)}(f) = -1,\\ a_{(-3\varphi+1)}(f) = -1,\ a_{(-3\varphi+2)}(f) = -6, \cdots, a_{(200\varphi-101)}(f) = 168$$

and determined, by examining the L-function, that the sign of the functional equation should be -1. (In fact, we do not really need to know the sign of the functional equation, as we would quickly determine that $+1$ is wrong when attempting to find the mixed periods.) Computing the sets of characters described above, and choosing the first 3 of each, we have

$$S^{--} = \{\chi_{(\varphi+6)}, \chi_{(7)}, \chi_{(7\varphi-4)}\}, \quad S^{-+} = \{\chi_{(-3\varphi+1)}, \chi_{(5\varphi-2)}, \chi_{(\varphi-9)}\}\\ S^{+-} = \{\chi_{(-4\varphi+3)}, \chi_{(5\varphi-3)}, \chi_{(-2\varphi+13)}\} \quad S^{++} = \{\chi_{(\varphi+9)}, \chi_{(9\varphi-5)}, \chi_{(\varphi+13)}\}.$$

By using the 5133 eigenvalues above as input to Rubinstein's `lcalc` [Rub11], we compute the lists of approximate values

$$\begin{aligned} \mathcal{L}^{--} &= \{-33.5784397862407, -3.73093775400387, -18.6546887691646\} \\ \mathcal{L}^{-+} &= \{18.2648617736017i, 32.8767511924831i, 3.65297235421633i\} \\ \mathcal{L}^{+-} &= \{41.4805656925342i, 8.29611313850694i, 41.4805677827298i\} \\ \mathcal{L}^{++} &= \{32.4909970742969, 162.454985515474, 162.454973589303\}. \end{aligned}$$

Note that `lcalc` will warn us that we do not have enough coefficients to obtain good accuracy, and we make no claim as far as the accuracy of these values is concerned. Hoping that the ends will justify the means, we proceed forward.

Dividing each list by the first entry, and recognizing the quotients as rational numbers, we get the lists

$$\begin{aligned} \{1.000, 9.00000000005519, 1.80000000009351\} &\approx \{1, 9, 9/5\} \\ \{1.000, 0.555555555555555, 5.00000000068986\} &\approx \{1, 5/9, 5\} \\ \{1.000, 4.99999999999994, 0.999999949610245\} &\approx \{1, 5, 1\} \\ \{1.000, 0.199999999822733, 0.200000014505165\} &\approx \{1, 1/5, 1/5\}, \end{aligned}$$

which may give an indication of the accuracy of our values. We now proceed with the guesses

$$\begin{aligned}
\Omega_{E,\text{guess}}^{--} &\approx -33.5784397862407/9 &&\approx -3.73093775402141\\
\Omega_{E,\text{guess}}^{-+} &\approx 18.2648617736017i/5 &&\approx 3.65297235472034i\\
\Omega_{E,\text{guess}}^{+-} &\approx 41.4805656925342i/5 &&\approx 8.29611313850683i\\
\Omega_{E,\text{guess}}^{++} &\approx 32.4909970742969 &&= 32.4909970742969.
\end{aligned}$$

These cannot possibly be all correct, as $\Omega_E^{--}\Omega_E^{++} = \Omega_E^{-+}\Omega_E^{+-}$. Still, we can choose any three and get a reasonable guess, and in fact we may choose all possible triples, dividing some of the guesses by small rational numbers, and choosing the fourth guess to be consistent with the first three; we build a list of possible embeddings of $j(E)$, which will contain the possibility $\sigma_1(j(E)) \approx 1.365554233954 \times 10^{12}$, $\sigma_2(j(E)) \approx 221270.95861123$, which is a possibility if

$$\Omega_E^{-+} = \Omega_{E,\text{guess}}^{-+},\quad \Omega_E^{+-} = \Omega_{E,\text{guess}}^{+-},\quad \Omega_E^{-+} = \frac{\Omega_{E,\text{guess}}^{-+}}{2},\quad \Omega_E^{++} = \frac{\Omega_{E,\text{guess}}^{++}}{8}.$$

Cycling through many discriminants, we eventually try

$$\Delta_{\text{guess}} = \varphi \cdot 2^5 \cdot (19\varphi - 13),$$

which leads us to the guess

$$\begin{aligned}
\sigma_1(c_{4,\text{guess}}) &= (\sigma_1(j(E))\sigma_1(\Delta_{\text{guess}}))^{1/3} \approx 107850.372979378\\
\sigma_2(c_{4,\text{guess}}) &= (\sigma_2(j(E))\sigma_2(\Delta_{\text{guess}}))^{1/3} \approx 476.625892034286.
\end{aligned}$$

We have enough precision to easily recognize this as

$$c_{4,\text{guess}} = \frac{108327 + 48019\sqrt{5}}{2} = 48019\varphi + 30154,$$

and

$$\sqrt{c_{4,\text{guess}}^3 - 1728\Delta_{\text{guess}}}$$

does in fact have two square roots: $\pm(15835084\varphi + 9796985)$. We try both of them, and the choice with the minus sign gives the curve

$$y^2 + \varphi xy + \varphi y = x^3 + (\varphi - 1)\,x^2 + (-1001\varphi - 628)\,x + (17899\varphi + 11079)\,,$$

which has the correct conductor. We compute a few values of $a_{\mathfrak{p}}$ for this curve, and it turns out to be the one that we are looking for.

4. Enumerating the curves in an isogeny class

Given an elliptic curve E/F, we wish to find representative isomorphism classes for all elliptic curves E'/F that are isogenous to E via an isogeny defined over F. The analogue of this problem over $\mathbb{Q}$ has an algorithmic solution as explained in [Cre97, §3.8]; it relies on:

(1) Mazur's theorem [Maz78] that if $\psi : E \to E'$ is a $\mathbb{Q}$-rational isogeny of prime degree, then $\deg(\psi) \leq 163$.

(2) Formulas of Vélu [Vél71] that provide a way to explicitly enumerate all p-isogenies (if any) with domain E. Vélu's formulas are valid for any number field, but so far there has not been an explicit generalization of Mazur's theorem for any number field other than $\mathbb{Q}$.

Remark 4.1. Assume the generalized Riemann hypothesis. Then work of Larson-Vaintrob from [LV] implies that there is an effectively computable constant C_F such that any prime degree isogeny over F has degree at most C_F.

Since we are interested in specific isogeny classes, we can use the algorithm described in [Bil11] that takes as input a specific non-CM elliptic curve E over a number field K, and outputs a provably finite list of primes p such that E might have a p-isogeny. The algorithm is particularly easy to implement in the case when K is a quadratic field, as explained in [Bil11, §2.3.4]. Using this algorithm combined with Vélu's formulas, we were able to enumerate *all* isomorphism classes of curves isogenous to the curves we found via the methods of Section 3, and thus divide our curves up into isogeny classes.

5. CM elliptic curves over F

In this section we make some general remarks about CM elliptic curves over F. The main surprise is that there are 31 distinct $\overline{\mathbb{Q}}$-isomorphism classes of CM elliptic curves defined over F, more than for any other quadratic field.

Proposition 5.1. *The field F has more isomorphism classes of CM elliptic curves than any other quadratic field.*

Proof. Let K be a quadratic extension of $\mathbb{Q}$. Let H_D denote the Hilbert class polynomial of the CM order $\mathcal{O}_D$ of discriminant D, so $H_D \in \mathbb{Q}[X]$ is the minimal polynomial of the j-invariant j_D of any elliptic curve $E = E_D$ with CM by $\mathcal{O}_D$. Since K is Galois, we have $j_D \in K$ if and only if H_D is either linear or quadratic with both roots in K. The D for which H_D is linear are the thirteen values $-3, -4, -7, -8, -11, -12, -16, -19, -27, -28, -43, -67, -163$. According to [Cre92], the D for which H_D is quadratic are the following 29 discriminants:

$$\begin{aligned}&-15, -20, -24, -32, -35, -36, -40, -48, -51, -52, -60,\\ &-64, -72, -75, -88, -91, -99, -100, -112, -115, -123,\\ &-147, -148, -187, -232, -235, -267, -403, -427.\end{aligned}$$

By computing discriminants of these Hilbert class polynomials, we obtain the following table:

Field	D so H_D has roots in field	Field	D so H_D has roots in field
$\mathbb{Q}(\sqrt{2})$	$-24, -32, -64, -88$	$\mathbb{Q}(\sqrt{21})$	-147
$\mathbb{Q}(\sqrt{3})$	$-36, -48$	$\mathbb{Q}(\sqrt{29})$	-232
$\mathbb{Q}(\sqrt{5})$	$-15, -20, -35, -40, -60,$	$\mathbb{Q}(\sqrt{33})$	-99
	$-75, -100, -115, -235$	$\mathbb{Q}(\sqrt{37})$	-148
$\mathbb{Q}(\sqrt{6})$	-72	$\mathbb{Q}(\sqrt{41})$	-123
$\mathbb{Q}(\sqrt{7})$	-112	$\mathbb{Q}(\sqrt{61})$	-427
$\mathbb{Q}(\sqrt{13})$	$-52, -91, -403$	$\mathbb{Q}(\sqrt{89})$	-267
$\mathbb{Q}(\sqrt{17})$	$-51, -187$		

The claim follows because the $\mathbb{Q}(\sqrt{5})$ row is largest, containing 9 entries. There are thus $31 = 2 \cdot 9 + 13$ distinct CM j-invariants in $\mathbb{Q}(\sqrt{5})$.

□

6. Related future projects

It would be natural to extend the tables to the first known curve of rank 3 over F, which may be the curve $y^2 + y = x^3 - 2x + 1$ of norm conductor $163^2 = 26569$. It would also be interesting to make a table in the style of [SW02], and compute analytic ranks of the large number of curves that we would find; this would benefit from Sutherland's `smalljac` program, which has very fast code for computing L-series coefficients. Some aspects of the tables could also be generalized to modular abelian varieties A_f attached to Hilbert modular newforms with not-necessarily-rational Hecke eigenvalues; in particular, we could enumerate the A_f up to some norm conductor, and numerically compute their analytic ranks.

7. Tables

As explained in Sections 3 and 4, assuming Conjecture 1.1, we found the complete list of elliptic curves with norm conductor up to 1831, which is the first norm conductor of a rank 2 curve over F. The complete dataset can be downloaded from [BDKM$^+$12].

In each of the following tables #isom refers to the number of curves, #isog refers to the number of classes, $\mathfrak{n}$ refers to the conductor of the given elliptic curve, and Weierstrass equations are given in the form $[a_1, a_2, a_3, a_4, a_6]$.

Table 7.1 gives the number of curves and isogeny classes we found. Note that in these counts we do not exclude conjugate curves, i.e., if σ denotes the nontrivial element of $\mathrm{Gal}(F/\mathbb{Q})$, then we count E and E^σ separately if they are not isomorphic.

Table 7.1. Curves over $\mathbb{Q}(\sqrt{5})$

rank	**#isog**	**#isom**	**smallest** Norm($\mathfrak{n}$)
0	745	2174	31
1	667	1192	199
2	2	2	1831
total	1414	3368	-

Table 7.2 gives counts of the number of isogeny classes of curves in our data of each size; note that we find some isogeny classes of cardinality 10, which is bigger than what one observes with elliptic curves over $\mathbb{Q}$.

Table 7.2. Number of Isogeny classes of a given size

	size							
bound	1	2	3	4	6	8	10	**total**
199	2	21	3	20	8	9	1	64
1831	498	530	36	243	66	38	3	1414

Table 7.3 gives the number of curves and classes up to a given norm conductor bound. Note that the first curve of rank 1 has norm conductor 199, and there are no curves of norm conductor 200.

TABLE 7.3. Counts of classes and curves with bounded norm conductors and specified ranks

	#isog				**#isom**			
	rank				**rank**			
bound	0	1	2	**total**	0	1	2	**total**
200	62	2	0	64	257	6	0	263
400	151	32	0	183	580	59	0	639
600	246	94	0	340	827	155	0	982
800	334	172	0	506	1085	285	0	1370
1000	395	237	0	632	1247	399	0	1646
1200	492	321	0	813	1484	551	0	2035
1400	574	411	0	985	1731	723	0	2454
1600	669	531	0	1200	1970	972	0	2942
1800	729	655	0	1384	2128	1178	0	3306
1831	745	667	2	1414	2174	1192	2	3368

Table 7.4 gives the number of curves and classes with isogenies of each degree; note that we do not see all possible isogeny degrees. For example, the elliptic curve $X_0(19)$ has rank 1 over F, so there are infinitely many curves over F with degree 19 isogenies (unlike over $\mathbb{Q}$ where $X_0(19)$ has rank 0). We also give an example curve (that need not have minimal conductor) with an isogeny of the given degree.

TABLE 7.4. Isogeny degrees

degree	**#isog**	**#isom**	**example curve**	Norm($\mathfrak{n}$)
None	498	498	$[\varphi+1, 1, 1, 0, 0]$	991
2	652	2298	$[\varphi, -\varphi+1, 0, -4, 3\varphi-5]$	99
3	289	950	$[\varphi, -\varphi, \varphi, -2\varphi-2, 2\varphi+1]$	1004
5	65	158	$[1, 0, 0, -28, 272]$	900
7	19	38	$[0, \varphi+1, \varphi+1, \varphi-1, -3\varphi-3]$	1025

Table 7.5 gives the number of curves with each torsion structure, along with an example curve (again, not necessarily with minimal conductor) with that torsion structure.

Table 7.5. Torsion subgroups

structure	**#isom**	**example curve**	Norm($\mathfrak{n}$)
1	296	$[0,-1,1,-8,-7]$	225
$\mathbb{Z}/2\mathbb{Z}$	1453	$[\varphi,-1,0,-\varphi-1,\varphi-3]$	164
$\mathbb{Z}/3\mathbb{Z}$	202	$[1,0,1,-1,-2]$	100
$\mathbb{Z}/4\mathbb{Z}$	243	$[\varphi+1,\varphi-1,\varphi,0,0]$	79
$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$	312	$[0,\varphi+1,0,\varphi,0]$	256
$\mathbb{Z}/5\mathbb{Z}$	56	$[1,1,1,22,-9]$	100
$\mathbb{Z}/6\mathbb{Z}$	183	$[1,\varphi,1,\varphi-1,0]$	55
$\mathbb{Z}/7\mathbb{Z}$	13	$[0,\varphi-1,\varphi+1,0,-\varphi]$	41
$\mathbb{Z}/8\mathbb{Z}$	21	$[1,\varphi+1,\varphi,\varphi,0]$	31
$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/4\mathbb{Z}$	51	$[\varphi+1,0,0,-4,-3\varphi-2]$	99
$\mathbb{Z}/9\mathbb{Z}$	6	$[\varphi,-\varphi+1,1,-1,0]$	76
$\mathbb{Z}/10\mathbb{Z}$	12	$[\varphi+1,\varphi,\varphi,0,0]$	36
$\mathbb{Z}/12\mathbb{Z}$	6	$[\varphi,\varphi+1,0,2\varphi-3,-\varphi+2]$	220
$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/6\mathbb{Z}$	11	$[0,1,0,-1,0]$	80
$\mathbb{Z}/15\mathbb{Z}$	1	$[1,1,1,-3,1]$	100
$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/8\mathbb{Z}$	2	$[1,1,1,-5,2]$	45

We computed the invariants in the Birch and Swinnerton-Dyer conjecture for our curves, and solved for the conjectural order of Ш; Table 7.6 gives the number of curves in our data having each order of Ш as well as a minimal conductor curve exhibiting each of these orders.

Table 7.6. Ш

#Ш	**#isom**	**first curve having #Ш**	Norm($\mathfrak{n}$)
1	3191	$[1,\varphi+1,\varphi,\varphi,0]$	31
4	84	$[1,1,1,-110,-880]$	45
9	43	$[\varphi+1,-\varphi,1,-54686\varphi-35336, -7490886\varphi-4653177]$	76
16	16	$[1,\varphi,\varphi+1,-4976733\varphi-3075797, -6393196918\varphi-3951212998]$	45
25	2	$[0,-1,1,-7820,-263580]$	121
36	2	$[1,-\varphi+1,\varphi,1326667\varphi-2146665, 880354255\varphi-1424443332]$	1580

References

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* **Q**: *wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic), `http://math.stanford.edu/~conrad/papers/tswfinal.pdf`. MR 2002d:11058

[BCP97] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265, Computational algebra and number theory (London, 1993). MR 1 484 478

[BDG04] Massimo Bertolini, Henri Darmon, and Peter Green, *Periods and points attached to quadratic algebras*, Heegner points and Rankin L-series, Math. Sci. Res. Inst. Publ., vol. 49, Cambridge Univ. Press, Cambridge, 2004, pp. 323–367. MR 2083218 (2005e:11062)

[BDKM$^+$12] Jon Bober, Alyson Deines, Ariah Klages-Mundt, Ben LeVeque, R. Andrew Ohana, Ashwath Rabindranath, Paul Sharaba, and William Stein, *A Database of Elliptic Curves over* $\mathbb{Q}(\sqrt{5})$, 2012, `http://wstein.org/papers/sqrt5`.

[Bil11] Nicolas Billerey, *Critères d'irréductibilité pour les représentations des courbes elliptiques*, Int. J. Number Theory **7** (2011), no. 4, 1001–1032. MR 2812649

[BK75] B. J. Birch and W. Kuyk (eds.), *Modular functions of one variable. IV*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 476.

[BMSW07] Baur Bektemirov, Barry Mazur, William Stein, and Mark Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 2, 233–254 (electronic). MR 2291676

[BS10] Hatice Boylan and Nils-Peter Skoruppa, *Explicit formulas for Hecke Gauss sums in quadratic number fields*, Abh. Math. Semin. Univ. Hambg. **80** (2010), no. 2, 213–226. MR 2734687 (2012c:11163)

[CL07] J. E. Cremona and M. P. Lingham, *Finding all elliptic curves with good reduction outside a given set of primes*, Experiment. Math. **16** (2007), no. 3, 303–312. MR 2367320 (2008k:11057)

[Cre] J. E. Cremona, *Elliptic Curves Data*, `http://www.warwick.ac.uk/~masgaj/ftp/data/`.

[Cre92] ———, *Abelian varieties with extra twist, cusp forms, and elliptic curves over imaginary quadratic fields*, J. London Math. Soc. (2) **45** (1992), no. 3, 404–416. MR 1180252 (93h:11056)

[Cre97] ———, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, `http://www.warwick.ac.uk/~masgaj/book/fulltext/`.

[Dem05] Lassina Dembélé, *Explicit computations of Hilbert modular forms on* $\mathbb{Q}(\sqrt{5})$, Experiment. Math. **14** (2005), no. 4, 457–466. MR 2193808

[Dem08] ———, *An algorithm for modular elliptic curves over real quadratic fields*, Experiment. Math. **17** (2008), no. 4, 427–438. MR 2484426 (2010a:11119)

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149, `http://arxiv.org/abs/math/0207280`. MR 2068888 (2005f:11128)

[Fis12] Tom Fisher, *On Families of n-congruent Elliptic Curves*, Preprint (2012).

[GV11] Matthew Greenberg and John Voight, *Computing systems of Hecke eigenvalues associated to Hilbert modular forms*, Math. Comp. **80** (2011), no. 274, 1071–1092, `http://www.cems.uvm.edu/~voight/articles/heckefun-021910.pdf`. MR 2772112 (2012c:11103)

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320, `http://wstein.org/papers/bib/Gross-Zagier_Heegner_points_and_derivatives_of_Lseries.pdf`. MR 87j:11057

[Kat81] N. M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. **62** (1981), no. 3, 481–502. MR 82d:14025

[KN12] Sheldon Kamienny and Filip Najman, *Torsion groups of elliptic curves over quadratic fields*, Acta. Arith. **152** (2012), 291–305.

[Kol91] V. A. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 429–436. MR 93c:11046

[Kub76] Daniel Sion Kubert, *Universal bounds on the torsion of elliptic curves*, Proceedings of the London Mathematical Society **s3-33** (1976), no. 2, 193–237.

[LV] E. Larson and D. Vaintrob, *Determinants of subquotients of Galois representations associated to abelian varieties*, arXiv:1110.0255.

[Maz78] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[Rub11] M. O. Rubinstein, *Lcalc*, 2011, `http://oto.math.uwaterloo.ca/~mrubinst/l_function_public/CODE/`.

[S+12] W. A. Stein et al., *Sage Mathematics Software (Version 4.8)*, The Sage Development Team, 2012, `http://www.sagemath.org`.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[SW02] William Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, `http://wstein.org/ecdb`, pp. 267–275. MR 2041090 (2005h:11113)

[Vél71] Jacques Vélu, *Isogénies entre courbes elliptiques*, C. R. Acad. Sci. Paris Sér. A-B **273** (1971), A238–A241.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551, `http://users.tpg.com.au/nanahcub/flt.pdf`.

[Zha01] Shou-Wu Zhang, *Heights of Heegner points on Shimura curves*, Ann. of Math. (2) **153** (2001), no. 1, 27–147. MR 1826411 (2002g:11081)

46 Non-commutative Iwasawa theory for modular forms, with J. Coates, T. Dokchitser, Z. Liang, and R. Sujatha

NON-COMMUTATIVE IWASAWA THEORY FOR MODULAR FORMS

J. COATES, T. DOKCHITSER, Z. LIANG, W. STEIN, R. SUJATHA

Abstract. The aim of the present paper is to give evidence, largely numerical, in support of the non-commutative main conjecture of Iwasawa theory for the motive of a primitive modular form of weight $k > 2$ over the Galois extension of $\mathbb{Q}$ obtained by adjoining to $\mathbb{Q}$ all p-power roots of unity, and all p-power roots of a fixed integer $m > 1$. The predictions of the main conjecture are rather intricate in this case because there is more than one critical point, and also there is no canonical choice of periods. Nevertheless, our numerical data agrees perfectly with all aspects of the main conjecture, including Kato's mysterious congruence between the cyclotomic Manin p-adic L-function, and the cyclotomic p-adic L-function of a twist of the motive by a certain non-abelian Artin character of the Galois group of this extension.

1. Introduction

Let z be a variable in the upper half complex plane, and put $q = e^{2\pi i z}$. Let

$$f(z) = \sum_{n=1}^{\infty} a_n q^n, \tag{1}$$

be a primitive cusp form of conductor N (in the sense of [18]), with trivial character, and weight $k > 2$. For simplicity, we shall always assume that the Fourier coefficients a_n $(n \geq 1)$ of f are in $\mathbb{Q}$. Let p be an odd prime number. The aim of the present paper is to provide some evidence, largely numerical, for the validity of the non-commutative main conjecture of Iwasawa theory for the motive $M(f)$ attached to f over the p-adic Lie extension

$$F_\infty = \mathbb{Q}(\mu_{p^\infty}, m^{1/p^n}, n = 1, 2, \dots),$$

which is obtained by adjoining to $\mathbb{Q}$ the group μ_{p^∞} of all p-power roots of unity, and all p-power roots of some fixed integer $m > 1$. In this case, the analytic continuation and functional equation for the complex L-function $L(f, \phi, s)$ of f twisted by any Artin character ϕ of the Galois group of F_∞ over $\mathbb{Q}$ are well-known consequences of the theory of automorphic base change. The points $s = 1, \dots, k-1$ are critical for all of the complex L-functions $L(f, \phi, s)$, and we show that essentially the same arguments as in [1] enable one to prove the expected algebraicity statement at these points. Moreover, these values

2000 *Mathematics Subject Classification.* Primary 11F67; Secondary 11F33, 11R23.

T. Dokchitser was supported by a Royal Society University Research Fellowship, Z. Liang by the National Natural Sciences Foundation of China (Grant Nos. 11001183 and 11171231), W. Stein from NSF Grant DMS-1015114, and R. Sujatha from NSERC Grant 22R65339.

are all non-zero, except perhaps for the central value $s = k/2$; in particular, there is always at least one non-zero critical value since $k > 2$.

In [2], a precise main conjecture was formulated for an elliptic curve over any p-adic Lie extension of a number field F containing the cyclotomic $\mathbb{Z}_p$-extension of F, and under the assumption that the elliptic curve is ordinary at the prime p. This was generalized to arbitrary ordinary motives in [8], and it is a special case of the main conjecture of [8] which we consider here. Thus we assume that p is an odd prime number such that $(p, a_p) = (p, N) = 1$. One of the underlying ideas of the non-commutative main conjecture is to prove the existence of a p-adic L-function, which interpolates a canonical normalization of the critical values $L(f, \phi, n)$, where $n = 1, \ldots, k-1$, and ϕ runs over all Artin representations of the Galois group

$$G = \mathrm{Gal}(F_\infty/\mathbb{Q}).$$

We denote these normalized L-values by $\mathcal{L}_p^{\mathrm{can}}(f, \phi, n)$ (for the precise definition, see formulae (68), (69) and (71) in §5). The definition of these normalized L-values requires making a choice of canonical periods for the form f, and, until such a time as the main conjectures of non-commutative Iwasawa theory are fully proven, we are only able to make an educated guess at present as to what these canonical periods should be. However, as we explain in §5, Manin's work on the construction of the p-adic L-function for our modular form f over the field $\mathbb{Q}(\mu_{p^\infty})$ gives some information about these canonical periods, which is relevant for our numerical examples.

As we explain in more detail in §5, the existence of a p-adic L-function attached to f over the non-abelian extension F_∞ of $\mathbb{Q}$, when combined with the work of Kato [14], implies the existence of the following mysterious congruence between two p-adic L-functions attached to f over certain abelian sub-extensions of $F_\infty/\mathbb{Q}$. We are very grateful to M. Kakde for explaining to us how this congruence follows from Kato's work. Let σ denote the $(p-1)$-dimensional representation of G given by the direct sum of the irreducible representations of $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$. Let ρ be the unique irreducible representation of dimension $p-1$ of the Galois group of the field

$$F = \mathbb{Q}(\mu_p, m^{1/p})$$

over $\mathbb{Q}$, where we now assume that $m > 1$ is p-power free. Write $\mathbb{Q}^{\mathrm{cyc}}$ for the cyclotomic $\mathbb{Z}_p$-extension of $\mathbb{Q}$, and Ξ for the group of irreducible characters of finite order of $\Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{cyc}}/\mathbb{Q})$. Further, let χ_p denote the character giving the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on μ_{p^∞}. We fix a topological generator γ of Γ, and put $u = \chi_p(\gamma)$. The work of Manin [16] proves that there exists a unique power series $H(\sigma, T)$ in the ring $R = \mathbb{Z}_p[[T]]$ such that

$$H(\sigma, \psi(\gamma)u^r - 1) = \mathcal{L}_p^{\mathrm{can}}(f, \sigma\psi, k/2 + r), \tag{2}$$

for all ψ in Ξ, and all integers r with $-k/2 + 1 \leq r \leq k/2 - 1$. On the other hand, the conjectural existence of a good p-adic L-function for f over the field F_∞ would imply, in particular, the existence of a power series $H(\rho, T)$ in the ring R such that

$$H(\rho, \psi(\gamma)u^r - 1) = \mathcal{L}_p^{\mathrm{can}}(f, \rho\psi, k/2 + r), \tag{3}$$

for all ψ in Ξ, and all integers r with $-k/2+1 \leq r \leq k/2-1$. Then Kato's work [14] implies the following conjectural congruence between formal power series

$$H(\rho, T) \equiv H(\sigma, T) \mod pR. \tag{4}$$

This conjectural congruence in R has the following consequences for our critical L-values. Firstly, on evaluation of our power series at the relevant point in $p\mathbb{Z}_p$, we deduce from (2) and (3) that the congruence

$$\mathcal{L}_p^{\mathrm{can}}(f, \rho, n) \equiv \mathcal{L}_p^{\mathrm{can}}(f, \sigma, n) \bmod p\mathbb{Z}_p \tag{5}$$

should hold for $n = 1, \ldots, k-1$. Secondly, if we assume the additional property that

$$L(f, \sigma, k/2) = L(f, \rho, k/2) = 0, \tag{6}$$

then we would have that $H(\rho, T)$ and $H(\sigma, T)$ both belong to the ideal TR. It is then clear from (2), (3) and (4) that the stronger congruence

$$\mathcal{L}_p^{\mathrm{can}}(f, \rho, n) \equiv \mathcal{L}_p^{\mathrm{can}}(f, \sigma, n) \bmod p^2\mathbb{Z}_p \tag{7}$$

should hold for $n = 1, \ldots, k-1$.

Our numerical computations (see §6) verify the first congruence (5) for the prime $p = 3$ and a substantial range of cube free integers $m > 1$, for three forms f of weight 4 and conductors 5, 7, 121, and one form f of weight 6 and conductor 5, all of which are ordinary at 3. These computations require us to determine numerically the Fourier coefficients a_n of these forms f for n in the range $1 \leq n \leq 10^8$. In addition, for the two forms of weight 4 and conductors 7 and 121, we prove that (6) holds for all integers $m > 1$, and happily, our numerical results show that the sharper congruence (7) holds for these two forms and the prime $p = 3$ for a good range of cube free integers $m > 1$. When f is a complex mutliplication form, some cases of the congruence (4) have already been established theoretically by Delbourgo and Ward [3] and Kim [15]. However, when f is not a complex multiplication form, our numerical data seems to provide the first hard evidence in support of the mysterious non-abelian congruence (4) between abelian p-adic L-functions.

We warmly thank T. Bouganis, M. Kakde, and D. Kim for very helpful advice on the writing of this paper.

2. Algebraicity of L-values

As in the Introduction, let f given by (1) be a primitive cusp form of conductor $N \geq 1$ with trivial character and weight $k > 2$ (thus k is necessarily even). For simplicity, we always assume that the Fourier coefficients $a_n\,(n \geq 1)$ of f belong to $\mathbb{Q}$. The complex L-function attached to f is

$$L(f, s) = \sum_{n=1}^{\infty} a_n/n^s. \tag{8}$$

This L-function has the following Euler product. For any prime p, let

$$\tau_p : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_{\mathbb{Q}_p}(V_p) \tag{9}$$

be the p-adic Galois representation attached to f; here V_p is a two dimensional vector space over the field $\mathbb{Q}_p$ of p-adic numbers. If q is any prime distinct from p, define the polynomial

$$P_q(f, X) = \det(1 - \tau_p(\mathrm{Frob}_q^{-1})X \mid V_p^{I_q}), \tag{10}$$

where I_q is the inertial subgroup of the decomposition group of any fixed prime of $\bar{\mathbb{Q}}$ above q, and Frob_q denote the Frobenius automorphism of q. Moreover, if $(q, N) = 1$, we have

$$P_q(f, X) = 1 - a_q X + q^{k-1}X^2. \tag{11}$$

Then

$$L(f, s) = \prod_q P_q(f, q^{-s})^{-1} \tag{12}$$

when $\mathrm{Re}(s) > 1 + (k-1)/2$. Defining

$$\Lambda(f, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f, s), \tag{13}$$

we know, since Hecke, that $\Lambda(f, s)$ is entire and satisfies the functional equation

$$\Lambda(f, s) = w(f)\Lambda(f, k - s) \tag{14}$$

where $w(f) = \pm 1$ is the sign in the functional equation. The critical values of $L(f, s)$ are at the points $s = 1, \dots, k-1$. Following Shimura [23], [24], we introduce the following naive periods for f, which we have normalized in view of our later numerical calculations. Define

$$\Omega_-(f) = iw(f)(2\pi)^{-1}L(f, 1). \tag{15}$$

Since the Euler product for $L(f, s)$ converges to a positive real number when s is real and $s > 1 + (k-1)/2$, it is clear from the functional equation (14), that $\Omega_-(f)$ is purely imaginary in the upper half plane. Motivated again by numerical calculations, we assume throughout the following simplifying hypothesis (see [7] for examples in which this hypothesis fails).

Hypothesis H1: $L(f, 2) \neq 0$ when $k = 4$.

We then define

$$\Omega_+(f) = w(f)(2\pi)^{-2}\, L(f, 2). \tag{16}$$

Again, $\Omega_+(f)$ is always a positive real number when $k > 4$, and presumably (it would, of course, be implied by the generalized Riemann Hypothesis) this remains true even when $k = 4$, although this value is outside the region of convergence of the Euler product.

Theorem 2.1. (See [23],[24]) *(i) If n is an odd integer such that $1 \leq n \leq k-1$, then*

$$(2\pi i)^{-n} L(f,n)/\Omega_-(f) \in \mathbb{Q};$$

(ii) If n is an even integer such that $1 \leq n \leq k-1$, then

$$(2\pi i)^{-n} L(f,n)/\Omega_+(f) \in \mathbb{Q}.$$

In what follows, we shall mainly be interested in the L-functions of f twisted by Artin characters. We rapidly recall the definitions of these L-functions. By an Artin representation, we mean a homomorphism

$$\phi\,:\,\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_{\bar{\mathbb{Q}}}(W) \tag{17}$$

which factors through the Galois group of a finite extension of $\mathbb{Q}$; here W is a vector space of finite dimension over $\bar{\mathbb{Q}}$. Put

$$d(\phi) = \dim_{\bar{\mathbb{Q}}}(W).$$

For each prime p, let

$$M_p(f) = V_p \otimes_{\mathbb{Q}_p} \bar{\mathbb{Q}}_p,\;\; M_p(\phi) = W \otimes_{\bar{\mathbb{Q}}} \bar{\mathbb{Q}}_p.$$

Then

$$L(f,\phi,s) = \prod_q P_q(f,\phi,q^{-s})^{-1}, \tag{18}$$

where

$$P_q(f,\phi,X) = \det\left((1-\mathrm{Frob}_q^{-1}X) \mid (M_p(f) \otimes_{\bar{\mathbb{Q}}_p} M_p(\phi))^{I_q}\right) \;\; (q \neq p) \tag{19}$$

is the Euler product attached to the tensor product Galois representation $M_p(f) \otimes_{\bar{\mathbb{Q}}_p} M_p(\phi)$. The Euler product (18) converges in the region $\mathrm{Re}(s) > 1 + (k-1)/2$. It is one of the fundamental problems of number theory to prove the analytic continuation and the following conjectural functional equation for $L(f,\phi,s)$. Let $N(f,\phi)$ be the conductor of the family of p-adic representations $M_p(f) \otimes_{\bar{\mathbb{Q}}_p} M_p(\phi)$, and define

$$\Lambda(f,\phi,s) = N(f,\phi)^{s/2} \left((2\pi)^{-s}\Gamma(s)\right)^{d(\phi)} L(f,\phi,s). \tag{20}$$

Then conjecturally

$$\Lambda(f,\phi,s) = w(f,\phi)\Lambda(f,\hat{\phi},k-s), \tag{21}$$

where $w(f,\phi)$ is an algebraic number of complex absolute value 1, and $\hat{\phi}$ is the contragredient representation of ϕ. There is one important case in which this result is known.

Theorem 2.2. *Let K be any finite Galois extension of $\mathbb{Q}$ with Galois group* $\mathrm{Gal}(K/\mathbb{Q})$ *abelian. Let ψ be an abelian character of K and define ϕ to be the induced character of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. *Then $\Lambda(f,\phi,s)$ is entire and satisfies the functional equation* (21).

Proof. Since K is an abelian extension of $\mathbb{Q}$, the base change of f to K, which we denote by $\pi_K(f)$, exists as a cuspidal automorphic representation of GL_2/K. The results of Jacquet-Langlands then establish the analytic continuation and functional equation for the automorphic L-function of $\pi_K(f)$, twisted by the abelian character ψ of K, which we view as a Hecke character of GL_1/K. We denote this automorphic L-function by $L(\pi_K(f), \psi, s)$. On the other hand, by the theory of base change, and the local Langlands correspondence for GL_2, $L(\pi_K(f), \psi, s)$ coincides with $L(f, \phi, s)$ defined by the Euler product (18). This completes the proof on noting that the functional equation (21) coincides with the automorphic functional equation. □

The following conjectural generalisation of Theorem 2.1 is folklore. Given an Artin representation ϕ as in (17), define $d^+(\phi)$ (resp. $d^-(\phi)$) to be the dimension of the subspaces of W on which complex conjugation acts like $+1$, (resp. as -1). If n is any integer, we write

$$d_n^+(\phi) = d^{(-1)^n}(\phi), \; d_n^-(\phi) = d^{(-1)^{n+1}}(\phi). \tag{22}$$

Conjecture 2.3. *For every Artin representation ϕ of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, *and all integers $n = 1, \dots, k-1$, we have*

$$\frac{L(f, \phi, n)}{\left((2\pi i)^{nd(\phi)} \times \Omega_+(f)^{d_n^+(\phi)} \times \Omega_-(f)^{d_n^-(\phi)}\right)} \in \bar{\mathbb{Q}}. \tag{23}$$

Of course, when ϕ has dimension 1, this conjecture is a well known consequence of the theory of higher weight modular symbols. However, as in [1], we shall study special cases of this conjecture by using the work of Shimura [23] on the special values of Rankin products of Hilbert modular forms for totally real number fields. Let K be an arbitrary totally real field, which is Galois over $\mathbb{Q}$, with $\mathrm{Gal}(K/\mathbb{Q})$ abelian. Take $\mathfrak{g}$ to be any Hilbert modular form relative to K, which corresponds to an Artin representation θ of dimension 2 of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$. The form $\mathfrak{g}$ has parallel weight 1 and level equal to the conductor of θ. We denote the Artin L-series of θ by

$$L(\theta, s) = \sum_{\mathfrak{a}} c(\mathfrak{a})(N\mathfrak{a})^{-s},$$

where $\mathfrak{a}$ runs over all integral ideals of K. Further, let $L(f/K, s)$ be the complex L-function attached to the restriction of the Galois representation (13) to $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$, and write

$$L(f/K, s) = \sum_{\mathfrak{a}} b(\mathfrak{a})(N\mathfrak{a})^{-s},$$

for its corresponding Dirichlet series. Since we have assumed K to be an abelian extension of $\mathbb{Q}$, the base change to K of our modular form f also exists as a primitive cusp form for the Hilbert modular group of K. We denote this base change by $\mathfrak{f}_K$. It has parallel weight k, trivial character, and level dividing $N\mathcal{O}_K$, where $\mathcal{O}_K$ is the ring of integers of K. In what follows, we will be primarily interested in the complex L-series defined by the tensor product of the Artin representation θ and the Galois representation (13) of f restricted to $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$. We denote this L-series by $L(\mathfrak{f}_K, \theta, s)$, and recall that it is defined by the Euler product

$$L(\mathfrak{f}_K, \theta, s) = \prod_v P_v(\mathfrak{f}_K, \theta, (Nv)^{-s})^{-1},$$

where v runs over all finite places of K, and

$$P_v(\mathfrak{f}_K, \theta, X) = \det\left(1 - \mathrm{Frob}_v^{-1} X \mid \left(M_p(f) \otimes_{\bar{\mathbb{Q}}_p} W_\theta\right)^{I_v}\right); \tag{24}$$

here W_θ is a two dimensional $\bar{\mathbb{Q}}_p$-vector space realizing θ, and I_v is the inertial subgroup of a place of $\bar{\mathbb{Q}}$ above v. Of course, by the inductive property of L-functions, we also have

$$L(\mathfrak{f}_K, \theta, s) = L(f, \phi_\theta, s), \tag{25}$$

where ϕ_θ is the Artin representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ induced from the representation θ of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$.

On the other hand, the classical theory of Rankin products (see [25, §4]) considers instead the complex L-series $\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s)$ defined by

$$\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s) = L_{\mathfrak{n}}(\psi, 2s - k + 1) \times \sum_{\mathfrak{a}} c(\mathfrak{a}) b(\mathfrak{a}) N(\mathfrak{a})^{-s} \tag{26}$$

with $\mathfrak{a}$ running over all integral ideals of K; here $\mathfrak{n}$ is the least common multiple of the levels of $\mathfrak{f}_K$ and $\mathfrak{g}$, ψ is the character of $\mathfrak{g}$, and $L_{\mathfrak{n}}(\psi, s)$ is the imprimitive L-series of ψ where the Euler factors at the primes dividing $\mathfrak{n}$ have been omitted. A well-known classical argument shows that $\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s)$ has the Euler product expansion

$$\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s) = \prod_v D_v(\mathfrak{f}_K, \mathfrak{g}, (Nv)^{-s})^{-1} \tag{27}$$

where

$$D_v(\mathfrak{f}_K, \mathfrak{g}, X) = \det\left(1 - \mathrm{Frob}_v^{-1} X \mid \left(M_p(f)^{I_v} \otimes_{\bar{\mathbb{Q}}_p} W_\theta^{I_v}\right)\right). \tag{28}$$

Thus the complex L-functions $L(\mathfrak{f}_K, \theta, s)$ and $\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s)$ coincide, except for the possible finite set of Euler factors at places v for which

$$\left(M_p(f) \otimes_{\bar{\mathbb{Q}}_p} W_\theta\right)^{I_v} \neq \left(M_p(f)^{I_v} \otimes_{\bar{\mathbb{Q}}_p} W_\theta^{I_v}\right). \tag{29}$$

To avoid this technical difficulty, we impose an additional simplifying hypothesis.

Lemma 2.4. *Assume that for each prime q such that $q^2 \mid N$, that q does not divide the conductor of the representation of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ *induced from θ. Then for every prime number p, and every finite place v of $\mathcal{K}$ which does not lie above p, we have*

$$\left(M_p(f) \otimes_{\bar{\mathbb{Q}}_p} W_\theta)\right)^{I_v} = (M_p(f))^{I_v} \otimes_{\bar{\mathbb{Q}}_p} (W_\theta)^{I_v}, \tag{30}$$

where I_v denotes the inertial subgroup at v. In particular, $\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s) = L(\mathfrak{f}_K, \theta, s)$.

Proof. Suppose that v lies above a prime q, where $q \neq p$. Assume first that $(q, N) = 1$. Then I_q, and hence also I_v, acts trivially on $M_p(f)$, and so (30) is plain. Suppose next that q divides N but q^2 does not divide N. Then it is well known that the image of I_q, hence also that of I_v, in the automorphism group of $M_p(f)$ is infinite, and that $M_p(f)^{I_q}$ has dimension one over $\bar{\mathbb{Q}}_p$. Clearly the same assertions remain valid if we replace I_q by any open subgroup I'_q of I_q. Thus we must have $M_p(f)^{I_v} = M_p(f)^{I'_v}$ for every open subgroup I'_v of I_v. Since some open subgroup of I_v acts trivially on W_θ, (30) follows immediately. Finally, if q^2 divides N, the hypothesis of the lemma shows that I_v acts trivially on W_θ, whence (30) is again clearly true.

□

Our next result relates the automorphic period of $\mathfrak{f}_K$ to the periods $\Omega^+(f)$ and $\Omega^-(f)$. We normalize the Petersson inner product on the space of cusp forms of level dividing $N\mathcal{O}_K$ for the Hilbert modular group of K as in [25] (see formula (2.7) on p. 651).

Proposition 2.5. *Let K be a real abelian field, and write $\mathfrak{f}_K$ for the base change of f to K. Then*

$$\frac{(2\pi i)^{(1-k)\beta}\pi^{\beta k}\langle \mathfrak{f}_K, \mathfrak{f}_K\rangle_K}{(\Omega_+(f) \times \Omega_-(f))^\beta} \in \mathbb{Q}, \tag{31}$$

where $\beta = [K : \mathbb{Q}]$.

We shall use the following notation in the proof of this proposition. If ψ is any abelian character of K, write $L(f/K, \psi, s)$ for the primitive L-function attached to the tensor product of ψ with the restriction of (13) to $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$. Also, for any abelian character χ of $\mathbb{Q}$, we write χ_K for the restriction of χ to $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$.

Lemma 2.6. *Let K be any real abelian extension of $\mathbb{Q}$ and η any abelian character of $\mathbb{Q}$. Then there exists an abelian character χ of $\mathbb{Q}$ as follows. For all σ in $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, we have*

1. $L(f/K, \chi_K^\sigma, k/2) \neq 0$ and $L(f/K, \chi_K^\sigma\eta_K, k/2) \neq 0$;
2. *$L(f/K, \chi_K^\sigma, s)$ (resp. $L(f/K, \chi_K^\sigma\eta_K, s)$) has Euler factor equal to 1 at all places of K where χ_K^σ (resp. $\chi_K^\sigma\eta_K$) is ramified.*

Proof. Let Σ be any finite set of primes of $\mathbb{Q}$ containing the primes dividing N, the primes dividing the conductor of η, and the primes which ramify in K. By an important theorem of Rohrlich [22], there exists a finite abelian extension M of $\mathbb{Q}$, unramified outside Σ, such that $L(f, \lambda, k/2) \neq 0$ for every abelian character λ of $\mathbb{Q}$ that is unramified outside Σ, and which does not factor through $\mathrm{Gal}(M/\mathbb{Q})$. By enlarging M if necessary, we can assume that $M \supset K$. Let $\mathcal{N}_K$ be the conductor of f/K, and $\Delta_{M/K}$ the relative discriminant of M over K. Also, if ξ is an abelian character of K, write $\mathcal{N}(\xi)$ for its conductor. Let χ be any abelian character of $\mathbb{Q}$ such that, for every prime v of K above Σ, we have

$$\mathrm{ord}_v(\mathcal{N}(\chi_K)) > \max\bigl\{\mathrm{ord}_v(\Delta_{M/K}), \mathrm{ord}_v(\mathcal{N}(\eta)), \mathrm{ord}_v(\mathcal{N}_K)\bigr\}. \tag{32}$$

Such a character can always be found by taking a character of $\mathrm{Gal}(\mathbb{Q}(\mu_m)/\mathbb{Q})$ for sufficiently large m. Let v be any place of K above Σ, and put

$$t_v = \max\bigl\{\mathrm{ord}_v(\Delta_{M/K}), \mathrm{ord}_v(\mathcal{N}_K)\bigr\}.$$

Thanks to (32), it is clear, that for each σ in $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, we have

$$\operatorname{ord}_v(\mathcal{N}(\chi_K^\sigma)) = \operatorname{ord}_v(\mathcal{N}(\chi_K^\sigma\eta)) > t_v. \tag{33}$$

In particular, none of these characters can factor through $\mathrm{Gal}(M/K)$. Moreover, it is also easily seen from (33) that

$$\left(M_p(f)\otimes_{\bar{\mathbb{Q}}_p}\chi_K^\sigma\right)^{I_v} = \left(M_p(f)\otimes_{\bar{\mathbb{Q}}_p}\chi_K^\sigma\eta\right)^{I_v} = 0,$$

whence the final assertion of the lemma is clear. □

We now prove that the left hand side of (31) is an algebraic number. Take $J = K(i)$, and let η be any abelian character of $\mathbb{Q}$ such that J is the fixed field of the kernel of η_K. Now let χ be an abelian character of $\mathbb{Q}$ having the properties specified in Lemma 2.6, and write χ_J for the restriction of χ to $\mathrm{Gal}(\bar{\mathbb{Q}}/J)$. Note that the representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ induced by χ_J is $\theta = \chi_K \oplus \chi_K\eta_K$. Write $\mathfrak{g}$ for the Hilbert modular form relative to K which corresponds to θ. Thus $\mathfrak{g}$ has parallel weight one, and character $\eta_K\chi_K^2$. Moreover, by the second assertion of Lemma 2.5, we have the exact equality of L-functions

$$\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s) = L(f/K, \theta, s). \tag{34}$$

On the other hand, since K is abelian over $\mathbb{Q}$, we also have the identity

$$L(f/K,\theta,s) = \prod_{j=1}^{\beta} L(f,\chi\zeta_j,s)L(f,\chi\eta\zeta_j,s), \tag{35}$$

where $\zeta_1,\dots,\zeta_d$ denote the characters of $\mathrm{Gal}(K/\mathbb{Q})$.

The desired algebraicity assertion follows by evaluating both sides of (35) at $s = k/2$, noting that this common value is non-zero by Lemma 2.6, and then applying Shimura's algebraicity results to each L-function. Indeed, Theorem 4.2 of [25] shows that

$$\frac{\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, k/2)}{(2\pi i)^\beta \pi^{\beta k}\langle \mathfrak{f}_K, \mathfrak{f}_K\rangle \tau_K(\eta_K\chi_K^2)} \in \bar{\mathbb{Q}}, \tag{36}$$

where $\tau_K(\eta_K\chi_K^2)$ denotes the Gauss sum for the character $\eta_K\chi_K^2$ of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ (see (3.9) of [25] for the definition of this Gauss sum). On the other hand, recalling that $\chi\zeta_j(-1) \neq \chi\zeta_j\eta(-1)$ for $j = 1,\dots,\beta$, it follows from [24, Theorem 1] that

$$\frac{\prod_{j=1}^{\beta} L(f,\chi\zeta_j,k/2)\times L(f,\chi\eta\zeta_j,k/2)}{(2\pi i)^{\beta k}\times(\Omega_+(f)\Omega_-(f))^\beta\times\prod_{j=1}^{\beta}\tau_{\mathbb{Q}}(\chi\zeta_j)\tau_{\mathbb{Q}}(\chi\eta\zeta_j)} \in \bar{\mathbb{Q}}; \tag{37}$$

here $\tau_{\mathbb{Q}}(\kappa)$ denotes the usual Gauss sum of an abelian character κ of $\mathbb{Q}$. Combining (36)and (37), it follows immediately that the left hand side of (31) is an algebraic number. Moreover, a more detailed analysis, exactly as in the proof of [1, Theorem 3.4] shows that this algebraic number is invariant under the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, completing the proof of Proposition 2.5. □

As a first application of Proposition 2.5, we establish the following case of Conjecture 2.3.

Theorem 2.7. *Assume F is an imaginary number field with* $\mathrm{Gal}(F/\mathbb{Q})$ *abelian. Let ψ be any abelian character of* $\mathrm{Gal}(\bar{\mathbb{Q}}/F)$*, and let ϕ be the induced character of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$*. Assume that, for every prime q such that q^2 divides N, q does not divide the conductor of ϕ. Then Conjecture 2.3 is valid for f and ϕ.*

Proof. Let K be the maximal real subfield of F, and let θ be the representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ induced from ψ. Thus θ is a two dimensional Artin representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$, and we let $\mathfrak{g}$ be the associated Hilbert modular form as above. Then, by Lemma 30,

$$\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, s) = L(\mathfrak{f}_K, \theta, s) = L(f, \phi, s). \tag{38}$$

But, assuming n is an integer with $1 \leq n \leq k-1$, it is shown in [25, (4.10)] that

$$\frac{(2\pi i)^{-2n\beta}\mathcal{D}(\mathfrak{f}_K, \mathfrak{g}, n)}{(2\pi i)^{\beta(1-k)}\pi^{\beta k}\langle \mathfrak{f}_K, \mathfrak{f}_k\rangle_K} \in \bar{\mathbb{Q}}. \tag{39}$$

Now making use of Lemma 2.6, and noting that

$$d(\phi) = 2\beta,\ d_n^+(\phi) = d_n^-(\phi) = \beta,$$

the algebraicity statement (23) follows on putting $s = n$ in (38). This completes the proof of Theorem 2.7. □

Again following the ideas of [1], we now prove a refined version of Conjecture 2.3 for Artin representations ϕ which factor through the Galois group over $\mathbb{Q}$ of the field

$$F_r = \mathbb{Q}(\mu_{p^r}, m^{1/p^r}); \tag{40}$$

here p is an odd prime number, $r \geq 1$ is an integer, μ_{p^r} is the group of p^r-th roots of unity, and m is an integer > 1. For simplicity, we shall always assume that m is not divisible by the p-th power of an integer > 1. In order to state the refinement of (23), we first recall the epsilon-factors of the Artin representation ϕ (for a fuller discussion, see [5, §6.2]). Fix the Haar measure μ on $\mathbb{Q}_p$ determined by $\mu(\mathbb{Z}_p) = 1$, and the additive character α of $\mathbb{Q}_p$ given by

$$\alpha(zp^{-t}) = e^{2i\pi z/p^t}, \text{ for } z \in \mathbb{Z}_p.$$

Write $\epsilon_p(\phi)$ for the local epsilon-factor of ϕ at the prime p, which is uniquely determined by this choice of μ and α. For each integer $n = 1, \ldots, k-1$, define

$$L_p^*(f, \phi, n) = \frac{L(f, \phi, n)\epsilon_p(\phi)}{\left((2\pi i)^{nd(\phi)} \times \Omega_+(f)^{d_n^+(\phi)} \times |\Omega_-(f)|^{d_n^-(\phi)}\right)} \tag{41}$$

Hypothesis H2: For all primes q such that q^2 divides N, we have $(q, mp) = 1$.

Theorem 2.8. *Assume that the Artin representation ϕ factors through* $\mathrm{Gal}(F_r/\mathbb{Q})$ *for some integer $r \geq 1$, where F_r is given by* (40)*. Suppose in addition that Hypotheses H1* *and H2 are valid. Then* $\Lambda(f, \phi, s)$ *is entire, and satisfies the functional equation* (21). *Moreover, for all integers* $n = 1, \ldots, k-1$, $L_p^*(f, \phi, n)$ *is an algebraic number satisfying*

$$L_p^*(f, \phi, n)^\sigma = L_p^*(f, \phi^\sigma, n) \tag{42}$$

for all σ *in* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$.

Proof. As remarked above, the proof we now give follows closely that given in [1], where f was assumed to have weight $k = 2$, and therefore corresponded to an isogeny class of elliptic curves defined over $\mathbb{Q}$. For each integer $r \geq 1$, define $\mathcal{K}_r = \mathbb{Q}(\mu_{p^r})$ and write K_r for its maximal real subfield. Note that $\mathrm{Gal}(F_r/\mathcal{K}_r)$ is cyclic of order p^r, since m is assumed to be p-power free. Put

$$F_\infty = \bigcup_{r \geq 1} F_r, \ \mathcal{K}_\infty = \bigcup_{r \geq 1} \mathcal{K}_r, \ G = \mathrm{Gal}(F_\infty/\mathbb{Q}). \tag{43}$$

For this proof, define ρ to be the representation of $\mathrm{Gal}(F_r/\mathbb{Q})$ induced by any character of exact order p^r of $\mathrm{Gal}(F_r/\mathcal{K}_r)$. It is then easy to see that ρ is irreducible, and that every irreducible Artin representation ϕ of G is of the form λ or $\rho\lambda$ for some integer $r \geq 1$,where λ is a one dimensional character of $\mathrm{Gal}(\mathcal{K}_\infty/\mathbb{Q})$. For the proof of Theorem 2.8, we may assume that ϕ is irreducible. Now it is clear from these remarks that every irreducible Artin representation ϕ of G is induced from an abelian character of $\mathcal{K}_r$ for some integer $r \geq 1$. Thus Theorem 2.2 implies that $\Lambda(f, \phi, s)$ is entire and satisfies the functional equation (21). Also, noting that $F_\infty/\mathbb{Q}$ is unramified outside of the primes dividing mp, we conclude from Theorem 2.7 that $L_p^*(f, \phi, n)$ is an algebraic number for all Artin characters ϕ of G and all integers $n = 1, \ldots, k-1$. Thus it remains to establish (42) for irreducible ϕ.

If $d(\phi) = 1$, one can easily deduce (42) from [23, Theorem 1]. Assuming $d(\phi) > 1$, it follows that for some integer $r \geq 1$, ϕ is induced by an abelian character of $\mathcal{K}_r$ of the form $\psi\lambda_{\mathcal{K}_r}$, where ψ is a character of $\mathrm{Gal}(F_r/\mathcal{K}_r)$ of exact order p^r, and $\lambda_{\mathcal{K}_r}$ is the restriction to $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathcal{K}_r)$ of a one dimensional character λ of $\mathrm{Gal}(\mathcal{K}_\infty/\mathbb{Q})$. We define θ to be the two dimensional Artin representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/K_r)$ induced by $\psi\lambda_{\mathcal{K}_r}$, and take $\mathfrak{g}$ to be the corresponding Hilbert modular form relative to K_r of parallel weight one. Let ν be the abelian character of K_r defining the quadratic extension $\mathcal{K}_r/K_r$, and let λ_{K_r} be the restriction of λ to $\mathrm{Gal}(\bar{\mathbb{Q}}/K_r)$. Since the determinant of θ is equal to $\nu\lambda_{K_r}^2$, $\mathfrak{g}$ will have character $\nu\lambda_{K_r}^2$. Moreover, noting that Hypothesis H2 is valid for $\mathfrak{f}$ and ϕ because the conductor of ϕ can only be divisible by primes dividing mp, we conclude from Lemma 2.4 that

$$\mathcal{D}(\mathfrak{f}_{K_r}, \mathfrak{g}, s) = L(f_{K_r}, \theta, s) = L(f, \phi, s). \tag{44}$$

Taking $s = n$ with $1 \leq n \leq k-1$, it then follows from [25, Theorem 4.2] that

$$\mathcal{A}(f, \phi, n) := \frac{L(f, \phi, n)}{(2\pi i)^{d(\phi)(1+2n-k)/2} \times \pi^{kd(\phi)/2} \times \langle \mathfrak{f}_K, \mathfrak{f}_K \rangle_{K_r} \times \tau_{K_r}(\nu\lambda_{K_r}^2)} \tag{45}$$

satisfies

$$\mathcal{A}(f, \phi, n)^\sigma = \mathcal{A}(f, \phi^\sigma, n)$$

for all σ in $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$; here $\tau_{K_r}(\nu\lambda_{K_r}^2)$ is the Gauss sum of the abelian character $\nu\lambda_{K_r}^2$ of K_r, as defined by [25, (3.9)]. Noting that

$$d(\phi) = 2[K_r : \mathbb{Q}],\ d^+(\phi) = d^-(\phi) = [K_r : \mathbb{Q}],$$

we conclude easily from (31) and [1, Proposition 4.5] that the last assertion of Theorem 2.8 will hold if and only if

$$\left(\prod_{q \neq p,\infty} \epsilon_q(\phi)\right)^{\sigma} = \prod_{q \neq p,\infty} \epsilon_q(\phi^{\sigma}), \tag{46}$$

for all σ in $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. But (46) is an immediate consequence of the fact that ρ can be realized over $\mathbb{Q}$, and the equation

$$\epsilon_q(\phi) = \epsilon_q(\lambda)^{\dim(\rho)} \operatorname{sgn}\left(\det(\rho)(q^{e_q(\rho)})\right),$$

which holds for all $q \neq p$, since λ is unramified at q; here $e_q(\rho) = \mathrm{ord}_q(\mathcal{N}(\rho))$, with $\mathcal{N}(\rho)$ denoting the conductor of ρ. This completes the proof. □

3. Interlude on root numbers

Recall that $F = \mathbb{Q}(\mu_p, m^{1/p})$. From now on, write ρ for the unique irreducible representation of $\mathrm{Gal}(F/\mathbb{Q})$ of dimension $p-1$; it is induced from any non-trivial character of $\mathrm{Gal}(F/\mathcal{K})$. We now describe the local root numbers

$$w_q(f,\rho) = \frac{\epsilon_q(f,\rho)}{|\epsilon_q(f,\rho)|}$$

and the corresponding global root number $w(f,\rho)$ under the hypothesis H2 above. This global root number is the sign in the functional equation of the twisted L-function $L(f,\rho,s)$. A similar computation in weight 2, i.e., for elliptic curves, was carried out by V. Dokchitser [6].

Theorem 3.1. *Let $f = \sum_n a_n e^{2\pi i n z}$ be a primitive cusp form of conductor N with trivial character, and weight $k \geq 2$. Assume that for all primes q such that $q^2 | N$, we have $(q, mp) = 1$. Then, for every finite prime q, the local root number $w_q(f,\rho)$ is given by*

$$w_q(f,\rho) = w_q(\rho)^2 \times \begin{cases} (\frac{q}{p})^{\mathrm{ord}_q(N)} & \text{if } (q,pm) = 1, \\ -\operatorname{sgn} a_p & \text{if } q = p,\ \mathrm{ord}_p(N) = 1 \text{ and } m^{p-1} \equiv 1 \mod p^2, \\ 1 & \text{otherwise.} \end{cases}$$

Further, the global root number is given by

$$w(f,\rho) = (-1)^{\frac{p-1}{2}}\, \delta \prod_{q \nmid pm} \left(\frac{q}{p}\right)^{\mathrm{ord}_q(N)},$$

where $\delta = -\operatorname{sgn} a_p$ when both $\mathrm{ord}_p(N) = 1$ and $m^{p-1} \equiv 1 \bmod p^2$, and 1 otherwise.

Proof. Let l be any prime distinct from q, and as before, let V_l be the l-adic Galois representation attached to f. Put $n(V) = \mathrm{ord}_q(N)$, and let $n(\rho)$ be such that $q^{n(\rho)}$ is the q-part of the conductor of ρ. We note that the determinant $\det\rho$ of ρ equals $(\frac{\cdot}{p})$, the non-trivial quadratic character of $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$. Recall that the inverse local Euler factors of $L(f,s)$ are

$$P_q(f,T) = \begin{cases} 1 - a_qT + q^{k-1}T^2 & \text{if } \ (q,N) = 1, \\ 1 - a_qT & \text{if } \ \mathrm{ord}_q(N) = 1, \\ 1 & \text{if } \ \mathrm{ord}_q(N) \geq 2. \end{cases}$$

The local root numbers $w_q(f,\rho)$ can be computed as follows:-

Case 1 ($(q,N) = 1$): In this case V_l is unramified, and we can use the unramified twist formula [26, 3.4.6],

$$w_q(f,\rho) = w_q(\rho)^{\dim V_l} \cdot \mathrm{sgn}((\det V_l)(q^{n(\rho)})).$$

Here $\mathrm{sgn}\, z = \frac{z}{|z|}$ for $z \in \mathbb{C}$, and we evaluate the one-dimensional character $\det V$ on a number $q^{n(\rho)} \in \mathbb{Q}_q^\times$ via the local reciprocity map. The second term is trivial since $\det V_l$ is a power of a cyclotomic character which takes positive values on $\mathbb{Q}_q^\times$. Thus $w_q(f,\rho) = w_q(\rho)^2$, as asserted.

Case 2 ($\mathrm{ord}_q(N) = 1$): Here $a_q \neq 0, \dim V_l^{I_q} = 1$, the action of inertia I_q is unipotent $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$, and the action of Frobenius is $\left(\begin{smallmatrix} a_q & * \\ 0 & a_q^{-1}p^{k/2} \end{smallmatrix}\right)$, where k is the weight of f; the top left corner can be seen e.g. from the local factor. Write $(V_l \otimes \rho)^{ss}$ for the semi-simplification of $V_l \otimes \rho$. Writing $\tau = \mathrm{Frob}_q$, the semi-simplification formula for ϵ-factors [26, 4.2.4] gives

$$\begin{aligned} w_q(f,\rho) &= w_q((V_l \otimes \rho)^{ss})\, \tfrac{\mathrm{sgn}\det(-\tau|((V_l\otimes\rho)^{ss})^{I_q})}{\mathrm{sgn}\det(-\tau|(V_l\otimes\rho)^{I_q})} \\ &= w_q(\rho \oplus \rho)\tfrac{\mathrm{sgn}\det(-a_q\tau|\rho^{I_q})\mathrm{sgn}\det(-a_q^{-1}p^{k/2}\tau|\rho^{I_q})}{\mathrm{sgn}\det(-a_q\tau|\rho^{I_q})} \\ &= \mathrm{sgn}\det(-a_q\tau|\rho^{I_q})^{-1} \\ &= w_q(\rho)^2\mathrm{sgn}(-a_q)^{d_q}\mathrm{sgn}\det(\tau|\rho^{I_q})^{-1}; \end{aligned}$$

here d_q denotes the dimension of ρ^{I_q}. It remains to determine ρ^{I_q} and the action of Frobenius on it. Let $J = \mathbb{Q}(m^{1/p})$. There is an equality of L-functions

$$\zeta_J(s) = \zeta(s)L(\rho,s).$$

By considering the ramification of q in $J/\mathbb{Q}$ and comparing the local factors at q, we find that

$$P_q(\rho,T) = \begin{cases} 1 + \cdots + (\frac{q}{p})T^{p-1} & \text{if } \ q \nmid pm, \\ 1 & \text{if } \ q|m, \\ 1 - T & \text{if } \ p = q \text{ and } m \text{ is a } p\text{th power in } \mathbb{Z}_p^\times, \\ 1 & \text{if } \ p = q \text{ and } m \text{ is not a } p\text{th power in } \mathbb{Z}_p^\times. \end{cases}$$

In particular, d_q is even (and so $\mathrm{sgn}\,(-a_q)^{d_q} = 1$) in all but the third case, and $\det(\tau|\rho^{I_q}) = 1$ in all but the first case; in the first case,

$$\det(\tau|\rho^{I_q}) = \det(\tau|\rho) = \Big(\frac{q}{p}\Big) = \Big(\frac{q}{p}\Big)^{\mathrm{ord}_q(N)},$$

as asserted by the formula. Finally, if $p \nmid m$, then it is easy to see by Hensel's lemma that m is a pth power in $\mathbb{Z}_p^\times$ if and only if it is a pth power in $(\mathbb{Z}/p^2\mathbb{Z})^\times$, which is in turn equivalent to the condition $m^{p-1} \equiv 1 \mod p^2$.

Case 3 ($\mathrm{ord}_q(N) \geq 2$): By assumption, $q \nmid mp$, so ρ is unramified. Then $w_q(\rho) = 1$, and by the unramified twist formula

$$w_q(f,\rho) \quad = \quad w_q(V_l)^{\dim \rho} \cdot \mathrm{sgn}((\det\rho)(q^{n(V)})) = (\pm 1)^{p-1}(\tfrac{q}{p})^{n(V)} = (\tfrac{q}{p})^{n(V)},$$

as claimed.

Turning to the global root number, we have

$$w(f,\rho) = \prod_v w_v(f,\rho),$$

the product being taken over all places v of $\mathbb{Q}$. As ρ is self-dual,

$$\prod_v w_v(\rho)^2 = w(\rho)^2 = 1,$$

and the remaining contribution from the real place is $(-1)^{\frac{p-1}{2}}$ (see e.g. [6]). This completes the proof. □

Example 3.2. We compute the global root numbers $w(f,\rho)$ when $p = 3$ and f is one of the primitive cusp forms with $(N,k) = (5,4), (5,6), (7,4)$ or $(121,4)$ that we will use in §6 to illustrate the congruences. In these cases, the answer does not actually depend on the weight.

- If f has level 5, then $\delta = 1$ as $(3,N) = 1$, whence
$$w(f,\rho) = (-1)^{\frac{3-1}{2}} \cdot 1 \cdot \begin{cases} (\frac{5}{3}) & \text{if} \quad (5,m) = 1 \\ 1 & \text{if} \quad \mathrm{ord}_5(m) \geq 1 \end{cases} = \begin{cases} 1 & \text{if} \quad (5,m) = 1 \\ -1 & \text{if} \quad \mathrm{ord}_5(m) \geq 1. \end{cases}$$
- Similarly, if f has level 7, then $(\frac{5}{3}) = -1$ is replaced by $(\frac{7}{3}) = +1$ and we get $w(f,\rho) = -1$ for every m. (cf. also [6, §7.1], first example).
- Finally, if f has level 121, then $(\frac{5}{3})$ is replaced by $(\frac{11}{3})^{\mathrm{ord}_{11}(121)} = +1$, and we again get $w(f,\rho) = -1$ for every m.

The congruence that we verify involves the twists of f by ρ and by the regular representation σ of $\mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$. It easy to check that the root numbers $w_q(f,\sigma)$ and $w(f,\sigma)$ are given by the formula in Theorem 3.1 with $m = 1$. When $p \nmid N$ the formula becomes

$$w(f,\sigma) = (-1)^{\frac{p-1}{2}} \prod_{q|N} \Big(\frac{q}{p}\Big)^{\mathrm{ord}_q(N)} = (-1)^{\frac{p-1}{2}} \Big(\frac{N}{p}\Big).$$

In particular, for $p = 3$ the global root number $w(f,\sigma)$ is $+1$ for the form of level 5, and -1 for the forms of level 7 and 121.

4. An analogue of a result of Hachimori-Matsuno

The aim of this section is to establish an analogue for our primitive cusp form f of results of Hachimori-Matsuno [10] for elliptic curves, over the fields

$$\mathcal{K}_\infty = \mathbb{Q}(\mu_{p^\infty}),\ F^{cyc} = \mathbb{Q}(\mu_{p^\infty}, m^{1/p}), \tag{47}$$

where again m is an integer > 1 which is p-power free. Such a result has already been established in [20], but we wish to give a slightly more explicit result in order to explain its connexion with the congruence (5). Write χ_p for the character giving the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on μ_{p^∞}. As usual, for each $n \in \mathbb{Z}$, write $\mathbb{Z}_p(n)$ for the free $\mathbb{Z}_p$-module of rank one on which $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts via χ_p^n. If W is any $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-module, which is also a $\mathbb{Z}_p$-module, define $W(n) = W \otimes_{\mathbb{Z}_p} \mathbb{Z}_p(n)$, endowed with the natural diagonal action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$.

Let V_p be the underlying $\mathbb{Q}_p$-vector space of the Galois representation τ_p attached to f. Fix once and for all a $\mathbb{Z}_p$-lattice T_p in V_p, which is stable under the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. We stress that we always view V_p as the cohomology group, not the homology group of the motive $M(f)$. We assume from now on that p and f satisfy:-

Hypothesis H3: The odd prime p is good ordinary for f, i.e., p is an odd prime such that $(p, N) = (p, a_p) = 1$.

As p is a good ordinary prime, it is shown in [17] that there exists a one dimensional subspace V_p^0 of V_p such that the inertial subgroup of $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ acts on V_p/V_p^0 by χ_p^{1-k}. Hence if we define

$$A_{p^\infty} = V_p(k-1)/T_p(k-1), \tag{48}$$

and define $A^0_{p^\infty}$ to be the image of $V_p^0(k-1)$ in A_{p^∞}, then $A_{p^\infty}/A^0_{p^\infty}$ is unramified at p. For each finite extension $\mathcal{F}$ of $\mathbb{Q}$, define $\mathcal{F}^{\mathrm{cyc}}$ to be the cyclotomic $\mathbb{Z}_p$-extension of $\mathcal{F}$, i.e., the compositum of $\mathcal{F}$ with the cyclotomic $\mathbb{Z}_p$-extension of $\mathbb{Q}$. We follow Greenberg and define the Selmer group of A_{p^∞} over $\mathcal{F}^{\mathrm{cyc}}$ by

$$\mathrm{Sel}(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}}) = \mathrm{Ker}\left(H^1(\mathcal{F}^{\mathrm{cyc}}, A_{p^\infty}) \to \prod_{w\nmid p} H^1(\mathcal{F}^{\mathrm{cyc}}_w, A_{p^\infty}) \times \prod_{w\mid p} H^1(\mathcal{F}^{\mathrm{cyc}}_w, A_{p^\infty}/A^0_{p^\infty})\right), \tag{49}$$

where w runs over all finite places of $\mathcal{F}^{\mathrm{cyc}}$, and $\mathcal{F}^{\mathrm{cyc}}_w$ denotes the union of the completions at w of the finite extensions of $\mathbb{Q}$ contained in $\mathcal{F}^{\mathrm{cyc}}$. Write

$$X(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}}) = \mathrm{Hom}(\mathrm{Sel}(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}}), \mathbb{Q}_p/\mathbb{Z}_p) \tag{50}$$

for the compact Pontryagin dual of $\mathrm{Sel}(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}})$. Assuming $\mathcal{F}$ is Galois over $\mathbb{Q}$, both $\mathrm{Sel}(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}})$ and $X(A_{p^\infty}/\mathcal{F}^{\mathrm{cyc}})$ are endowed with canonical left actions of $\mathrm{Gal}(\mathcal{F}^{\mathrm{cyc}}/\mathbb{Q})$, and these extend by continuity to left module structures over the Iwasawa algebra

$$\Lambda(\mathrm{Gal}(\mathcal{F}^{\mathrm{cyc}}/\mathbb{Q})) = \varprojlim \mathbb{Z}_p[\mathrm{Gal}(M/\mathbb{Q})],$$

where M runs over the finite Galois extensions of $\mathbb{Q}$ contained in $\mathcal{F}^{\mathrm{cyc}}$.

We shall need the following fundamental result of Kato (see [13]). Note that for $\mathcal{K} = \mathbb{Q}(\mu_p)$, we have $\mathcal{K}^{\mathrm{cyc}} = \mathbb{Q}(\mu_{p^\infty})$.

Theorem 4.1. *Assume Hypothesis H3. Then* $X(A_{p^\infty}/\mathcal{K}^{\mathrm{cyc}})$ *is a torsion* $\Lambda(\mathrm{Gal}(\mathcal{K}^{\mathrm{cyc}}/\mathbb{Q}))$*-module.*

Theorem 4.1 implies that the quotient

$$X(A_{p^\infty}/\mathcal{K}^{\mathrm{cyc}})/(X(A_{p^\infty}/\mathcal{K}^{\mathrm{cyc}})(p)) \tag{51}$$

is a finitely generated $\mathbb{Z}_p$-module, where $X(A_{p^\infty}/\mathcal{K}^{\mathrm{cyc}})(p)$ denotes the p-primary submodule. Define $\lambda(f/F^{\mathrm{cyc}})$ to be the $\mathbb{Z}_p$-rank of (51). We shall also need to consider the Euler factors of the complex L-function $L(f/\mathcal{K}, s)$ at places v with $(v, Np) = 1$. Let q_v denote the characteristic of the residue field of v, and write $q_v^{r_v}$ for the absolute norm of v. Then these Euler factors are given explicitly by

$$P_v(f/\mathcal{K}, X) = \det(1 - \mathrm{Frob}_v^{-1}X \mid V_p) = 1 - b_vX + q_v^{r_v(k-1)}X^2, \tag{52}$$

where $\mathrm{Frob}_v = \mathrm{Frob}_{q_v}^{r_v}$, and $b_v \in \mathbb{Z}$. Since $q_v^{r_v} \equiv 1 \bmod p$, it is clear that for all integers n, we have

$$P_v(f/\mathcal{K}, q_v^{-r_vn}) \equiv 2 - b_v \bmod p, \tag{53}$$

when both sides are viewed as elements of $\mathbb{Z}_p$. In particular the question whether or not the left hand side lies in $p\mathbb{Z}_p$ is independent of n. Define $\mathcal{P}_2$ to be the set of all places w of $\mathcal{K}^{\mathrm{cyc}}$ such that, writing $v = w \mid \mathcal{K}$, we have

$$\mathcal{P}_2 = \{w : (q_v, Np) = 1,\ q_v \mid m,\ \text{and } \mathrm{ord}_p(2 - b_v) > 0\}. \tag{54}$$

Similarly, suppose v is a place of $\mathcal{K}$, with residue characteristic $q_v \neq p$ and $\mathrm{ord}_{q_v}N = 1$. Then the Euler factor $P_v(f/\mathcal{K}, X)$ is given explicitly by

$$P_v(f/\mathcal{K}, X) = \det(1 - \mathrm{Frob}_v^{-1}X \mid V_p^{I_v}) = 1 - b_vX, \tag{55}$$

where $b_v = a_{q_v}^{r_v}$, with $q_v^{r_v}$ again being the absolute norm of v. Note again that

$$P_v(f/\mathcal{K}, q_v^{-r_vn}) \equiv 1 - b_v \bmod p$$

for all integers n. Also, since $a_{q_v}^2 = q_v^{k-2}$ and $q_v^{r_v} \equiv 1 \bmod p$, we always have $b_v^2 \equiv 1 \bmod p$. Define $\mathcal{P}_1$ to be the set of all places w of $\mathcal{K}^{\mathrm{cyc}}$ such that, writing $v = w \mid \mathcal{K}$, we have

$$\mathcal{P}_1 = \{w : \mathrm{ord}_{q_v}N = 1,\ q_v \mid m \text{ and } b_v \equiv 1 \text{ mod } p\} \tag{56}$$

To establish an analogue for f of the theorem of Hachimori-Matsuno, we shall need the following additional hypothesis.

Hypothesis H4: $X(A_{p^\infty}/\mathcal{K}^{\mathrm{cyc}})$ is a finitely generated $\mathbb{Z}_p$-module.

Recall that $\mathcal{K}^{\mathrm{cyc}} = \mathbb{Q}(\mu_{p^\infty})$ and $\mathcal{F}^{\mathrm{cyc}} = \mathbb{Q}(\mu_{p^\infty}, m^{1/p})$.

Theorem 4.2. *Assume Hypotheses H2, H3 and H4. Then $X(A_{p^\infty}/F^{\mathrm{cyc}})$ is also a finitely generated $\mathbb{Z}_p$-module and*

$$\lambda(f/F^{\mathrm{cyc}}) = p\lambda(f/\mathcal{K}^{\mathrm{cyc}}) + \sum_{w \in \mathcal{P}_2} 2(p-1) + \sum_{w \in \mathcal{P}_1} (p-1). \tag{57}$$

Proof. Put $\Delta = \mathrm{Gal}(F/\mathcal{K}) = \mathrm{Gal}(F^{\mathrm{cyc}}/\mathcal{K}^{\mathrm{cyc}})$. If B is any Δ-module, we recall that the Herbrand quotient $h_\Delta(B)$ is defined by

$$h_\Delta(B) = \frac{\# H^2(\Delta, B)}{\# H^1(\Delta, B)},$$

whenever the cohomology groups are both finite.

Entirely similar arguments to those given for elliptic curves in [10] show that, under the hypotheses H2, H3 and H4, $X(A_{p^\infty}/F^{\mathrm{cyc}})$ is indeed a finitely generated $\mathbb{Z}_p$-module, and we have

$$\lambda(f/F^{\mathrm{cyc}}) = p\lambda(f/\mathcal{K}^{\mathrm{cyc}}) + (p-1)\mathrm{ord}_p(h_\Delta(\mathrm{Sel}(A_{p^\infty}/F^{\mathrm{cyc}})) \tag{58}$$

where $h_\Delta(A_{p^\infty}/F^{\mathrm{cyc}})$ is finite.

Let Σ denote the set of primes of $\mathcal{K}^{\mathrm{cyc}}$ lying above the rational primes dividing Nmp. As in [10, §4], well-known arguments from Galois cohomology show that

$$h_\Delta(\mathrm{Sel}(A_{p^\infty}/F^{\mathrm{cyc}})) = \prod_{w \in \Sigma} h_\Delta\left(\prod_{u \mid w} H^1(F_u^{\mathrm{cyc}}, C_w)\right) \tag{59}$$

where u runs over the places of F^{cyc} above w, and

$$C_w = A_{p^\infty}, \text{ or } A_{p^\infty}/A_{p^\infty}^0 \tag{60}$$

according as w does not or does lie above p. Moreover, since a prime w of $\mathcal{K}^{\mathrm{cyc}}$ either splits completely or has a unique prime above it in F^{cyc}, it is clear that the right hand side of (59) simplifies to a product of the $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w))$, where w now runs over the primes in Σ which do not split completely in F^{cyc}. Assume from now on that w is a prime of $\mathcal{K}^{\mathrm{cyc}}$ which does not split in F^{cyc}. In particular, this means that the residue characteristic q_w of w must divide pm. Since F_u^{cyc} and $\mathcal{K}_w^{\mathrm{cyc}}$ contain μ_{p^∞}, their absolute Galois groups have p-cohomological dimension at most 1. As Δ is cyclic of order p, it then follows easily from the Hochschild-Serre spectral sequence that

$$H^i(\Delta, H^1(F_u^{\mathrm{cyc}}, C_w)) \simeq H^i(\Delta, C_w(F_u^{\mathrm{cyc}})), \tag{61}$$

where $C_w(F_u^{\mathrm{cyc}}) = H^0(F_u^{\mathrm{cyc}}, C_w)$.

Lemma 4.3. *Assume there is a unique prime u of F^{cyc} above p, and put $w = u \mid \mathcal{K}^{\mathrm{cyc}}$. Then we have*

$$h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = 1.$$

Proof. Since C_w is unramified, and F_u^{cyc} is a totally ramified extension of $\mathbb{Q}_p$, we have

$$C_w(F_u^{\mathrm{cyc}}) = H^0(\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_p), C_w).$$

But Frob_p acts on C_w by multiplication by the p-adic unit root of $1 - a_pX + p^{k-1}X^2$. However, this unit root cannot be equal to 1 as it has complex absolute value $p^{(k-1)/2}$. Hence $C_w(F_u^{\mathrm{cyc}})$ must be finite, and thus has Herbrand quotient equal to 1. □

Write q_w for the residue characteristic of w.

Lemma 4.4. *Assume that w is a prime of $\mathcal{K}^{\mathrm{cyc}}$ such that $(q_w, Np) = 1$, and q_w divides m. Then there is a unique prime u of F^{cyc} above w, and $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = p^{-2}$ if $w \in \mathcal{P}_2$, and $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = 1$ otherwise.*

Proof. The first assertion of the lemma is clear since w must ramify in F^{cyc} because q_w divides m. As $(q_w, Np) = 1$, we know that the inertial subgroup I_w of the absolute Galois group of $\mathbb{Q}_{q_w}$ acts trivially on V_p. We claim that I_w also acts trivially on $C_w = A_{p^\infty}$. Indeed, we have an exact sequence of Galois modules

$$0 \to T_p \to V_p \to C_w \to 0,$$

whence one obtains the long exact sequence

$$0 \to T_p \to V_p \to C_w^{I_w} \to H^1(I_w, T_p) \to H^1(I_w, V_p). \tag{62}$$

As the inertial action is trivial on T_p and V_p, and $q_w \neq p$, we see that

$$H^1(I_w, T_p) = \mathrm{Hom}(J_w, T_p),\ H^1(I_w, V_p) = \mathrm{Hom}(J_w, V_p),$$

where J_w is the Galois group of the unique tamely ramified $\mathbb{Z}_p$-extension of $\mathcal{K}_w^{\mathrm{cyc}}$. Thus the last map in (62) is injective, and so $C_w^{I_w} = C_w$ as claimed.

Let v be the restriction of w to $\mathcal{K}$. We next show that $C_v(\mathcal{K}_v) \neq 0$ if and only if $v \in \mathcal{P}_2$. Since C_v is unramified, we have the commutative diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & T_p \arrow[r] \arrow[d, "\delta_v"'] & V_p \arrow[r] \arrow[d, "\delta_v"'] & C_w \arrow[r] \arrow[d, "\delta_v"'] & 0 \\
0 \arrow[r] & T_p \arrow[r] & V_p \arrow[r] & C_w \arrow[r] & 0
\end{tikzcd}$$

where δ_v is the map given by applying $\mathrm{Frob}_v - 1$. The characteristic polynomial of Frob_v acting on V_p is $X^2P_v(f/\mathcal{K}, X^{-1})$. The roots of this polynomial have complex absolute value $q_v^{r_v(k-1)/2}$, and thus are distinct from 1. Hence the middle vertical map in the above diagram is an isomorphism. It follows from the snake lemma that $C_w(K_v)$ has order equal to the cokernel of the left hand vertical map, which is equal to the exact power of p dividing $P_v(f/\mathcal{K}, 1)$. But

$$P_v(f/\mathcal{K}, 1) = 1 - b_v + q_v^{r_v(k-1)} \equiv (2 - b_v) \bmod p,$$

showing that $C_w(K_v) \neq 0$ if and only if $w \in \mathcal{P}_2$.

Our next claim is that $C_w = C_w(\mathcal{K}^{\mathrm{cyc}})$ if and only if $C_w(\mathcal{K}_v) \neq 0$. As $\mathrm{Gal}(\mathcal{K}_w^{\mathrm{cyc}}/\mathcal{K}_v)$ is pro-p, Nakayama's lemma shows that $C_w(\mathcal{K}_v) = 0$ implies that $C_w(\mathcal{K}_w^{\mathrm{cyc}}) = 0$. Conversely, assume that $C_w(\mathcal{K}_v) \neq 0$. We then assert that the extension $\mathcal{K}_v(C_w)$ is a pro-p extension of $\mathcal{K}_v$. To prove this, let $(C_w)_p$ be the kernel of multiplication by p on C_w. It is easily seen that the extension $\mathcal{K}_v(C_w)/\mathcal{K}_v((C_w)_p)$ is pro-p. On the other hand, choosing an $\mathbb{F}_p$-basis of $(C_w)_p$ in which the first element belongs to $C_w(\mathcal{K}_v)$, and noting that the determinant of $(C_w)_p$ is trivial because it is equal to $\omega^{r_v(k-1)}$, where ω is the cyclotomic character mod p, it follows that the extension $\mathcal{K}_v((C_w)_p)/\mathcal{K}_v$ is a p-extension. Thus $\mathcal{K}_v(C_w)$ is a pro-p extension of $\mathcal{K}_v$, and it is unramified as inertia acts trivially on C_w. Hence we must have $\mathcal{K}_v(C_w) = \mathcal{K}_w^{\mathrm{cyc}}$.

It is now clear from (60) that $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = 0$ if $w \notin \mathcal{P}_2$, and $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = p^{-2}$ if $w \in \mathcal{P}_2$. This completes the proof of the lemma. □

Lemma 4.5. *Assume that w is a prime of $\mathcal{K}^{\mathrm{cyc}}$ such that* $\mathrm{ord}_{q_w} N = 1$ *and q_w divides m. Then there is a unique prime u of F^{cyc} above w, and $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = p^{-1}$ if $w \in \mathcal{P}_1$, and $h_\Delta(H^1(F_u^{\mathrm{cyc}}, C_w)) = 1$ otherwise.*

Proof. The first assertion is clear, since w must ramify in F^{cyc}, because q_v divides m. Again, let v be the restriction of w to $\mathcal{K}$. Since $\mathrm{ord}_{q_w} N = 1$, we have

$$P_v(f/\mathcal{K}, X) = 1 - b_v X$$

where we recall that $b_v^2 \equiv 1 \bmod p$. Let W_p be the subspace $V_p^{I_p}$ of V_p, so that $\mathrm{Gal}(\bar{\mathbb{Q}}_{q_v}/F_v)$ acts on W_p via the unramified character η with $\eta(\mathrm{Frob}_v) = b_v$. Choosing a basis of V_p with the first basis element being a basis of W_p, the representation of $\mathrm{Gal}(\bar{\mathbb{Q}}_{q_v}/F_v)$ on V_p must be of the form $\begin{pmatrix} \eta & * \\ 0 & \lambda \end{pmatrix}$, where λ is a character of $\mathrm{Gal}(\bar{\mathbb{Q}}_{q_v}/F_v)$. As the determinant of V_p is the cyclotomic character to the power $(k-1)$, we conclude that λ is also unramified. Moreover, the image of the restriction of this representation to $\mathrm{Gal}(\bar{\mathbb{Q}}_{q_v}/\mathcal{K}_v^{\mathrm{nr}})$ is infinite, where $\mathcal{K}_v^{\mathrm{nr}}$ is the maximal unramified extension of $\mathcal{K}_v$. Since η takes values in $\mathbb{Z}_p^\times$, it is clear that the restriction of η to $\mathrm{Gal}(\mathcal{K}_v^{\mathrm{nr}}/\mathcal{K}_w^{\mathrm{cyc}})$ is the trivial character if and only if $w \in \mathcal{P}_1$. Similarly, writing v' for the restriction of u to F, and recalling that $F_{v'}/\mathcal{K}_v$ is totally ramified, it follows that the restriction of η to $\mathrm{Gal}(F_{v'}^{\mathrm{nr}}/F_u^{\mathrm{cyc}})$ is the trivial character if and only if $w \in \mathcal{P}_1$. One concludes easily that, if $w \notin \mathcal{P}_1$, then $C_w(F_u^{\mathrm{cyc}})$ must be finite, and if $w \in \mathcal{P}_1$, then the divisible subgroup of $C_w(F_u^{\mathrm{cyc}})$ has $\mathbb{Z}_p$-corank 1. In view of (60), the assertion of the lemma is now clear.

□

Combining (58), (59), and Lemmas 3.3, 3.4 and 3.5, the proof of Theorem 4.2 is now complete. □

5. The congruence from non-commutative Iwasawa theory

As before, let

$$F = \mathbb{Q}(\mu_p, m^{1/p}), \ \mathcal{K} = \mathbb{Q}(\mu_p) \tag{63}$$

where p is an odd prime, and $m > 1$ is an integer which is not divisible by the p-th power of an integer > 1. Assume throughout this section that Hypotheses H1, H2, and H3 are valid.

Let ϕ be an Artin representation of $\mathrm{Gal}(F_\infty/\mathbb{Q})$. For each integer $n = 1, \ldots, k-1$, we recall that $L_p^*(f, \phi, n)$ is defined by (41). By Theorem 2.8, we know that $L_p^*(f, \phi, n)$ is an algebraic number. Very roughly speaking, the non-commutative p-adic L-function seeks to interpolate the numbers $L_p^*(f, \phi, n)$, as ϕ, and n both vary. While there has been important recent progress on the study of these non-commutative p-adic L-functions for the Tate motive over totally real number fields (see [12],[21]), very little is still known about their existence for other motives, including the motive attached to our modular form f. In the present paper, we shall only discuss what is perhaps the simplest congruence between abelian p-adic L-functions, which would follow from the existence of a non-commutative p-adic L-function for the motive of f over the field F_∞. A specialization of this congruence for elliptic curves has been studied in the earlier paper [5].

To state this congruence, we must first make a canonical modification of the values $L_p^*(f, \phi, n)$, given by (41) following [2], [8]. Recall that since $(p, a_pN) = 1$, the Euler factor

$$P_p(f, X) = 1 - a_pX + p^{k-1}X^2$$

can be written as

$$P_p(f, X) = (1 - \alpha X)(1 - \beta X), \tag{64}$$

where α is a unit in $\mathbb{Z}_p$, and $\mathrm{ord}_p(\beta) = k-1$. We shall also need the Euler factors of the complex L-series $L(\phi, s)$ of the Artin representation ϕ, which are defined by

$$P_q(\phi, X) = \det\left(1 - \mathrm{Frob}_q^{-1}X \mid M_l(\phi)^{I_q}\right) \tag{65}$$

where l is any prime distinct from q. As before, let $d(\phi)$ be the dimension of ϕ. Moreover, writing $\mathcal{N}(\phi)$ for the conductor of ϕ, define

$$e_p(\phi) = \mathrm{ord}_p(\mathcal{N}(\phi)). \tag{66}$$

Recall that $P_q(f, \phi, X)$ defined by (17) is the Euler factor at the prime q of the complex L-function $L(f, \phi, s)$. Recall also that, for $n = 1, \ldots k-1$,

$$L_p^*(f, \phi, n) = \frac{L(f, \phi, n)\epsilon_p(\phi)}{\left((2\pi i)^{nd(\phi)} \times \Omega_+(f)^{d_n^+(\phi)} \times |\Omega_-(f)|^{d_n^-(\phi)}\right)}. \tag{67}$$

We then define

$$M_p(f, \phi, n) = \Gamma(n)^{d(\phi)} \times L_p^*(f, \phi, n) \times P_p(f, \phi, p^{-n}) \times \frac{P_p(\hat{\phi}, p^{n-1}/\alpha)}{P_p(\phi, \alpha/p^n)} \times (p^{n-1}/\alpha)^{e_p(\phi)}, \tag{68}$$

and

$$\mathcal{L}_p(f, \phi, n) = M_p(f, \phi, n) \prod_{q \neq p,\, q|m} P_q(f, \phi, q^{-n}), \tag{69}$$

where q runs over the prime factors of m distinct from p. It is these modified L-values, defined using the naive periods $\Omega_+(f)$ and $\Omega_-(f)$, which we shall actually compute in a number of numerical examples.

Secondly, in order to obtain p-adic L-functions which will in the end satisfy the main conjectures of Iwasawa theory, we may also have to adjust the naive periods $\Omega^+(f)$ and $\Omega^-(f)$ by certain non-zero rational numbers. Writing $\Omega_+^{\mathrm{can}}(f)$ and $\Omega_-^{\mathrm{can}}(f)$ for these canonical periods, we will have

$$\Omega_+^{\mathrm{can}}(f) = c_+(f)\Omega_+(f),\ \Omega_-^{\mathrm{can}}(f) = c_-(f)\Omega_-(f) \tag{70}$$

for certain non-zero rational numbers $c_+(f)$ and $c_-(f)$. It is then natural to define

$$\mathcal{L}_p^{\mathrm{can}}(f,\phi,n) = c_+(f)^{-d_n^+(\phi)}c_-(f)^{-d_n^-(\phi)}\mathcal{L}_p(f,\phi,n). \tag{71}$$

It is these modified values $\mathcal{L}_p^{\mathrm{can}}(f,\phi,n)$ which should satisfy the non-abelian congruences for the p-adic L-functions arising in the main conjectures. However, in our present state of knowledge, we do not know in general how to determine $c_+(f)$ and $c_-(f)$ precisely. Nevertheless, as we shall now explain, the work of Manin on the p-adic L-function of f for the extension $\mathcal{K}_\infty/\mathbb{Q}$ provides some partial information on this question.

Theorem 5.1. *Let σ be the sum of the irreducible characters of $\mathrm{Gal}(\mathcal{K}_1/\mathbb{Q})$, where $\mathcal{K}_1 = \mathbb{Q}(\mu_p)$. If $L(f,\sigma,k/2) = 0$, then $\mathcal{L}_p^{\mathrm{can}}(f,\sigma,n)$ belongs to $p\mathbb{Z}_p$ for $n = 1,\dots,k-1$.*

Proof. Let χ_p be the character giving the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on μ_{p^∞}. Fix a topological generator γ of $\mathrm{Gal}(\mathcal{K}^{cyc}/K_1)$ and put $u = \chi_p(\gamma)$. The work of Manin then shows [16] that there exists a power series $g(T)$ in $\mathbb{Z}_p[[T]]$ such that

$$g(u^r - 1) = M_p^{\mathrm{can}}(f,\sigma,k/2+r), \tag{72}$$

for all integers r with $-k/2+1 \le r \le k/2-1$, and where

$$M_p^{\mathrm{can}}(f,\sigma,n) = (c_+(f)c_-(f))^{(1-p)/2}\,M_p(f,\sigma,n).$$

Here it is understood that the canonical periods are those for which we expect $g(T)$ to be a characteristic power series for the dual Selmer group of f over $\mathcal{K}_\infty$. Assuming that $L(f,\sigma,k/2) = 0$, it follows that

$$g(0) = 0,$$

and so $g(u^n - 1) \in p\mathbb{Z}_p$ for all integers n. The assertion of the theorem then follows on noting that

$$\prod_{q\neq p,\, q|m} P_q(f,\sigma,q^{-n})$$

lies in $\mathbb{Z}_p$ for all $n \in \mathbb{Z}$. This completes the proof. □

Example 5.2. Take f to be the unique primitive eigenform of level 7 and weight 4, and $p = 3$. Then $L(f,\sigma,2) = 0$. Moreover, we see from Table II in §6 that $\mathcal{L}_3(f,\sigma,1) \in 3\mathbb{Z}_3$. In view of Theorem 5.1, this strongly suggests that in this case, we must have $\mathrm{ord}_3(c_+(f)) = \mathrm{ord}_3(c_-(f)) = 0$.

Example 5.3. Take f to be the complex multiplication form of level 121 and weight 4, which is attached to the cube of the Grössencharacter of the elliptic curve over E over $\mathbb{Q}$ given by the equation (79), of conductor 121 and with complex multiplication by the full ring of integers of the field $L = \mathbb{Q}(\sqrt{-11})$, and again take $p = 3$. Then $L(f, \sigma, 2) = 0$. However, we see from Table III in §6 that $\mathcal{L}_3(f, \sigma, 1)$ is a 3-adic unit when $m = 3$, 7 or 11. Hence the naive periods $\Omega^+(f)$ and $\Omega^-(f)$ cannot be the good periods, and at least one of $\mathrm{ord}_3(c_+(f))$ or $\mathrm{ord}_3(c_-(f))$ must be strictly less than zero. In fact, in this case we do know the canonical periods for f, since, for all good ordinary primes p for f, we know the periods for which the relevant cyclotomic main conjecture for f over $\mathcal{K}^{cyc}$ is valid. This is because this cyclotomic main conjecture can easily be deduced from the main conjecture for E over the field obtained by adjoining to L the coordinates of all p-power division points on E; and this latter main conjecture is proven for all good ordinary primes p for E by the work of Yager and Rubin. Invoking the Chowla-Selberg formula, we see easily that the explicit values of these canonical periods can be taken as follows. Let

$$\Theta = \Gamma(1/11)\Gamma(3/11)\Gamma(4/11)\Gamma(5/11)\Gamma(9/11).$$

Then

$$\Omega_+^{\mathrm{can}}(f) = \sqrt{11} \times \Theta^3/(2\pi)^9,\ \Omega_-^{\mathrm{can}}(f) = i\Theta^3/(2\pi)^9. \tag{73}$$

Direct computations show that

$$\Omega_+(f)/\Omega_+^{\mathrm{can}}(f) = 1/22,\ \Omega_-(f)/\Omega_-^{\mathrm{can}}(f) = 3, \tag{74}$$

whence

$$\mathrm{ord}_3(c_+(f)) = 0,\ \mathrm{ord}_3(c_-(f)) = -1, \tag{75}$$

precisely as required.

As in the Introduction, let σ be the Artin representation of dimension $(p-1)$ given by the direct sum of the one dimensional characters of $\mathrm{Gal}(\mathcal{K}/\mathbb{Q})$. Define ρ to be the representation of $\mathrm{Gal}(F/\mathbb{Q})$ induced from any non-trivial degree one character of $\mathrm{Gal}(F/\mathcal{K})$. Thus ρ also has dimension $(p-1)$, and is easily seen to be irreducible (cf. [5]). Moreover, both σ and ρ are self-dual, can be realized over $\mathbb{Z}$, and their reductions modulo p are isomorphic. Let $R = \mathbb{Z}_p[[T]]$ be the ring of formal power series in an indeterminate T with coefficients in $\mathbb{Z}_p$. As explained in the Introduction, the work of Manin [16] establishes the existence of a power series $H(\sigma, T)$ in R satisfying the interpolation property (2) It is conjectured that there exists a power series $H(\rho, T)$ in R satisfying the interpolation condition (3).

Conjecture 5.4. *(Congruence Conjecture). Assume Hypotheses H1, H2, H3. Then there exists a power series $H(\rho, T)$ in R satisfying the interpolation property* (3)*, and we have the congruence of power series*

$$H(\rho, T) \ \equiv\ H(\sigma, T) \bmod pR. \tag{76}$$

We are grateful to M. Kakde for pointing out to us that the congruence (76) is simply a special case of the congruences predicted by Kato in [13], and we now briefly explain why this is the case. Assume for simplicity that Hypothesis H4 is also valid. Recall that G denotes the Galois group of F_∞ over $\mathbb{Q}$, and write $\Lambda(G)$ for the Iwasawa algebra of G, S for the canonical Ore set in $\Lambda(G)$, which is defined in [2], and $\Lambda(G)_S$ for its localization at S. In addition, define $\mathfrak{G}_0 = \mathrm{Gal}(\mathcal{K}_\infty/\mathbb{Q})$, and for each integer $n \geq 1$, let $\mathfrak{G}_n$ be the unique closed subgroup of index p^{n-1} in $\mathrm{Gal}(\mathbb{Q}^{\mathrm{cyc}}/\mathbb{Q})$. Write S_n for the canonical Ore set of [2] in the Iwasawa algebra $\Lambda(\mathfrak{G}_n)$. In [13], Kato defines a canonical map

$$\theta_{G,S} : K_1(\Lambda(G)_S) \to \prod_{n\geq 0} K_1(\Lambda(\mathfrak{G}_n)_{S_n}),$$

and characterizes its image by a remarkable set of congruences which we do not state in detail here. In particular, writing $\theta_{G,S}(\alpha) = (\alpha_n)$ for any element α of $K_1(\Lambda(G)_S)$, we always have

$$N(\alpha_0) \equiv \alpha_1 \bmod p, \tag{77}$$

where N denotes the norm map from $K_1(\Lambda(\mathfrak{G}_0)_{S_0})$ to $K_1(\Lambda(\mathfrak{G}_1)_{S_1})$. Now take α to be the conjectural p-adic L-function for f over F_∞, which we denote by $\zeta(f/F_\infty)$. Let us also identify $\Lambda(\mathfrak{G}_1)$ with the formal power series ring $R = \mathbb{Z}_p[[T]]$ by mapping the fixed topological generator γ of $\mathfrak{G}_1$ to $1 + T$. Then it follows essentially from the construction of the map $\theta_{G,S}$ and the interpolation properties of these p-adic L-functions that we will have

$$N(\zeta(f/F_\infty)_0) = H(\sigma, T),\ \zeta(f/F_\infty)_1 = H(\rho, T).$$

Thus the congruence (76) is indeed just a special case of the congruence (77) of Kato, as claimed.

As was pointed out in the Introduction, if we evaluate both sides of the congruence (76) at the appropriate points in $p\mathbb{Z}_p$, we deduce the following congruence of normalized L-values from (2) and (3):-

Conjecture 5.5. *Assume Hypotheses H1, H2 and H3. Then for all integers* $n = 1, \ldots, k-1$*, we have*

$$\mathcal{L}_p^{\mathrm{can}}(f, \rho, n) \equiv \mathcal{L}_p^{\mathrm{can}}(f, \sigma, n) \bmod p. \tag{78}$$

We end this section by explaining how this latter congruence is intimately connected with Theorem 4.2. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be the set of places of $\mathcal{K}_\infty = \mathbb{Q}(\mu_{p^\infty})$ defined by (56) and (54) respectively.

Lemma 5.6. *Let q be any rational prime with q dividing m and $(q, Np) = 1$. Then all primes of $\mathcal{K}^{\mathrm{cyc}}$ above q belong to $\mathcal{P}_2$ if and only if* $\mathrm{ord}_p(P_q(f, \sigma, q^{-n})) > 0$ *for some integer n.*

Proof. Let q have exact order r_q modulo p, and let v be a prime of $\mathcal{K}$ above q. Then one sees immediately that

$$P_q(f, \sigma, X) = (1 - b_v X^{r_q} + q^{(k-1)r_q} X^{2r_q})^{\frac{p-1}{r_q}},$$

where b_v is defined by (52). Since $q^{r_q} \equiv 1 \bmod p$, the assertion of the lemma is now plain from the definition of $\mathcal{P}_2$. □

Lemma 5.7. *Let q be any rational prime not equal to p such that q divides m and $\mathrm{ord}_q(N) = 1$ .Then all primes of $\mathcal{K}^{\mathrm{cyc}}$ above q belong to $\mathcal{P}_1$ if and only if $\mathrm{ord}_p(P_q(f, \sigma, q^{-n})) > 0$ for some integer n.*

Proof. Let q have exact order r_q and let v be a prime of $\mathcal{K}$ above q. Since σ is unramified at q, one sees easily that

$$P_q(f, \sigma, X) = (1 - b_v X^{r_q})^{\frac{p-1}{r_q}},$$

where b_v is defined by (55), and hence the assertion of the lemma is clear. □

By the work of Manin, we always have $M_p^{\mathrm{can}}(f, \sigma, n)$ is in $\mathbb{Z}_p$ for $n = 1, \ldots, k-1$. Hence we conclude from Lemmas 5.6 and 5.7 that $\mathcal{L}_p^{\mathrm{can}}(f, \sigma, n) \in p\mathbb{Z}_p$ if either $\mathcal{P}_1$ or $\mathcal{P}_2$ is non-empty. On the other hand, assuming Hypotheses H1-H4, Theorem 4.2 shows that $X(A_{p^\infty}/F^{\mathrm{cyc}})$ is infinite if either $\mathcal{P}_1$ or $\mathcal{P}_2$ is non-empty. But $X(A_{p^\infty}/F^{\mathrm{cyc}})$ is infinite if and only if its characteristic element as a $\Lambda(\mathrm{Gal}(F^{\mathrm{cyc}}/F))$-module is not a unit in the Iwasawa algebra. But the main conjecture for $X(A_{p^\infty})$ predicts that the $\mathcal{L}_p^{\mathrm{can}}(f, \rho, n)$ are all values of the characteristic power series of $X(A_{p^\infty}/F^{\mathrm{cyc}})$. Thus it would follow that $\mathcal{L}_p^{\mathrm{can}}(f, \rho, n) \in p\mathbb{Z}_p$ if either $\mathcal{P}_1$ or $\mathcal{P}_2$ is non-empty, in accord with the Congruence Conjecture 5.5.

6. Numerical data

We refer the reader to Section 6 of [5] for a detailed discussion of how the computations are carried out in the case of a primitive form of weight 2. Entirely similar arguments (see [4]) apply to the calculation of the numerical values $\mathcal{L}_p(f, \phi, n)$, for $n = 1, \ldots, k-1$, for our given primitive modular form $f = \sum_{n=1}^{\infty} a_n q^n$ of conductor N. We do not enter into the details here, apart from listing the explicit Euler factors which occur for the primes dividing pm. As before, let

$$\mathcal{K} = \mathbb{Q}(\mu_p), \ F = \mathbb{Q}(\mu_p, m^{1/p}),$$

where m is a p-power free integer > 1. As earlier, we write ϕ for either the direct sum σ of the one dimensional characters of $\mathrm{Gal}(\mathcal{K}/\mathbb{Q})$ or the unique irreducible representation ρ of dimension $p-1$ of $\mathrm{Gal}(F/\mathbb{Q})$, and note that both of these Artin representations are self-dual. We suppose that p is an odd prime number satisfying $(p, a_p) = (p, N) = 1$. In addition, we assume that Hypothesis H2 holds. As earlier, let $P_p(\phi, X)$ denote the polynomial in X giving the inverse Euler factor at p of the complex L-series $L(\phi, s)$ of the Artin representation ϕ, and $P_q(f, \phi, X)$ the polynomial giving the inverse Euler factor at a prime q of the complex L-series $L(f, \phi, s)$.

Lemma 6.1. *We have that $P_p(\sigma, X) = 1 - X$, and $P_p(f, \sigma, X) = P_p(f, X)$. If $m \equiv \pm 1 \bmod p^2$, then $P_p(\rho, X) = 1 - X$, and $P_p(f, \rho, X) = P_p(f, X)$. Otherwise, both $P_p(\rho, X)$ and $P_p(f, \rho, X)$ are equal to 1.*

□

Lemma 6.2. *Let q be any prime factor of m distinct from p, and write r_q for the order of q modulo p. Then we have:-*

1. $P_q(f, \sigma, X) = P_v\left(f/\mathcal{K}, X^{r_q}\right)^{\frac{p-1}{r_q}}$ *, where $P_v(f/\mathcal{K}, X)$ is the Euler factor of f over $\mathcal{K}$ at any prime v of $\mathcal{K}$ above q if $(q, N) = 1$.*
2. $P_q(f, \sigma, X) = \left(1 - a_q^{r_q} X^{r_q}\right)^{\frac{p-1}{r_q}}$ *if* $\mathrm{ord}_q(N) = 1$.
3. $P_q(f, \rho, X) = 1$.

□

We remark that the computations require knowledge of the Fourier coefficients a_n of f for n ranging from 1 up to approximately the square root of the conductor of the complex L-function $L(f, \phi, s)$. Since these conductors are very large even for small N, this explains why we need to know explicitly the a_n for $1 \leq n \leq 10^8$, and why we are essentially restricted to the case of the prime $p = 3$. For our primitive cusp form f of small conductor, we computed these Fourier coefficients a_n using [SAGE] as follows. We use linear algebra to express f explicitly as a polynomial in terms of Eisenstein series (we only used small conductor forms f where this was possible), then we evaluate this expression using arithmetic with polynomials of large degree over the integers. This high precision evaluation took about 1 day of CPU time in some cases, and relies on the fast FFT-based polynomial arithmetic from `http://flintlib.org`, and optimized code for computing coefficients of Eisenstein series due to Craig Citro, along with other optimizations specific to this problem. For evaluation of the CM form of level 121 and weight 4, we computed the Fourier coefficients d_p for the corresponding elliptic curve of weight 2 (using [19]), then obtained the coefficients a_p of the weight 4 form as the sum of the cubes of the roots of $X^2 - d_p X + p$, and finally extended these multiplicatively to obtain all of the coefficients a_n.

For even $k \geq 2$, let $E_k(q) \in \mathbb{Q}[[q]]$ denote the weight k Eisenstein series of level 1, normalized so that the coefficient of q is 1. For integers $t \geq 1$, define $E_2^*(q^t) = E_2(q) - tE_2(q^t)$, which is a holomorphic modular form of level t and weight 2. We consider 4 explicit primitive forms; 3 have expressions in terms of Eisenstein series, and the fourth in terms of an elliptic curve with complex multiplication. The first 3 are the unique primitive forms on $\Gamma_0(p)$ with given weight. The fourth form f is the complex multiplication form of conductor 121 which is attached to the cube of the Grossencharacter of the elliptic curve

$$y^2 + y = x^3 - x^2 - 7x + 10. \tag{79}$$

This curve has complex multiplication by the full ring of integers of $\mathbb{Q}(\sqrt{-11})$, and has conductor 121 when viewed as a curve over $\mathbb{Q}$. The following table gives the first few terms of the q-expansion of these four forms, and note that, in each case, 3 is an ordinary prime because the coefficient of q^3 is not divisible by 3.

Conductor	**Weight**	**Primitive form**
5	4	$-\frac{250}{3}E_4(q^5) - \frac{10}{3}E_4(q) + 13E_2^*(q^5)^2 = q - 4q^2 + 2q^3 + 8q^4 - 5q^5 + \cdots$
7	4	$-\frac{147}{2}E_4(q^7) - \frac{3}{2}E_4(q) + 5E_2^*(q^7)^2 = q - q^2 - 2q^3 - 7q^4 + 16q^5 + \cdots$
5	6	$\frac{521}{6}E_6(q^5) - \frac{1}{30}E_6(q) + 248E_2^*(q^5)E_4(q^5) = q + 2q^2 - 4q^3 - 28q^4 + 25q^5 + \cdots$
121	4	$q + 8q^3 - 8q^4 + 18q^5 + 37q^9 - 64q^{12} + 144q^{15} + 64q^{16} + \cdots$

The first two tables below provide numerical evidence in support of the congruences (5), and the third and fourth table below provides evidence in support of the stronger congruence (7). The notation used in these four tables is as follows. We have taken $p = 3$, and assume that ϕ denotes either σ or ρ, so that $d(\phi) = 2$. For each integer $n = 1, \ldots, k-1$, put

(80)
$$L_3^*(f,\phi,n) = L(f,\phi,n)\epsilon_3(\phi)(2\pi i)^{-2n}(\Omega^+(f)|\Omega^-(f)|)^{-1}, \;\; \mathcal{P}_3(f,\phi,n) = \prod_{q|3m} P_q(f,\phi,q^{-n}),$$

and define

$$\mathcal{L}_3(f,\phi,n) = \Gamma(n)^2 \times L_3^*(f,\phi,n) \times \mathcal{P}_3(f,\phi,n) \times \frac{P_3(\phi,3^{n-1}/\alpha)}{P_3(\phi,\alpha/3^n)} \times (3^{n-1}/\alpha)^{e_3(\phi)}.$$

We also write $N(f,\phi)$ for the conductor of the complex L-function $L(f,\phi,s)$. it is easily seen that $\epsilon_3(\sigma)$ is equal to the positive square root of 3. Moreover, $\epsilon_3(\rho) = 3^5$ when $\mathrm{ord}_3(m) \geq 1$. When $(3,m) = 1$, we have that $\epsilon_3(\rho)$ is equal to 3 when $m \equiv \pm 1 \bmod 3^2$, and is equal to 3^3 otherwise. If r is any integer ≥ 1, and w is an integer, the symbol $w + O(3^r)$ will denote a 3-adic integer which is congruent to w modulo 3^r.

The reader should also bear in mind the following comments about the signs of the values $L_3^*(f,\phi,n)$ given in our tables below. Since ϕ can be realized over $\mathbb{Q}$, it follows from the convergence of the Euler product that $L(f,\phi,n)$ is strictly positive for $n = k/2+1, \ldots, k-1$; in addition, the generalized Riemann hypothesis would also imply that the value at $n = k/2$ should either be zero or strictly positive (and this is the case in all of our numerical examples) Thus, by Theorem (2.8), $L_3^*(f,\phi,n)$ is a rational number,which will have the sign $(-1)^n w(f,\phi)$ for $n = 1, \ldots, k/2-1$ by the functional equation (21); and the sign of $L_3^*(f,\phi,k/2)$ should be $(-1)^{k/2}$ if it is non-zero.

Finally, we recall (see Example 5.3 in section 5) that, for the form f of conductor 121 and weight 4, the periods in Table IV are the naive periods, and that they must be replaced by the canonical periods defined in Example 5.3 to deduce the stronger congruence (7) in this case.

Table I: form f of conductor 5 and weight 4.
$L_3^*(f,\sigma,1)=-100,\ L_3^*(f,\sigma,2)=\frac{13}{3}.$

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
	$n=1$					
2	$-2^5\cdot5^3\cdot7\cdot13$	$\frac{2\cdot5^2}{3}$	1	$2^4\cdot3^6\cdot5^2$	$1+O(3)$	$1+O(3)$
3	$-2^4\cdot5^4\cdot13\cdot41$	$\frac{2\cdot5}{3}$	1	$3^{10}\cdot5^2$	$2+O(3)$	$2+O(3)$
5	$2^5\cdot3\cdot5^2\cdot13\cdot17$	0	1	$3^6\cdot5^4$	0	$2\cdot3^1+O(3^2)$
6	$-2^5\cdot5^3\cdot13\cdot1801$	$\frac{2\cdot5^2}{3}$	1	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
7	$-\frac{2^8\cdot5^3\cdot13\cdot23\cdot41}{7}$	$\frac{2^3\cdot5^5}{3\cdot7^2}$	1	$3^6\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
10	$2^3\cdot3\cdot5^3\cdot13$	0	$\frac{2\cdot5}{3}$	$2^4\cdot3^2\cdot5^4$	0	$1\cdot3^1+O(3^2)$
11	$-\frac{2^{10}\cdot5^3\cdot13\cdot2311}{11}$	$\frac{2^5\cdot5^3\cdot41}{3\cdot11^2}$	1	$3^6\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
12	$-2^6\cdot5^3\cdot13\cdot839$	$\frac{2\cdot5^2}{3}$	1	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
13	$-2^4\cdot5^3\cdot11\cdot13\cdot43\cdot53$	$\frac{2^5\cdot5^3\cdot11^2}{3\cdot13^2}$	1	$3^6\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
14	$-2^5\cdot5^4\cdot13^2\cdot251$	$\frac{2^3\cdot5^6}{3\cdot7^2}$	1	$2^4\cdot3^6\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
15	$2^9\cdot3\cdot5^2\cdot13\cdot281$	0	1	$3^{10}\cdot5^4$	0	$2\cdot3^1+O(3^2)$
17	$-\frac{2^4\cdot5^3\cdot13\cdot31\cdot167}{17}$	$\frac{2^6\cdot5^2\cdot7\cdot83}{3\cdot17^2}$	$\frac{2\cdot5}{3}$	$3^2\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
19	$-\frac{2^4\cdot5^4\cdot13\cdot43^2}{19}$	$\frac{2^7\cdot5^3\cdot7^2}{3\cdot19^2}$	$\frac{2\cdot5}{3}$	$3^2\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
20	$2^3\cdot3\cdot5^2\cdot13^2\cdot97$	0	1	$2^4\cdot3^6\cdot5^4$	0	$1\cdot3^1+O(3^2)$
21	$-\frac{2^4\cdot5^3\cdot13\cdot3425341}{7}$	$\frac{2^3\cdot5^5}{3\cdot7^2}$	1	$3^{10}\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
22	$-\frac{2^6\cdot5^3\cdot13\cdot43\cdot13841}{11}$	$\frac{2^5\cdot5^4\cdot41}{3\cdot11^2}$	1	$2^4\cdot3^6\cdot5^2\cdot11^4$	$2+O(3)$	$2+O(3)$
23	$-\frac{2^8\cdot3^5\cdot5^3\cdot13\cdot1409}{23}$	$\frac{2^3\cdot3^2\cdot5^2\cdot7\cdot79}{23^2}$	1	$3^6\cdot5^2\cdot23^4$	$1\cdot3^3+O(3^4)$	$2\cdot3^5+O(3^6)$
26	$-2^5\cdot5^3\cdot13\cdot887$	$\frac{2^5\cdot5^4\cdot11^2}{3\cdot13^2}$	$\frac{2\cdot5}{3}$	$2^4\cdot3^2\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$
28	$-\frac{2^5\cdot5^3\cdot13\cdot503}{7}$	$\frac{2^3\cdot5^6}{3\cdot7^2}$	$\frac{2\cdot5}{3}$	$2^4\cdot3^2\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
29	$-\frac{2^4\cdot5^3\cdot11\cdot13\cdot1678031}{29}$	$\frac{2^6\cdot5^3\cdot23\cdot41}{3\cdot29^2}$	1	$3^6\cdot5^2\cdot29^4$	$1+O(3)$	$1+O(3)$
30	$2^3\cdot3\cdot5^2\cdot7^2\cdot13\cdot61\cdot97$	0	1	$2^4\cdot3^{10}\cdot5^4$	0	$1\cdot3^1+O(3^2)$
31	$-\frac{2^4\cdot5^4\cdot13\cdot79\cdot62351}{31}$	$\frac{2^5\cdot5^5\cdot11^2}{3\cdot31^2}$	1	$3^6\cdot5^2\cdot31^4$	$2+O(3)$	$2+O(3)$
33	$-2^4\cdot5^4\cdot11\cdot13\cdot19\cdot2879$	$\frac{2^5\cdot5^3\cdot41}{3\cdot11^2}$	1	$3^{10}\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
34	$-2^7\cdot5^3\cdot13\cdot142427$	$\frac{2^6\cdot5^3\cdot7\cdot83}{3\cdot17^2}$	1	$2^4\cdot3^6\cdot5^2\cdot17^4$	$2+O(3)$	$2+O(3)$
35	$\frac{2^6\cdot3\cdot5^2\cdot13\cdot653}{7}$	0	$\frac{2\cdot5}{3}$	$3^2\cdot5^4\cdot7^4$	0	$2\cdot3^1+O(3^2)$
37	$-\frac{2^8\cdot3^2\cdot5^4\cdot13\cdot367}{37}$	$\frac{2^5\cdot3\cdot5^3\cdot19^2}{37^2}$	$\frac{2\cdot5}{3}$	$3^2\cdot5^2\cdot37^4$	$2\cdot3^2+O(3^3)$	$2\cdot3^2+O(3^3)$
39	$-2^{10}\cdot5^3\cdot71\cdot17489$	$\frac{2^5\cdot5^3\cdot11^2}{3\cdot13^2}$	1	$3^{10}\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
41	$-\frac{2^4\cdot5^3\cdot13\cdot17\cdot31\cdot211\cdot941}{41}$	$\frac{2^7\cdot5^3\cdot17\cdot109}{3\cdot41^2}$	1	$3^6\cdot5^2\cdot41^4$	$1+O(3)$	$1+O(3)$
42	$-\frac{2^5\cdot5^3\cdot13\cdot19\cdot859\cdot1801}{7}$	$\frac{2^3\cdot5^6}{3\cdot7^2}$	1	$2^4\cdot3^{10}\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
43	$-\frac{2^6\cdot5^3\cdot7\cdot13\cdot19\cdot251\cdot491}{43}$	$\frac{2^3\cdot5^5\cdot29^2}{3\cdot43^2}$	1	$3^6\cdot5^2\cdot43^4$	$2+O(3)$	$2+O(3)$
44	$-2^5\cdot5^3\cdot13\cdot421$	$\frac{2^5\cdot5^4\cdot41}{3\cdot11^2}$	$\frac{2\cdot5}{3}$	$2^4\cdot3^2\cdot5^2\cdot11^4$	$2+O(3)$	$2+O(3)$
45	$2^5\cdot3\cdot5^3\cdot7^2\cdot13\cdot19$	0	1	$3^{10}\cdot5^4$	0	$2\cdot3^1+O(3^2)$
46	$-\frac{2^5\cdot3^3\cdot5^3\cdot13\cdot7283}{23}$	$\frac{2^3\cdot3^2\cdot5^3\cdot7\cdot79}{23^2}$	$\frac{2\cdot5}{3}$	$2^4\cdot3^2\cdot5^2\cdot23^4$	$2\cdot3^3+O(3^4)$	$2\cdot3^3+O(3^4)$
47	$\frac{2^8\cdot5^3\cdot13\cdot23^2\cdot22567}{47}$	1	$\frac{2^3\cdot5^2\cdot13\cdot67\cdot277}{3\cdot47^2}$	$3^6\cdot5^2\cdot47^4$	$1+O(3)$	$1+O(3)$

Table I: form f of conductor 5 and weight 4.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
50	$2^3\cdot 3\cdot 5^2\cdot 13^2\cdot 97$	0	1	$2^4\cdot 3^6\cdot 5^4$	0	$1\cdot 3^1+O(3^2)$
51	$-\frac{2^8\cdot 5^4\cdot 13^2\cdot 278591}{17}$	$\frac{2^6\cdot 5^2\cdot 7\cdot 83}{3\cdot 17^2}$	1	$3^{10}\cdot 5^2\cdot 17^4$	$1+O(3)$	$1+O(3)$
52	$-2^5\cdot 5^3\cdot 2513617$	$\frac{2^5\cdot 5^4\cdot 11^2}{3\cdot 13^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 13^4$	$1+O(3)$	$1+O(3)$
53	$-\frac{2^4\cdot 5^4\cdot 13\cdot 290161}{53}$	$\frac{2^7\cdot 5^2\cdot 11\cdot 13\cdot 179}{3\cdot 53^2}$	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^2\cdot 53^4$	$1+O(3)$	$1+O(3)$
55	$\frac{2^5\cdot 3\cdot 5^2\cdot 13\cdot 12799}{11}$	0	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^4\cdot 11^4$	0	$1\cdot 3^1+O(3^2)$
57	$-\frac{2^6\cdot 5^3\cdot 13\cdot 61\cdot 503\cdot 4241}{19}$	$\frac{2^7\cdot 5^3\cdot 7^2}{3\cdot 19^2}$	1	$3^{10}\cdot 5^2\cdot 19^4$	$2+O(3)$	$2+O(3)$
58	$-\frac{2^9\cdot 5^3\cdot 13\cdot 9208039}{29}$	$\frac{2^6\cdot 5^4\cdot 23\cdot 41}{3\cdot 29^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 29^4$	$2+O(3)$	$2+O(3)$
59	$-\frac{2^8\cdot 5^4\cdot 13\cdot 23\cdot 397\cdot 853}{59}$	$\frac{2^9\cdot 5^3\cdot 19\cdot 101}{3\cdot 59^2}$	1	$3^6\cdot 5^2\cdot 59^4$	$1+O(3)$	$1+O(3)$
60	$2^5\cdot 3\cdot 5^2\cdot 13\cdot 19\cdot 3499$	0	1	$2^4\cdot 3^{10}\cdot 5^4$	0	$1\cdot 3^1+O(3^2)$
61	$\frac{2^6\cdot 5^3\cdot 13\cdot 179702101}{61}$	1	$\frac{2^5\cdot 5^5\cdot 43^2}{3\cdot 61^2}$	$3^6\cdot 5^2\cdot 61^4$	$2+O(3)$	$2+O(3)$
62	$-\frac{2^5\cdot 5^3\cdot 13\cdot 307\cdot 2879}{31}$	$\frac{2^5\cdot 5^6\cdot 11^2}{3\cdot 31^2}$	$\frac{2\cdot 5}{3}$	$2^4\cdot 3^2\cdot 5^2\cdot 31^4$	$1+O(3)$	$1+O(3)$
66	$-\frac{2^8\cdot 5^4\cdot 13\cdot 6458773}{11}$	$\frac{2^5\cdot 5^4\cdot 41}{3\cdot 11^2}$	1	$2^4\cdot 3^{10}\cdot 5^2\cdot 11^4$	$2+O(3)$	$2+O(3)$
67	$-\frac{2^6\cdot 5^3\cdot 13^3\cdot 19^2\cdot 4759}{67}$	$\frac{2^3\cdot 5^3\cdot 443^2}{3\cdot 67^2}$	1	$3^6\cdot 5^2\cdot 67^4$	$2+O(3)$	$2+O(3)$
68	$-\frac{2^5\cdot 5^3\cdot 13\cdot 10484557}{17}$	$\frac{2^6\cdot 5^3\cdot 7\cdot 83}{3\cdot 17^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 17^4$	$2+O(3)$	$2+O(3)$
69	$-\frac{2^6\cdot 3^3\cdot 5^3\cdot 13\cdot 857\cdot 15733}{23}$	$\frac{2^3\cdot 3^2\cdot 5^2\cdot 7\cdot 79}{23^2}$	1	$3^{10}\cdot 5^2\cdot 23^4$	$1\cdot 3^3+O(3^4)$	$2\cdot 3^3+O(3^4)$
70	$\frac{2^9\cdot 3\cdot 5^4\cdot 13\cdot 89\cdot 131}{7}$	0	1	$2^4\cdot 3^6\cdot 5^4\cdot 7^4$	0	$1\cdot 3^1+O(3^2)$
71	$-\frac{2^6\cdot 5^3\cdot 7\cdot 13\cdot 31\cdot 79\cdot 101}{71}$	$\frac{2^5\cdot 5^3\cdot 47\cdot 1381}{3\cdot 71^2}$	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^2\cdot 71^4$	$1+O(3)$	$1+O(3)$
73	$-\frac{2^4\cdot 5^4\cdot 13\cdot 17\cdot 47\cdot 1831}{73}$	$\frac{2^7\cdot 5^3\cdot 157^2}{3\cdot 73^2}$	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^2\cdot 73^4$	$2+O(3)$	$2+O(3)$
74	$-\frac{2^5\cdot 3^4\cdot 5^3\cdot 11\cdot 13\cdot 523\cdot 1031}{37}$	$\frac{2^5\cdot 3\cdot 5^4\cdot 19^2}{37^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 37^4$	$1\cdot 3^2+O(3^3)$	$1\cdot 3^4+O(3^5)$
75	$2^5\cdot 3\cdot 5^3\cdot 7^2\cdot 13\cdot 19$	0	1	$3^{10}\cdot 5^4$	0	$2\cdot 3^1+O(3^2)$
76	$-\frac{2^8\cdot 5^3\cdot 13\cdot 311\cdot 7297}{19}$	$\frac{2^7\cdot 5^4\cdot 7^2}{3\cdot 19^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 19^4$	$1+O(3)$	$1+O(3)$
77	$-\frac{2^8\cdot 5^3\cdot 13\cdot 2377\cdot 60913}{7\cdot 11}$	$\frac{2^7\cdot 5^7\cdot 41}{3\cdot 7^2\cdot 11^2}$	1	$3^6\cdot 5^2\cdot 7^4\cdot 11^4$	$1+O(3)$	$1+O(3)$
82	$-\frac{2^6\cdot 5^4\cdot 13\cdot 73\cdot 4817}{41}$	$\frac{2^7\cdot 5^4\cdot 17\cdot 109}{3\cdot 41^2}$	$\frac{2\cdot 5}{3}$	$2^4\cdot 3^2\cdot 5^2\cdot 41^4$	$2+O(3)$	$2+O(3)$
84	$-2^5\cdot 5^3\cdot 13\cdot 431\cdot 10259$	$\frac{2^3\cdot 5^6}{3\cdot 7^2}$	1	$2^4\cdot 3^{10}\cdot 5^2\cdot 7^4$	$1+O(3)$	$1+O(3)$
89	$-\frac{2^6\cdot 3^5\cdot 5^3\cdot 13\cdot 71\cdot 293}{89}$	$\frac{2^8\cdot 3\cdot 5^3\cdot 17\cdot 131}{89^2}$	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^2\cdot 89^4$	$1\cdot 3^2+O(3^3)$	$2\cdot 3^5+O(3^6)$
90	$2^5\cdot 3\cdot 5^2\cdot 13\cdot 151\cdot 463$	0	1	$2^4\cdot 3^{10}\cdot 5^4$	0	$1\cdot 3^1+O(3^2)$
91	$-\frac{2^4\cdot 5^4\cdot 4519393}{7}$	$\frac{2^7\cdot 5^7\cdot 11^2}{3\cdot 7^2\cdot 13^2}$	$\frac{2\cdot 5}{3}$	$3^2\cdot 5^2\cdot 7^4\cdot 13^4$	$2+O(3)$	$2+O(3)$
92	$-\frac{2^5\cdot 3^2\cdot 5^3\cdot 13\cdot 157\cdot 31019}{23}$	$\frac{2^3\cdot 3^2\cdot 5^3\cdot 7\cdot 79}{23^2}$	1	$2^4\cdot 3^6\cdot 5^2\cdot 23^4$	$2\cdot 3^3+O(3^4)$	$1\cdot 3^2+O(3^3)$
	$n=2$					
2	$\frac{2\cdot 5^2\cdot 13}{3^3}$	1	$\frac{5^2}{2\cdot 3^2}$	$2^4\cdot 3^6\cdot 5^2$	$1+O(3)$	$1+O(3)$
3	$\frac{2^2\cdot 5^2\cdot 13}{3^5}$	1	$\frac{2\cdot 5}{3^2}$	$3^{10}\cdot 5^2$	$2+O(3)$	$2+O(3)$
5	0	0	$\frac{2^4}{3\cdot 5}$	$3^6\cdot 5^4$	$1\cdot 3^1+O(3^2)$	0
6	$\frac{2\cdot 5^2\cdot 7^2\cdot 13}{3^5}$	1	$\frac{5^2}{2\cdot 3^2}$	$2^4\cdot 3^{10}\cdot 5^2$	$1+O(3)$	$1+O(3)$
7	$\frac{2^6\cdot 5^2\cdot 13}{3^3\cdot 7^3}$	1	$\frac{2^3\cdot 5^5}{3^2\cdot 7^4}$	$3^6\cdot 5^2\cdot 7^4$	$2+O(3)$	$2+O(3)$
10	0	0	$\frac{2^2}{3}$	$2^4\cdot 3^2\cdot 5^4$	$2\cdot 3^1+O(3^2)$	0
11	$\frac{2^6\cdot 5^2\cdot 13}{3^3\cdot 11}$	1	$\frac{2^5\cdot 5^3\cdot 41}{3^2\cdot 11^4}$	$3^6\cdot 5^2\cdot 11^4$	$1+O(3)$	$1+O(3)$
12	$\frac{2^5\cdot 5^2\cdot 13}{3^5}$	1	$\frac{5^2}{2\cdot 3^2}$	$2^4\cdot 3^{10}\cdot 5^2$	$1+O(3)$	$1+O(3)$
13	$\frac{2^2\cdot 5^2}{3^3\cdot 13^2}$	1	$\frac{2^5\cdot 5^3\cdot 11^2}{3^2\cdot 13^4}$	$3^6\cdot 5^2\cdot 13^4$	$2+O(3)$	$2+O(3)$

Table I: form f of conductor 5 and weight 4.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
14	$\frac{2\cdot5^4\cdot13}{3^3\cdot7^3}$	1	$\frac{2\cdot5^6}{3^2\cdot7^4}$	$2^4\cdot3^6\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
15	0	0	$\frac{2^4}{3\cdot5}$	$3^{10}\cdot5^4$	$1\cdot3^1+O(3^2)$	0
17	$\frac{2^2\cdot5^4\cdot13}{3\cdot17^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^6\cdot5^2\cdot7\cdot83}{3^2\cdot17^4}$	$3^2\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
19	$\frac{2^2\cdot5^2\cdot13^3}{3\cdot19^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^7\cdot5^3\cdot7^2}{3^2\cdot19^4}$	$3^2\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
20	0	0	$\frac{2^2}{3}$	$2^4\cdot3^6\cdot5^4$	$2\cdot3^1+O(3^2)$	0
21	$\frac{2^2\cdot5^2\cdot13\cdot67^2}{3^5\cdot7^3}$	1	$\frac{2^3\cdot5^5}{3^2\cdot7^4}$	$3^{10}\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
22	$\frac{2^3\cdot5^2\cdot13}{3^3\cdot11}$	1	$\frac{2^3\cdot5^4\cdot41}{3^2\cdot11^4}$	$2^4\cdot3^6\cdot5^2\cdot11^4$	$2+O(3)$	$2+O(3)$
23	$\frac{2^6\cdot5^4\cdot13}{3\cdot23^3}$	1	$\frac{2^3\cdot3\cdot5^2\cdot7\cdot79}{23^4}$	$3^6\cdot5^2\cdot23^4$	$1\cdot3^3+O(3^4)$	$1\cdot3^2+O(3^3)$
26	$\frac{2\cdot5^2}{3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^3\cdot5^4\cdot11^2}{3^2\cdot13^4}$	$2^4\cdot3^2\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$
28	$\frac{2\cdot5^2\cdot13}{3\cdot7^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2\cdot5^6}{3^2\cdot7^4}$	$2^4\cdot3^2\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
29	$\frac{2^2\cdot5^2\cdot13\cdot109^2}{3^3\cdot29^3}$	1	$\frac{2^6\cdot5^3\cdot23\cdot41}{3^2\cdot29^4}$	$3^6\cdot5^2\cdot29^4$	$1+O(3)$	$1+O(3)$
30	0	0	$\frac{2^2}{3}$	$2^4\cdot3^{10}\cdot5^4$	$2\cdot3^1+O(3^2)$	0
31	$\frac{2^2\cdot5^2\cdot13\cdot151^2}{3^3\cdot31^3}$	1	$\frac{2^5\cdot5^5\cdot11^2}{3^2\cdot31^4}$	$3^6\cdot5^2\cdot31^4$	$2+O(3)$	$2+O(3)$
33	$\frac{2^2\cdot5^4\cdot7^2\cdot13}{3^5\cdot11^3}$	1	$\frac{2^5\cdot5^3\cdot41}{3^2\cdot11^4}$	$3^{10}\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
34	$\frac{2^3\cdot5^2\cdot13}{3^3\cdot17^3}$	1	$\frac{2^4\cdot5^3\cdot7\cdot83}{3^2\cdot17^4}$	$2^4\cdot3^6\cdot5^2\cdot17^4$	$2+O(3)$	$2+O(3)$
35	0	0	$\frac{2^6\cdot5^3}{3\cdot7^4}$	$3^2\cdot5^4\cdot7^4$	$1\cdot3^1+O(3^2)$	0
37	$\frac{2^6\cdot3\cdot5^2\cdot13}{37^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^5\cdot5^3\cdot19^2}{37^4}$	$3^2\cdot5^2\cdot37^4$	$2\cdot3^2+O(3^3)$	$2\cdot3^2+O(3^3)$
38	$\frac{2\cdot5^2\cdot13\cdot41^2}{3^3\cdot19^3}$	1	$\frac{2^5\cdot5^4\cdot7^2}{3^2\cdot19^4}$	$2^4\cdot3^6\cdot5^2\cdot19^4$	$1+O(3)$	$1+O(3)$
39	$\frac{2^6\cdot5^2\cdot11^2}{3^5\cdot13^2}$	1	$\frac{2^5\cdot5^3\cdot11^2}{3^2\cdot13^4}$	$3^{10}\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
41	$\frac{2^2\cdot5^2\cdot11^2\cdot13}{3^3\cdot41^3}$	1	$\frac{2^7\cdot5^3\cdot17\cdot109}{3^2\cdot41^4}$	$3^6\cdot5^2\cdot41^4$	$1+O(3)$	$1+O(3)$
42	$\frac{2\cdot5^2\cdot13\cdot149^2}{3^5\cdot7^3}$	1	$\frac{2\cdot5^6}{3^2\cdot7^4}$	$2^4\cdot3^{10}\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
43	$\frac{2^4\cdot5^6\cdot13}{3^3\cdot43^3}$	1	$\frac{2^3\cdot5^5\cdot29^2}{3^2\cdot43^4}$	$3^6\cdot5^2\cdot43^4$	$2+O(3)$	$2+O(3)$
44	$\frac{2\cdot5^2\cdot7^2\cdot13}{3\cdot11^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^3\cdot5^4\cdot41}{3^2\cdot11^4}$	$2^4\cdot3^2\cdot5^2\cdot11^4$	$2+O(3)$	$2+O(3)$
45	0	0	$\frac{2^4}{3\cdot5}$	$3^{10}\cdot5^4$	$1\cdot3^1+O(3^2)$	0
46	$\frac{2\cdot3\cdot5^2\cdot13}{23^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2\cdot3\cdot5^3\cdot7\cdot79}{23^4}$	$2^4\cdot3^2\cdot5^2\cdot23^4$	$2\cdot3^3+O(3^4)$	$2\cdot3^2+O(3^3)$
47	$\frac{2^{10}\cdot5^2\cdot13^3}{3^3\cdot47^3}$	1	$\frac{2^3\cdot5^2\cdot13\cdot67\cdot277}{3^2\cdot47^4}$	$3^6\cdot5^2\cdot47^4$	$1+O(3)$	$1+O(3)$
50	0	0	$\frac{2^2}{3}$	$2^4\cdot3^6\cdot5^4$	$2\cdot3^1+O(3^2)$	0
51	$\frac{2^6\cdot5^2\cdot13\cdot101^2}{3^5\cdot17^3}$	1	$\frac{2^6\cdot5^2\cdot7\cdot83}{3^2\cdot17^4}$	$3^{10}\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
52	$\frac{2\cdot5^2}{3^3\cdot13^2}$	1	$\frac{2^3\cdot5^4\cdot11^2}{3^2\cdot13^4}$	$2^4\cdot3^6\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$
53	$\frac{2^2\cdot5^2\cdot13\cdot29^2}{3\cdot53^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^7\cdot5^2\cdot11\cdot13\cdot179}{3^2\cdot53^4}$	$3^2\cdot5^2\cdot53^4$	$1+O(3)$	$1+O(3)$
55	0	0	$\frac{2^8\cdot5\cdot41}{3\cdot11^4}$	$3^2\cdot5^4\cdot11^4$	$2\cdot3^1+O(3^2)$	0
57	$\frac{2^4\cdot5^2\cdot7^4\cdot13}{3^5\cdot19^3}$	1	$\frac{2^7\cdot5^3\cdot7^2}{3^2\cdot19^4}$	$3^{10}\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
58	$\frac{2^{11}\cdot5^4\cdot13}{3^3\cdot29^3}$	1	$\frac{2^4\cdot5^4\cdot23\cdot41}{3^2\cdot29^4}$	$2^4\cdot3^6\cdot5^2\cdot29^4$	$2+O(3)$	$2+O(3)$
59	$\frac{2^6\cdot5^2\cdot11^4\cdot13}{3^3\cdot59^3}$	1	$\frac{2^9\cdot5^3\cdot19\cdot101}{3^2\cdot59^4}$	$3^6\cdot5^2\cdot59^4$	$1+O(3)$	$1+O(3)$
60	0	0	$\frac{2^2}{3}$	$2^4\cdot3^{10}\cdot5^4$	$2\cdot3^1+O(3^2)$	0
61	$\frac{2^4\cdot5^2\cdot7^2\cdot13\cdot17^2}{3^3\cdot61^3}$	1	$\frac{2^5\cdot5^5\cdot43^2}{3^2\cdot61^4}$	$3^6\cdot5^2\cdot61^4$	$2+O(3)$	$2+O(3)$
62	$\frac{2\cdot5^4\cdot7^2\cdot13}{3\cdot31^3}$	$\frac{2\cdot5}{3^2}$	$\frac{2^3\cdot5^6\cdot11^2}{3^2\cdot31^4}$	$2^4\cdot3^2\cdot5^2\cdot31^4$	$1+O(3)$	$1+O(3)$
66	$\frac{2^5\cdot5^2\cdot13}{3^5\cdot11}$	1	$\frac{2^3\cdot5^4\cdot41}{3^2\cdot11^4}$	$2^4\cdot3^{10}\cdot5^2\cdot11^4$	$2+O(3)$	$2+O(3)$

Table I: form f of conductor 5 and weight 4.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
67	$\frac{2^4 \cdot 5^2 \cdot 13 \cdot 103^2}{3^3 \cdot 67^3}$	1	$\frac{2^3 \cdot 5^3 \cdot 443^2}{3^2 \cdot 67^4}$	$3^6 \cdot 5^2 \cdot 67^4$	$2 + O(3)$	$2 + O(3)$
68	$\frac{2 \cdot 5^2 \cdot 7^2 \cdot 13^3}{3^3 \cdot 17^3}$	1	$\frac{2^4 \cdot 5^3 \cdot 7 \cdot 83}{3^2 \cdot 17^4}$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 17^4$	$2 + O(3)$	$2 + O(3)$
69	$\frac{2^4 \cdot 5^2 \cdot 13 \cdot 31^2}{3^3 \cdot 23^3}$	1	$\frac{2^3 \cdot 3 \cdot 5^2 \cdot 7 \cdot 79}{23^4}$	$3^{10} \cdot 5^2 \cdot 23^4$	$1 \cdot 3^3 + O(3^4)$	$1 \cdot 3^2 + O(3^3)$
70	0	0	$\frac{2^4 \cdot 5^4}{3 \cdot 7^4}$	$2^4 \cdot 3^6 \cdot 5^4 \cdot 7^4$	$2 \cdot 3^1 + O(3^2)$	0
71	$\frac{2^4 \cdot 5^2 \cdot 13 \cdot 29^2}{3 \cdot 71^3}$	$\frac{2 \cdot 5}{3^2}$	$\frac{2^5 \cdot 5^3 \cdot 47 \cdot 1381}{3^2 \cdot 71^4}$	$3^2 \cdot 5^2 \cdot 71^4$	$1 + O(3)$	$1 + O(3)$
73	$\frac{2^2 \cdot 5^2 \cdot 13 \cdot 43^2}{3 \cdot 73^3}$	$\frac{2 \cdot 5}{3^2}$	$\frac{2^7 \cdot 5^3 \cdot 157^2}{3^2 \cdot 73^4}$	$3^2 \cdot 5^2 \cdot 73^4$	$2 + O(3)$	$2 + O(3)$
74	$\frac{2 \cdot 3^3 \cdot 5^2 \cdot 13}{37^3}$	1	$\frac{2^3 \cdot 5^4 \cdot 19^2}{37^4}$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 37^4$	$1 \cdot 3^2 + O(3^3)$	$1 \cdot 3^6 + O(3^7)$
75	0	0	$\frac{2^4}{3 \cdot 5}$	$3^{10} \cdot 5^4$	$1 \cdot 3^1 + O(3^2)$	0
76	$\frac{2^5 \cdot 5^2 \cdot 13^3}{3^3 \cdot 19^3}$	1	$\frac{2^5 \cdot 5^4 \cdot 7^2}{3^2 \cdot 19^4}$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 19^4$	$1 + O(3)$	$1 + O(3)$
77	$\frac{2^6 \cdot 5^2 \cdot 13 \cdot 101^2}{3^3 \cdot 7^3 \cdot 11^3}$	1	$\frac{2^7 \cdot 5^7 \cdot 41}{3^2 \cdot 7^4 \cdot 11^4}$	$3^6 \cdot 5^2 \cdot 7^4 \cdot 11^4$	$1 + O(3)$	$1 + O(3)$
82	$\frac{2^3 \cdot 5^2 \cdot 11^2 \cdot 13}{3 \cdot 41^3}$	$\frac{2 \cdot 5}{3^2}$	$\frac{2^5 \cdot 5^4 \cdot 17 \cdot 109}{3^2 \cdot 41^4}$	$2^4 \cdot 3^2 \cdot 5^2 \cdot 41^4$	$2 + O(3)$	$2 + O(3)$
83	$\frac{2^6 \cdot 3 \cdot 5^2 \cdot 7^2 \cdot 13}{83^3}$	1	$\frac{2^3 \cdot 3 \cdot 5^2 \cdot 13 \cdot 31 \cdot 223}{83^4}$	$3^6 \cdot 5^2 \cdot 83^4$	$1 \cdot 3^3 + O(3^4)$	$1 \cdot 3^4 + O(3^5)$
84	$\frac{2 \cdot 5^2 \cdot 13 \cdot 199^2}{3^5 \cdot 7^3}$	1	$\frac{2 \cdot 5^6}{3^2 \cdot 7^4}$	$2^4 \cdot 3^{10} \cdot 5^2 \cdot 7^4$	$1 + O(3)$	$1 + O(3)$
89	$\frac{2^4 \cdot 3^3 \cdot 5^4 \cdot 13}{89^3}$	$\frac{2 \cdot 5}{3^2}$	$\frac{2^8 \cdot 5^3 \cdot 17 \cdot 131}{89^4}$	$3^2 \cdot 5^2 \cdot 89^4$	$1 \cdot 3^2 + O(3^3)$	$1 \cdot 3^4 + O(3^5)$
90	0	0	$\frac{2^2}{3}$	$2^4 \cdot 3^{10} \cdot 5^4$	$2 \cdot 3^1 + O(3^2)$	0
91	$\frac{2^2 \cdot 5^2}{3 \cdot 7^3 \cdot 13^2}$	$\frac{2 \cdot 5}{3^2}$	$\frac{2^7 \cdot 5^7 \cdot 11^2}{3^2 \cdot 7^4 \cdot 13^4}$	$3^2 \cdot 5^2 \cdot 7^4 \cdot 13^4$	$2 + O(3)$	$2 + O(3)$
92	$\frac{2 \cdot 3 \cdot 5^2 \cdot 13}{23^3}$	1	$\frac{2 \cdot 3 \cdot 5^3 \cdot 7 \cdot 79}{23^4}$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 23^4$	$2 \cdot 3^3 + O(3^4)$	$2 \cdot 3^4 + O(3^5)$
93	$\frac{2^4 \cdot 5^2 \cdot 13 \cdot 31^3 \cdot 179^2}{3^5}$	1	$\frac{2^5 \cdot 5^5 \cdot 11^2}{3^2 \cdot 31^4}$	$3^{10} \cdot 5^2 \cdot 31^4$	$2 + O(3)$	$2 + O(3)$
94	$\frac{2 \cdot 5^2 \cdot 11^2 \cdot 13 \cdot 19^2}{3^3 \cdot 47^3}$	1	$\frac{2 \cdot 5^3 \cdot 13 \cdot 67 \cdot 277}{3^2 \cdot 47^4}$	$2^4 \cdot 3^6 \cdot 5^2 \cdot 47^4$	$2 + O(3)$	$2 + O(3)$
97	$\frac{2^6 \cdot 3 \cdot 5^6 \cdot 13}{97^3}$	1	$\frac{2^{11} \cdot 5^3 \cdot 19^2}{97^4}$	$3^6 \cdot 5^2 \cdot 97^4$	$2 \cdot 3^2 + O(3^3)$	$2 \cdot 3^4 + O(3^5)$

Table II: form f of conductor 7 and weight 4.
$L_3^*(f,\sigma,1) = 49, L_3^*(f,\sigma,2) = L_3^*(f,\rho,2) = 0.$

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
2	$2^3 \cdot 3 \cdot 5 \cdot 7^4$	1	$\frac{5 \cdot 7^2}{2 \cdot 3}$	$2^4 \cdot 3^6 \cdot 7^2$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
3	$2^2 \cdot 3 \cdot 5 \cdot 7^3 \cdot 13^2$	1	$\frac{2 \cdot 7}{3}$	$3^{10} \cdot 7^2$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
5	$2^7 \cdot 3 \cdot 7^3 \cdot 71$	1	$\frac{2^3 \cdot 7^2 \cdot 23}{3 \cdot 5^2}$	$3^6 \cdot 5^4 \cdot 7^2$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
6	$2^4 \cdot 3 \cdot 5 \cdot 7^4 \cdot 113$	1	$\frac{5 \cdot 7^2}{2 \cdot 3}$	$2^4 \cdot 3^{10} \cdot 7^2$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
7	$2^2 \cdot 3 \cdot 5 \cdot 7^2 \cdot 223$	1	$\frac{2^3 \cdot 7}{3}$	$3^6 \cdot 7^4$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
10	$2^3 \cdot 7^3 \cdot 239$	$\frac{2 \cdot 7}{3}$	$\frac{2 \cdot 7^3 \cdot 23}{3 \cdot 5}$	$2^4 \cdot 3^2 \cdot 5^4 \cdot 7^2$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
11	$\frac{2^2 \cdot 3 \cdot 5 \cdot 7^3 \cdot 211 \cdot 499}{11}$	1	$\frac{2^5 \cdot 5 \cdot 7^2 \cdot 31}{3 \cdot 11^2}$	$3^6 \cdot 7^2 \cdot 11^4$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
12	$2^3 \cdot 3 \cdot 5 \cdot 7^4 \cdot 241$	1	$\frac{5 \cdot 7^2}{2 \cdot 3}$	$2^4 \cdot 3^{10} \cdot 7^2$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
13	$\frac{2^8 \cdot 3 \cdot 5^2 \cdot 7^3 \cdot 773}{13}$	1	$\frac{2^3 \cdot 7^3 \cdot 11^2}{3 \cdot 13^2}$	$3^6 \cdot 7^2 \cdot 13^4$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
14	$2^3 \cdot 3 \cdot 5 \cdot 7^2 \cdot 41 \cdot 59$	1	$\frac{2 \cdot 5 \cdot 7^2}{3}$	$2^4 \cdot 3^6 \cdot 7^4$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
15	$2^2 \cdot 3 \cdot 5 \cdot 7^3 \cdot 13 \cdot 43 \cdot 179$	1	$\frac{2^3 \cdot 7^2 \cdot 23}{3 \cdot 5^2}$	$3^{10} \cdot 5^4 \cdot 7^2$	$1 \cdot 3^1 + O(3^2)$	$1 \cdot 3^1 + O(3^2)$
17	$\frac{2^2 \cdot 3^2 \cdot 5 \cdot 7^3 \cdot 1223}{17}$	$\frac{2 \cdot 7}{3}$	$\frac{2^6 \cdot 3^3 \cdot 5 \cdot 7^2}{17^2}$	$3^2 \cdot 7^2 \cdot 17^4$	$2 \cdot 3^5 + O(3^6)$	$1 \cdot 3^3 + O(3^4)$
19	$\frac{2^{11} \cdot 5 \cdot 7^3 \cdot 37}{19}$	$\frac{2 \cdot 7}{3}$	$\frac{2^3 \cdot 5^2 \cdot 7^5}{3 \cdot 19^2}$	$3^2 \cdot 7^2 \cdot 19^4$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
20	$2^7 \cdot 3 \cdot 7^3 \cdot 1213$	1	$\frac{2 \cdot 7^3 \cdot 23}{3 \cdot 5}$	$2^4 \cdot 3^6 \cdot 5^4 \cdot 7^2$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$
21	$2^{13} \cdot 3 \cdot 5 \cdot 7^2 \cdot 29$	1	$\frac{2^3 \cdot 7}{3}$	$3^{10} \cdot 7^4$	$2 \cdot 3^1 + O(3^2)$	$2 \cdot 3^1 + O(3^2)$

Table II: form f of conductor 7 and weight 4.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
22	$\frac{2^4\cdot3\cdot5\cdot7^3\cdot19\cdot28277}{11}$	1	$\frac{2^3\cdot5^2\cdot7^3\cdot31}{3\cdot11^2}$	$2^4\cdot3^6\cdot7^2\cdot11^4$	$2\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
23	$\frac{2^2\cdot3^3\cdot5\cdot7^3\cdot47\cdot10463}{23}$	1	$\frac{2^7\cdot3^2\cdot5^2\cdot7^2}{23^2}$	$3^6\cdot7^2\cdot23^4$	$2\cdot3^4+O(3^5)$	$1\cdot3^3+O(3^4)$
26	$\frac{2^3\cdot5\cdot7^4\cdot3917}{13}$	$\frac{2\cdot7}{3}$	$\frac{2\cdot5\cdot7^4\cdot11^2}{3\cdot13^2}$	$2^4\cdot3^2\cdot7^2\cdot13^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
28	$2^4\cdot5\cdot7^2\cdot13$	$\frac{2\cdot7}{3}$	$\frac{2\cdot5\cdot7^2}{3}$	$2^4\cdot3^2\cdot7^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
29	$\frac{2^2\cdot3\cdot5\cdot7^4\cdot1904647}{29}$	1	$\frac{2^6\cdot5^2\cdot7^3\cdot19}{3\cdot29^2}$	$3^6\cdot7^2\cdot29^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
30	$2^3\cdot3\cdot7^3\cdot19\cdot266839$	1	$\frac{2\cdot7^3\cdot23}{3\cdot5}$	$2^4\cdot3^{10}\cdot5^4\cdot7^2$	$2\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
31	$\frac{2^8\cdot3\cdot5\cdot7^3\cdot307267}{31}$	1	$\frac{2^5\cdot5^2\cdot7^5}{3\cdot31^2}$	$3^6\cdot7^2\cdot31^4$	$2\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
33	$\frac{2^7\cdot3\cdot5\cdot7^3\cdot849221}{11}$	1	$\frac{2^5\cdot5\cdot7^2\cdot31}{3\cdot11^2}$	$3^{10}\cdot7^2\cdot11^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
34	$\frac{2^{10}\cdot3^3\cdot5\cdot7^3\cdot83\cdot101}{17}$	1	$\frac{2^4\cdot3^3\cdot5^2\cdot7^3}{17^2}$	$2^4\cdot3^6\cdot7^2\cdot17^4$	$1\cdot3^5+O(3^6)$	$1\cdot3^3+O(3^4)$
35	$2^2\cdot7^2\cdot43\cdot191$	$\frac{2\cdot7}{3}$	$\frac{2^5\cdot7^2\cdot23}{3\cdot5^2}$	$3^2\cdot5^4\cdot7^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
37	$\frac{2^7\cdot5\cdot7^3\cdot15937}{37}$	$\frac{2\cdot7}{3}$	$\frac{2^5\cdot7^3\cdot59^2}{3\cdot37^2}$	$3^2\cdot7^2\cdot37^4$	$2\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
38	$\frac{2^5\cdot3\cdot5^2\cdot7^3\cdot864947}{19}$	1	$\frac{2\cdot5^3\cdot7^6}{3\cdot19^2}$	$2^4\cdot3^6\cdot7^2\cdot19^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
39	$\frac{2^8\cdot3\cdot5\cdot7^3\cdot957811}{13}$	1	$\frac{2^3\cdot7^3\cdot11^2}{3\cdot13^2}$	$3^{10}\cdot7^2\cdot13^4$	$2\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
41	$\frac{2^4\cdot3\cdot5^2\cdot7^3\cdot13\cdot173\cdot1693}{41}$	1	$\frac{2^7\cdot5\cdot7^3\cdot11\cdot17}{3\cdot41^2}$	$3^6\cdot7^2\cdot41^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
42	$2^3\cdot3\cdot5\cdot7^2\cdot37\cdot15601$	1	$\frac{2\cdot5\cdot7^2}{3}$	$2^4\cdot3^{10}\cdot7^4$	$1\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
43	$\frac{2^{11}\cdot3^3\cdot5\cdot7^3\cdot20639}{43}$	1	$\frac{2^5\cdot3^3\cdot7^5}{43^2}$	$3^6\cdot7^2\cdot43^4$	$2\cdot3^5+O(3^6)$	$2\cdot3^3+O(3^4)$

Table III: form f of conductor 5 and weight 6.
$L_3^*(f,\sigma,1)=-400, L_3^*(f,\sigma,2)=62/15, L_3^*(f,\sigma,3)=-31/1125.$

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
	$n=1$					
2	$-2^{12}\cdot5^2\cdot31\cdot661$	1	$\frac{2^7\cdot5\cdot11}{3}$	$2^4\cdot3^6\cdot5^2$	$1+O(3)$	$1+O(3)$
3	$-2^{10}\cdot5^3\cdot13\cdot31\cdot2953$	1	$\frac{2^3\cdot11}{3}$	$3^{10}\cdot5^2$	$2+O(3)$	$2+O(3)$
5	$2^8\cdot3\cdot5\cdot31\cdot193\cdot211$	1	$-2^6\cdot11$	$3^6\cdot5^4$	$2\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
6	$-2^{11}\cdot5^3\cdot31\cdot137\cdot39323$	1	$\frac{2^7\cdot5\cdot11}{3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
7	$-2^{12}\cdot5^2\cdot31\cdot14230919$	1	$\frac{2^9\cdot11\cdot277^2}{3\cdot7^2}$	$3^6\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
10	$2^7\cdot3\cdot5\cdot31\cdot1097$	$\frac{2^3\cdot11}{3}$	$-2^{10}\cdot5\cdot11$	$2^4\cdot3^2\cdot5^4$	$1\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
11	$-\frac{2^{13}\cdot5^3\cdot31\cdot971\cdot592759}{11}$	1	$\frac{2^{10}\cdot5^2\cdot7^2\cdot37^2}{3\cdot11}$	$3^6\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
12	$-2^{11}\cdot5^2\cdot7^2\cdot31\cdot533063$	1	$\frac{2^7\cdot5\cdot11}{3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
13	$-\frac{2^{10}\cdot5^2\cdot7\cdot11\cdot31\cdot211\cdot6591061}{13}$	1	$\frac{2^{17}\cdot11\cdot17^2}{3}$	$3^6\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
14	$-\frac{2^{12}\cdot5^4\cdot31\cdot1082124649}{7}$	1	$\frac{2^{13}\cdot5\cdot11\cdot277^2}{3\cdot7^2}$	$2^4\cdot3^6\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
15	$2^{10}\cdot3\cdot5\cdot31\cdot13697\cdot15101$	1	$-2^6\cdot11$	$3^{10}\cdot5^4$	$2\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
17	$-\frac{2^{10}\cdot5^5\cdot7\cdot31\cdot65777}{17}$	$\frac{2^3\cdot11}{3}$	$\frac{2^{10}\cdot5\cdot11\cdot2663\cdot4093}{3\cdot17^2}$	$3^2\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
18	$-2^{11}\cdot5^2\cdot7^2\cdot31\cdot533063$	1	$\frac{2^7\cdot5\cdot11}{3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
19	$-\frac{2^{10}\cdot5^2\cdot11\cdot31\cdot14243891}{19}$	$\frac{2^3\cdot11}{3}$	$\frac{2^{19}\cdot5^2\cdot11\cdot101^2}{3\cdot19^2}$	$3^2\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
20	$2^7\cdot3\cdot5\cdot31\cdot59\cdot387077$	1	$-2^{10}\cdot5\cdot11$	$2^4\cdot3^6\cdot5^4$	$1\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
21	$-\frac{2^{10}\cdot5^2\cdot29\cdot31\cdot104789\cdot2583353}{7}$	1	$\frac{2^9\cdot11\cdot277^2}{3\cdot7^2}$	$3^{10}\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
23	$-\frac{2^{12}\cdot3^3\cdot5^3\cdot31\cdot32517200203}{23}$	1	$\frac{2^9\cdot3\cdot5\cdot11\cdot83\cdot139\cdot2357}{23^2}$	$3^6\cdot5^2\cdot23^4$	$1\cdot3^2+O(3^3)$	$1\cdot3^3+O(3^4)$
26	$-\frac{2^{11}\cdot5^2\cdot31\cdot699507967}{13}$	$\frac{2^3\cdot11}{3}$	$\frac{2^{21}\cdot5\cdot11\cdot17^2}{3}$	$2^4\cdot3^2\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$

Table III: form f of conductor 5 and weight 6.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
28	$-\frac{2^{13}\cdot5^3\cdot29\cdot31\cdot41\cdot113}{7}$	$\frac{2^3\cdot11}{3}$	$\frac{2^{13}\cdot5\cdot11\cdot277^2}{3\cdot7^2}$	$2^4\cdot3^2\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
29	$-\frac{2^{10}\cdot5^3\cdot19\cdot31\cdot37\cdot41633381443}{29}$	1	$\frac{2^{11}\cdot5^3\cdot11\cdot2221\cdot7039}{3\cdot29^2}$	$3^6\cdot5^2\cdot29^4$	$1+O(3)$	$1+O(3)$
30	$2^7\cdot3\cdot5\cdot13^2\cdot17$ $\cdot31\cdot53\cdot1051\cdot2713$	1	$-2^{10}\cdot5\cdot11$	$2^4\cdot3^{10}\cdot5^4$	$1\cdot3^1+O(3^2)$	$2\cdot3^1+O(3^2)$
31	$-2^{10}\cdot5^4\cdot1597\cdot25447$ $\cdot254627$	1	$\frac{2^{11}\cdot5^6\cdot11\cdot463^2}{3\cdot31^2}$	$3^6\cdot5^2\cdot31^4$	$2+O(3)$	$2+O(3)$
33	$-\frac{2^{10}\cdot5^2\cdot31\cdot79\cdot5727093605801}{11}$	1	$\frac{2^{10}\cdot5^2\cdot7^2\cdot37^2}{3\cdot11}$	$3^{10}\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
34	$-2^{12}\cdot5^2\cdot7\cdot23$ $\cdot31\cdot227\cdot130914857$	1	$\frac{2^{14}\cdot5^2\cdot11\cdot2663\cdot4093}{3\cdot17^2}$	$2^4\cdot3^6\cdot5^2\cdot17^4$	$2+O(3)$	$2+O(3)$
35	$2^{16}\cdot3\cdot5\cdot31\cdot46049$	$\frac{2^3\cdot11}{3}$	$-\frac{2^{12}\cdot11\cdot277^2}{7^2}$	$3^2\cdot5^4\cdot7^4$	$2\cdot3^1+O(3^2)$	$1\cdot3^1+O(3^2)$
37	$-\frac{2^{12}\cdot3^2\cdot5^3\cdot31\cdot181\cdot199\cdot9743}{37}$	$\frac{2^3\cdot11}{3}$	$\frac{2^{13}\cdot3^9\cdot11\cdot241^2}{37^2}$	$3^2\cdot5^2\cdot37^4$	$2\cdot3^{10}+O(3^{11})$	$2\cdot3^2+O(3^3)$
38	$-\frac{2^{14}\cdot5^4\cdot7\cdot31\cdot61\cdot27077\cdot185057}{19}$	1	$\frac{2^{23}\cdot5^3\cdot11\cdot101^2}{3\cdot19^2}$	$2^4\cdot3^6\cdot5^2\cdot19^4$	$1+O(3)$	$1+O(3)$
39	$-\frac{2^{13}\cdot5^2\cdot31\cdot2957\cdot86182236263}{13}$	1	$\frac{2^{17}\cdot11\cdot17^2}{3}$	$3^{10}\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
41	$-\frac{2^{10}\cdot5^2\cdot31\cdot3303519970879679}{41}$	1	$\frac{2^{13}\cdot5^2\cdot11\cdot443\cdot704101}{3\cdot41^2}$	$3^6\cdot5^2\cdot41^4$	$1+O(3)$	$1+O(3)$
42	$-2^{12}\cdot5^2\cdot7^2\cdot31\cdot267139$ $\cdot5797783$	1	$\frac{2^{13}\cdot5\cdot11\cdot277^2}{3\cdot7^2}$	$2^4\cdot3^{10}\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
43	$-\frac{2^{13}\cdot5^2\cdot19\cdot638839\cdot52230109}{31\cdot43}$	1	$\frac{2^9\cdot7^2\cdot11\cdot157^2\cdot389^2}{3\cdot43^2}$	$3^6\cdot5^2\cdot43^4$	$2+O(3)$	$2+O(3)$
	$n=2$					
2	$\frac{2^3\cdot31\cdot1759}{3^3}$	1	$\frac{2\cdot5^2\cdot7}{3^2}$	$2^4\cdot3^6\cdot5^2$	$1+O(3)$	$1+O(3)$
3	$\frac{2^8\cdot5\cdot31\cdot1223}{3^5}$	1	$\frac{2^3\cdot5}{3^2}$	$3^{10}\cdot5^2$	$2+O(3)$	$2+O(3)$
5	$-\frac{2^6\cdot13\cdot31\cdot37}{3^2\cdot5}$	1	0	$3^6\cdot5^4$	0	$2\cdot3^1+O(3^2)$
6	$\frac{2^3\cdot19\cdot31\cdot47\cdot5531}{3^5}$	1	$\frac{2\cdot5^2\cdot7}{3^2}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
7	$\frac{2^8\cdot31\cdot47\cdot53813}{3^3\cdot7^3}$	1	$\frac{2^9\cdot5^5}{3^2\cdot7^4}$	$3^6\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
10	$-\frac{2^3\cdot31^2}{5}$	$\frac{2^3\cdot5}{3^2}$	0	$2^4\cdot3^2\cdot5^4$	0	$1\cdot3^1+O(3^2)$
11	$\frac{2^9\cdot31\cdot28000571}{3^3\cdot11^3}$	1	$\frac{2^{12}\cdot5^3\cdot163}{3^2\cdot11^4}$	$3^6\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
12	$\frac{2^5\cdot7\cdot31\cdot145543}{3^5}$	1	$\frac{2\cdot5^2\cdot7}{3^2}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
13	$\frac{2^9\cdot31\cdot112051757}{3^3\cdot13^3}$	1	$\frac{2^{13}\cdot5^3}{3^2\cdot13^2}$	$3^6\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
14	$\frac{2^3\cdot5^2\cdot31\cdot65780839}{3^3\cdot7^3}$	1	$\frac{2^7\cdot5^6}{3^2\cdot7^3}$	$2^4\cdot3^6\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
15	$-\frac{2^9\cdot31\cdot103\cdot1559}{3^4\cdot5}$	1	0	$3^{10}\cdot5^4$	0	$2\cdot3^1+O(3^2)$
17	$\frac{2^8\cdot31\cdot491\cdot971}{3\cdot17^3}$	$\frac{2^3\cdot5}{3^2}$	$\frac{2^{10}\cdot5^2\cdot43\cdot881}{3^2\cdot17^4}$	$3^2\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
18	$\frac{2^5\cdot7\cdot31\cdot145543}{3^5}$	1	$\frac{2\cdot5^2\cdot7}{3^2}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
19	$\frac{2^7\cdot5\cdot31\cdot418273}{3\cdot19^3}$	$\frac{2^3\cdot5}{3^2}$	$\frac{2^{11}\cdot5^3\cdot7^2\cdot11^2}{3^2\cdot19^4}$	$3^2\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
20	$-\frac{2^3\cdot31\cdot96457}{3^2\cdot5}$	1	0	$2^4\cdot3^6\cdot5^4$	0	$1\cdot3^1+O(3^2)$
21	$\frac{2^9\cdot19^2\cdot31\cdot647\cdot11827}{3^5\cdot7^3}$	1	$\frac{2^9\cdot5^5}{3^2\cdot7^4}$	$3^{10}\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
23	$\frac{2^9\cdot31\cdot1117\cdot156733}{23^3}$	1	$\frac{2^9\cdot3^4\cdot5^2\cdot653}{23^4}$	$3^6\cdot5^2\cdot23^4$	$2\cdot3^6+O(3^7)$	$2\cdot3^3+O(3^4)$
26	$\frac{2^3\cdot7\cdot31\cdot6916561}{3\cdot13^3}$	$\frac{2^3\cdot5}{3^2}$	$\frac{2^{11}\cdot5^4\cdot7}{3^2\cdot13^2}$	$2^4\cdot3^2\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$
28	$\frac{2^3\cdot17\cdot31\cdot39383}{3\cdot7^3}$	$\frac{2^3\cdot5}{3^2}$	$\frac{2^7\cdot5^6}{3^2\cdot7^3}$	$2^4\cdot3^2\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
29	$\frac{2^8\cdot31\cdot283\cdot3323\cdot64067}{3^3\cdot29^3}$	1	$\frac{2^{10}\cdot5^3\cdot179\cdot1091}{3^2\cdot29^4}$	$3^6\cdot5^2\cdot29^4$	$1+O(3)$	$1+O(3)$

Table III: form f of conductor 5 and weight 6.						
m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
30	$-\frac{2^3\cdot7^2\cdot31\cdot3320281}{3^4\cdot5}$	1	0	$2^4\cdot3^{10}\cdot5^4$	0	$1\cdot3^1+O(3^2)$
31	$\frac{2^8\cdot5\cdot73\cdot219638621}{3^3\cdot31^2}$	1	$\frac{2^{11}\cdot5^5\cdot83^2}{3^2\cdot31^4}$	$3^6\cdot5^2\cdot31^4$	$2+O(3)$	$2+O(3)$
33	$\frac{2^8\cdot5^2\cdot31\cdot43\cdot109868293}{3^5\cdot11^3}$	1	$\frac{2^{12}\cdot5^3\cdot163}{3^2\cdot11^4}$	$3^{10}\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
34	$\frac{2^7\cdot31\cdot487\cdot122916679}{3^3\cdot17^3}$	1	$\frac{2^8\cdot5^3\cdot7\cdot43\cdot881}{3^2\cdot17^4}$	$2^4\cdot3^6\cdot5^2\cdot17^4$	$2+O(3)$	$2+O(3)$
35	$-\frac{2^{10}\cdot31\cdot10729}{5\cdot7^3}$	$\frac{2^3\cdot5}{3^2}$	0	$3^2\cdot5^4\cdot7^4$	0	$2\cdot3^1+O(3^2)$
37	$\frac{2^{10}\cdot3\cdot5\cdot13\cdot31\cdot59\cdot829}{37^3}$	$\frac{2^3\cdot5}{3^2}$	$\frac{2^{17}\cdot3^6\cdot5^3}{37^4}$	$3^2\cdot5^2\cdot37^4$	$2\cdot3^8+O(3^9)$	$2\cdot3^2+O(3^3)$
38	$\frac{2^3\cdot31\cdot1657646829583}{3^3\cdot19^3}$	1	$\frac{2^9\cdot5^4\cdot7^3\cdot11^2}{3^2\cdot19^4}$	$2^4\cdot3^6\cdot5^2\cdot19^4$	$1+O(3)$	$1+O(3)$
39	$\frac{2^9\cdot31\cdot1039\cdot3011\cdot62311}{3^5\cdot13^3}$	1	$\frac{2^{13}\cdot5^3}{3^2\cdot13^2}$	$3^{10}\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
41	$\frac{2^7\cdot31\cdot53\cdot709\cdot36628831}{3^3\cdot41^3}$	1	$\frac{2^{12}\cdot5^5\cdot11\cdot13\cdot107}{3^2\cdot41^4}$	$3^6\cdot5^2\cdot41^4$	$1+O(3)$	$1+O(3)$
42	$\frac{2^3\cdot31\cdot59\cdot59147190533}{3^5\cdot7^3}$	1	$\frac{2^7\cdot5^6}{3^2\cdot7^3}$	$2^4\cdot3^{10}\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
43	$\frac{2^9\cdot31\cdot482148655367}{3^3\cdot43^3}$	1	$\frac{2^9\cdot5^5\cdot7^2\cdot59^2}{3^2\cdot43^4}$	$3^6\cdot5^2\cdot43^4$	$2+O(3)$	$2+O(3)$
	$n=3$					
2	$-\frac{2\cdot31}{3^6\cdot5}$	1	$\frac{5^2\cdot7}{2\cdot3^3}$	$2^4\cdot3^6\cdot5^2$	$1+O(3)$	$1+O(3)$
3	$-\frac{2^8\cdot31}{3^{10}\cdot5}$	1	$\frac{2^3\cdot5}{3^3}$	$3^{10}\cdot5^2$	$2+O(3)$	$2+O(3)$
5	0	0	$\frac{2^6}{3^2\cdot5}$	$3^6\cdot5^4$	$1\cdot3^1+O(3^2)$	0
6	$-\frac{2\cdot31\cdot59^2}{3^{10}\cdot5}$	1	$\frac{5^2\cdot7}{2\cdot3^3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
7	$-\frac{2^6\cdot31\cdot47^2}{3^6\cdot5\cdot7^5}$	1	$\frac{2^9\cdot5^5}{3^3\cdot7^6}$	$3^6\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
10	0	0	$\frac{2^2\cdot7}{3^2}$	$2^4\cdot3^2\cdot5^4$	$2\cdot3^1+O(3^2)$	0
11	$-\frac{2^6\cdot31\cdot181^2}{3^6\cdot5\cdot11^5}$	1	$\frac{2^{12}\cdot5^3\cdot163}{3^3\cdot11^6}$	$3^6\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
12	$-\frac{2^7\cdot31}{3^{10}\cdot5}$	1	$\frac{5^2\cdot7}{2\cdot3^3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
13	$-\frac{2^6\cdot31^3}{3^6\cdot5\cdot13^3}$	1	$\frac{2^{13}\cdot5^3}{3^3\cdot13^4}$	$3^6\cdot5^2\cdot13^4$	$2+O(3)$	$2+O(3)$
14	$-\frac{2\cdot5^3\cdot19^2\cdot31}{3^6\cdot7^5}$	1	$\frac{2^5\cdot5^6}{3^3\cdot7^5}$	$2^4\cdot3^6\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
15	0	0	$\frac{2^6}{3^2\cdot5}$	$3^{10}\cdot5^4$	$1\cdot3^1+O(3^2)$	0
17	$-\frac{2^6\cdot5\cdot31}{3^2\cdot17^5}$	$\frac{2^3\cdot5}{3^3}$	$\frac{2^{10}\cdot5^2\cdot43\cdot881}{3^3\cdot17^6}$	$3^2\cdot5^2\cdot17^4$	$1+O(3)$	$1+O(3)$
18	$-\frac{2^7\cdot31}{3^{10}\cdot5}$	1	$\frac{5^2\cdot7}{2\cdot3^3}$	$2^4\cdot3^{10}\cdot5^2$	$1+O(3)$	$1+O(3)$
19	$-\frac{2^{10}\cdot31}{3^2\cdot5\cdot19^5}$	$\frac{2^3\cdot5}{3^3}$	$\frac{2^{11}\cdot5^3\cdot7^2\cdot11^2}{3^3\cdot19^6}$	$3^2\cdot5^2\cdot19^4$	$2+O(3)$	$2+O(3)$
20	0	0	$\frac{2^2\cdot7}{3^2}$	$2^4\cdot3^6\cdot5^4$	$2\cdot3^1+O(3^2)$	0
21	$-\frac{2^8\cdot31\cdot191^2}{3^{10}\cdot5\cdot7^5}$	1	$\frac{2^9\cdot5^5}{3^3\cdot7^6}$	$3^{10}\cdot5^2\cdot7^4$	$2+O(3)$	$2+O(3)$
23	$-\frac{2^6\cdot5\cdot31}{3^4\cdot23^5}$	1	$\frac{2^9\cdot3^3\cdot5^2\cdot653}{23^6}$	$3^6\cdot5^2\cdot23^4$	$2\cdot3^6+O(3^7)$	$1\cdot3^2+O(3^3)$
26	$-\frac{2\cdot31\cdot47^2}{3^2\cdot5\cdot13^5}$	$\frac{2^3\cdot5}{3^3}$	$\frac{2^9\cdot5^4\cdot7}{3^3\cdot13^4}$	$2^4\cdot3^2\cdot5^2\cdot13^4$	$1+O(3)$	$1+O(3)$
28	$-\frac{2\cdot19^2\cdot31}{3^2\cdot5\cdot7^5}$	$\frac{2^3\cdot5}{3^3}$	$\frac{2^5\cdot5^6}{3^3\cdot7^5}$	$2^4\cdot3^2\cdot5^2\cdot7^4$	$1+O(3)$	$1+O(3)$
29	$-\frac{2^8\cdot31\cdot757^2}{3^6\cdot5\cdot29^5}$	1	$\frac{2^{10}\cdot5^3\cdot179\cdot1091}{3^3\cdot29^6}$	$3^6\cdot5^2\cdot29^4$	$1+O(3)$	$1+O(3)$
30	0	0	$\frac{2^2\cdot7}{3^2}$	$2^4\cdot3^{10}\cdot5^4$	$2\cdot3^1+O(3^2)$	0
31	$-\frac{2^8\cdot967^2}{3^6\cdot5\cdot31^4}$	1	$\frac{2^{11}\cdot5^5\cdot83^2}{3^3\cdot31^6}$	$3^6\cdot5^2\cdot31^4$	$2+O(3)$	$2+O(3)$
33	$-\frac{2^6\cdot5\cdot31^3}{3^{10}\cdot11^5}$	1	$\frac{2^{12}\cdot5^3\cdot163}{3^3\cdot11^6}$	$3^{10}\cdot5^2\cdot11^4$	$1+O(3)$	$1+O(3)$
34	$-\frac{2^3\cdot31\cdot1867^2}{3^6\cdot5\cdot17^5}$	1	$\frac{2^6\cdot5^3\cdot7\cdot43\cdot881}{3^3\cdot17^6}$	$2^4\cdot3^6\cdot5^2\cdot17^4$	$2+O(3)$	$2+O(3)$
35	0	0	$\frac{2^{12}\cdot5^3}{3^2\cdot7^6}$	$3^2\cdot5^4\cdot7^4$	$1\cdot3^1+O(3^2)$	0
37	$-\frac{2^6\cdot7^4\cdot31}{5\cdot37^5}$	$\frac{2^3\cdot5}{3^3}$	$\frac{2^{17}\cdot3^5\cdot5^3}{37^6}$	$3^2\cdot5^2\cdot37^4$	$2\cdot3^8+O(3^9)$	$2\cdot3^2+O(3^3)$

Table III: form f of conductor 5 and weight 6.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
38	$-\frac{2\cdot 31^3\cdot 149^2}{3^6\cdot 5\cdot 19^5}$	1	$\frac{2^7\cdot 5^4\cdot 7^3\cdot 11^2}{3^3\cdot 19^6}$	$2^4\cdot 3^6\cdot 5^2\cdot 19^4$	$1+O(3)$	$1+O(3)$
39	$-\frac{2^6\cdot 7^2\cdot 31\cdot 97^2}{3^{10}\cdot 5\cdot 13^5}$	1	$\frac{2^{13}\cdot 5^3}{3^3\cdot 13^4}$	$3^{10}\cdot 5^2\cdot 13^4$	$2+O(3)$	$2+O(3)$
41	$-\frac{2^6\cdot 31\cdot 881^2}{3^6\cdot 5\cdot 41^5}$	1	$\frac{2^{12}\cdot 5^5\cdot 11\cdot 13\cdot 107}{3^3\cdot 41^6}$	$3^6\cdot 5^2\cdot 41^4$	$1+O(3)$	$1+O(3)$
42	$-\frac{2\cdot 31\cdot 4919^2}{3^{10}\cdot 5\cdot 7^5}$	1	$\frac{2^5\cdot 5^6}{3^3\cdot 7^5}$	$2^4\cdot 3^{10}\cdot 5^2\cdot 7^4$	$1+O(3)$	$1+O(3)$
43	$-\frac{2^6\cdot 5\cdot 13^2\cdot 31\cdot 67^2}{3^6\cdot 43^5}$	1	$\frac{2^9\cdot 5^5\cdot 7^2\cdot 59^2}{3^3\cdot 43^6}$	$3^6\cdot 5^2\cdot 43^4$	$2+O(3)$	$2+O(3)$

Table IV: form f of conductor 121 and weight 4.
$L_3^*(f,\sigma,s)(n=1,2)=(176,0)$ and $L_3(f,\sigma,2)=L_3(f,\rho,2)=0$.

m	$L_3^*(f,\rho,n)$	$\mathcal{P}_3(f,\rho,n)$	$\mathcal{P}_3(f,\sigma,n)$	$N(f,\rho)$	$\mathcal{L}_3(f,\sigma,n)$	$\mathcal{L}_3(f,\rho,n)$
2	$2^5\cdot 3\cdot 11\cdot 17\cdot 37$	1	$2^2\cdot 3$	$2^4\cdot 3^6\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$1\cdot 3^1+O(3^2)$
3	$2^5\cdot 5\cdot 11\cdot 4373$	1	$\frac{2^2}{3}$	$3^{10}\cdot 11^4$	$2+O(3)$	$2+O(3)$
5	$2^5\cdot 3^3\cdot 11\cdot 2069$	1	$\frac{2^8\cdot 3}{5^2}$	$3^6\cdot 5^4\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$1\cdot 3^3+O(3^4)$
6	$2^3\cdot 3^2\cdot 11\cdot 83\cdot 2297$	1	$2^2\cdot 3$	$2^4\cdot 3^{10}\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$2\cdot 3^2+O(3^3)$
7	$\frac{2^5\cdot 5\cdot 11\cdot 349\cdot 863}{7}$	1	$\frac{2^8}{3}$	$3^6\cdot 7^4\cdot 11^4$	$2+O(3)$	$2+O(3)$
10	$2^3\cdot 3^2\cdot 11\cdot 13\cdot 211$	$\frac{2^2}{3}$	$\frac{2^8\cdot 3^3}{5^2}$	$2^4\cdot 3^2\cdot 5^4\cdot 11^4$	$2\cdot 3^4+O(3^5)$	$1\cdot 3^2+O(3^3)$
11	$2^8\cdot 11^2$	1	$\frac{2^2}{3}$	$3^6\cdot 11^4$	$2+O(3)$	$2+O(3)$
12	$2^5\cdot 3\cdot 11\cdot 13\cdot 31\cdot 367$	1	$2^2\cdot 3$	$2^4\cdot 3^{10}\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$2\cdot 3^1+O(3^2)$
13	$\frac{2^5\cdot 5\cdot 11\cdot 29\cdot 230281}{13}$	1	$\frac{2^4\cdot 7^2}{3}$	$3^6\cdot 11^4\cdot 13^4$	$2+O(3)$	$2+O(3)$
14	$\frac{2^7\cdot 3\cdot 5\cdot 11\cdot 439\cdot 1129}{7}$	1	$2^8\cdot 3$	$2^4\cdot 3^6\cdot 7^4\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$1\cdot 3^1+O(3^2)$
15	$\frac{2^5\cdot 3\cdot 11\cdot 23234851}{5}$	$\frac{2^8\cdot 3}{5^2}$	1	$3^{10}\cdot 5^4\cdot 11^4$	$2\cdot 3^2+O(3^3)$	$1\cdot 3^1+O(3^2)$
17	$\frac{2^5\cdot 3\cdot 11\cdot 29\cdot 8219}{17}$	$\frac{2^2}{3}$	$2^4\cdot 3^3$	$3^2\cdot 11^4\cdot 17^4$	$2\cdot 3^4+O(3^5)$	$2\cdot 3^1+O(3^2)$
19	$\frac{2^5\cdot 5^3\cdot 11^2\cdot 29\cdot 31}{19}$	$\frac{2^2}{3}$	$\frac{2^6\cdot 5^2}{3}$	$3^2\cdot 11^4\cdot 19^4$	$2+O(3)$	$2+O(3)$
20	$\frac{2^7\cdot 3^2\cdot 11\cdot 156241}{5}$	1	$\frac{2^8\cdot 3^3}{5^2}$	$2^4\cdot 3^6\cdot 5^4\cdot 11^4$	$2\cdot 3^4+O(3^5)$	$1\cdot 3^2+O(3^3)$

In the remaining four tables, we give some intriguing integrality and squareness assertions for the L-values computed in the previous four tables. Although we do not enter into any detailed discussion in the present paper, it seems highly likely that these phenomena can be explained via the Bloch-Kato conjecture, and Flach's motivic generalization of the Cassels-Tate pairing. We define M to be the product of the distinct primes dividing m, but excluding the prime 3. Let N denote the conductor of our primitive form f. For $n = 1, \ldots, k-1$, we define

$$A_n(f) = |L_3^*(f,\rho,n)|M^n\epsilon_3(\rho)^{(n-1)}/4 \tag{81}$$

In Table V, for the form f of conductor 5 and weight 4, define

$$B_1(f) = A_1(f)/(2^2 \times 5^3 \times 13),$$
$$B_2(f) = A_2(f)/(5^2 \times 13).$$

Table V: form f of conductor 5 and weight 4.		
m	$B_1(f)$	$\sqrt{B_2(f)}$
2	$2^2 \cdot 7$	2
3	$5 \cdot 41$	1
6	$2^2 \cdot 1801$	$2 \cdot 7$
7	$2^4 \cdot 23 \cdot 41$	2^2
11	$2^6 \cdot 2311$	$2^2 \cdot 11$
12	$2^3 \cdot 839$	2^3
13	$11 \cdot 13 \cdot 43 \cdot 53$	1
14	$2^2 \cdot 5 \cdot 7 \cdot 13 \cdot 251$	$2 \cdot 5$
17	$31 \cdot 167$	5
19	$5 \cdot 43^2$	13
21	3425341	67
22	$2^3 \cdot 43 \cdot 13841$	$2^2 \cdot 11$
23	$2^4 \cdot 3^5 \cdot 1409$	$2^2 \cdot 3 \cdot 5$
26	$2^2 \cdot 13 \cdot 887$	$2 \cdot 13$
28	$2^2 \cdot 503$	2
29	$11 \cdot 1678031$	109
31	$5 \cdot 79 \cdot 62351$	151
33	$5 \cdot 11^2 \cdot 19 \cdot 2879$	$5 \cdot 7$
34	$2^4 \cdot 17 \cdot 142427$	2^2
37	$2^4 \cdot 3^2 \cdot 5 \cdot 367$	$2^2 \cdot 3$
39	$2^6 \cdot 71 \cdot 17489$	$2^2 \cdot 11$
41	$17 \cdot 31 \cdot 211 \cdot 941$	11
42	$2^2 \cdot 19 \cdot 859 \cdot 1801$	$2 \cdot 149$
43	$2^2 \cdot 7 \cdot 19 \cdot 251 \cdot 491$	$2 \cdot 5^2$
44	$2^2 \cdot 11 \cdot 421$	$2 \cdot 7$
46	$2^2 \cdot 3^3 \cdot 7283$	$2 \cdot 3$
47	$2^4 \cdot 23^2 \cdot 22567$	$2^4 \cdot 13$
51	$2^4 \cdot 5 \cdot 13 \cdot 278591$	$2^2 \cdot 101$
52	$2^2 \cdot 2513617$	2
53	$5 \cdot 290161$	29
57	$2^2 \cdot 61 \cdot 503 \cdot 4241$	$2 \cdot 7^2$
58	$2^6 \cdot 9208039$	$2^6 \cdot 5$
59	$2^4 \cdot 5 \cdot 23 \cdot 397 \cdot 853$	$2^2 \cdot 11^2$
62	$2^2 \cdot 307 \cdot 2879$	$2 \cdot 5 \cdot 7$
66	$2^5 \cdot 5 \cdot 6458773$	$2^3 \cdot 11$
67	$2^2 \cdot 13^2 \cdot 19^2 \cdot 4759$	$2 \cdot 103$

Table V: form f of conductor 5 and weight 4.		
m	$B_1(f)$	$\sqrt{B_2(f)}$
68	$2^2 \cdot 10484557$	$2 \cdot 7 \cdot 13$
69	$2^2 \cdot 3^3 \cdot 857 \cdot 15733$	$2 \cdot 3 \cdot 31$
71	$2^2 \cdot 7 \cdot 31 \cdot 79 \cdot 101$	$2 \cdot 29$
73	$5 \cdot 17 \cdot 47 \cdot 1831$	43
74	$2^2 \cdot 3^4 \cdot 11 \cdot 523 \cdot 1031$	$2 \cdot 3^3$
76	$2^5 \cdot 311 \cdot 7297$	$2^3 \cdot 13$
77	$2^4 \cdot 7 \cdot 11 \cdot 2377 \cdot 60913$	$2^2 \cdot 101$
82	$2^3 \cdot 5 \cdot 73 \cdot 4817$	$2^2 \cdot 11$
83	?	$2^2 \cdot 3^2 \cdot 7$
84	$2^2 \cdot 7 \cdot 431 \cdot 10259$	$2 \cdot 199$
89	$2^2 \cdot 3^5 \cdot 71 \cdot 293$	$2 \cdot 3^2 \cdot 5$
91	$5 \cdot 4519393$	1
92	$2^2 \cdot 3^2 \cdot 157 \cdot 31019$	$2 \cdot 3^2$
93	?	$2 \cdot 31^3 \cdot 179$
94	?	$2 \cdot 11 \cdot 19$
97	?	$2^2 \cdot 3^2 \cdot 5^2$

In Table VI, for the form f of conductor 7 and weight 4, define

$$B_1(f) = A_1(f)/(7^3 \times 5).$$

Table VI: form f of conductor 7 and weight 4.	
m	$B_1(f)$
2	$2^2 \cdot 3 \cdot 7$
3	$3 \cdot 13^2$
5	$2^5 \cdot 3 \cdot 71$
6	$2^3 \cdot 3 \cdot 7 \cdot 113$
7	$3 \cdot 223$
10	$2^2 \cdot 239$
11	$3 \cdot 211 \cdot 499$
12	$2^2 \cdot 3 \cdot 7 \cdot 241$
13	$2^6 \cdot 3 \cdot 5 \cdot 773$
14	$2^2 \cdot 3 \cdot 41 \cdot 59$
15	$3 \cdot 5 \cdot 13 \cdot 43 \cdot 179$
17	$3^2 \cdot 1223$
19	$2^9 \cdot 37$
20	$2^6 \cdot 3 \cdot 1213$
21	$2^{11} \cdot 3 \cdot 29$
22	$2^3 \cdot 3 \cdot 19 \cdot 28277$

Table VI: form f of conductor 7 and weight 4.	
m	$B_1(f)$
23	$3^3 \cdot 47 \cdot 10463$
26	$2^2 \cdot 7 \cdot 3917$
28	$2^3 \cdot 13$
29	$3 \cdot 7 \cdot 1904647$
30	$2^2 \cdot 3 \cdot 19 \cdot 266839$
31	$2^6 \cdot 3 \cdot 307267$
33	$2^5 \cdot 3 \cdot 849221$
34	$2^9 \cdot 3^3 \cdot 83 \cdot 101$
35	$43 \cdot 191$
37	$2^5 \cdot 15937$
38	$2^4 \cdot 3 \cdot 5 \cdot 864947$
39	$2^6 \cdot 3 \cdot 957811$
41	$2^2 \cdot 3 \cdot 5 \cdot 13 \cdot 173 \cdot 1693$
42	$2^2 \cdot 3 \cdot 37 \cdot 15601$

In Table VII, for the form f of conductor 5 and weight 6, define

$$B_1(f) = A_1(f)/(2^6 \cdot 31 \cdot 5^2),$$
$$B_2(f) = A_2(f)/(2^4 \cdot 31),$$
$$B_3(f) = A_3(f) \times 5/(2^4 \cdot 31).$$

Table VII: form f of conductor 5 and weight 6.			
m	$B_1(f)$	$B_2(f)$	$\sqrt{B_3(f)}$
2	$2^5 \cdot 661$	1759	1
3	$2^2 \cdot 5 \cdot 13 \cdot 2953$	$2^2 \cdot 5 \cdot 1223$	2
5	$3 \cdot 193 \cdot 211$	$3 \cdot 5^2 \cdot 13 \cdot 37$	0
6	$2^4 \cdot 5 \cdot 137 \cdot 39323$	$19 \cdot 47 \cdot 5531$	59
7	$2^4 \cdot 7 \cdot 14230919$	$2^2 \cdot 47 \cdot 53813$	47
10	$3 \cdot 1097$	$3 \cdot 5^2 \cdot 31$	0
11	$2^5 \cdot 5 \cdot 971 \cdot 592759$	$2^3 \cdot 28000571$	181
12	$2^4 \cdot 7^2 \cdot 533063$	$2^2 \cdot 7 \cdot 145543$	2^3
13	$2^2 \cdot 7 \cdot 11 \cdot 211 \cdot 6591061$	$2^3 \cdot 112051757$	$13 \cdot 31$
14	$2^5 \cdot 5^2 \cdot 1082124649$	$5^2 \cdot 65780839$	$5^2 \cdot 19$
15	$2^2 \cdot 3 \cdot 13697 \cdot 15101$	$2^3 \cdot 3 \cdot 5^2 \cdot 103 \cdot 1559$	0
17	$2^2 \cdot 5^3 \cdot 7 \cdot 65777$	$2^2 \cdot 491 \cdot 971$	5
18	$2^4 \cdot 7^2 \cdot 533063$	$2^2 \cdot 7 \cdot 145543$	2^3
19	$2^2 \cdot 11 \cdot 14243891$	$2 \cdot 5 \cdot 418273$	2^2
20	$3 \cdot 59 \cdot 387077$	$3 \cdot 5^2 \cdot 96457$	0
21	$2^2 \cdot 29 \cdot 104789 \cdot 2583353$	$2^3 \cdot 19^2 \cdot 647 \cdot 11827$	$2 \cdot 191$

Table VII: form f of conductor 5 and weight 6.			
m	$B_1(f)$	$B_2(f)$	$\sqrt{B_3(f)}$
23	$2^4 \cdot 3^3 \cdot 5 \cdot 32517200203$	$2^3 \cdot 3^3 \cdot 1117 \cdot 156733$	$3 \cdot 5$
26	$2^4 \cdot 699507967$	$7 \cdot 6916561$	47
28	$2^6 \cdot 5 \cdot 29 \cdot 41 \cdot 113$	$17 \cdot 39383$	19
29	$2^2 \cdot 5 \cdot 19 \cdot 37 \cdot 41633381443$	$2^2 \cdot 283 \cdot 3323 \cdot 64067$	$2 \cdot 757$
30	$3 \cdot 13^2 \cdot 17 \cdot 53 \cdot 1051 \cdot 2713$	$3 \cdot 5^2 \cdot 7^2 \cdot 3320281$	0
31	$2^2 \cdot 5^2 \cdot 1597 \cdot 25447 \cdot 254627$	$2^2 \cdot 5 \cdot 73 \cdot 219638621$	$2 \cdot 967$
33	$2^2 \cdot 79 \cdot 5727093605801$	$2^2 \cdot 5^2 \cdot 43 \cdot 109868293$	$5 \cdot 31$
34	$2^5 \cdot 7 \cdot 17 \cdot 23 \cdot 227 \cdot 130914857$	$2^4 \cdot 487 \cdot 122916679$	$2 \cdot 1867$
35	$2^8 \cdot 3 \cdot 7 \cdot 46049$	$2^4 \cdot 3 \cdot 5^2 \cdot 10729$	0
37	$2^4 \cdot 3^2 \cdot 5 \cdot 181 \cdot 199 \cdot 9743$	$2^4 \cdot 3^2 \cdot 5 \cdot 13 \cdot 59 \cdot 829$	$3 \cdot 7^2$
38	$2^7 \cdot 5^2 \cdot 7 \cdot 61 \cdot 27077 \cdot 185057$	1657646829583	$31 \cdot 149$
39	$2^5 \cdot 2957 \cdot 86182236263$	$2^3 \cdot 1039 \cdot 3011 \cdot 62311$	$7 \cdot 97$
41	$2^2 \cdot 3303519970879679$	$2 \cdot 53 \cdot 709 \cdot 36628831$	881
42	$2^5 \cdot 7^3 \cdot 267139 \cdot 5797783$	$59 \cdot 59147190533$	4919
43	$2^5 \cdot 19 \cdot 638839 \cdot 52230109$	$2^3 \cdot 482148655367$	$5 \cdot 13 \cdot 67$

In Table VIII, for the CM form f of conductor 121 and weight 4, define

$$B_1(f) = A_1(f)/(2^2 \cdot 11).$$

Table VIII: form f of conductor 121 and weight 4.	
m	$B_1(f)$
2	$2^2 \cdot 3 \cdot 17 \cdot 37$
3	$2 \cdot 5 \cdot 4373$
5	$2 \cdot 3^3 \cdot 5 \cdot 2069$
6	$3^2 \cdot 83 \cdot 2297$
7	$2 \cdot 5 \cdot 349 \cdot 863$
10	$3^2 \cdot 5 \cdot 13 \cdot 211$
11	$2^4 \cdot 11^2$
12	$2^2 \cdot 3 \cdot 13 \cdot 31 \cdot 367$
14	$2^4 \cdot 3 \cdot 5 \cdot 439 \cdot 1129$
17	$2 \cdot 3 \cdot 29 \cdot 8219$
19	$2 \cdot 5^3 \cdot 11 \cdot 29 \cdot 31$
20	$2^4 \cdot 3^2 \cdot 156241$

References

[1] T. Bouganis, V. Dokchitser, *Algebraicity of L-values for elliptic curves in a false Tate curve tower*, Proc. Camb. Phil. Soc. **142** (2007), 193–204.

[2] J. Coates, T. Fukaya, K. Kato, R. Sujatha, O. Venjakob, *The GL_2 main conjecture for elliptic curves without complex multiplication*, Publ. Math. Inst. Hautes Études Sci. **101** (2005).

[3] D. Delbourgo, T. Ward, *The growth of CM periods over false Tate extensions*, Experiment. Math. **19** (2010), 195-210.

[4] T. Dokchitser, *Computing Special Values of Motivic L-functions*, Experiment. Math. **13** (2004), 137-150.

[5] T. Dokchitser, V. Dokchitser, *Computations in non-commutative Iwasawa theory*, Proc. London Math. Soc. **94** (2006), 211–272.

[6] V. Dokchitser, *Root numbers of non-abelian twists of elliptic curves (with appendix by Tom Fisher)*, Proc. London Math. Soc. **3** 91 (2005), 300–324.

[7] N. Dummigan, W. A. Stein, and M. Watkins, *Constructing Elements in Shafarevich-Tate Groups of Modular Motives*, Number theory and algebraic geometry, ed. by Miles Reid and Alexei Skorobogatov **303** (2003), 91–118.

[8] T. Fukaya, K.Kato, *A formulation of conjectures on p-adic zeta functions in noncommutative Iwasawa theory*, Proceedings of the St. Petersburg Mathematical Society. Vol. XII, 185, Amer. Math. Soc. Transl. Ser. 2, **219**,Amer. Math. Soc., Providence, RI, 2006.

[9] R. Greenberg, *Iwasawa theory for p-adic representations*, Algebraic Number Theory, Advanced Studies in Pure Mathematics, Academic Press, **17** (1989), 97-137.

[10] Y. Hachimori, K. Matsuno, *An analogue of Kida's formula for the Selmer groups of elliptic curves*, J. Algebraic Geom. **8** (1999), 581–601.

[11] B. Hart et.al. *FLINT: Fast library for number theory*, `http://www.flintlib.org`

[12] M. Kakde, *The main conjecture of Iwasawa theory for totally real fields*, arXiv:1008.0142.

[13] K. Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, in Cohomologies p-adiques et applications arithmétiques III, Astérisque **295** (2004), 117–290.

[14] K. Kato, *K_1 of some non-commutative completed group rings*, K-theory **34** (2005), 99–140.

[15] D. Kim, *p-adic L-functions over false Tate extensions*, preprint.

[16] Y. Manin, Values of p-adic Hecke series at lattice points of the critical strip, (Russian) Mat. Sb. (N.S.) **93** (135) (1974), 621626, 631.

[17] B. Mazur, A. Wiles, *On p-adic analytic families of Galois representations*, Compositio Math. **59** (1986), 231–264.

[18] T. Miyake, *Modular Forms*, Springer (1989).

[19] PARI, A computer algebra system designed for fast computations in number theory, `http://pari.math.u-bordeaux.fr/`.

[20] R. Pollack, T. Weston, *Kida's formula and congruences*, Documenta Mathematica, Special Volume (2006), 615-630.

[21] J. Ritter, A. Weiss, *On the 'main conjecture' of equivariant Iwasawa theory*, J. Amer. Math. Soc. **24**(2011), 1015-1050.

[22] D. Rohrlich, *L-functions and division towers*, Math. Ann. **281** (1988), 611–632.

[SAGE] W. A. Stein et al., *Sage Mathematics Software (Version 4.7.2)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[23] G. Shimura, *On the special values of the zeta functions associated with cusp forms*, Comm. Pure Appl. Math. **29** (1976), 783–804.

[24] G. Shimura, *On the periods of modular forms*, Math. Ann. **229** (1977), 211–221.

[25] G. Shimura, *The special values of zeta functions associated with Hilbert modular forms*, Duke Math. Jour. **45** (1978), 637–679.

[26] J. Tate, *Number theoretic background*, in: Automorphic forms, representations and L-functions, Part 2 (ed. A. Borel and W. Casselman), Proc. Symp. in Pure Math. Math. Ann. **281** (1988), 611–632.

J. Coates, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, England

T. Dokchitser, Department of Mathematics, University Walk, Bristol BS8 1TW, United Kingdom

Z. Liang, School of Mathematical Sciences, Capital Normal University, Xisanhuanbeilu 105, Haidan District, Beijing, China

W. Stein, Department of Mathematics, University of Washington, Seattle, Box 354350 WA 98195, USA

R. Sujatha, Mathematics Department, 1984, Mathematics Road, University of British Columbia, Vancouver BC, Canada V6T 1Z2

E-mail address: J.H.Coates@dpmms.cam.ac.uk, tim.dokchitser@bristol.ac.uk, liangzhb@gmail.com, wstein@uw.edu, sujatha@math.ubc.ca