Lectures on Modular Forms and Hecke Operators

Kenneth A. Ribet William A. Stein

December 2, 2011

Contents

Preface

This book began when the second author typed notes for the first author's 1996 Berkeley course on modular forms with a view toward explaining some of the key ideas in Wiles's celebrated proof of Fermat's Last Theorem. The second author then expanded and rewrote the notes while teaching a course at Harvard in 2003 on modular abelian varieties.

The intended audience of this book is advanced graduate students and mathematical researchers. This book is more advanced than [LR11, Ste07b, DS05], and at a relatively similar level to [DI95], though with more details. It should be substantially more accessible than a typical research paper in the area.

Some things about how this book is (or will be!) fully "modern" in that it takes into account:

- The full modularity theorem.
- The proof of the full Serre conjecture
- Computational techniques (algorithms, Sage)

Notation

$\cong$ an isomorphism
$\approx$ a noncanonical isomorphism

Acknowledgement

Joe Wetherell wrote the first version of Section 2.5.

Contact

Kenneth A. Ribet (ribet@math.berkeley.edu)
William A. Stein (wstein@gmail.com)

1
The Main Objects

1.1 Torsion points on elliptic curves

The main geometric objects that we will study in this book are elliptic curves, which are curves of genus one equipped with a distinguished point. More generally, we consider certain algebraic curves of larger genus called modular curves, which in turn give rise via the Jacobian construction to higher-dimensional abelian varieties from which we will obtain representations of the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of the rational numbers.

It is convenient to view the group of complex points $E(\mathbf{C})$ on an elliptic curve E over the complex numbers $\mathbf{C}$ as a quotient $\mathbf{C}/L$. Here

$$L = \left\{ \int_\gamma \omega \ : \ \gamma \in H_1(E(\mathbf{C}), \mathbf{Z}) \right\}$$

is a lattice attached to a nonzero holomorphic differential ω on E, and the homology $H_1(E(\mathbf{C}), \mathbf{Z}) \approx \mathbf{Z} \times \mathbf{Z}$ is the abelian group of smooth closed paths on $E(\mathbf{C})$ modulo the homology relations.

Viewing E as $\mathbf{C}/L$ immediately gives us information about the structure of the group of torsion points on E, which we exploit in the next section to construct two-dimensional representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

1.1.1 The Tate module

In the 1940s, Andre Weil studied the analogous situation for elliptic curves defined over a finite field k. He desperately wanted to find an algebraic way to describe the above relationship between elliptic curves and lattices. He found an algebraic definition of L/nL, when n is prime to the characteristic of k.

Let

$$E[n] := \{P \in E(\overline{k}) : nP = 0\}.$$

When E is defined over $\mathbf{C}$,

$$E[n] = \left(\frac{1}{n}L\right)/L \cong L/nL \approx (\mathbf{Z}/n\mathbf{Z}) \times (\mathbf{Z}/n\mathbf{Z}),$$

so $E[n]$ is a purely algebraic object canonically isomorphic to L/nL.

Now suppose E is defined over an arbitrary field k. For any prime ℓ, let

$$\begin{aligned} E[\ell^\infty] &:= \{P \in E(\overline{k}) : \ell^\nu P = 0, \text{ some } \nu \geq 1\} \\ &= \bigcup_{\nu=1}^{\infty} E[\ell^\nu] = \varinjlim E[\ell^\nu]. \end{aligned}$$

In an analogous way, Tate constructed a rank 2 free $\mathbf{Z}_\ell$-module

$$T_\ell(E) := \varprojlim E[\ell^\nu],$$

where the map $E[\ell^\nu] \to E[\ell^{\nu-1}]$ is multiplication by ℓ. The $\mathbf{Z}/\ell^\nu\mathbf{Z}$-module structure of $E[\ell^\nu]$ is compatible with the maps $E[\ell^\nu] \xrightarrow{\ell} E[\ell^{\nu-1}]$ (see, e.g., [Sil92, III.7]). If ℓ is coprime to the characteristic of the base field k, then $T_\ell(E)$ is free of rank 2 over $\mathbf{Z}_\ell$, and

$$V_\ell(E) := T_\ell(E) \otimes \mathbf{Q}_\ell$$

is a 2-dimensional vector space over $\mathbf{Q}_\ell$.

1.2 Galois representations

Number theory is largely concerned with the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, which is often studied by considering continuous linear representations

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(K)$$

where K is a field and n is a positive integer, usually 2 in this book. Artin, Shimura, Taniyama, and Tate pioneered the study of such representations.

Let E be an elliptic curve defined over the rational numbers $\mathbf{Q}$. Then $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on the set $E[n]$, and this action respects the group operations, so we obtain a representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(E[n]) \approx \mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z}).$$

Let K be the field cut out by the $\ker(\rho)$, i.e., the fixed field of $\ker(\rho)$. Then K is a finite Galois extension of $\mathbf{Q}$. Since

$$\mathrm{Gal}(K/\mathbf{Q}) \cong \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})/\ker\rho \cong \mathrm{Im}\,\rho \hookrightarrow \mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$$

we obtain, in this way, subgroups of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ as Galois groups.

Shimura showed that if we start with the elliptic curve E defined by the equation $y^2 + y = x^3 - x^2$ then for "most" n the image of ρ is all of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$. More generally, the image is "most" of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ when E does not have complex multiplication. (We say E has *complex multiplication* if its endomorphism ring over $\mathbf{C}$ is strictly larger than $\mathbf{Z}$.)

1.3 Modular forms

Many spectacular theorems and deep conjectures link Galois representations with modular forms. Modular forms are extremely symmetric analytic objects, which we will first view as holomorphic functions on the complex upper half plane that behave well with respect to certain groups of transformations.

Let $\mathrm{SL}_2(\mathbf{Z})$ be the group of 2×2 integer matrices with determinant 1. For any positive integer N, consider the subgroup

$$\Gamma_1(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \,:\, a \equiv d \equiv 1,\ c \equiv 0 \pmod{N} \right\}$$

of matrices in $\mathrm{SL}_2(\mathbf{Z})$ that are of the form $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$ when reduced modulo N.

The space $S_k(N)$ of *cusp forms* of weight k and level N for $\Gamma_1(N)$ consists of all holomorphic functions $f(z)$ on the complex upper half plane

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

that vanish at the cusps (see below) and satisfy the equation

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z) \text{ for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N) \text{ and } z \in \mathfrak{h}.$$

Thus $f(z+1) = f(z)$, so f determines a function F of $q(z) = e^{2\pi i z}$ such that $F(q) = f(z)$. Viewing F as a function on $\{z : 0 < |z| < 1\}$, the condition that $f(z)$ is holomorphic and vanishes at infinity is that $F(z)$ extends to a holomorphic function on $\{z : |z| < 1\}$ and $F(0) = 0$. In this case, f is determined by its *Fourier expansion*

$$f(q) = \sum_{n=1}^{\infty} a_n q^n.$$

It is also useful to consider the space $M_k(N)$ of *modular forms* of level N, which is defined in the same way as $S_k(N)$, except that the condition that $F(0) = 0$ is relaxed, and we require only that F extends to a holomorphic function at 0 (and there is a similar condition at the cusps other than ∞).

The spaces $M_k(N)$ and $S_k(N)$ are finite dimensional.

Example 1.3.1. We compute $\dim(M_5(30))$ and $\dim(S_5(30))$ in Sage:

```
sage: ModularForms(Gamma1(30),5).dimension()
112
sage: CuspForms(Gamma1(30),5).dimension()
80
```

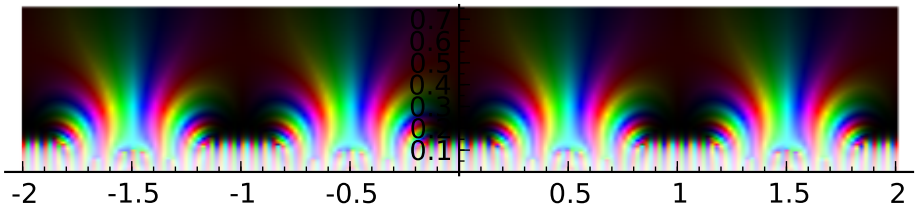
For example, the space $S_{12}(1)$ has dimension one and is spanned by the famous cusp form

$$\Delta = q \prod_{n=1}^{\infty} (1-q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n.$$

The coefficients $\tau(n)$ define the *Ramanujan τ-function*. A non-obvious fact is that τ is multiplicative and for every prime p and positive integer ν, we have

$$\tau(p^{\nu+1}) = \tau(p)\tau(p^{\nu}) - p^{11}\tau(p^{\nu-1}).$$

Example 1.3.2. We draw a plot of the Δ function (using 20 terms of the q-expansion) on the upper half plane. Notice the symmetry $\Delta(z) = \Delta(z+1)$:




```
sage: z = var('z'); q = exp(2*pi*i*z)
sage: D = delta_qexp(20)(q)
sage: complex_plot(D, (-2,2), (0,.75), plot_points=200)
```

1.4 Hecke operators

Mordell defined operators T_n, $n \geq 1$, on $S_k(N)$ which are called *Hecke operators*. These proved very fruitful. The set of such operators forms a commuting family of endomorphisms and is hence "almost" simultaneously diagonalizable.

Often there is a basis $f_1, \ldots, f_r$ of $S_k(N)$ such that each $f = f_i = \sum_{n=1}^{\infty} a_n q^n$ is a simultaneous eigenvector for all the Hecke operators T_n normalized so that $T_n f = a_n f$, i.e., so the coefficient of q is 1. In this situation, the eigenvalues a_n are necessarily algebraic integers and the field

$$K = K_f = \mathbf{Q}(\ldots, a_n, \ldots)$$

generated by all a_n is finite over $\mathbf{Q}$.

The a_n exhibit remarkable properties. For example,

$$\tau(n) \equiv \sum_{d|n} d^{11} \pmod{691}.$$

We can check this congruence for $n = 30$ in Sage as follows:

```
sage: n=30; t = delta_qexp(n+1)[n]; t
-29211840
sage: sigma(n,11)
17723450167663752
sage: (sigma(n,11) - t)%691
0
```

The key to studying and interpreting the a_n is to understand the deep connections between Galois representations and modular forms that were discovered by Serre, Shimura, Eichler and Deligne.

2

Modular Representations and Algebraic Curves

2.1 Modular forms and Arithmetic

Consider a cusp form

$$f = \sum_{n=1}^{\infty} a_n q^n \in S_k(N)$$

which is an eigenform for all of the Hecke operators T_p, and assume f is normalized so $a_1 = 1$. Then the the L-function

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

Hecke proved that $L(f, s)$ extends uniquely to a holomorphic function on $\mathbf{C}$ that satisfies a functional equation. He did this using the (Mellin) integral transform

$$\int_0^{\infty} f(it) t^s \frac{dt}{t} = N^{s/2} (2\pi)^{-s} \Gamma(s) L(f, s) = \Lambda(f, s).$$

There is a complex number c with absolute value 1 such that the function $\Lambda(f, s)$ satisfies the functional equation

$$\Lambda(f, s) = c\Lambda(\tilde{f}, k - s),$$

where $\tilde{f} = \sum_{n=1}^{\infty} \overline{a}_n q^n \in S_k(N)$ is obtained from f by complex conjugation of coefficients (see Section 6.5).

Let $K = \mathbf{Q}(a_1, a_2, \ldots)$ be the number field generated by the Fourier coefficients of f. One can show that the a_n are algebraic integers and that K is a number field. When $k = 2$, Shimura associated to f an abelian variety A_f over $\mathbf{Q}$ of dimension $[K : \mathbf{Q}]$ on which $\mathbf{Z}[a_1, a_2, \ldots]$ acts [Shi94, Theorem 7.14].

Example 2.1.1 (Modular Elliptic Curves). Suppose now that all coefficients a_n of f lie in $\mathbf{Q}$ so that $[K : \mathbf{Q}] = 1$ and hence A_f is a one dimensional abelian variety, i.e., an elliptic curve. An elliptic curve isogenous to one arising via this construction is called *modular*.

Definition 2.1.2. Elliptic curves E_1 and E_2 are *isogenous* if there is a morphism $E_1 \to E_2$ of algebraic groups, having finite kernel.

The following theorem motivates much of the theory discussed in this course. It is a theorem of Breuil, Conrad, Diamond, Taylor, and Wiles (see [BCDT01]).

Theorem 2.1.3 (Modularity Theorem)**.** *Every elliptic curve over* $\mathbf{Q}$ *is modular, that is, isogenous to a curve constructed in the above way.*

For $k \geq 2$ Serre and Deligne discovered a way to associate to f a family of ℓ-adic representations. Let ℓ be a prime number and $K = \mathbf{Q}(a_1, a_2, \ldots)$ be as above. Then it is well known that

$$K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell \cong \prod_{\lambda|\ell} K_\lambda.$$

One can associate to f a representation

$$\rho_{\ell,f} : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell)$$

unramified at all primes $p \nmid \ell N$. Let $\overline{\mathbf{Z}}$ be the ring of all algebraic integers. For $\rho_{\ell,f}$ to be unramified at p we mean that for all primes P of $\overline{\mathbf{Z}}$ lying over p, the inertia subgroup of the decomposition group at P is contained in the kernel of $\rho_{\ell,f}$. The decomposition group D_P at P is the set of those $g \in G$ which fix P. Let $\overline{k}$ be the residue field $\overline{\mathbf{Z}}/P$ and $k = \mathbf{F}_p$. Then the inertia group I_P is the kernel of the surjective map $D_P \to \mathrm{Gal}(\overline{k}/k)$.

Now $I_P \subset D_P \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and D_P/I_P is pro-cyclic (being isomorphic to the Galois group $\mathrm{Gal}(\overline{k}/k)$), so it is generated by a Frobenious automorphism Frob_p lying over p. One has

$$\mathrm{tr}(\rho_{\ell,f}(\mathrm{Frob}_p)) = a_p \in K \subset K \otimes \mathbf{Q}_\ell$$
$$\text{and}$$
$$\det(\rho_{\ell,f}) = \chi_\ell^{k-1}\varepsilon$$

where, as explained below, χ_ℓ is the ℓth cyclotomic character and ε is the Dirichlet character associated to f.

There is an incredible amount of "abuse of notation" packed into the above statement. Let $M = \overline{\mathbf{Q}}^{\ker(\rho_{\ell,f})}$ be the field fixed by the kernel of $\rho_{\ell,f}$. Then the Frobenius element Frob_P (note P not p) is well defined as an element of $\mathrm{Gal}(M/\mathbf{Q})$, and the element Frob_p is then only well defined up to conjugacy. But this works out since $\rho_{\ell,f}$ is well-defined on $\mathrm{Gal}(M/\mathbf{Q})$ (it kills $\mathrm{Gal}(\overline{\mathbf{Q}}/M)$) and trace is well-defined on conjugacy classes ($\mathrm{tr}(AB) = \mathrm{tr}(BA)$ so $\mathrm{tr}(ABA^{-1}) = Tr(B)$).

2.2 Characters

Let $f \in S_k(N)$ be an eigenform for all Hecke operators. Then for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ with $c \equiv 0 \mod N$ we have

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k \varepsilon(d) f(z),$$

where $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ is a Dirichlet character mod N. If f is also normalized so that $a_1 = 1$, as in Section 1.4, then ε actually takes values in K^*.

Let $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and let φ_N be the mod N cyclotomic character so that $\varphi_N : G \to (\mathbf{Z}/N\mathbf{Z})^*$ takes $g \in G$ to the automorphism induced by g on the Nth cyclotomic extension $\mathbf{Q}(\mu_N)$ of $\mathbf{Q}$ (where we identify $\mathrm{Gal}(\mathbf{Q}(\mu_N)/\mathbf{Q})$ with $(\mathbf{Z}/N\mathbf{Z})^*$). Then what we called ε above in the formula $\det(\rho_\ell) = \chi_\ell^{k-1}\varepsilon$ is really the composition

$$G \xrightarrow{\varphi_N} (\mathbf{Z}/N\mathbf{Z})^* \xrightarrow{\varepsilon} \mathbf{C}^*.$$

For each positive integer ν we consider the ℓ^νth cyclotomic character on G,

$$\varphi_{\ell^\nu} : G \to (\mathbf{Z}/\ell^\nu\mathbf{Z})^*.$$

Putting these together gives the ℓ-adic cyclotomic character

$$\chi_\ell : G \to \mathbf{Z}_\ell^*.$$

2.3 Parity conditions

Let $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be complex conjugation. Then $\varphi_N(c) = -1$ so $\varepsilon(c) = \varepsilon(-1)$ and $\chi_\ell^{k-1}(c) = (-1)^{k-1}$. Letting $\left(\begin{smallmatrix} a & b \\ c' & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, for $f \in S_k(N)$, we have

$$f(z) = (-1)^k \varepsilon(-1) f(z),$$

so $(-1)^k\varepsilon(-1) = 1$. Thus

$$\det(\rho_{\ell,f}(c)) = \varepsilon(-1)(-1)^{k-1} = -1.$$

We say a representation is *odd* if the determinant of complex conjugation is -1. Thus the representation $\rho_{\ell,f}$ is odd.

Remark 2.3.1 (Vague Question). How can one recognize representations like $\rho_{\ell,f}$ "in nature"? Fontaine and Mazur have made relevant conjectures. The modularity theorem can be reformulated by saying that for any representation $\rho_{\ell,E}$ coming from an elliptic curve E there is an f so that $\rho_{\ell,E} \cong \rho_{\ell,f}$.

2.4 Conjectures of Serre (mod ℓ version)

Suppose f is a modular form, $\ell \in \mathbf{Z}$ prime, λ a prime lying over ℓ, and the representation

$$\rho_{\lambda,f} : G \to \mathrm{GL}_2(K_\lambda)$$

(constructed by Serre-Deligne) is irreducible. Then $\rho_{\lambda,f}$ is conjugate to a representation with image in $\mathrm{GL}_2(\mathcal{O}_\lambda)$, where $\mathcal{O}_\lambda$ is the ring of integers of K_λ (see Section 2.5 below). Reducing mod λ gives a representation

$$\overline{\rho}_{\lambda,f} : G \to \mathrm{GL}_2(\mathbf{F}_\lambda)$$

which has a well-defined trace and det, i.e., the det and trace do not depend on the choice of conjugate representation used to obtain the reduced representation. One knows from representation theory (the Brauer-Nesbitt theorem – see [CR62]) that if such a representation is semisimple then it is completely determined by its trace and det (more precisely, it is determined by the characteristic polynomials of all of its elements). Thus if $\overline{\rho}_{\lambda,f}$ is irreducible (and hence semisimple) then it is unique in the sense that it does not depend on the choice of conjugate.

2.5 General remarks on mod p Galois representations

First, what are semisimple and irreducible representations? Remember that a representation ρ is a map from a group G to the endomorphisms of some vector space W (or a free module M if we are working over a ring instead of a field, but let's not worry about that for now). A subspace W' of W is said to be invariant under ρ if ρ takes W' back into itself. (The point is that if W' is invariant, then ρ induces representations on both W' and W/W'.) An irreducible representation is one whose only invariant subspaces are $\{0\}$ and W. A semisimple representation is one where for every invariant subspace W' there is a complementary invariant subspace W'' – that is, you can write ρ as the direct sum of $\rho|_{W'}$ and $\rho|_{W''}$.

Another way to say this is that if W' is an invariant subspace then we get a short exact sequence

$$0 \to \rho|_{W/W'} \to \rho \to \rho|_{W'} \to 0.$$

Furthermore ρ is semisimple if and only if every such sequence splits.

Note that irreducible representations are semisimple. As mentioned above, two-dimensional semisimple Galois representations are uniquely determined (up to isomorphism class) by their trace and determinant. In the case we are considering, $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact, so the image of any Galois representation ρ into $\mathrm{GL}_2(K_\lambda)$ is compact. Thus we can conjugate it into $\mathrm{GL}_2(\mathcal{O}_\lambda)$. Irreducibility is not needed for this.

Now that we have a representation into $\mathrm{GL}_2(\mathcal{O}_\lambda)$, we can reduce to get a representation $\overline{\rho}$ to $\mathrm{GL}_2(\mathbf{F}_\lambda)$. This reduced representation is not uniquely determined by ρ, since we made a choice of basis (via conjugation) so that ρ would have image in $\mathrm{GL}_2(\mathcal{O}_\lambda)$, and a different choice may lead to a non-isomorphic representation mod λ. However, the trace and determinant of a matrix are invariant under conjugation, so the trace and determinant of the reduced representation $\overline{\rho}$ are uniquely determined by ρ.

So we know the trace and determinant of the reduced representation. If we also knew that it was semisimple, then we would know its isomorphism class, and we would be done. So we would be happy if the reduced representation is irreducible. And in fact, it is easy to see that if the reduced representation is irreducible, then ρ must also be irreducible. Most ρ of interest to us *in this book* will be irreducible. Unfortunately, the opposite implication does not hold: ρ irreducible need not imply that $\overline{\rho}$ is irreducible.

2.6 Serre's conjecture

Serre has made the following conjecture which is *now a theorem* (see [KW08]).

Conjecture 2.6.1 (Serre)**.** *All 2-dimensional irreducible representation of G over a finite field which are odd, i.e., $det(\sigma(c)) = -1$, c complex conjugation, are of the form $\overline{\rho}_{\lambda,f}$ for some representation $\rho_{\lambda,f}$ constructed as above.*

Example 2.6.2. Let $E/\mathbf{Q}$ be an elliptic curve and let $\sigma_\ell : G \to \mathrm{GL}_2(\mathbf{F}_\ell)$ be the representation induced by the action of G on the ℓ-torsion of E. Then $\det \sigma_\ell = \varphi_\ell$ is odd and σ_ℓ is usually irreducible, so Serre's conjecture implies that σ_ℓ is modular. From this one can, assuming Serre's conjecture, prove that E is itself modular (see [Rib92]).

Definition 2.6.3 (Modular representation)**.** Let $\sigma : G \to \mathrm{GL}_2(\mathbf{F})$ ($\mathbf{F}$ is a finite field) be an irreducible representation of the Galois group G. Then we say that the representation σ is

Definition 2.6.4. modular if there is a modular form f, a prime λ, and an embedding $\mathbf{F} \hookrightarrow \overline{\mathbf{F}}_\lambda$ such that $\sigma \cong \overline{\rho}_{\lambda,f}$ over $\overline{\mathbf{F}}_\lambda$.

For more details, see Chapter 21 and [RS01].

2.7 Wiles's perspective

Suppose $E/\mathbf{Q}$ is an elliptic curve and $\rho_{\ell,E} : G \to \mathrm{GL}_2(\mathbf{Z}_\ell)$ the associated ℓ-adic representation on the Tate module T_ℓ. Then by reducing we obtain a mod ℓ representation

$$\overline{\rho}_{\ell,E} = \sigma_{\ell,E} : G \to \mathrm{GL}_2(\mathbf{F}_\ell).$$

If we can show this representation is modular for infinitely many ℓ then we will know that E is modular.

Theorem 2.7.1 (Langlands and Tunnel)**.** *If $\sigma_{2,E}$ and $\sigma_{3,E}$ are irreducible, then they are modular.*

This is proved by using that $\mathrm{GL}_2(\mathbf{F}_2)$ and $\mathrm{GL}_2(\mathbf{F}_3)$ are solvable so we may apply something called "base change for GL_2."

Theorem 2.7.2 (Wiles)**.** *If ρ is an ℓ-adic representation which is irreducible and modular mod ℓ with $\ell > 2$ and certain other reasonable hypothesis are satisfied, then ρ itself is modular.*

3
Modular Forms of Level 1

In this chapter, we view modular forms of level 1 both as holomorphic functions on the upper half plane and functions on lattices. We then define Hecke operators on modular forms, and derive explicit formulas for the action of Hecke operators on q-expansions. An excellent reference for the theory of modular forms of level 1 is Serre [Ser73, Ch. 7].

3.1 The Definition

Let k be an integer. The space $S_k = S_k(1)$ of cusp forms of level 1 and weight k consists of all functions f that are holomorphic on the upper half plane $\mathfrak{h}$ and such that for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ one has

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^k f(\tau), \tag{3.1.1}$$

and f vanishes at infinity, in a sense which we will now make precise. The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is in $\mathrm{SL}_2(\mathbf{Z})$, so $f(\tau + 1) = f(\tau)$. Thus f passes to a well-defined function of $q(\tau) = e^{2\pi i \tau}$. Since for $\tau \in \mathfrak{h}$ we have $|q(\tau)| < 1$, we may view $f(z) = F(q)$ as a function of q on the punctured open unit disc $\{q : 0 < |q| < 1\}$. The condition that $f(\tau)$ vanishes at infinity means that $F(q)$ extends to a holomorphic function on the open disc $\{q : |q| < 1\}$ so that $F(0) = 0$. Because holomorphic functions are represented by power series, there is a neighborhood of 0 such that

$$f(q) = \sum_{n=1}^{\infty} a_n q^n,$$

so for all $\tau \in \mathfrak{h}$ with sufficiently large imaginary part (but see Remark 3.1.1 below), $f(\tau) = \sum_{n=1}^{\infty} a_n e^{2\pi i n \tau}$.

It will also be useful to consider the slightly larger space $M_k(1)$ of holomorphic functions on $\mathfrak{h}$ that transform as above and are merely required to be holomorphic at infinity.

Remark 3.1.1. In fact, the series $\sum_{n=1}^{\infty} a_n e^{2\pi i n\tau}$ converges for all $\tau \in \mathfrak{h}$. This is because the Fourier coefficients a_n are $O(n^{k/2})$ (see [Miy89, Cor. 2.1.6, pg. 43]).

Remark 3.1.2. In [Ser73, Ch. 7], the weight is defined in the same way, but in the notation our k is twice his k.

3.2 Some examples and conjectures

The space $S_k(1)$ of cusp forms is a finite-dimensional complex vector space. For k even we have $\dim S_k(1) = \lfloor k/12 \rfloor$ if $k \not\equiv 2 \pmod{12}$ and $\lfloor k/12 \rfloor - 1$ if $k \equiv 2 \pmod{12}$, except when $k = 2$ in which case the dimension is 0. For even k, the space $M_k(1)$ has dimension 1 more than the dimension of $S_k(1)$, except when $k = 2$ when both have dimension 0. (For proofs, see, e.g., [Ser73, Ch. 7, §3].)

By the dimension formula mentioned above, the first interesting example is the space $S_{12}(1)$, which is a 1-dimensional space spanned by

$$\begin{aligned}\Delta(q) &= q\prod_{n=1}^{\infty}(1-q^n)^{24} \\ &= q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 - 16744q^7 + 84480q^8 \\ &\quad - 113643q^9 - 115920q^{10} + 534612q^{11} - 370944q^{12} - 577738q^{13} + \cdots\end{aligned}$$

That Δ lies in $S_{12}(1)$ is proved in [Ser73, Ch. 7, §4.4] by expressing Δ in terms of elements of $M_4(1)$ and $M_6(1)$, and computing the q-expansion of the resulting expression.

Example 3.2.1. We compute the q-expansion of Δ in Sage:

```
sage: delta_qexp(7)
q - 24*q^2 + 252*q^3 - 1472*q^4 + 4830*q^5 - 6048*q^6 + O(q^7)
```

In Sage, computing `delta_qexp(10^6)` only takes a few seconds, and computing up to precision `10^8` is even reasonable. Sage does not use the formula $q\prod_{n=1}^{\infty}(1-q^n)^{24}$ given above, which would take a very long time to directly evaluate, but instead uses the identity

$$\Delta(q) = \left(\sum_{n\geq 0}(-1)^n(2n+1)q^{n(n+1)/2}\right)^8,$$

and computes the 8th power using asymptotically fast polynomial arithmetic in $\mathbf{Z}[q]$, which involves a discrete fast Fourier transform (implemented in [HH]).

The Ramanujan τ function $\tau(n)$ assigns to n the nth coefficient of $\Delta(q)$.

Conjecture 3.2.2 (Lehmer)**.** *$\tau(n) \neq 0$ for all $n \geq 1$.*

This conjecture has been verified for $n \leq 22796241520242687999$ (Bosman, 2007 – see `http://en.wikipedia.org/wiki/Tau-function`).

Theorem 3.2.3 (Edixhoven et al.)**.** *Let p be a prime. There is a polynomial time algorithm to compute $\tau(p)$, polynomial in the number of digits of p.*

Edixhoven's idea is to use ℓ-adic cohomology and Arakelov theory to find an analogue of the Schoof-Elkies-Atkin algorithm (which counts the number N_q of points on an elliptic curves over a finite field $\mathbf{F}_q$ by computing $N_q \mod \ell$ for many primes ℓ). Here's some of what Edixhoven has to say about his result:

> "You need to compute on varying curves such as $X_1(\ell)$ for ℓ up to $\log(p)$ say. An important by-product of my method is the computation of the mod ℓ Galois representations associated to Δ in time polynomial in ℓ. So, it should be seen as an attempt to make the Langlands correspondence for GL_2 over $\mathbf{Q}$ available computationally."

If $f \in M_k(1)$ and $g \in M_{k'}(1)$, then it is easy to see from the definitions that $fg \in M_{k+k'}(1)$. Moreover, $\bigoplus_{k\geq 0} M_k(1)$ is a commutative graded ring generated freely by $E_4 = 1+240\sum_{n=1}^{\infty}\sigma_3(n)q^n$ and $E_6 = 1-504\sum_{n=1}^{\infty}\sigma_5(n)q^n$, where $\sigma_d(n)$ is the sum of the dth powers of the positive divisors of n (see [Ser73, Ch.7, §3.2]).

Example 3.2.4. Because E_4 and E_6 generate, it is straightforward to write down a basis for any space $M_k(1)$. For example, the space $M_{36}(1)$ has basis

$$\begin{aligned} f_1 &= 1 + 6218175600q^4 + 15281788354560q^5 + \cdots \\ f_2 &= q + 57093088q^4 + 37927345230q^5 + \cdots \\ f_3 &= q^2 + 194184q^4 + 7442432q^5 + \cdots \\ f_4 &= q^3 - 72q^4 + 2484q^5 + \cdots \end{aligned}$$

3.3 Modular forms as functions on lattices

In order to define Hecke operators, it will be useful to view modular forms as functions on lattices in $\mathbf{C}$.

Definition 3.3.1 (Lattice)**.** A *lattice* $L \subset \mathbf{C}$ is a subgroup $L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ for which $\omega_1, \omega_2 \in \mathbf{C}$ are linearly independent over $\mathbf{R}$.

We may assume that $\omega_1/\omega_2 \in \mathfrak{h} = \{z \in \mathbf{C} : \mathrm{Im}(z) > 0\}$. Let $\mathcal{R}$ be the set of all lattices in $\mathbf{C}$. Let $\mathcal{E}$ be the set of isomorphism classes of pairs (E, ω), where E is an elliptic curve over $\mathbf{C}$ and $\omega \in \Omega^1_E$ is a nonzero holomorphic differential 1-form on E. Two pairs (E, ω) and (E', ω') are isomorphic if there is an isomorphism $\varphi : E \to E'$ such that $\varphi^*(\omega') = \omega$.

Proposition 3.3.2. *There is a bijection between $\mathcal{R}$ and $\mathcal{E}$ under which $L \in \mathcal{R}$ corresponds to $(\mathbf{C}/L,\, dz) \in \mathcal{E}$.*

Proof. We describe the maps in each direction, but leave the proof that they induce a well-defined bijection as an exercise for the reader [[add ref to actual exercise]]. Given $L \in \mathcal{R}$, by Weierstrass theory the quotient $\mathbf{C}/L$ is an elliptic curve, which is equipped with the distinguished differential ω induced by the differential dz on $\mathbf{C}$. [[TODO: See [Kat73, Appendix A1.1] where the exact curve, weierstrass $\wp$ and that $\omega = dx/y$ corresponds to dz is explained nicely.]]

Conversely, if E is an elliptic curve over $\mathbf{C}$ and $\omega \in \Omega^1_E$ is a nonzero differential, we obtain a lattice L in $\mathbf{C}$ by integrating homology classes:

$$L = L_\omega = \left\{ \int_\gamma \omega : \gamma \in \mathrm{H}_1(E(\mathbf{C}), \mathbf{Z}) \right\}.$$

□

Let

$$\mathcal{B} = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \mathbf{C},\, \omega_1/\omega_2 \in \mathfrak{h}\},$$

be the set of ordered basis of lattices in $\mathbf{C}$, ordered so that $\omega_1/\omega_2 \in \mathfrak{h}$. There is a left action of $\mathrm{SL}_2(\mathbf{Z})$ on $\mathcal{B}$ given by

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(\omega_1, \omega_2) \mapsto (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$$

and $\mathrm{SL}_2(\mathbf{Z})\backslash\mathcal{B} \cong \mathcal{R}$. (The action is just the left action of matrices on column vectors, except we write (ω_1, ω_2) as a row vector since it takes less space.)

Give a modular form $f \in M_k(1)$, associate to f a function $F : \mathcal{R} \to \mathbf{C}$ as follows. First, on lattices of the special form $\mathbf{Z}\tau + \mathbf{Z}$, for $\tau \in \mathfrak{h}$, let $F(\mathbf{Z}\tau + \mathbf{Z}) = f(\tau)$.

In order to extend F to a function on all lattices, note that F satisfies the homogeneity condition $F(\lambda L) = \lambda^{-k} F(L)$, for any $\lambda \in \mathbf{C}$ and $L \in \mathcal{R}$. Then

$$F(\mathbf{Z}\omega_1 + \mathbf{Z}\omega_2) = \omega_2^{-k} F(\mathbf{Z}\omega_1/\omega_2 + \mathbf{Z}) := \omega_2^{-k} f(\omega_1/\omega_2).$$

That F is well-defined exactly amounts to the transformation condition (3.1.1) that f satisfies.

Lemma 3.3.3. *The lattice function $F : \mathcal{R} \to \mathbf{C}$ associated to $f \in M_k(1)$ is well defined.*

Proof. Suppose $\mathbf{Z}\omega_1 + \mathbf{Z}\omega_2 = \mathbf{Z}\omega_1' + \mathbf{Z}\omega_2'$ with ω_1/ω_2 and ω_1'/ω_2' both in $\mathfrak{h}$. We must verify that $\omega_2^{-k} f(\omega_1/\omega_2) = (\omega_2')^{-k} f(\omega_1'/\omega_2')$. There exists $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ such that $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$. Dividing, we see that $\omega_1'/\omega_2' = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(\omega_1/\omega_2)$. Because f is a weight k modular form, we have

$$f\left(\frac{\omega_1'}{\omega_2'}\right) = f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\left(\frac{\omega_1}{\omega_2}\right)\right) = \left(c\frac{\omega_1}{\omega_2} + d\right)^k f\left(\frac{\omega_1}{\omega_2}\right).$$

Multiplying both sides by ω_2^k yields

$$\omega_2^k f\left(\frac{\omega_1'}{\omega_2'}\right) = (c\omega_1 + d\omega_2)^k f\left(\frac{\omega_1}{\omega_2}\right).$$

Observing that $\omega_2' = c\omega_1 + d\omega_2$ and dividing again completes the proof. □

Since $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$, we can recover f from F, so the map $f \mapsto F$ is injective. Moreover, it is surjective in the sense that if F is homogeneous of degree $-k$, then F arises from a function $f : \mathfrak{h} \to \mathbf{C}$ that transforms like a modular form. More precisely, if $F : \mathcal{R} \to \mathbf{C}$ satisfies the homogeneity condition $F(\lambda L) = \lambda^{-k} F(L)$, then the function $f : \mathfrak{h} \to \mathbf{C}$ defined by $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$ transforms like a modular

form of weight k, as the following computation shows: For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ and $\tau \in \mathfrak{h}$, we have

$$\begin{aligned} f\left(\frac{a\tau+b}{c\tau+d}\right) &= F\left(\mathbf{Z}\frac{a\tau+b}{c\tau+d} + \mathbf{Z}\right) \\ &= F((c\tau+d)^{-1}\left(\mathbf{Z}(a\tau+b) + \mathbf{Z}(c\tau+d)\right)) \\ &= (c\tau+d)^k F\left(\mathbf{Z}(a\tau+b) + \mathbf{Z}(c\tau+d)\right) \\ &= (c\tau+d)^k F(\mathbf{Z}\tau + \mathbf{Z}) \\ &= (c\tau+d)^k f(\tau). \end{aligned}$$

Say that a function $F : \mathcal{R} \to \mathbf{C}$ is holomorphic on $\mathfrak{h} \cup \{\infty\}$ if the function $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$ is. We summarize the above discussion in a proposition.

Proposition 3.3.4. *There is a bijection between $M_k(1)$ and functions $F : \mathcal{R} \to \mathbf{C}$ that are homogeneous of degree $-k$ and holomorphic on $\mathfrak{h} \cup \{\infty\}$. Under this bijection $F : \mathcal{R} \to \mathbf{C}$ corresponds to $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$.*

3.4 Hecke operators

Define a map T_n from the free abelian group generated by all $\mathbf{C}$-lattices into itself by

$$T_n(L) = \sum_{\substack{L' \subset L \\ [L:L']=n}} L',$$

where the sum is over all sublattices $L' \subset L$ of index n. For any function $F : \mathcal{R} \to \mathbf{C}$ on lattices, define $T_n(F) : \mathcal{R} \to \mathbf{C}$ by

$$(T_n(F))(L) = n^{k-1} \sum_{\substack{L' \subset L \\ [L:L']=n}} F(L').$$

If F is homogeneous of degree $-k$, then $T_n(F)$ is also homogeneous of degree $-k$.

In the next Section, we will show that $\gcd(n, m) = 1$ implies $T_n T_m = T_{nm}$ and that T_{p^k} is a polynomial in $\mathbf{Z}[T_p]$.

Suppose $L' \subset L$ with $[L : L'] = n$. Then every element of L/L' has order dividing n, so $nL \subset L' \subset L$ and

$$L'/nL \subset L/nL \approx (\mathbf{Z}/n\mathbf{Z})^2.$$

Thus the subgroups of $(\mathbf{Z}/n\mathbf{Z})^2$ of order n correspond to the sublattices L' of L of index n. When $n = \ell$ is prime, there are $\ell + 1$ such subgroups, since the subgroups correspond to nonzero vectors in $\mathbf{F}_\ell$ modulo scalar equivalence, and there are $(\ell^2 - 1)/(\ell - 1) = \ell + 1$ of them.

Recall from Proposition 3.3.2 that there is a bijection between the set $\mathcal{R}$ of lattices in $\mathbf{C}$ and the set $\mathcal{E}$ of isomorphism classes of pairs (E, ω), where E is an elliptic curve over $\mathbf{C}$ and ω is a nonzero differential on E.

Suppose $F : \mathcal{R} \to \mathbf{C}$ is homogeneous of degree $-k$, so $F(\lambda L) = \lambda^{-k} F(L)$. Then we may also view T_ℓ as a sum over lattices that contain L with index ℓ, as follows.

Suppose $L' \subset L$ is a sublattice of index ℓ and set $L'' = \ell^{-1}L'$. Then we have a chain of inclusions

$$\ell L \subset L' \subset L \subset \ell^{-1}L' = L''.$$

Since $[\ell^{-1}L' : L'] = \ell^2$ and $[L : L'] = \ell$, it follows that $[L'' : L] = \ell$. Because F is homogeneous of degree $-k$,

$$T_\ell(F)(L) = \ell^{k-1} \sum_{[L:L']=\ell} F(L') = \frac{1}{\ell} \sum_{[L'':L]=\ell} F(L''). \tag{3.4.1}$$

3.4.1 Relations Between Hecke Operators

In this section we show that the Hecke operators T_n, viewed as functions on the free abelian group on lattices via

$$T_n(L) = \sum_{\substack{L' \subset L \\ [L:L']=n}} L'$$

are multiplicative satisfy a recurrence for prime powers. Let $R_p(L) = pL \in \mathcal{R}$ be the lattice L scaled by p (this is not p copies of L in the free abelian group).

Proposition 3.4.1. *If* $\gcd(n, m) = 1$, *then*

$$T_{nm} = T_n T_m. \tag{3.4.2}$$

If $r \geq 1$ *and* p *is a prime, then*

$$T_{p^{r+1}} = T_{p^r} T_p - p R_p T_{p^{r-1}}. \tag{3.4.3}$$

Proof. (Compare [Ser73, Cor. 1, pg. 99].)

Proving relation (3.4.2) is equivalent to showing that for every sublattice L'' of L of index nm, there is a unique sublattice $L' \subset L$ with $L'' \subset L'$ such that $[L : L'] = m$ and $[L' : L''] = n$. To see this, note that the abelian group L/L'' is of order nm, so it decomposes *uniquely* as a product of subgroups of orders n and m. The unique subgroup of L/L'' of order m corresponds to the unique sublattice $L' \subset L$ such that $[L : L'] = m$ and $[L' : L''] = n$.

To prove (3.4.3), let L be a lattice and note that

$$T_{p^r} T_p(L), \qquad T_{p^{r+1}}(L), \qquad \text{and} \qquad R_p T_{p^{r-1}}(L)$$

are all linear combinations of lattices of index p^{r+1} in L (note that R_p commutes with $T_{p^{r-1}}$ and $[L : R_p(L)] = p^2$). Let L'' be a lattice of index p^{r+1} in L. In the linear combination L'' appears with coefficients $a, b, c \in \mathbf{Z}$ (say), and our goal is to prove that $a = b + pc$. Note that $b = 1$, since the lattices in the sum $T_{p^{r+1}}(L)$ each appear exactly once, so in fact we must prove that $a = 1 + pc$. The lattices appearing in $R_p T_{p^{r-1}}(L) = T_{p^{r-1}}(R_p L) = T_{p^{r-1}}(pL)$ are exactly those of index p^{r-1} in pL, each with multiplicity 1. We consider two cases, depending on whether or not L'' is in that sum.

- **Case $L'' \subset pL$, so $c = 1$:** We must show that $a = 1 + p$. Every single one of the $p+1$ sublattices of L of index p must contain pL, so they also all contain $L'' \subset pL$. Thus $a = p + 1$, as claimed.

- **Case $L'' \not\subset pL$, so $c = 0$:** We must show that $a = 1$. We have that a is the number of lattices L' of index p in L with $L'' \subset L' \subset L$. Let L' be such a lattice. Since $[L : L'] = p$, we have $pL \subset L'$, and the image of L' in L/pL is of order p. Since $L'' \not\subset pL$, the image of L'' in L/pL is also of order p (the image is nonzero and contained in the image of L'), and since $L'' \subset L'$, we thus must have that the images of L'' and L' in L/pL are the same subgroup of order p. Hence L' is completely determined by L'', so there is exactly one L' that works, hence $a = 1$.

□

Corollary 3.4.2. *The Hecke operators T_n commute with each other, for all n.*

3.5 Hecke operators directly on q-expansions

Recall that the nth Hecke operator T_n of weight k on lattice functions is given by

$$T_n(F)(L) = n^{k-1} \sum_{\substack{L' \subset L \\ [L:L']=n}} F(L'). \tag{3.5.1}$$

Modular forms of weight k correspond to holomorphic functions of degree $-k$ on lattices, and each T_n extends to an operator on these functions on lattices, so T_n defines on operator on $M_k(1)$.

Extending F linearly to a function on the free abelian group on lattices, we have

$$T_n(F)(L) = n^{k-1} F(T_n(L)),$$

which allows us to apply Proposition 3.4.1.

Proposition 3.5.1. *The above action of the Hecke operators on homomogenous lattice function $F : \mathcal{R} \to \mathbf{C}$ of degree $-k$ (equivalently, on $M_k(1)$) satisfies the following relations:*

$$\begin{aligned} T_{nm} &= T_n T_m && \text{if } \gcd(n,m) = 1, \\ T_{p^{r+1}} &= T_{p^r} T_p - p^{k-1} T_{p^{r-1}} && \text{if } r \geq 1 \text{ and } p \text{ prime.} \end{aligned}$$

Proof. We compute F of both sides of the formulas in Proposition 3.4.1, applied to a lattice L. The first relation is immediate since $(nm)^{k-1} = n^{k-1}m^{k-1}$. For the second, note that having extended F linearly to the free abelian group on lattices, we have

$$F((pR_p)(L)) = p \cdot F(pL) = p \cdot p^{-k} F(L) = p^{1-k} F(L).$$

Thus (3.4.3) implies that if L is a lattice, then

$$F(T_{p^{r+1}}(L)) = F(T_{p^r} T_p(L)) - p^{1-k} F(T_{p^{r-1}}(L)). \tag{3.5.2}$$

Unwinding definitions, our goal is to prove that

$$(p^{r+1})^{k-1} F(T_{p^{r+1}}(L)) = (p^r)^{k-1} p^{k-1} F(T_{p^r} T_p(L)) - p^{k-1} (p^{r-1})^{k-1} F(T_{p^{r-1}}(L)).$$

Dividing both sides by $p^{(r+1)(k-1)}$ and using that (3.5.2) holds, we see that this follows from the fact that

$$\frac{p^{k-1}(p^{r-1})^{k-1}}{p^{(r+1)(k-1)}} = p^{1-k}.$$

□

A holomorphic function on the unit disk is determined by its Fourier expansion, so Fourier expansion defines an injective map $M_k(1) \hookrightarrow \mathbf{C}[[q]]$. In this section, we describe $T_n(\sum a_m q^m)$ explicitly as a q-expansion.

3.5.1 Explicit description of sublattices

In order to describe T_n more explicitly, we enumerate the sublattices $L' \subset L$ of index n. More precisely, we give a basis for each L' in terms of a basis for L. Note that L/L' is a group of order n and

$$L'/nL \subset L/nL = (\mathbf{Z}/n\mathbf{Z})^2.$$

Write $L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$, let Y_2 be the cyclic subgroup of L/L' generated by ω_2 and let $d = \#Y_2$. If $Y_1 = (L/L')/Y_2$, then Y_1 is generated by the image of ω_1, so it is a cyclic group of order $a = n/d$. Our goal is to exhibit a basis of L'. Let $\omega_2' = d\omega_2 \in L'$ and use that Y_1 is generated by the image of ω_1 to write $a\omega_1 = \omega_1' - b\omega_2$ for some integer b and some $\omega_1' \in L'$. Since b is only well-defined modulo d we may assume $0 \le b \le d-1$. Thus

$$\begin{pmatrix}\omega_1' \\ \omega_2'\end{pmatrix} = \begin{pmatrix}a & b \\ 0 & d\end{pmatrix}\begin{pmatrix}\omega_1 \\ \omega_2\end{pmatrix}$$

and the change of basis matrix has determinant $ad = n$. Since

$$\mathbf{Z}\omega_1' + \mathbf{Z}\omega_2' \subset L' \subset L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$$

and $[L : \mathbf{Z}\omega_1' + \mathbf{Z}\omega_2'] = n$ (since the change of basis matrix has determinant n) and $[L : L'] = n$ we see that $L' = \mathbf{Z}\omega_1' + \mathbf{Z}\omega_2'$.

Proposition 3.5.2. *Let n be a positive integer. There is a one-to-one correspondence between sublattices $L' \subset L$ of index n and matrices $\left(\begin{smallmatrix} a & b \\ 0 & d\end{smallmatrix}\right)$ with $ad = n$ and $0 \le b \le d-1$.*

Proof. The correspondence is described above. To check that it is a bijection, we just need to show that if $\gamma = \left(\begin{smallmatrix} a & b \\ 0 & d\end{smallmatrix}\right)$ and $\gamma' = \left(\begin{smallmatrix} a' & b' \\ 0 & d'\end{smallmatrix}\right)$ are two matrices satisfying the listed conditions, and

$$\mathbf{Z}(a\omega_1 + b\omega_2) + \mathbf{Z}d\omega_2 = \mathbf{Z}(a\omega_1' + b\omega_2') + \mathbf{Z}d\omega_2',$$

then $\gamma = \gamma'$. Equivalently, if $\sigma \in \mathrm{SL}_2(\mathbf{Z})$ and $\sigma\gamma = \gamma'$, then $\sigma = 1$. To see this, we compute

$$\sigma = \gamma'\gamma^{-1} = \frac{1}{n}\begin{pmatrix} a'd & ab' - a'b \\ 0 & ad'\end{pmatrix}.$$

Since $\sigma \in \mathrm{SL}_2(\mathbf{Z})$, we have $n \mid a'd$, and $n \mid ad'$, and $aa'dd' = n^2$. If $a'd > n$, then because $aa'dd' = n^2$, we would have $ad' < n$, which would contradict the fact

that $n \mid ad'$; also, $a'd < n$ is impossible since $n \mid a'd$. Thus $a'd = n$ and likewise $ad' = n$. Since $ad = n$ as well, it follows that $a' = a$ and $d' = d$, so $\sigma = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$ for some $t \in \mathbf{Z}$. Then $\sigma\gamma = \left(\begin{smallmatrix} a & b+dt \\ 0 & d \end{smallmatrix}\right)$, which implies that $t = 0$, since $0 \leq b \leq d-1$ and $0 \leq b + dt \leq d - 1$. □

Remark 3.5.3. As mentioned earlier, when $n = \ell$ is prime, there are $\ell+1$ sublattices of index ℓ. In general, the number of such sublattices is the sum of the positive divisors of n (exercise)[1] . 1

3.5.2 Hecke operators on q-expansions

Recall that if $f \in M_k(1)$, then f is a holomorphic function on $\mathfrak{h} \cup \{\infty\}$ such that

$$f(\tau) = f\left(\frac{a\tau + b}{c\tau + d}\right)(c\tau + d)^{-k}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$. Using Fourier expansion we write

$$f(\tau) = \sum_{m=0}^{\infty} c_m e^{2\pi i \tau m},$$

and say f is a cusp form if $c_0 = 0$. Also, there is a bijection between modular forms f of weight k and holomorphic lattice functions $F : \mathcal{R} \to \mathbf{C}$ that satisfy $F(\lambda L) = \lambda^{-k} F(L)$; under this bijection F corresponds to $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$.

Now assume $f(\tau) = \sum_{m=0}^{\infty} c_m q^m$ is a modular form with corresponding lattice function F. Using the explicit description of sublattices from Section 3.5.1 above, we can describe the action of the Hecke operator T_n on the Fourier expansion of $f(\tau)$, as follows:

$$\begin{aligned}
T_n F(\mathbf{Z}\tau + \mathbf{Z}) &= n^{k-1} \sum_{\substack{a,b,d \\ ad=n \\ 0 \leq b \leq d-1}} F((a\tau + b)\mathbf{Z} + d\mathbf{Z}) \\
&= n^{k-1} \sum d^{-k} F\left(\frac{a\tau + b}{d}\mathbf{Z} + \mathbf{Z}\right) \\
&= n^{k-1} \sum d^{-k} f\left(\frac{a\tau + b}{d}\right) \\
&= n^{k-1} \sum_{a,d,b,m} d^{-k} c_m e^{2\pi i \left(\frac{a\tau+b}{d}\right)m} \\
&= n^{k-1} \sum_{a,d,m} d^{1-k} c_m e^{\frac{2\pi i a m \tau}{d}} \frac{1}{d} \sum_{b=0}^{d-1} \left(e^{\frac{2\pi i m}{d}}\right)^b \\
&= n^{k-1} \sum_{\substack{ad=n \\ m' \geq 0}} d^{1-k} c_{dm'} e^{2\pi i a m' \tau} \\
&= \sum_{\substack{ad=n \\ m' \geq 0}} a^{k-1} c_{dm'} q^{am'}.
\end{aligned}$$

[1] Put reference to actual exercise

In the second to the last expression we let $m = dm'$ for $m' \geq 0$, then use that the sum $\frac{1}{d}\sum_{b=0}^{d-1}(e^{\frac{2\pi i m}{d}})^b$ is only nonzero if $d \mid m$, in which case the sum equals 1.

Thus

$$T_n f(q) = \sum_{\substack{ad=n \\ m\geq 0}} a^{k-1} c_{dm} q^{am}.$$

Put another way, if μ is a nonnegative integer, then the coefficient of q^μ is

$$\sum_{\substack{a|n \\ a|\mu}} a^{k-1} c_{\frac{n\mu}{a^2}}.$$

(To see this, let $m = \mu/a$ and $d = n/a$ and substitute into the formula above.)

Remark 3.5.4. When $k \geq 1$ the coefficients of q^μ for all μ belong to the $\mathbf{Z}$-module generated by the c_m.

Remark 3.5.5. Setting $\mu = 0$ gives the constant coefficient of $T_n f$ which is

$$\sum_{a|n} a^{k-1} c_0 = \sigma_{k-1}(n) c_0.$$

Thus if f is a cusp form so is $T_n f$. ($T_n f$ is holomorphic since its original definition is as a finite sum of holomorphic functions.)

Remark 3.5.6. Setting $\mu = 1$ shows that the coefficient of q in $T_n f$ is $\sum_{a|1} a^{k-1} c_n = c_n$. As an immediate corollary we have the following important result.

Corollary 3.5.7. *If f is a cusp form such that $T_n f$ has 0 as coefficient of q for all $n \geq 1$, then $f = 0$.*

In the special case when $n = p$ is prime, the action action of T_p on the q-expansion of f is given by the following formula:

$$T_p f = \sum_{\mu\geq 0} \sum_{\substack{a|p \\ a|\mu}} a^{k-1} c_{\frac{n\mu}{a^2}} q^\mu.$$

Since p is prime, either $a = 1$ or $a = p$. When $a = 1$, $c_{p\mu}$ occurs in the coefficient of q^μ and when $a = p$, we can write $\mu = p\lambda$ and we get terms $p^{k-1}c_\lambda$ in $q^{p\lambda}$. Thus

$$T_p f = \sum_{\mu\geq 0} c_{p\mu} q^\mu + p^{k-1} \sum_{\lambda\geq 0} c_\lambda q^{p\lambda}.$$

3.5.3 The Hecke algebra and eigenforms

Definition 3.5.8 (Hecke Algebra)**.** The *Hecke algebra* $\mathbf{T}$ associated to $M_k(1)$ is the subring of $\mathrm{End}(M_k(1))$ generated by the operators T_n for all n. Similarly, the *Hecke algebra* associated to $S_k(1)$ is the subring of $\mathrm{End}(S_k(1))$ generated by all Hecke operators T_n.

The Hecke algebra is commutative because T_{p^ν} is a polynomial in T_p and when $\gcd(n, m) = 1$ we have $T_n T_m = T_{nm} = T_{mn} = T_m T_n$ (see Section 3.4.1). Also, $\mathbf{T}$ is of finite rank over $\mathbf{Z}$, because of Remark 3.5.4 and that the finite dimensional space $S_k(1)$ has a basis with q-expansions in $\mathbf{Z}[[q]]$.

Definition 3.5.9 (Eigenform)**.** An *eigenform* $f \in M_k(1)$ is a nonzero element such that f is an eigenvector for every Hecke operator T_n. If $f \in S_k(1)$ is an eigenform, then f is *normalized* if the coefficient of q in the q-expansion of f is 1. We sometimes called a normalized cuspidal eigenform a *newform*.

If $f = \sum_{n=1}^{\infty} c_n q^n$ is a normalized eigenform, then Remark 3.5.6 implies that $T_n(f) = c_n f$. Thus the coefficients of a newform are exactly the system of eigenvalues of the Hecke operators acting on the newform.

Remark 3.5.10. It follows from Victor Miller's thesis [[ref my modform book??]] that $T_1, \ldots, T_n$ generate $\mathbf{T} \subset \mathrm{End}(S_k(1))$, where $n = \dim S_k(1)$.

3.5.4 Examples

We compute the space of weight 12 modular forms of level 1, along with its cuspidal subspace:

```
sage: M = ModularForms(1,12, prec=3)
sage: M.basis()
[
q - 24*q^2 + O(q^3),
1 + 65520/691*q + 134250480/691*q^2 + O(q^3)
]
sage: M.hecke_matrix(2)
[ -24    0]
[   0 2049]
sage: S = M.cuspidal_subspace()
sage: S.hecke_matrix(2)
[-24]
sage: factor(M.hecke_polynomial(2))
(x - 2049) * (x + 24)
```

We also compute the space of forms of weight 40:

```
sage: M = ModularForms(1,40)
sage: M.basis()
[
q + 19291168*q^4 + 37956369150*q^5 + O(q^6),
q^2 + 156024*q^4 + 57085952*q^5 + O(q^6),
q^3 + 168*q^4 - 12636*q^5 + O(q^6),
1 + 1082400/2610827184964491220 51*q + ...
]
sage: M.hecke_matrix(2)
[             0   549775105056 14446985236992              0]
[             1         156024     1914094476              0]
[             0            168         392832              0]
[             0              0              0   549755813889]
sage: factor(M.hecke_polynomial(2))
(x - 549755813889) *
     (x^3 - 548856*x^2 - 810051757056*x + 213542160549543936)
```

3.6 Two Conjectures about Hecke operators on level 1 modular forms

3.6.1 Maeda's conjecture

Conjecture 3.6.1 (Maeda)**.** *Let k be a positive integer such that $S_k(1)$ has positive dimension and let $T \subset \mathrm{End}(S_k(1))$ be the Hecke algebra. Then there is only one $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ orbit of normalized eigenforms of level 1.*

There is some numerical evidence for this conjecture. It is true for $k \leq 2000$, according to [FJ02]. The MathSciNet reviewer of [FJ02] said "In the present paper the authors take a big step forward towards proving Maeda's conjecture in the affirmative by establishing that the Hecke polynomial $T_{p,k}(x)$ is irreducible and has full Galois group over $\mathbb{Q}$ for $k \leq 2000$ and $p < 2000, p$ prime." Using Sage, Alex Ghitza verified the conjecture for $k \leq 4096$ (see [Ghi]). Buzzard shows in [Buz96] that for the weights $k \leq 228$ with $k/12$ a prime, the Galois group of the characteristic polynomial of T_2 is the full symmetric group, and is, in particular, irreducible.

3.6.2 The Gouvea-Mazur conjecture

Fix a prime p, and let $F_{p,k} \in \mathbf{Z}[x]$ be the characteristic polynomial of T_p acting on $M_k(1)$. The *slopes* of $F_{p,k}$ are the p-adic valuations $\mathrm{ord}_p(\alpha) \in \mathbf{Q}$ of the roots $\alpha \in \overline{\mathbf{Q}}_p$ of $F_{p,k}$. They can be computed easily using Newton polygons.[2] For example, the $p = 5$ slopes for $F_{5,12}$ are $0, 1, 1$, for $F_{5,12+4\cdot 5}$ they are $0, 1, 1, 4, 4$, and for $F_{5,12+4\cdot 5^2}$ they are $0, 1, 1, 5, 5, 5, 5, 5, 5, 10, 10, 11, 11, 14, 14, 15, 15, 16, 16$. 2

```
sage: def s(k,p):
...       M = ModularForms(1,k)
...       v = M.hecke_polynomial(p).newton_slopes(p)
...       return list(sorted(v))
sage: s(12,5)
[0, 1]
sage: s(12 + 4*5, 5)
[0, 1, 4]
sage: s(12 + 4*5^2, 5)
[0, 1, 5, 5, 5, 10, 11, 14, 15, 16]
sage: s(12 + 4*5^3, 5)          # long time
!! WAY TOO SLOW -- TODO -- see trac 9749 !!
```

Instead, we compute the slopes more directly as follows (this is fast):

```
sage: def s(k,p):
...       d = dimension_modular_forms(1, k)
...       B = victor_miller_basis(k, p*d+1)
...       T = hecke_operator_on_basis(B, p, k)
...       return list(sorted(T.charpoly().newton_slopes(p)))
sage: s(12,5)
```

[2] Jared Weinstein suggests we add some background explaining newton polygons and why they are helpful.

```
[0, 1]
sage: s(12 + 4*5, 5)
[0, 1, 4]
sage: s(12 + 4*5^2, 5)
[0, 1, 5, 5, 5, 10, 11, 14, 15, 16]
sage: s(12 + 4*5^3, 5)
[0, 1, 5, 5, 5, 10, 11, 14, 15, 16, 20, 21, 24, 25, 27,
 30, 31, 34, 36, 37, 40, 41, 45, 46, 47, 50, 51, 55, 55,
 55, 59, 60, 63, 64, 65, 69, 70, 73, 74, 76, 79, 80, 83]
```

Let $d(k, \alpha, p)$ be the multiplicity of α as a slope of $F_{p,k}$.

Conjecture 3.6.2 (Gouvea-Mazur, 1992)**.** *Fix a prime p and a nonnegative rational number α. Suppose k_1 and k_2 are integers with $k_1, k_2 \geq 2\alpha + 2$, and $k_1 \equiv k_2 \pmod{p^n(p-1)}$ for some integer $n \geq \alpha$. Then $d(k_1, \alpha, p) = d(k_2, \alpha, p)$.*

Notice that the above examples, with $p = 5$ and $k_1 = 12$, are consistent with this conjecture. However, it came as a huge surprise that the conjecture is false in general!

Frank Calegari and Kevin Buzzard [BC04] found the first counterexample, when $p = 59$, $k_1 = 16$, $\alpha = 1$, and $k_2 = 16 + 59 \cdot 58 = 3438$. We have $d(16, 0, 59) = 0$, $d(16, 1, 59) = 1$, and $d(16, \alpha, 59) = 0$ for all other α. However, initial computations strongly suggest (but do not prove!) that $d(3438, 1, 59) = 2$. It is a finite, but difficult, computation to decide what $d(3438, 1, 59)$ really is (see Section 3.7). Using a trace formula, Calegari and Buzzard at least showed that either $d(3438, 1, 59) \geq 2$ or there exists $\alpha < 1$ such that $d(3438, \alpha, 59) > 0$, both of which contradict Conjecture 3.6.2.

There are many theorems about more general formulations of the Gouvea-Mazur conjecture, and a whole geometric theory "the Eigencurve" [CM98] that helps explain it, but discussing this further is beyond the scope of this book.

3.7 An Algorithm for computing characteristic polynomials of Hecke operators

In computational investigations, it is frequently useful to compute the characteristic polynomial of the Hecke operator $T_{p,k}$ of T_p acting on $S_k(1)$. This can be accomplished in several ways, each of which has advantages. The Eichler-Selberg trace formula (see Zagier's appendix to [Lan95, Ch. III]), can be used to compute the trace of $T_{n,k}$, for $n = 1, p, p^2, \ldots, p^{d-1}$, where $d = \dim S_k(1)$, and from these traces it is straightforward to recover the characteristic polynomial of $T_{p,k}$. Using the trace formula, the time required to compute $\mathrm{Tr}(T_{n,k})$ grows "very quickly" in n (though *not* in k), so this method becomes unsuitable when the dimension is large or p is large, since p^{d-1} is huge. Another alternative is to use modular symbols of weight k, as in [Mer94], but if one is only interested in characteristic polynomials, little is gained over more naive methods (modular symbols are most useful for investigating special values of L-functions).

In this section, we describe an algorithm to compute the characteristic polynomial of the Hecke operator $T_{p,k}$, which is adapted for the case when $p > 2$. It could be generalized to modular forms for $\Gamma_1(N)$, given a method to compute a basis

of q-expansions to "low precision" for the space of modular forms of weight k and level N. By "low precision" we mean to precision $O(q^{dp+1})$, where $T_1, T_2, \ldots, T_d$ generate the Hecke algebra $\mathbf{T}$ as a ring. The algorithm described here uses nothing more than the basics of modular forms and some linear algebra; in particular, no trace formulas or modular symbols are involved.

3.7.1 Review of basic facts about modular forms

We briefly recall the background for this section. Fix an even integer k. Let $M_k(1)$ denote the space of weight k modular forms for $\mathrm{SL}_2(\mathbf{Z})$ and $S_k(1)$ the subspace of cusp forms. Thus $M_k(1)$ is a $\mathbf{C}$-vector space that is equipped with a ring

$$\mathbf{T} = \mathbf{Z}[\ldots T_{p,k} \ldots] \subset \mathrm{End}(M_k(1))$$

of Hecke operators. Moreover, there is an injective q-expansion map $M_k(1) \hookrightarrow \mathbf{C}[[q]]$. For example, when $k \geq 4$ there is an Eisenstein series E_k, which lies in $M_k(1)$. The first two Eisenstein series are

$$E_4(q) = \frac{1}{240} + \sum_{n\geq 1} \sigma_3(n) q^n \quad \text{and} \quad E_6(q) = \frac{1}{504} + \sum_{n\geq 1} \sigma_5(n) q^n,$$

where $q = e^{2\pi i z}$, $\sigma_{k-1}(n)$ is the sum of the $k-1$st power of the positive divisors. For every prime number p, the *Hecke operator* $T_{p,k}$ acts on $M_k(1)$ by

$$T_{p,k}\left(\sum_{n\geq 0} a_n q^n\right) = \sum_{n\geq 0} a_{np} q^n + p^{k-1} a_n q^{np}. \tag{3.7.1}$$

Proposition 3.7.1. *The set of modular forms $E_4^a E_6^b$ is a basis for $M_k(1)$, where a and b range through nonnegative integers such that $4a+6b = k$. Moreover, $S_k(1)$ is the subspace of $M_k(1)$ of elements whose q-expansions have constant coefficient 0.*

3.7.2 The Naive approach

Let k be an even positive integer and p be a prime. Our goal is to compute the characteristic polynomial of the Hecke operator $T_{p,k}$ acting on $S_k(1)$. In practice, when k and p are both reasonably large, e.g., $k = 886$ and $p = 59$, then the coefficients of the characteristic polynomial are huge (the roots of the characteristic polynomial are $O(p^{k/2-1})$). A naive way to compute the characteristic polynomial of $T_{p,k}$ is to use (3.7.1) to compute the matrix $[T_{p,k}]$ of $T_{p,k}$ on the basis of Proposition 3.7.1, where E_4 and E_6 are computed to precision $p \dim M_k(1)$, and to then compute the characteristic polynomial of $[T_{p,k}]$ using, e.g., a modular algorithm (compute the characteristic polynomial modulo many primes, and use the Chinese Remainder Theorem). The difficulty with this approach is that the coefficients of the q-expansions of $E_4^a E_6^b$ to precision $p \dim M_k(1)$ quickly become enormous, so both storing them and computing with them is costly, and the components of $[T_{p,k}]$ are also huge so the characteristic polynomial is difficult to compute. See Example 3.2.4 above, where the coefficients of the q-expansions are already large.

3.7.3 The Eigenform method

We now describe another approach to computing characteristic polynomials, which gets just the information required. Recall Maeda's conjecture from Section 3.6.1, which asserts that $S_k(1)$ is spanned by the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of a single eigenform $f = \sum b_n q^n$. For simplicity of exposition below, we assume this conjecture, though the algorithm can probably be modified to deal with the general case. We will refer to this eigenform f, which is well-defined up to $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugacy, as *Maeda's eigenform*.

Lemma 3.7.2. *The characteristic polynomial of the pth coefficient b_p of Maeda's eigenform f, in the field $\mathbf{Q}(b_1, b_2, \ldots)$, is equal to the characteristic polynomial of $T_{p,k}$ acting on $S_k(1)$.*

Proof. The map $\mathbf{T} \otimes \mathbf{Q} \to \mathbf{Q}(b_1, b_2, \ldots)$ that sends $T_n \to b_n$ is an isomorphism of $\mathbf{Q}$-algebras. □

Victor Miller shows in his thesis that $S_k(1)$ has a unique basis $f_1, \ldots, f_d \in \mathbf{Z}[[q]]$ with $a_i(f_j) = \delta_{ij}$, i.e., the first $d \times d$ block of coefficients is the identity matrix. Again, in the general case, the requirement that there is such a basis can be avoided, but for simplicity of exposition we assume there is such a basis. We refer to the basis $f_1, \ldots, f_d$ as *Miller's basis*.

Algorithm 3.7.3. We assume in the algorithm that the characteristic polynomial of T_2 has no multiple roots (this is easy to check, and if false, then you have found an interesting counterexample to the conjecture that the characteristic polynomial of T_2 has Galois group the full symmetric group).

1. Using Proposition 3.7.1 and Gauss elimination, we compute Miller's basis $f_1, \ldots, f_d$ to precision $O(q^{2d+1})$, where $d = \dim S_k(1)$. This is exactly the precision needed to compute the matrix of T_2.

2. Using Definition 3.7.1, we compute the matrix $[T_2]$ of T_2 with respect to Miller's basis $f_1, \ldots, f_d$. We compute the matrix with respect to the Miller basis mainly because it makes the linear algebra much simpler.

3. Using Algorithm 3.7.5 below we write down an eigenvector $\mathbf{e} = (e_1, \ldots, e_d) \in K^d$ for $[T_2]$. In practice, the components of T_2 are not very large, so the numbers involved in computing $\mathbf{e}$ are also not very large.

4. Since $e_1 f_1 + \cdots + e_d f_d$ is an eigenvector for T_2, our assumption that the characteristic polynomial of T_2 is square free (and the fact that $\mathbf{T}$ is commutative) implies that $e_1 f_1 + \cdots + e_d f_d$ is also an eigenvector for T_p. Normalizing, we see that up to Galois conjugacy,

$$b_p = \sum_{i=1}^{d} \frac{e_i}{e_1} \cdot a_p(f_i),$$

where the b_p are the coefficients of Maeda's eigenform f. For example, since the f_i are Miller's basis, if $p \leq d$ then

$$b_p = \frac{e_p}{e_1} \qquad \text{if } p \leq d,$$

since $a_p(f_i) = 0$ for all $i \neq p$ and $a_p(f_p) = 1$.

5. Finally, once we have computed b_p, we can compute the characteristic polynomial of T_p, because it is the minimal polynomial of b_p. We spend the rest of this section discussing how to make this step practical.

Computing b_p directly in step 4 is extremely costly because the divisions e_i/e_1 lead to massive coefficient explosion, and the same remark applies to computing the minimal polynomial of b_p. Instead we compute the reductions $\overline{b}_p$ modulo ℓ and the characteristic polynomial of $\overline{b}_p$ modulo ℓ for many primes ℓ, then recover *only* the characteristic polynomial of b_p using the Chinese Remainder Theorem. Deligne's bound on the magnitude of Fourier coefficients tells us how many primes we need as moduli (we leave this analysis to the reader)[3] . 3

More precisely, the reduction modulo ℓ steps are as follows. The field K can be viewed as $\mathbf{Q}[x]/(f(x))$ where $f(x) \in \mathbf{Z}[x]$ is the characteristic polynomial of T_2. We work only modulo primes such that

1. $f(x)$ has no repeated roots modulo ℓ,
2. ℓ does not divide any denominator involved in our representation of $\mathbf{e}$, and
3. the image of e_1 in $\mathbf{F}_\ell[x]/(f(x))$ is invertible.

For each such prime, we compute the image $\overline{b}_p$ of b_p in the reduced Artin ring $\mathbf{F}_\ell[x]/(f(x))$. Then the characteristic polynomial of T_p modulo ℓ equals the characteristic polynomial of $\overline{b}_p$. This modular arithmetic is fast and requires negligible storage. Most of the time is spent doing the Chinese Remainder Theorem computations, which we do each time we do a few computations of the characteristic polynomial of T_p modulo ℓ.

Remark 3.7.4. If k is really large, so that steps 1 and 2 of the algorithm take too long or require too much memory, steps 1 and 2 can be performed modulo the prime ℓ. Since the characteristic polynomial of $T_{p,k}$ modulo ℓ does not depend on any choices, we will still be able to recover the original characteristic polynomial.

3.7.4 How to write down an eigenvector over an extension field

The following algorithm, which was suggested to the author by H. Lenstra, produces an eigenvector defined over an extension of the base field.

Algorithm 3.7.5. 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

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

[3] Say more later.

We 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 [Lan93, 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 subspace $\ker(g(A))$. □

3.7.5 Simple example: weight 36*,* $p = 3$

We compute the characteristic polynomial of T_3 acting on $S_{36}(1)$ using the algorithm described above. A basis for $M_{36}(1)$ to precision $6 = 2\dim(S_{36}(1))$ is

$$\begin{aligned}
E_4^9 &= 1 + 2160q + 2093040q^2 + 1198601280q^3 + 449674832880q^4 \\
&\quad + 115759487504160q^5 + 20820305837344320q^6 + O(q^7) \\
E_4^6E_6^2 &= 1 + 432q - 353808q^2 - 257501376q^3 - 19281363984q^4 \\
&\quad + 28393576094880q^5 + 11565037898063424q^6 + O(q^7) \\
E_4^3E_6^4 &= 1 - 1296q + 185328q^2 + 292977216q^3 - 52881093648q^4 \\
&\quad - 31765004621280q^5 + 1611326503499328q^6 + O(q^7) \\
E_6^6 &= 1 - 3024q + 3710448q^2 - 2309743296q^3 + 720379829232q^4 \\
&\quad - 77533149038688q^5 - 8759475843314112q^6 + O(q^7)
\end{aligned}$$

The reduced row-echelon form (Miller) basis is:

$$\begin{aligned}
f_0 &= 1 + 6218175600q^4 + 15281788354560q^5 + 9026867482214400q^6 + O(q^7) \\
f_1 &= q + 57093088q^4 + 37927345230q^5 + 5681332472832q^6 + O(q^7) \\
f_2 &= q^2 + 194184q^4 + 7442432q^5 - 197264484q^6 + O(q^7) \\
f_3 &= q^3 - 72q^4 + 2484q^5 - 54528q^6 + O(q^7)
\end{aligned}$$

The matrix of T_2 with respect to the basis f_1, f_2, f_3 is

$$[T_2] = \begin{pmatrix} 0 & 34416831456 & 5681332472832 \\ 1 & 194184 & -197264484 \\ 0 & -72 & -54528 \end{pmatrix}$$

This matrix has (irreducible) characteristic polynomial

$$g = x^3 - 139656x^2 - 59208339456x - 1467625047588864.$$

If a is a root of this polynomial, then one finds that

$$\mathbf{e} = (2a + 108984, \quad 2a^2 + 108984a, \quad a^2 - 394723152a + 11328248114208)$$

is an eigenvector with eigenvalue a. The characteristic polynomial of T_3 is then the characteristic polynomial of e_3/e_1, which we can compute modulo ℓ for any prime ℓ such that $\overline{g} \in \mathbf{F}_\ell[x]$ is square free. For example, when $\ell = 11$,

$$\frac{e_3}{e_1} = \frac{a^2 + a + 3}{2a^2 + 7} = 9a^2 + 2a + 3,$$

which has characteristic polynomial

$$x^3 + 10x^2 + 8x + 2.$$

If we repeat this process for enough primes ℓ and use the Chinese remainder theorem, we find that the characteristic polynomial of T_3 acting on $S_{36}(1)$ is

$$x^3 + 104875308x^2 - 144593891972573904x - 211752921051049840043944 32.$$

4

Duality, Rationality, and Integrality

4.1 Modular forms for $\mathrm{SL}_2(\mathbf{Z})$ and Eisenstein series

Let $\Gamma = \Gamma_1(1) = \mathrm{SL}_2(\mathbf{Z})$ and for $k \geq 0$ let

$$M_k = \left\{ f = \sum_{n=0}^{\infty} a_n q^n : f \text{ is a modular form of weight } k \text{ for } \Gamma \right\}$$
$$\supset S_k = \left\{ f = \sum_{n=1}^{\infty} a_n q^n : f \in M_k \right\}$$

These are finite dimensional $\mathbf{C}$-vector spaces whose dimensions are easily computed. Furthermore, they are generated by familiar elements (see Serre [Ser73] or Lang [Lan95].) The main tool is the formula

$$\sum_{p \in D \cup \{\infty\}} \frac{1}{e(p)} \operatorname{ord}_p(f) = \frac{k}{12}$$

where D is the standard fundamental domain for Γ and

$$e(p) = \begin{cases} 1 & \text{otherwise} \\ 2 & \text{if } p = i \\ 3 & \text{if } p = \rho = e^{2\pi i/3} \end{cases}$$

One can alternatively define $e(p)$ as follows. If $p = \tau$ and $E = \mathbf{C}/(\mathbf{Z}\tau + \mathbf{Z})$ then $e(p) = \frac{1}{2} \# \operatorname{Aut}(E)$.

For $k \geq 4$ we define the *Eisenstein series* G_k by

$$G_k(q) = \frac{1}{2}\zeta(1-k) + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n.$$

The map

$$\tau \mapsto \sum_{\substack{(m,n)\neq(0,0)\\ m,n\in\mathbf{Z}}} \frac{1}{(m\tau+n)^k}$$

differs from G_k by a constant (see [Ser73, §VII.4.2]). Also, $\zeta(1-k) \in \mathbf{Q}$ and one may say, *symbolically* at least, "$\zeta(1-k) = \sum_{d=1}^{\infty} d^{k-1} = \sigma_{k-1}(0)$." The *nth Bernoulli number* B_n is defined by the equation

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n x^n}{n!}.$$

One can show, using the functional equation for ζ and [Ser73, §VII.4.1], that $\zeta(1-k) = -\frac{B_k}{k}$ so the constant coefficient of G_k is $-\frac{B_k}{2k}$ which is rational.

4.2 Pairings between Hecke algebras and modular forms

In what follows we assume $k \geq 2$ to avoid trivialities.. The Hecke operators T_n acts on the space M_k. Fix a subspace $V \subset M_k$ which is stable under the action of the T_n. Let $\mathbf{T} = \mathbf{T}_{\mathbf{C}}(V) = \mathbf{C}[\ldots, (T_n)_{|V}, \ldots]$ be the $\mathbf{C}$-algebra generated by the endomorphism T_n acting on V and note that $\mathbf{T}(V)$ is a finite dimensional $\mathbf{C}$-vector space since it is a subspace of $\mathrm{End}(V)$ and V is finite dimensional. Recall that $\mathbf{T}$ is commutative.

For any Hecke operator $T \in \mathbf{T}$ and $f \in V$, to make certain arguments notationally more natural, we will often write $T(f) = f|T$. The ring $\mathbf{T}$ is commutative, so it is harmless to use this notation, which suggests both a left and right module structure.

Let $\mathbf{T} = \mathbf{T}_{\mathbf{C}}(V)$, and define the bilinear pairing

$$\begin{aligned} V \times \mathbf{T} &\to \mathbf{C} \\ \langle f, T\rangle &= a_1(f|T) \end{aligned}$$

where $f|T = \sum_{n=0}^{\infty} a_n(f|T)q^n$. We thus obtain maps

$$\begin{aligned} V &\to \mathrm{Hom}(\mathbf{T}, \mathbf{C}) = \mathbf{T}^* \\ \mathbf{T} &\to \mathrm{Hom}(V, \mathbf{C}) = V^*. \end{aligned}$$

Theorem 4.2.1. *The above maps are isomorphisms.*

Proof. It suffices to show that each map is injective. Then since a finite dimensional vector space and its dual have the same dimension the result follows. First suppose $f \mapsto 0 \in \mathrm{Hom}(\mathbf{T}, \mathbf{C})$. Then $a_1(f|T) = 0$ for all $T \in \mathbf{T}$, so $a_n = a_1(f|T_n) = 0$ for all $n \geq 1$. Thus f is a constant; since $k \geq 2$ this implies $f = 0$.

Next suppose $T \mapsto 0 \in \mathrm{Hom}(V, \mathbf{C})$, then $a_1(f|T) = 0$ for all $f \in V$. Substituting $f|T_n$ for f and using the commutativity of $\mathbf{T}$ we have

$$\begin{aligned} a_1((f|T_n)|T) &= 0 && \text{for all } f,\ n \geq 1 \\ a_1((f|T)|T_n) &= 0 && \text{by commutativity} \\ a_n(f|T) &= 0 && n \geq 1 \\ f|T &= 0 && \text{since } k \geq 2, \text{ as above} \end{aligned}$$

Thus $T = 0$ which completes the proof. □

Remark 4.2.2. The above isomorphisms are $\mathbf{T}$-*equivariant*, in the following sense. We endow $\mathrm{Hom}(\mathbf{T}, \mathbf{C})$ with a $\mathbf{T}$-module structure by letting $T \in \mathbf{T}$ act on $\varphi \in \mathrm{Hom}(\mathbf{T}, \mathbf{C})$ by $(T \cdot \varphi)(T') = \varphi(TT')$. If $\alpha : V \to \mathrm{Hom}(\mathbf{T}, \mathbf{C})$ is the above isomorphism (so $\alpha : f \mapsto \varphi_f := (T' \mapsto a_1(f|T'))$), then equivariance is the statement that $\alpha(Tf) = T\alpha(f)$. This follows since

$$\begin{aligned}\alpha(Tf)(T') = \varphi_{Tf}(T') = a_1(Tf|T') = a_1(f|T'T) \\ = \varphi_f(T'T) = T\varphi_f(T') = T\alpha(f)(T').\end{aligned}$$

Similar remarks hold for $T \to V^*$.

4.3 Eigenforms

We continue to assume that $k \geq 2$.

Definition 4.3.1. A modular form $f \in M_k$ is an *eigenform for* $\mathbf{T}$ if $f|T_n = \lambda_n f$ for all $n \geq 1$ and some complex numbers λ_n.

Suppose f is an eigenform, so $a_n(f) = a_1(f|T_n) = \lambda_n a_1(f)$. Thus if $a_1(f) = 0$, then $a_n(f) = 0$ for all $n \geq 1$, so since $k \geq 2$ this implies that $f = 0$. Thus $a_1(f) \neq 0$, and we may as well divide through by $a_1(f)$ to obtain the *normalized eigenform* $\frac{1}{a_1(f)} f$. We thus assume that $a_1(f) = 1$, then the formula becomes $a_n(f) = \lambda_n$, so $f|T_n = a_n(f)f$, for all $n \geq 1$.

Theorem 4.3.2. *Let $f \in V$ and let ψ be the image of f in $\mathrm{Hom}(\mathbf{T}, \mathbf{C})$, so $\psi(T) = a_1(f|T)$. Then f is a normalized eigenform if and only if ψ is a ring homomorphism.*

Proof. First suppose f is a normalized eigenform, so $f|T_n = a_n(f)f$. Then

$$\begin{aligned}\psi(T_nT_m) = a_1(f|T_nT_m) = a_m(f|T_n) \\ = a_m(a_n(f)f) = a_m(f)a_n(f) \\ = \psi(T_n)\psi(T_m),\end{aligned}$$

so ψ is a homomorphism.

Conversely, assume ψ is a homomorphism. Then $f|T_n = \sum a_m(f|T_n)q^m$, so to show that $f|T_n = a_n(f)f$ we must show that $a_m(f|T_n) = a_n(f)a_m(f)$. Recall that $\psi(T_n) = a_1(f|T_n) = a_n$, so

$$\begin{aligned}a_n(f)a_m(f) = a_1(f|T_n)a_1(f|T_m) = \psi(T_n)\psi(T_m) \\ = \psi(T_nT_m) = a_1(f|T_n|T_m) \\ = a_m(f|T_n)\end{aligned}$$

as desired.

□

4.4 Integrality

In the previous sections, we looked at subspaces $V \subset M_k \subset \mathbf{C}[[q]]$, with $k \geq 2$, and considered the space $\mathbf{T} = \mathbf{T}(V) = \mathbf{C}[\ldots, T_n, \ldots] \subset \mathrm{End}_{\mathbf{C}} V$ of Hecke operators on V. We defined a pairing $\mathbf{T} \times V \to \mathbf{C}$ by $(T, f) \mapsto a_1(f|T)$ and showed this pairing is nondegenerate and that it induces isomorphisms $\mathbf{T} \cong \mathrm{Hom}(V, \mathbf{C})$ and $V \cong \mathrm{Hom}(\mathbf{T}, \mathbf{C})$.

Fix $k \geq 4$ and let $S = S_k$ be the space of weight k cusp forms with respect to the action of $\mathrm{SL}_2(\mathbf{Z})$. Let

$$S(\mathbf{Q}) = S_k \cap \mathbf{Q}[[q]]$$
$$S(\mathbf{Z}) = S_k \cap \mathbf{Z}[[q]].$$

Theorem 4.4.1. *There is a $\mathbf{C}$-basis of M_k consisting of forms with integer coefficients.*

Proof. We see this by exhibiting a basis. Recall that for all $k \geq 4$

$$G_k = -\frac{B_k}{2k} + \sum_{k=1}^{\infty} \sum_{d|k} d^{k-1} q^n$$

is the kth Eisenstein series, which is a modular form of weight k, and

$$E_k = -\frac{2k}{B_k} \cdot G_k = 1 + \cdots$$

is the normalization that starts with 1. Since the Bernoulli numbers $B_2, \ldots, B_8$ have 1 as numerator (this isn't always the case, $B_{10} = \frac{5}{66}$) we see that E_4 and E_6 have coefficients in $\mathbf{Z}$ and constant term 1. Furthermore one shows (see [Ser73, §VII.3.2]) by dimension and independence arguments that the modular forms

$$\{E_4^a E_6^b \,:\, 4a + 6b = k\}$$

form a basis for M_k. □

4.5 A Result from Victor Miller's thesis

Set $d = \dim_{\mathbf{C}} S_k$. Victor Miller showed in his thesis (see [Lan95], Ch. X, Theorem 4.4) that there exists

$$f_1, \ldots, f_d \in S_k(\mathbf{Z}) \quad \text{such that} \quad a_i(f_j) = \delta_{ij}$$

for $1 \leq i, j \leq d$. The f_i then form a basis for $S_k(\mathbf{Z})$.

Example 4.5.1. The space $S_{36}(\mathbf{Z})$ has basis

$$f_1 = q + 57093088q^4 + 37927345230q^5 + 5681332472832q^6 + \cdots$$
$$f_2 = q^2 + 194184q^4 + 7442432q^5 - 197264484q^6 + 722386944q^7 \cdots$$
$$f_3 = q^3 - 72q^4 + 2484q^5 - 54528q^6 + 852426q^7 - 10055232q^8 + \cdots$$

Let $\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_k)$ be the Hecke algebra associated to S_k. Miller's thesis implies the following result about $\mathbf{T}$.

Proposition 4.5.2. *We have* $\mathbf{T} = \bigoplus_{i=1}^{d} \mathbf{Z}T_i$, *as* $\mathbf{Z}$*-modules.*

Proof. To see that $T_1, \ldots, T_d \in \mathbf{T} = \mathbf{T}(S_k)$ are linearly independent over $\mathbf{C}$, suppose $\sum_{i=1}^{d} c_i T_i = 0$. Then

$$0 = a_1\left(f_j | \sum c_i T_i\right) = \sum_i c_i a_i(f_j) = \sum_i c_i \delta_{ij} = c_j.$$

From the isomorphism $\mathbf{T} \cong \mathrm{Hom}(S_k, \mathbf{C})$ we know that $\dim_{\mathbf{C}} \mathbf{T} = d$, so we can write any T_n as a $\mathbf{C}$-linear combination

$$T_n = \sum_{i=1}^{d} c_{n_i} T_i, \quad c_{n_i} \in \mathbf{C}.$$

But

$$\mathbf{Z} \ni a_n(f_j) = a_1(f_j | T_n) = \sum_{i=1}^{d} c_{n_i} a_1(f_j | T_i) = \sum_{i=1}^{d} c_{n_i} a_i(f_j) = c_{n_j}$$

so the c_{n_i} all lie in $\mathbf{Z}$, which completes the proof. □

Thus $\mathbf{T}$ is an integral Hecke algebra of finite rank d over $\mathbf{Z}$. We have a map

$$\begin{aligned} S(\mathbf{Z}) \times \mathbf{T} &\to \mathbf{Z} \\ (f, T) &\mapsto a_1(f|T) \end{aligned}$$

which induces an embedding

$$S(\mathbf{Z}) \hookrightarrow \mathrm{Hom}(\mathbf{T}, \mathbf{Z}) \approx \mathbf{Z}^d.$$

Remark 4.5.3. The map $S(\mathbf{Z}) \hookrightarrow \mathrm{Hom}(\mathbf{T}, \mathbf{Z})$ is an isomorphism of $\mathbf{T}$-modules, since if $\varphi \in \mathrm{Hom}(\mathbf{T}, \mathbf{Z})$, then $f = \sum_{j=1}^{d} \varphi(T_j) f_j \in S(\mathbf{Z})$ maps to φ.

4.6 The Petersson inner product

The main theorem is

Theorem 4.6.1. *The* $T_n \in \mathbf{T}(S_k)$ *are all diagonalizable over* $\mathbf{C}$.

To prove this, we note that S_k supports a nondegenerate positive definite Hermitean inner product (the Petersson inner product)

$$(f, g) \mapsto \langle f, g \rangle \in \mathbf{C}$$

such that

$$\langle f|T_n, g\rangle = \langle f, g|T_n\rangle. \tag{4.6.1}$$

We need some background facts.

Definition 4.6.2. An operator T is *normal* if it commutes with its adjoint, thus $TT^* = T^*T$.

It follows from (4.6.1) that $T_n^* = T_n$, so T_n is clearly normal.

Theorem 4.6.3. *A normal operator is diagonalizable.*

Thus each T_n is diagonalizable.

Theorem 4.6.4. *A commuting family of semisimple (=diagonalizable) operators can be simultaneously diagonalized.*

Since the T_n commute this implies S_k has a basis consisting of normalized eigenforms f. Their eigenvalues are real since

$$a_n(f)\langle f, f\rangle = \langle a_n(f)f, f\rangle = \langle f|T_n, f\rangle \tag{4.6.2}$$

$$= \langle f, a_n(f)f\rangle = \overline{a_n(f)}\langle f, f\rangle. \tag{4.6.3}$$

Proposition 4.6.5. *The coefficients a_n of the eigenforms are totally real algebraic integers.*

Proof. Theorem 4.4.1 implies that there is a basis for S_k with integer coefficients. With respect to a basis for $S_k(\mathbf{Z}) = S_k \cap \mathbf{Z}[[q]]$, the matrices of the Hecke opeators T_n all have integer entries, so their characteristic polynomials are monic polynomials with integer coefficients. The roots of these polynomials are all real (by (4.6.2)), so the roots are totally real algebraic integers. □

Let

$$\mathfrak{h} = \{x + iy : x, y \in \mathbf{R}, \text{ and } y > 0\}$$

be the upper half plane.

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbf{R})$ then $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ acts on $\mathfrak{h}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \quad z \mapsto \frac{az+b}{cz+d}.$$

Lemma 4.6.6.

$$\mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\det(\alpha)}{|cz+d|^2} y.$$

Proof. Apply the identity $\mathrm{Im}(w) = \frac{1}{2i}(w - \overline{w})$ to the left hand side and simplify. □

Proposition 4.6.7. *The volume form $\frac{dx \wedge dy}{y^2}$ is invariant under the action of*

$$\mathrm{GL}_2^+(\mathbf{R}) = \{\alpha \in \mathrm{GL}_2(\mathbf{R}) \,:\, \det(\alpha) > 0\}.$$

Proof. Differentiating $\alpha = \frac{az+b}{cz+d}$ gives

$$\begin{aligned} d\left(\frac{az+b}{cz+d}\right) &= \frac{a(cz+d)dz - c(az+b)dz}{(cz+d)^2} \\ &= \frac{(ad-bc)dz}{(cz+d)^2} \\ &= \frac{\det(\alpha)}{(cz+d)^2} dz \end{aligned}$$

Thus, under the action of α, $dz \wedge d\overline{z}$ takes on a factor of

$$\frac{\det(\alpha)^2}{(cz+d)^2(c\overline{z}+d)^2} = \Big(\frac{\det(\alpha)}{|cz+d|^2}\Big)^2.$$

□

Definition 4.6.8. The *Petersson inner product* of forms $f, g \in S_k$ is defined by

$$\langle f, g\rangle = \int_{\Gamma\backslash\mathfrak{h}} (f(z)\overline{g(z)}y^k)\frac{dx \wedge dy}{y^2},$$

where $\Gamma = \mathrm{SL}_2(\mathbf{Z})$.

Integrating over $\Gamma\backslash\mathfrak{h}$ can be taken to mean integrating over a fundamental domain for the action of $\mathfrak{h}$. Showing that the operators T_n are self-adjoint with respect to the Petersson inner product is a harder computation than one might think (see [Lan95, §III.4]).

5
Analytic Theory of Modular Curves

5.1 The Modular group

This section very closely follows Sections 1.1–1.2 of [Ser73]. We introduce the modular group $G = \mathrm{PSL}_2(\mathbf{Z})$, describe a fundamental domain for the action of G on the upper half plane, and use it to prove that G is generated by

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

5.1.1 The Upper half plane

Let

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

be the open complex upper half plane. The group

$$\mathrm{SL}_2(\mathbf{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,:\, a, b, c, d \in \mathbf{R} \text{ and } ad - bc = 1 \right\}$$

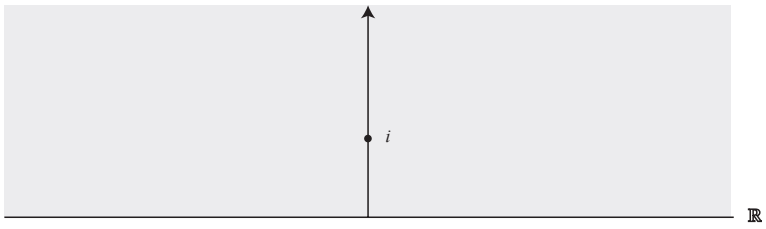


FIGURE 5.1.1. The upper half plane $\mathfrak{h}$

acts by linear fractional transformations $(z \mapsto (az+b)/(cz+d))$ on $\mathbf{C} \cup \{\infty\}$. Suppose $g \in \mathrm{SL}_2(\mathbf{R})$ and $z \in \mathfrak{h}$. Then Lemma 4.6.6 implies that $\mathrm{Im}(gz) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$, so $\mathrm{SL}_2(\mathbf{R})$ acts on $\mathfrak{h}$.

The only element of $\mathrm{SL}_2(\mathbf{R})$ that acts trivially on $\mathfrak{h}$ is -1, so

$$G = \mathrm{PSL}_2(\mathbf{Z}) = \mathrm{SL}_2(\mathbf{Z})/\langle -1 \rangle$$

acts faithfully on $\mathfrak{h}$. Let S and T be as above and note that S and T induce the linear fractional transformations $z \mapsto -1/z$ and $z \mapsto z+1$, respectively. In fact, S and T generate G.

5.2 Points on modular curves parameterize elliptic curves with extra structure

The classical theory of the Weierstass $\wp$-function sets up a bijection between isomorphism classes of elliptic curves over $\mathbf{C}$ and isomorphism classes of one-dimensional complex tori $\mathbf{C}/\Lambda$. Here Λ is a lattice in $\mathbf{C}$, i.e., a free abelian group $\mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ of rank 2 such that $\mathbf{R}\omega_1 + \mathbf{R}\omega_2 = \mathbf{C}$.

Any homomorphism φ of complex tori $\mathbf{C}/\Lambda_1 \to \mathbf{C}/\Lambda_2$ is determined by a $\mathbf{C}$-linear map $T : \mathbf{C} \to \mathbf{C}$ that sends Λ_1 into Λ_2.

Lemma 5.2.1. *Suppose $\varphi : \mathbf{C}/\Lambda_1 \to \mathbf{C}/\Lambda_2$ is nonzero. Then the kernel of φ is isomorphic to $\Lambda_2/T(\Lambda_1)$.*

Lemma 5.2.2. *Two complex tori $\mathbf{C}/\Lambda_1$ and $\mathbf{C}/\Lambda_2$ are isomorphic if and only if there is a complex number α such that $\alpha\Lambda_1 = \Lambda_2$.*

Proof. Any $\mathbf{C}$-linear map $\mathbf{C} \to \mathbf{C}$ is multiplication by a scalar $\alpha \in \mathbf{C}$. □

Suppose $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ is a lattice in $\mathbf{C}$, and let $\tau = \omega_1/\omega_2$. Then $\Lambda_\tau = \mathbf{Z}\tau + \mathbf{Z}$ defines an elliptic curve that is isomorphic to the elliptic curve determined by Λ. By replacing ω_1 by $-\omega_1$, if necessary, we may assume that $\tau \in \mathfrak{h}$. Thus every elliptic curve is of the form $E_\tau = \mathbf{C}/\Lambda_\tau$ for some $\tau \in \mathfrak{h}$ and each $\tau \in \mathfrak{h}$ determines an elliptic curve.

Proposition 5.2.3. *Suppose $\tau, \tau' \in \mathfrak{h}$. Then $E_\tau \approx E_{\tau'}$ if and only if there exists $g \in \mathrm{SL}_2(\mathbf{Z})$ such that $\tau = g(\tau')$. Thus the set of isomorphism classes of elliptic curves over $\mathbf{C}$ is in natural bijection with the orbit space $\mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{h}$.*

Proof. Suppose $E_\tau \approx E_{\tau'}$. Then there exists $\alpha \in \mathbf{C}$ such that $\alpha\Lambda_\tau = \Lambda_{\tau'}$, so $\alpha\tau = a\tau' + b$ and $\alpha 1 = c\tau' + d$ for some $a, b, c, d \in \mathbf{Z}$. The matrix $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has determinant ± 1 since $a\tau' + b$ and $c\tau' + d$ form a basis for $\mathbf{Z}\tau' + \mathbf{Z}$; this determinant is positive because $g(\tau') = \tau$ and $\tau, \tau' \in \mathfrak{h}$. Thus $\det(g) = 1$, so $g \in \mathrm{SL}_2(\mathbf{Z})$.

Conversely, suppose $\tau, \tau' \in \mathfrak{h}$ and $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ is such that

$$\tau = g(\tau') = \frac{a\tau' + b}{c\tau' + d}.$$

Let $\alpha = c\tau' + d$, so $\alpha\tau = a\tau' + b$. Since $\det(g) = 1$, the scalar α defines an isomorphism from Λ_τ to $\Lambda_{\tau'}$, so $E_\tau \approx E'_\tau$, as claimed. □

Let $E = \mathbf{C}/\Lambda$ be an elliptic curve over $\mathbf{C}$ and N a positive integer. Using Lemma 5.2.1, we see that

$$E[N] := \{x \in E \,:\, Nx = 0\} \cong \left(\frac{1}{N}\Lambda\right)/\Lambda \cong (\mathbf{Z}/N\mathbf{Z})^2.$$

If $\Lambda = \Lambda_\tau = \mathbf{Z}\tau + \mathbf{Z}$, this means that τ/N and $1/N$ are a basis for $E[N]$.

Suppose $\tau \in \mathfrak{h}$ and recall that $E_\tau = \mathbf{C}/\Lambda_\tau = \mathbf{C}/(\mathbf{Z}\tau + \mathbf{Z})$. To τ, we associate three "level N structures". First, let C_τ be the subgroup of E_τ generated by $1/N$. Second, let P_τ be the point of order N in E_τ defined by $1/N \in \frac{1}{N}\Lambda_\tau$. Third, let Q_τ be the point of order N in E_τ defined by τ/N, and consider the basis (P_τ, Q_τ) for $E[N]$.

In order to describe the third level structure, we introduce the *Weil pairing*

$$e : E[N] \times E[N] \to \mathbf{Z}/N\mathbf{Z}$$

as follows. If $E = \mathbf{C}/(\mathbf{Z}\omega_1 + \mathbf{Z}\omega_2)$ with $\tau = \omega_1/\omega_2 \in \mathfrak{h}$, and $P = a\omega_1/N + b\omega_2/N$, $Q = c\omega_1/N + d\omega_2/N$, then

$$e(P, Q) = ad - bc \in \mathbf{Z}/N\mathbf{Z}.$$

Note that e does not depend on choice of basis ω_1, ω_2 for Λ. Also if $\mathbf{C}/\Lambda \cong \mathbf{C}/\Lambda'$ via multiplication by α, and $P, Q \in (\mathbf{C}/\Lambda)[N]$, then we have $e(\alpha(P), \alpha(Q)) = e(P, Q)$, so e does not depend on the choice of Λ. In particular, P_τ and Q_τ map to P, Q via the map $E_\tau \to E$ given by multiplication by ω_2, so $e(P_\tau, Q_\tau) = -1 \in \mathbf{Z}/N\mathbf{Z}$.

Remark 5.2.4. There is a canonical Nth root of 1 in $\mathbf{C}$, namely $\zeta = e^{2\pi i/N}$. Using ζ as a canonical generator of μ_N, we can view the transcendental Weil pairing indifferently as a map with values in $\mathbf{Z}/N\mathbf{Z}$ or as a map with values in μ_N. However, for generalizations it is important to use μ_N rather than $\mathbf{Z}/N\mathbf{Z}$. There are several intrinsic algebraic definitions of the Weil pairing on N-division points for an elliptic curve (or, more generally, an abelian variety) over a field k whose characteristic is prime to N. In all cases, the Weil pairing takes values in the group of Nth roots of unity with values in the algebraic closure of k. The various definitions all coincide "up to sign" in the sense that any two of them either coincide or are inverse to each other. There are discussions of the Weil pairing in [Kat81, §5.2] and [Sil92, III.8]. The Weil pairing is bilinear, alternating, non-degenerate, Galois invariant, and maps surjectively onto μ_N.

We next consider the three modular curves $X_0(N)$, $X_1(N)$, and $X(N)$, associated to the congruence subgroups $\Gamma_0(N)$, $\Gamma_1(N)$ and $\Gamma(N)$. Recall that $\Gamma_0(N)$ is the subgroup of $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ congruent to $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$ modulo N, that $\Gamma_1(N)$ consists of matrices congruent to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ modulo N, and $\Gamma(N)$ consists of matrices congruent to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ modulo N.

Theorem 5.2.5. *Let N be a positive integer.*

1. *The non-cuspidal points on $X_0(N)$ correspond to isomorphism classes of pairs (E, C) where C is a cyclic subgroup of E of order N. (Two pairs (E, C), (E', C') are isomorphic if there is an isomorphism $\varphi : E \to E'$ such that $\varphi(C) = C'$.)*

2. *The non-cuspidal points on $X_1(N)$ correspond to isomorphism classes of pairs (E, P) where P is a point on E of exact order N. (Two pairs (E, P)*

and (E', P') are isomorphic if there is an isomorphism $\varphi : E \to E'$ such that $\varphi(P) = P'$.)

3. *The non-cuspidal points on $X(N)$ correspond to isomorphism classes of triples (E, P, Q) where P, Q is a basis for $E[N]$ such that $e(P, Q) = -1 \in \mathbf{Z}/N\mathbf{Z}$. (Triples (E, P, Q) and (E, P', Q') are isomorphic if there is an isomorphism $\varphi : E \to E'$ such that $\varphi(P) = P'$ and $\varphi(Q) = Q'$.)*

This theorem follows from Propositions 5.2.8 and 5.2.9 below, whose proofs make extensive use of the following lemma.

Lemma 5.2.6. *The reduction map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective.*

Proof. By considering the bottom two entries of an element of $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ and using the extended Euclidean algorithm, it suffices to prove that if $c, d \in \mathbf{Z}$ and $\gcd(c, d, N) = 1$, then there exists $\alpha \in \mathbf{Z}$ such that $\gcd(c, d + \alpha N) = 1$. This suffices because of $g_1 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $g_2 = \left(\begin{smallmatrix} a' & b' \\ c & d \end{smallmatrix}\right)$ are both in $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$, then a simple calculation shows that $g_1 g_2^{-1} = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$, and every element of $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ is in the image of the reduction map.

Let $c_0 = \prod_{\ell \mid \gcd(c,d)} \ell^{\mathrm{ord}_\ell(c)}$ and $c_1 = c/c_0$. We have constructed c_0 and c_1 so that $\gcd(d, c_1) = 1$ and $\gcd(c_0, c_1) = 1$; also, $\gcd(c_0, N) = 1$ since if $\ell \mid \gcd(c_0, N)$, then $\ell \mid \gcd(c, d, N) = 1$. Because $\gcd(Nc_1, c_0) = 1$, the class of Nc_1 generates the additive group $\mathbf{Z}/c_0\mathbf{Z}$, so there exists m such that $m \cdot Nc_1 \equiv 1 - d \pmod{c_0}$. Thus $d + mc_1N \equiv 1 \pmod{c_0}$ and $d + mc_1N \equiv d \pmod{c_1}$, so $d + mc_1N$ is a unit modulo both c_0 and c_1, hence $\gcd(d + mc_1N, c) = 1$, which proves the lemma. □

Remark 5.2.7. The analogue of Lemma 5.2.6 is also true for $\mathrm{SL}_m(\mathbf{Z}) \to \mathrm{SL}_m(\mathbf{Z}/N\mathbf{Z})$ for any integer m. See [Shi94, ?].

Proposition 5.2.8. *Let E be an elliptic curve over $\mathbf{C}$. If C is a cyclic subgroup of E of order N, then there exists $\tau \in \mathfrak{h}$ such that (E, C) is isomorphic to (E_τ, C_τ). If P is a point on E of order N, then there exists $\tau \in \mathbf{C}$ such that (E, P) is isomorphic to (E_τ, P_τ). If P, Q is a basis for $E[N]$ and $e(P, Q) = -1 \in \mathbf{Z}/N\mathbf{Z}$, then there exists $\tau \in \mathbf{C}$ such that (E, P, Q) is isomorphic to (E_τ, P_τ, Q_τ).*

Proof. Write $E = \mathbf{C}/\Lambda$ with $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ and $\omega_1/\omega_2 \in \mathfrak{h}$.

Suppose $P = a\omega_1/N + b\omega_2/N$ is a point of order N. Then $\gcd(a, b, N) = 1$, since otherwise P would have order strictly less than N, a contradiction. As in Lemma 5.2.6, we can modify a and b by adding multiples of N to them, so that $P = a\omega_1/N + b\omega_2/N$ and $\gcd(a, b) = 1$. There exists $c, d \in \mathbf{Z}$ such that $ad - bc = 1$, so $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$ form a basis for Λ, and C is generated by $P = \omega_1'/N$. If necessary, replace ω_2' by $-\omega_2'$ so that $\tau = \omega_2'/\omega_1' \in \mathfrak{h}$. Then (E, P) is isomorphic to (E_τ, P_τ). Also, if C is the subgroup generated by P, then (E, C) is isomorphic to (E_τ, C_τ).

Suppose $P = a\omega_1/N + b\omega_2/N$ and $Q = c\omega_1/N + d\omega_2/N$ are a basis for $E[N]$ with $e(P, Q) = -1$. Then the matrix $\left(\begin{smallmatrix} a & b \\ -c & -d \end{smallmatrix}\right)$ has determinant 1 modulo N, so because the map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective, we can replace a, b, c, d by integers that are equivalent to them modulo N (so P and Q are unchanged) so that $ad - bc = -1$. Thus $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$ form a basis for Λ. Let

$$\tau = \omega_2'/\omega_1' = \frac{c\frac{\omega_1}{\omega_2} + d}{a\frac{\omega_1}{\omega_2} + b}.$$

Then $\tau \in \mathfrak{h}$ since $\omega_1/\omega_2 \in \mathfrak{h}$ and $\left(\begin{smallmatrix} c & d \\ a & b \end{smallmatrix}\right)$ has determinant $+1$. Finally, division by ω_1' defines an isomorphism $E \to E_\tau$ that sends P to $1/N$ and Q to τ/N. □

The following proposition completes the proof of Theorem 5.2.5.

Proposition 5.2.9. *Suppose $\tau, \tau' \in \mathfrak{h}$. Then (E_τ, C_τ) is isomorphic $(E_{\tau'}, C_{\tau'})$ if and only if there exists $g \in \Gamma_0(N)$ such that $g(\tau) = \tau'$. Also, (E_τ, P_τ) is isomorphic $(E_{\tau'}, P_{\tau'})$ if and only if there exists $g \in \Gamma_1(N)$ such that $g(\tau) = \tau'$. Finally, (E_τ, P_τ, Q_τ) is isomorphic $(E_{\tau'}, P_{\tau'}, Q_{\tau'})$ if and only if there exists $g \in \Gamma(N)$ such that $g(\tau) = \tau'$.*

Proof. We prove only the first assertion, since the others are proved in a similar way. Suppose (E_τ, C_τ) is isomorphic to (E'_τ, C'_τ). Then there is $\lambda \in \mathbf{C}$ such that $\lambda\Lambda_\tau = \Lambda_{\tau'}$, and multiplication by λ sends C_τ onto C'_τ. Thus $\lambda\tau = a\tau' + b$ and $\lambda 1 = c\tau' + d$ with $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ (as we saw in the proof of Proposition 5.2.3). Dividing the second equation by N we get $\lambda\frac{1}{N} = \frac{c}{N}\tau' + \frac{d}{N}$, which lies in $\Lambda_{\tau'} = \mathbf{Z}\tau' + \frac{1}{N}\mathbf{Z}$, by hypothesis. Thus $c \equiv 0 \pmod{N}$, so $g \in \Gamma_0(N)$, as claimed. For the converse, note that if $N \mid c$, then $\frac{c}{N}\tau' + \frac{d}{N} \in \Lambda_{\tau'}$. □

5.3 The Genus of $X(N)$

Let N be a positive integer. The aim of this section is to establish some facts about modular curves associated to congruence subgroups and compute the genus of $X(N)$. Similar methods can be used to compute the genus of $X_0(N)$ and $X_1(N)$ (see Chapter 10).

The groups $\Gamma_0(1)$, $\Gamma_1(1)$, and $\Gamma(1)$ are all equal to $\mathrm{SL}_2(\mathbf{Z})$, so $X_0(1) = X_1(1) = X(1) = \mathbf{P}^1$. Since $\mathbf{P}^1$ has genus 0, we know the genus for each of these three cases. For general N we obtain the genus by determining the ramification of the corresponding cover of $\mathbf{P}^1$ and applying the Hurwitz formula, which we assume the reader is familiar with, but which we now recall.

Suppose $f : X \to Y$ is a surjective morphism of Riemann surfaces of degree d. For each point $x \in X$, let e_x be the ramification exponent at x, so $e_x = 1$ precisely when f is unramified at x, which is the case for all but finitely many x. (There is a point over $y \in Y$ that is ramified if and only if the cardinality of $f^{-1}(y)$ is less than the degree of f.) Let $g(X)$ and $g(Y)$ denote the genera of X and Y, respectively.

Theorem 5.3.1 (Hurwitz Formula)**.** *Let $f : X \to Y$ be as above. Then*

$$2g(X) - 2 = d(2g(Y) - 2) + \sum_{x \in X} (e_x - 1).$$

If $X \to Y$ is Galois, so the e_x in the fiber over each fixed $y \in Y$ are all equal to a fixed value e_y, then this formula becomes

$$2g(X) - 2 = d\left(2g(Y) - 2 + \sum_{y \in Y}\left(1 - \frac{1}{e_y}\right)\right).$$

Let X be one of the modular curves $X_0(N)$, $X_1(N)$, or $X(N)$ corresponding to a congruence subgroup Γ, and let $Y = X(1) = \mathbf{P}^1$. There is a natural map $f : X \to Y$

got by sending the equivalence class of τ modulo the congruence subgroup Γ to the equivalence class of τ modulo $\mathrm{SL}_2(\mathbf{Z})$. This is "the" map $X \to \mathbf{P}^1$ that we mean everywhere below.

Because $\mathrm{PSL}_2(\mathbf{Z})$ acts faithfully on $\mathfrak{h}$, the degree of f is the index in $\mathrm{PSL}_2(\mathbf{Z})$ of the image of Γ in $\mathrm{PSL}_2(\mathbf{Z})$. Using Lemma 5.2.6 that the map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective, we compute these indices, and obtain the following:

Proposition 5.3.2. *Suppose $N > 2$. The degree of the map $X_0(N) \to \mathbf{P}^1$ is $N\prod_{p|N}(1+1/p)$. The degree of the map $X_1(N) \to \mathbf{P}^1$ is $\frac{1}{2}N^2\prod_{p|N}(1-1/p^2)$. The degree of the map from $X(N) \to \mathbf{P}^1$ is $\frac{1}{2}N^3\prod_{p|N}(1-1/p^2)$. If $N = 2$, then the degrees are 3, 3, and 6, respectively.*

Proof. This follows from the discussion above, and the observation that for $N > 2$ the groups $\Gamma(N)$ and $\Gamma_1(N)$ do not contain -1 and the group $\Gamma_0(N)$ does. □

Proposition 5.3.3. *Let X be $X_0(N)$, $X_1(N)$ or $X(N)$. Then the map $X \to \mathbf{P}^1$ is ramified at most over ∞ and the two points corresponding to elliptic curves with extra automorphisms (i.e., the two elliptic curves with j-invariants 0 and 1728).*

Proof. Since we have a tower $X(N) \to X_1(N) \to X_0(N) \to \mathbf{P}^1$, it suffices to prove the assertion for $X = X(N)$. Since we do not claim that there is no ramification over ∞, we may restrict to $Y(N)$. By Theorem 5.2.5, the points on $Y(N)$ correspond to isomorphism classes of triples (E, P, Q), where E is an elliptic curve over $\mathbf{C}$ and P, Q are a basis for $E[N]$. The map from $Y(N)$ to $\mathbf{P}^1$ sends the isomorphism class of (E, P, Q) to the isomorphism class of E. The equivalence class of (E, P, Q) also contains $(E, -P, -Q)$, since $-1 : E \to E$ is an isomorphism. The only way the fiber over E can have cardinality smaller than the degree is if there is an extra equivalence $(E, P, Q) \to (E, \varphi(P), \varphi(Q))$ with φ an automorphism of E not equal to ± 1. The theory of CM elliptic curves shows that there are only two isomorphism classes of elliptic curves E with automorphisms other than ± 1, and these are the ones with j-invariant 0 and 1728. This proves the proposition. □

Theorem 5.3.4. *For $N > 2$, the genus of $X(N)$ is*

$$g(X(N)) = 1 + \frac{N^2(N-6)}{24}\prod_{p|N}\left(1 - \frac{1}{p^2}\right),$$

where p runs through the prime divisors of N. For $N = 1, 2$, the genus is 0.

Thus if $g_N = g(X(N))$, then $g_1 = g_2 = g_3 = g_4 = g_5 = 0$, $g_6 = 1$, $g_7 = 3$, $g_8 = 5$, $g_9 = 10$, $g_{389} = 2414816$, and $g_{2003} = 333832500$.

Proof. Since $\Gamma(N)$ is a normal subgroup of $\mathrm{SL}_2(\mathbf{Z})$, it follows that $X(N)$ is a Galois covering of $X(1) = \mathbf{P}^1$, so the ramification indices e_x are all the same for x over a fixed point $y \in \mathbf{P}^1$; we denote this common index by e_y. The fiber over the curve with j-invariant 0 has size one-third of the degree, since the automorphism group of the elliptic curve with j-invariant 0 has order 6, so the group of automorphisms modulo ± 1 has order three, hence $e_0 = 3$. Similarly, the fiber over the curve with j-invariant 1728 has size half the degree, since the automorphism group of the elliptic curve with j-invariant 1728 is cyclic of order 4, so $e_{1728} = 2$.

To compute the ramification degree e_∞ we use the orbit stabilizer theorem. The fiber of $X(N) \to X(1)$ over ∞ is exactly the set of $\Gamma(N)$ equivalence classes

of cusps, which is $\Gamma(N)\infty, \Gamma(N)g_2\infty, \ldots, \Gamma(N)g_r\infty$, where $g_1 = 1, g_2, \ldots, g_r$ are coset representatives for $\Gamma(N)$ in $\mathrm{SL}_2(\mathbf{Z})$. By the orbit-stabilizer theorem, the number of cusps equals $\#(\Gamma(1)/\Gamma(N))/\#S$, where S is the stabilizer of $\Gamma(N)\infty$ in $\Gamma(1)/\Gamma(N) \cong \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. Thus S is the subgroup $\{\pm \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : 0 \leq n < N-1\}$, which has order $2N$. Since the degree of $X(N) \to X(1)$ equals $\#(\Gamma(1)/\Gamma(N))/2$, the number of cusps is the degree divided by N. Thus $e_\infty = N$.

The Hurwitz formula for $X(N) \to X(1)$ with $e_0 = 3$, $e_{1728} = 2$, and $e_\infty = N$, is

$$2g(X(N)) - 2 = d\left(0 - 2 + \left(1 - \frac{1}{3} + 1 - \frac{1}{2} + 1 - \frac{1}{N}\right)\right),$$

where d is the degree of $X(N) \to X(1)$. Solving for $g(X(N))$ we obtain

$$2g(X(N)) - 2 = d\left(1 - \frac{5}{6} - \frac{1}{N}\right) = d\left(\frac{N-6}{6N}\right),$$

so

$$g(X(N)) = 1 + \frac{d}{2}\left(\frac{N-6}{6N}\right) = \frac{d}{12N}(N-6) + 1.$$

Substituting the formula for d from Proposition 5.3.2 yields the claimed formula. □

6
Modular Curves

6.1 Cusp Forms

Recall that if N is a positive integer we define the congruence subgroups $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$ by

$$\begin{aligned}\Gamma_0(N) &= \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : c \equiv 0 \pmod{N}\} \\ \Gamma_1(N) &= \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : a \equiv d \equiv 1, c \equiv 0 \pmod{N}\} \\ \Gamma(N) &= \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \pmod{N}\}.\end{aligned}$$

Let Γ be one of the above subgroups. One can give a construction of the space $S_k(\Gamma)$ of cusp forms of weight k for the action of Γ using the language of algebraic geometry. Let $X_\Gamma = \Gamma \backslash \mathcal{H}^*$ be the upper half plane (union the cusps) modulo the action of Γ. Then X_Γ can be given the structure of compact Riemann surface and $S_2(\Gamma) = H^0(X_\Gamma, \Omega^1)$ where Ω^1 is the sheaf of differential 1-forms on X_Γ. This is true since an element of $H^0(X_\Gamma, \Omega^1)$ is a differential form $f(z)dz$, holomorphic on $\mathcal{H}$ and the cusps, which is invariant with respect to the action of Γ. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ then

$$d(\gamma(z))/dz = (cz + d)^{-2}$$

so

$$f(\gamma(z))d(\gamma(z)) = f(z)dz$$

if and only if f satisfies the modular condition

$$f(\gamma(z)) = (cz + d)^2 f(z).$$

6.2 Modular curves

Recall from Section 5.2 that $\mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{h}$ parameterizes isomorphism classes of elliptic curves, and the other congruence subgroups also give rise to similar parameteriza-

tions. Thus $\Gamma_0(N)\backslash\mathfrak{h}$ parameterizes isomorphism classes of pairs (E, C) where E is an elliptic curve and C is a cyclic subgroup of order N, and $\Gamma_1(N)\backslash\mathcal{H}$ parameterizes isomorphism classes of pairs (E, P) where E is an elliptic curve and P is a point of exact order N.

We can specify a point of exact order N on an elliptic curve E by giving an injection $\mathbf{Z}/N\mathbf{Z} \hookrightarrow E[N]$, or equivalently, an injection $\mu_N \hookrightarrow E[N]$ where μ_N denotes the Nth roots of unity. Then $\Gamma(N)\backslash\mathfrak{h}$ parameterizes isomorphism classes of pairs $(E, \{\alpha, \beta\})$, where $\{\alpha, \beta\}$ is a basis for $E[N] \approx (\mathbf{Z}/N\mathbf{Z})^2$.

The above quotients spaces are called *moduli spaces* for the *moduli problem* of determining equivalence classes of pairs ($E+$ extra structure).

6.3 Classifying $\Gamma(N)$-structures

Definition 6.3.1. Let S be an arbitrary scheme. An **elliptic curve** E/S is a proper smooth curve

$$\begin{array}{c} E \\ {\scriptstyle f}\big\downarrow \\ S \end{array}$$

with geometrically connected fibers all of genus one, give with a section "0".

Loosely speaking, proper is a generalization of projective and smooth generalizes nonsingularity. See Hartshorne [Har77, III.10] for the precise definitions.

Definition 6.3.2. Let S be any scheme and E/S an elliptic curve. A **$\Gamma(N)$-structure** on E/S is a group homomorphism

$$\varphi : (\mathbf{Z}/N\mathbf{Z})^2 \to E[N](S)$$

whose image "generates" $E[N](S)$. (A good reference is [KM85, Ch. 3].)

Define a functor from the category of $\mathbf{Q}$-schemes to the category of sets by sending a scheme S to the set of isomorphism classes of pairs

$$(E, \Gamma(N)\text{-structure})$$

where E is an elliptic curve defined over S and isomorphisms (preserving the $\Gamma(N)$-structure) are taken over S. An isomorphism preserves the $\Gamma(N)$-structure if it takes the two distinguished generators to the two distinguished generators in the image (in the correct order).

Theorem 6.3.3. *For $N \geq 4$ the functor defined above is representable and the object representing it is the modular curve X corresponding to $\Gamma(N)$.*

What this means is that given a $\mathbf{Q}$-scheme S, the set

$$X(S) = \mathrm{Mor}_{\mathbf{Q}\text{-schemes}}(S, X)$$

is isomorphic to the image of the functor's value on S.

There is a natural way to map a pair $(E, \Gamma(N)\text{-structure})$ to an Nth root of unity. Recall from Section 5.2 that if P, Q are the distinguished basis of $E[N]$ we send the pair $(E, \Gamma(N)\text{-structure})$ to

$$e_N(P, Q) \in \mu_N$$

where $e_N : E[N] \times E[N] \to \mu_N$ is the Weil pairing.

6.4 More on integral Hecke operators

Consider the algebra of integral Hecke operators $\mathbf{T} = \mathbf{T}_{\mathbf{Z}}$ on the space of cusp forms $S_k(\mathbf{C})$ with respect to the action of the full modular group $\mathrm{SL}_2(\mathbf{Z})$. Our goal is to see why $\mathbf{T} \cong \mathbf{Z}^d$ where $d = \dim_{\mathbf{C}} S_k(\mathbf{C})$.

Suppose $A \subset \mathbf{C}$ is any *subring* of $\mathbf{C}$ and let

$$\mathbf{T}_A = A[\ldots, T_n, \ldots] \subset \mathrm{End}_{\mathbf{C}} S_k.$$

We have a natural map

$$\mathbf{T}_A \otimes_A \mathbf{C} \to \mathbf{T}_C$$

but we do not yet know that it is an isomorphism.

6.5 Complex conjugation

We have a conjugate linear map on functions

$$f(\tau) \mapsto \overline{f(-\overline{\tau})}.$$

Since $\overline{(e^{-2\pi i \overline{\tau}})} = e^{2\pi i \tau}$, it follows that under the above map,

$$\sum_{n=1}^{\infty} a_n q^n \mapsto \sum_{n=1}^{\infty} \overline{a_n} q^n,$$

so it is reasonable to call this map "complex conjugation". Furthermore, if we know that

$$S_k(\mathbf{C}) = \mathbf{C} \otimes_{\mathbf{Q}} S_k(\mathbf{Q}),$$

then it follows that complex conjugation takes $S_k(\mathbf{C})$ into $S_k(\mathbf{C})$. To see this note that if we have the above equality then every element of $S_k(\mathbf{C})$ is a $\mathbf{C}$-linear combination of elements of $S_k(\mathbf{Q})$; conversely, it is clear that the set of such $\mathbf{C}$-linear combinations is invariant under the action of complex conjugation.

6.6 Isomorphism in the real case

Proposition 6.6.1. $\mathbf{T}_{\mathbf{R}} \otimes_{\mathbf{R}} \mathbf{C} \cong \mathbf{T}_{\mathbf{C}}$, *as* $\mathbf{C}$*-vector spaces.*

Proof. Since $S_k(\mathbf{R}) = S_k(\mathbf{C}) \cap \mathbf{R}[[q]]$ and since Theorem 4.4.1 assures us that there is a $\mathbf{C}$-basis of $S_k(\mathbf{C})$ consisting of forms with integral coefficients, we see that $S_k(\mathbf{R}) \approx \mathbf{R}^d$ where $d = \dim_{\mathbf{C}} S_k(\mathbf{C})$. (Any element of $S_k(\mathbf{R})$ is a $\mathbf{C}$-linear combination of the integral basis, hence equating real and imaginary parts, an $\mathbf{R}$-linear combination of the integral basis, and the integral basis stays independent over $\mathbf{R}$.) By considering the explicit formula for the action of the Hecke operators T_n on S_k (see Section 3.5.2) we see that $\mathbf{T}_{\mathbf{R}}$ leaves $S_k(\mathbf{R})$ invariant, thus

$$\mathbf{T}_{\mathbf{R}} = \mathbf{R}[\ldots, T_d, \ldots] \subset \mathrm{End}_{\mathbf{R}} S_k(\mathbf{R}).$$

In Section 4.2 we defined a perfect pairing

$$\mathbf{T}_{\mathbf{C}} \times S_k(\mathbf{C}) \to \mathbf{C}.$$

By restricting to $\mathbf{R}$ we again obtain a perfect pairing, so we see that $\mathbf{T}_{\mathbf{R}} \approx S_k(\mathbf{R}) \approx \mathbf{R}^d$ which implies that $\mathbf{T}_R \otimes_{\mathbf{R}} \mathbf{C} \xrightarrow{\sim} \mathbf{T}_{\mathbf{C}}$. □

The above argument also shows that $S_k(\mathbf{C}) \cong S_k(\mathbf{R}) \otimes_{\mathbf{R}} \mathbf{C}$, so complex conjugation is defined over $\mathbf{R}$.

6.7 The Eichler-Shimura isomorphism

Our goal in this section is to briefly outline a homological interpretation of $S_k = S_k(\mathbf{C})$. For details see [Lan95, VI], the original paper [Shi59], or [Shi94, VIII].

How is the space $S_k(\mathbf{C})$ of cusp forms related to the cohomology group $H^1(X_\Gamma, \mathbf{R})$? Suppose $k = 2$ and $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ is a congruence subgroup, and let $X_\Gamma = \overline{\Gamma\backslash\mathcal{H}}$ be the Riemann surface obtained by compactifying the upper half plane modulo the action of Γ. Then $S_k(\mathbf{C}) = H^0(X_\Gamma, \Omega^1)$ so we have a pairing

$$H_1(X_\Gamma, \mathbf{Z}) \times S_k(\mathbf{C}) \to \mathbf{C}$$

given by integration

$$(\gamma, \omega) \mapsto \int_\gamma \omega.$$

This gives an embedding

$$\mathbf{Z}^{2d} \approx H_1(X_\Gamma, \mathbf{Z}) \hookrightarrow \mathrm{Hom}_{\mathbf{C}}(S_k(\mathbf{C}), \mathbf{C}) \cong \mathbf{C}^d$$

of a "lattice" in $\mathbf{C}^d$ (we write "lattice" because we have not shown that the image of $\mathbf{Z}^{2d}$ is really a lattice, i.e., has $\mathbf{R}$ span equal to $\mathbf{C}^d$). Passing to the quotient and compactifying, we obtain a complex torus called the Jacobian of X_Γ. Again, using the above pairing, we obtain an embedding

$$\mathbf{C}^d \approx S_k(\mathbf{C}) \hookrightarrow \mathrm{Hom}(H_1(X_\Gamma, \mathbf{Z}), \mathbf{C}) \cong \mathbf{C}^{2d}$$

which, upon taking the real part, gives

$$S_k(\mathbf{C}) \to \mathrm{Hom}(H_1(X_\Gamma, \mathbf{Z}), \mathbf{R}) \cong H^1(X_\Gamma, \mathbf{R}) \cong H^1_p(\Gamma, \mathbf{R})$$

where $H^1_p(\Gamma, \mathbf{R})$ denotes the *parabolic* group cohomology of Γ with respect to the trivial action. It is this result, that we may view $S_k(\mathbf{C})$ as the cohomology group $H^1_p(\Gamma, \mathbf{R})$, which was alluded to above.

Shimura generalized this for arbitrary $k \geq 2$, and showed that

$$S_k(\mathbf{C}) \cong H^1_p(\Gamma, V_k),$$

where V_k is a $k-1$ dimensional $\mathbf{R}$-vector space. The isomorphism is (approximately) the following: $f \in S_k(\mathbf{C})$ is sent to the cocycle map

$$\gamma \mapsto \mathrm{Re} \int_{\tau_0}^{\gamma\tau_0} f(\tau)\tau^i d\tau, \quad i = 0, \ldots, k-2.$$

Let $W = \mathbf{R} \oplus \mathbf{R}$, then Γ acts on W by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

so Γ acts on

$$V_k = \operatorname{Sym}^{k-2} W = W^{\otimes k-2}/S_{k-2}$$

where S_{k-2} is the symmetric group on $k-2$ symbols (note that $\dim V_k = k-1$). Let

$$L = H^1_p(\Gamma, \operatorname{Sym}^{k-2}(\mathbf{Z} \oplus \mathbf{Z})).$$

Under the isomorphism

$$S_k(\mathbf{C}) \cong H^1_p(\Gamma, \mathbf{R}),$$

L is a sublattice of $S_k(\mathbf{C})$ of $\mathbf{Z}$-rank $2d = 2\dim_{\mathbf{C}} S_k(\mathbf{C})$ which is T_n-stable for all n. Thus we have an embedding

$$\mathbf{T}_{\mathbf{Z}} = \mathbf{T} \hookrightarrow \operatorname{End} L,$$

so $\mathbf{T}_{\mathbf{R}} \subset \operatorname{End}_{\mathbf{R}}(L \otimes \mathbf{R})$ and $\mathbf{T}_{\mathbf{Z}} \otimes_{\mathbf{Z}} \mathbf{R} \cong \mathbf{T}_{\mathbf{R}}$, which has rank d.

7
Modular Symbols

This chapter is about how to explicitly compute the homology of modular curves using modular symbols.

We assume the reader is familiar with basic notions of algebraic topology, including homology groups of surfaces and triangulation. We also assume that the reader has read Chapter 5 about the fundamental domain for the action of $\mathrm{PSL}_2(\mathbf{Z})$ on the upper half plane and the analytic construction of modular curves.

Some standard references for modular symbols are [Man72] [Lan95, IV], [Cre97], and [Mer94]. Sections 7.1–7.2 below very closely follow Section 1 of Manin's paper [Man72].

For the rest of this chapter, let $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$ and let G be a subgroup of Γ of finite index. Note that we do not require G to be a congruence subgroup. The quotient $X(G) = G\backslash\mathfrak{h}^*$ of $\mathfrak{h}^* = \mathfrak{h} \cup \mathbf{P}^1(\mathbf{Q})$ by G has an induced structure of a compact Riemann surface. Let $\pi : \mathfrak{h}^* \to X(G)$ denote the natural projection. The matrices

$$s = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad t = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

together generate Γ; they have orders 2 and 3, respectively.

7.1 Modular symbols

Let $\mathrm{H}^0(X(G), \Omega^1)$ denote the complex vector space of holomorphic 1-forms on $X(G)$. Integration of differentials along homology classes defines a perfect pairing

$$\mathrm{H}_1(X(G), \mathbf{R}) \times \mathrm{H}^0(X(G), \Omega^1) \to \mathbf{C},$$

hence an isomorphism

$$\mathrm{H}_1(X(G), \mathbf{R}) \cong \mathrm{Hom}_{\mathbf{C}}(\mathrm{H}^0(X(G), \Omega^1), \mathbf{C}).$$

For more details, see [Lan95, §IV.1].

Given two elements $\alpha, \beta \in \mathfrak{h}^*$, integration from α to β induces a well-defined element of $\mathrm{Hom}_{\mathbf{C}}(\mathrm{H}^0(X(G), \Omega^1), \mathbf{C})$, hence an element

$$\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{R}).$$

Definition 7.1.1 (Modular symbol)**.** The homology class $\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{R})$ associated to $\alpha, \beta \in \mathfrak{h}^*$ is called the *modular symbol* attached to α and β.

Proposition 7.1.2. *The symbols $\{\alpha, \beta\}$ have the following properties:*

1. $\{\alpha, \alpha\} = 0$, $\{\alpha, \beta\} = -\{\beta, \alpha\}$, *and* $\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$.
2. $\{g\alpha, g\beta\} = \{\alpha, \beta\}$ *for all* $g \in G$
3. *If $X(G)$ has nonzero genus, then $\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{Z})$ if and only if $G\alpha = G\beta$ (i.e., we have $\pi(\alpha) = \pi(\beta)$).*

Remark 7.1.3. We only have $\{\alpha, \beta\} = \{\beta, \alpha\}$ if $\{\alpha, \beta\} = 0$, so the modular symbols notation, which suggests "unordered pairs," is actively misleading.

Proposition 7.1.4. *For any $\alpha \in \mathfrak{h}^*$, the map $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ that sends g to $\{\alpha, g\alpha\}$ is a surjective group homomorphism that does not depend on the choice of α.*

Proof. If $g, h \in G$ and $\alpha \in \mathfrak{h}^*$, then

$$\{\alpha, gh(\alpha)\} = \{\alpha, g\alpha\} + \{g\alpha, gh\alpha\} = \{\alpha, g\alpha\} + \{\alpha, h\alpha\},$$

so the map is a group homomorphism. To see that the map does not depend on the choice of α, suppose $\beta \in \mathfrak{h}^*$. By Proposition 7.1.2, we have $\{\alpha, \beta\} = \{g\alpha, g\beta\}$. Thus

$$\{\alpha, g\alpha\} + \{g\alpha, \beta\} = \{g\alpha, \beta\} + \{\beta, g\beta\},$$

so cancelling $\{g\alpha, \beta\}$ from both sides proves the claim.

The fact that the map is surjective follows from general facts from algebraic topology. Let $\mathfrak{h}^0$ be the complement of $\Gamma i \cup \Gamma\rho$ in $\mathfrak{h}$, fix $\alpha \in \mathfrak{h}^0$, and let $X(G)^0 = \pi(\mathfrak{h}^0)$. The map $\mathfrak{h}^0 \to X(G)^0$ is an unramified covering of (noncompact) Riemann surfaces with automorphism group G. Thus α determines a group homomorphism $\pi_1(X(G)^0, \pi(\alpha)) \to G$. When composed with the morphism $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ above, the composition

$$\pi_1(X(G)^0, \pi(\alpha)) \to G \to \mathrm{H}_1(X(G), \mathbf{Z})$$

is the canonical map from the fundamental group of $X(G)^0$ to the homology of the corresponding compact surface, which is surjective. This forces the map $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ to be surjective, which proves the claim. □

7.2 Manin symbols

We continue to assume that G is a finite-index subgroup of $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, so the set $G\backslash\Gamma = \{Gg_1, \ldots Gg_d\}$ of right cosets of G in Γ is finite. Manin symbols are a certain finite subset of modular symbols that are indexed by right cosets of G in Γ.

7.2.1 Using continued fractions to obtain surjectivity

Let $R = G\backslash\Gamma$ be the set of right cosets of G in Γ. Define

$$[\,] : R \to \mathrm{H}_1(X(G), \mathbf{R})$$

by $[r] = \{r0, r\infty\}$, where $r0$ means the image of 0 under any element of the coset r (it doesn't matter which). For $g \in \Gamma$, we also write $[g] = [gG]$.

Proposition 7.2.1. *Any element of* $\mathrm{H}_1(X(G), \mathbf{Z})$ *is a sum of elements of the form* $[r]$*, and the representation* $\sum n_r\{\alpha_r, \beta_r\}$ *of* $h \in \mathrm{H}_1(X(G), \mathbf{Z})$ *can be chosen so that* $\sum n_r(\pi(\beta_r) - \pi(\alpha_r)) = 0 \in \mathrm{Div}(X(G))$.

Proof. By Proposition 7.1.4, every element h of $\mathrm{H}_1(X(G), \mathbf{Z})$ is of the form $\{0, g(0)\}$ for some $g \in \mathbf{G}$. If $g(0) = \infty$, then $h = [G]$ and $\pi(\infty) = \pi(0)$, so we may assume $g(0) = a/b \neq \infty$, with a/b in lowest terms and $b > 0$. Also assume $a > 0$, since the case $a < 0$ is treated in the same way. Let

$$0 = \frac{p_{-2}}{q_{-2}} = \frac{0}{1},\ \frac{p_{-1}}{q_{-1}} = \frac{1}{0},\ \frac{p_0}{1} = \frac{p_0}{q_0},\ \frac{p_1}{q_1},\ \frac{p_2}{q_2}, \ldots,\ \frac{p_n}{q_n} = \frac{a}{b}$$

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

$$p_j q_{j-1} - p_{j-1} q_j = (-1)^{j-1} \qquad \text{for } -1 \leq j \leq n.$$

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

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

For the assertion about the divisor sum equaling zero, notice that the endpoints of the successive modular symbols cancel out, leaving the difference of 0 and $g(0)$ in the divisor group, which is 0. □

Lemma 7.2.2. *If* $x = \sum_{j=1}^{t} n_j\{\alpha_j, \beta_j\}$ *is a* $\mathbf{Z}$*-linear combination of modular symbols for* G *and* $\sum n_j(\pi(\beta_j) - \pi(\alpha_j)) = 0 \in \mathrm{Div}(X(G))$*, then* $x \in \mathrm{H}_1(X(G), \mathbf{Z})$.

Proof. We may assume that each n_j is ± 1 by allowing duplication. We may further assume that each $n_j = 1$ by using that $\{\alpha, \beta\} = -\{\beta, \alpha\}$. Next reorder the sum so $\pi(\beta_j) = \pi(\alpha_{j+1})$ by using that the divisor is 0, so every β_j must be equivalent to some $\alpha_{j'}$, etc. The lemma should now be clear. □

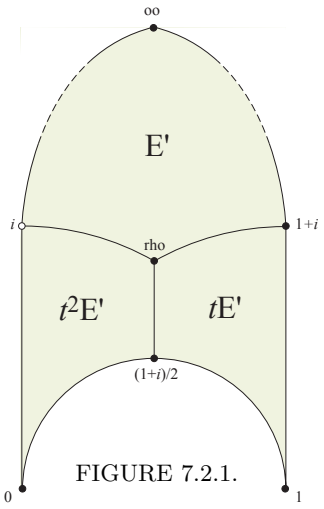


FIGURE 7.2.1.

7.2.2 Triangulating $X(G)$ to obtain injectivity

Let C be the abelian group generated by symbols (r) for $r \in G\backslash\Gamma$, subject to the relations

$$(r) + (rs) = 0, \qquad \text{and } (r) = 0 \quad if \quad r = rs.$$

For $(r) \in C$, define the boundary of (r) to be the difference $\pi(r\infty) - \pi(r0) \in \mathrm{Div}(X(G))$. Since s swaps 0 and ∞, the boundary map is a well-defined map on C. Let Z be its kernel.

Let B be the subgroup of C generated by symbols (r), for all $r \in G\backslash\Gamma$ that satisfy $r = rt$, and by $(r) + (rt) + (rt^2)$ for all other r. If $r = rt$, then $rt(0) = r(0)$, so $r(\infty) = r(0)$, so $(r) \in Z$. Also, using (7.2.1) below, we see that for any r, the element $(r) + (rt) + (rt^2)$ lies in Z.

The map $G\backslash\Gamma \to \mathrm{H}_1(X(G), \mathbf{R})$ that sends (r) to $[r]$ induces a homomorphism $C \to \mathrm{H}_1(X(G), \mathbf{R})$, so by Proposition 7.2.1 we obtain a surjective homomorphism

$$\psi : Z/B \to \mathrm{H}_1(X(G), \mathbf{Z}).$$

Theorem 7.2.3 (Manin)**.** *The map $\psi : Z/B \to \mathrm{H}_1(X(G), \mathbf{Z})$ is an isomorphism.*

Proof. We only have to prove that ψ is injective. Our proof follows the proof of [Man72, Thm. 1.9] very closely. We compute the homology $\mathrm{H}_1(X(G), \mathbf{Z})$ by triangulating $X(G)$ to obtain a simplicial complex L with homology Z_1/B_1, then embed Z/B in the homology Z_1/B_1 of $X(G)$. Most of our work is spent describing the triangulation L.

Let E denote the *interior* of the triangle with vertices 0, 1, and ∞, as illustrated in Figure 7.2.1. Let E' denote the union of the interior of the region bounded by the path from i to $\rho = e^{\pi i/3}$ to $1 + i$ to ∞ with the indicated path from i to ρ, not including the vertex i.

When reading the proof below, it will be helpful to look at the following table, which illustrates what $s = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $t = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, and t^2 do to the vertices in

Figure 7.2.1:

1	0	1	∞	i	$1+i$	$(1+i)/2$	ρ
s	∞	-1	0	i	$(-1+i)/2$	$-1+i$	$-\overline{\rho}$
t	∞	0	1	$1+i$	$(1+i)/2$	i	ρ
t^2	1	∞	0	$(1+i)/2$	i	$1+i$	ρ

(7.2.1)

Note that each of E', tE', and t^2E' is a fundamental domain for Γ, in the sense that every element of the upper half plane is conjugate to exactly one element in the closure of E' (except for identifications along the boundaries). For example, E' is obtained from the standard fundamental domain for Γ, which has vertices ρ^2, ρ, and ∞, by chopping it in half along the imaginary axis, and translating the piece on the left side horizontally by 1.

If $(0,\infty)$ is the path from 0 to ∞, then $t(0,\infty) = (\infty, 1)$ and $t^2(0,\infty) = (1,0)$. Also, $s(0,\infty) = (\infty, 0)$. Thus each half side of E is Γ-conjugate to the side from i to ∞. Also, each 1-simplex in Figure 7.2.1, i.e., the sides that connected two adjacent labeled vertices such as i and ρ, maps homeomorphically into $X(\Gamma)$. This is clear for the half sides, since they are conjugate to a path in the interior of the standard fundamental domain for Γ, and for the medians (lines from midpoints to ρ) since the path from i to ρ is on an edge of the standard fundamental domain with no self identifications.

We now describe our triangulation L of $X(G)$:

0-cells The 0 cells are the cusps $\pi(\mathbf{P}^1(\mathbf{Q}))$ and i-elliptic points $\pi(\Gamma i)$. Note that these are the images under π of the vertices and midpoints of sides of the triangles gE, for all $g \in \Gamma$.

1-cells The 1 cells are the images of the half-sides of the triangles gE, for $g \in \Gamma$, oriented from the edge to the midpoint (i.e., from the cusp to the i-elliptic point). For example, if $r = Gg$ is a right coset, then

$$e_1(r) = \pi(g(\infty), g(i)) \in X(G)$$

is a 1 cell in L. Since, as we observed above, every half side is Γ-conjugate to $e_1(G)$, it follows that every 1-cell is of the form $e(r)$ for some right coset $r \in G\backslash\Gamma$.

Next observe that if $r \neq r'$ then

$$e_1(r) = e_1(r') \qquad \text{implies} \qquad r' = rs. \tag{7.2.2}$$

Indeed, if $\pi(g(\infty), g(i)) = \pi(g'(\infty), g'(i))$, then $ri = r'i$ (note that the endpoints of a path are part of the definition of the path). Thus there exists $h, h' \in G$ such that $hg(i) = h'g'(i)$. Since the only nontrivial element of Γ that stabilizes i is s, this implies that $(hg)^{-1}h'g' = s$. Thus $h'g' = hgs$, so $Gg' = Ggs$, so $r' = rs$.

2-cells There are two types of 2-cells, those with 2 sides and those with 3.

2-sided: The 2-sided 2-cells $e_2(r)$ are indexed by the cosets $r = Gg$ such that $rt = r$. Note that for such an r, we have $\pi(rE') = \pi(rtE') = \pi(rt^2E')$. The 2-cell $e_2(r)$ is $\pi(gE')$. The image $g(\rho, i)$ of the half median maps to a

line from the center of $e_2(r)$ to the edge $\pi(g(i)) = \pi(g(1+i))$. Orient $e_2(r)$ in a way compatible with the e_1. Since $Ggt = Gg$,

$$\pi(g(1+i), g(\infty)) = \pi(gt^2(1+i), gt^2(\infty)) = \pi(g(i), g(0)) = \pi(gs(i), gs(\infty)),$$

so

$$e_1(r) - e_1(rs) = \pi(g(\infty), g(i)) + \pi(gs(i), gs(\infty)) = \pi(g(\infty), g(i)) + \pi(g(1+i), g(\infty)).$$

Thus

$$\partial e_2(r) = e_1(r) - e_1(rs).$$

Finally, note that if $r' \neq r$ also satisfies $r't = r'$, then $e_2(r) \neq e_2(r')$ (to see this use that E' is a fundamental domain for Γ).

3-sided: The 3-sided 2-cells $e_2(r)$ are indexed by the cosets $r = Gg$ such that $rt \neq r$. Note that for such an r, the three triangles rE', rtE', and rt^2E' are distinct (since they are nontrivial translates of a fundamental domain). Orient $e_2(r)$ in a way compatible with the e_1 (so edges go from cusps to midpoints). Then

$$\partial e_2(r) = \sum_{n=0}^{2} \left(e_1(rt^n) - e_1(rt^n s)\right).$$

We have now defined a complex L that is a triangulation of $X(G)$. Let C_1, Z_1, and B_1 be the group of 1-chains, 1-cycles, and 1-boundaries of the complex L. Thus C_1 is the abelian group generated by the paths $e_1(r)$, the subgroup Z_1 is the kernel of the map that sends $e_1(r) = \pi(r(\infty), r(0))$ to $\pi(r(0)) - \pi((\infty))$, and B_1 is the subgroup of Z_1 generated by boundaries of 2-cycles.

Let C, Z, B be as defined before the statement of the Theorem 7.2.3. We have $\mathrm{H}_1(X(G), \mathbf{Z}) \cong Z_1/B_1$, and would like to prove that $Z/B \cong Z_1/B_1$.

Define a map $\varphi : C \to C_1$ by $(r) \mapsto e_1(rs) - e_1(r)$. The map φ is well defined because if $r = rs$, then clearly $(r) \mapsto 0$, and $(r) + (rs)$ maps to $e_1(rs) - e_1(r) + e_1(r) - e_1(rs) = 0$. To see that f is injective, suppose $\sum n_r(r) \neq 0$. Since in C we have the relations $(r) = -(rs)$ and $(r) = 0$ if $rs = r$, we may assume that $n_r n_{rs} = 0$ for all r. We have

$$\varphi\left(\sum n_r(r)\right) = \sum n_r(e_1(rs) - e_1(r)).$$

If $n_r \neq 0$ then $r \neq rs$, so (7.2.2) implies that $e_1(r) \neq e_1(rs)$. If $n_r \neq 0$ and $n_{r'} \neq 0$ with $r' \neq r$, then $r \neq rs$ and $r' \neq r's$, so $e_1(r), e_1(rs), e_1(r'), e_1(r's)$ are all distinct. We conclude that $\sum n_r(e_1(rs) - e_1(r)) \neq 0$, which proves that φ is injective.

Suppose $(r) \in C$. Then

$$\varphi(r) + B_1 = \psi(r) = \{r(0), r(\infty)\} \in \mathrm{H}_1(X(G), \mathbf{Z}) = C_1/B_1,$$

since

$$\varphi(r) = e_1(rs) - e_1(r) = \pi(rs(\infty), rs(i)) - \pi(r(\infty), r(i)) = \pi(r(0), r(i)) - \pi(r(\infty), r(i))$$

belongs to the homology class $\{r(0), r(\infty)\}$. Extending linearly, we have, for any $z \in C$, that $\varphi(z) + B_1 = \psi(z)$.

The generators for B_1 are the boundaries of 2-cells $e_2(r)$. As we saw above, these have the form $\varphi(r)$ for all r such that $r = rt$, and $\varphi(r) + \varphi(rt) + \varphi(rt^2)$ for the r such that $rt \neq r$. Thus $B_1 = \varphi(B) \subset \varphi(Z)$, so the map φ induces an injection $Z/B \hookrightarrow Z_1/B_1$. This completes the proof of the theorem.

□

7.3 Hecke operators

In this section we will only consider the modular curve $X_0(N)$ associated to the subgroup $\Gamma_0(N)$ of $\mathrm{SL}_2(\mathbf{Z})$ of matrices that are upper triangular modulo N. Much of what we say will also be true, possibly with slight modification, for $X_1(N)$, but not for arbitrary finite-index subgroups.

There is a commutative ring

$$\mathbf{T} = \mathbf{Z}[T_1, T_2, T_3, \ldots]$$

of *Hecke operators* that acts on $\mathrm{H}_1(X_0(N), \mathbf{R})$. We will frequently revisit this ring, which also acts on the Jacobian $J_0(N)$ of $X_0(N)$, and on modular forms. The ring $\mathbf{T}$ is generated by T_p, for p prime, and as a free $\mathbf{Z}$-module $\mathbf{T}$ is isomorphic to $\mathbf{Z}^g$, where g is the genus of $X_0(N)$. We will not prove these facts here (see [1]). 1

Suppose

$$\{\alpha, \beta\} \in \mathrm{H}_1(X_0(N), \mathbf{R}),$$

is a modular symbol, with $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$. For $g \in \mathrm{M}_2(\mathbf{Z})$, write $g(\{\alpha, \beta\}) = \{g(\alpha), g(\beta)\}$. This is **not** a well-defined action of $\mathrm{M}_2(\mathbf{Z})$ on $\mathrm{H}_1(X_0(N), \mathbf{R})$, since $\{\alpha', \beta'\} = \{\alpha, \beta\} \in \mathrm{H}_1(X_0(N), \mathbf{R})$ does not imply that $\{g(\alpha'), g(\beta')\} = \{g(\alpha), g(\beta)\}$.

Example 7.3.1. Using MAGMA we see that the homology $\mathrm{H}_1(X_0(11), \mathbf{R})$ is generated by $\{-1/7, 0\}$ and $\{-1/5, 0\}$.

```
> M := ModularSymbols(11);   // Homology relative to cusps,
                             // with Q coefficients.
> S := CuspidalSubspace(M);  // Homology, with Q coefficients.
> Basis(S);
[ {-1/7, 0}, {-1/5, 0} ]
```

Also, we have $5\{0, \infty\} = \{-1/5, 0\}$.

```
> pi := ProjectionMap(S);    // The natural map M --> S.
> M.3;
{oo, 0}
> pi(M.3);
-1/5*{-1/5, 0}
```

Let $g = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $5\{g(0), g(\infty)\}$ is not equal to $\{g(-1/5), g(0)\}$, so g does not define a well-defined map on $\mathrm{H}_1(X_0(11), \mathbf{R})$.

```
> x := 5*pi(M!<1,[Cusps()|0,Infinity()]>);
> y := pi(M!<1,[-2/5,0]>);
> x;
```

[1] Add some references and pointers to other parts of this book.

```
{-1/5, 0}
> y;
-1*{-1/7, 0} + -1*{-1/5, 0}
> x eq y;
false
```

Definition 7.3.2 (Hecke operators)**.** We define the *Hecke operator* T_p on $\mathrm{H}_1(X_0(N), \mathbf{R})$ as follows. When p is a prime with $p \nmid N$, we have

$$T_p(\{\alpha,\beta\}) = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}(\{\alpha,\beta\}) + \sum_{r=0}^{p-1} \begin{pmatrix} 1 & r \\ 0 & p \end{pmatrix}(\{\alpha,\beta\}).$$

When $p \mid N$, the formula is the same, except that the first summand, which involves $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$, is omitted.

Example 7.3.3. We continue with Example 7.3.1. If we apply the Hecke operator T_2 to both $5\{0,\infty\}$ and $\{-1/5, 0\}$, the "non-well-definedness" cancels out.

```
> x := 5*pi(M!<1,[Cusps()|0,Infinity()]> +
     M!<1,[Cusps()|0,Infinity()]> + M!<1,[Cusps()|1/2,Infinity()]>);
> x;
-2*{-1/5, 0}
> y := pi(M!<1,[-2/5,0]>+ M!<1,[-1/10,0]> + M!<1,[2/5,1/2]>);
> y;
-2*{-1/5, 0}
```

Examples 7.3.1 shows that it is not clear that the definition of T_p given above makes sense. For example, if $\{\alpha,\beta\}$ is replaced by an equivalent modular symbol $\{\alpha',\beta'\}$, why does the formula for T_p give the same answer? We will not address this question further here, but will revisit it later[2] when we have a more natural and intrinsic definition of Hecke operators. We only remark that T_p is induced by a "correspondence" from $X_0(N)$ to $X_0(N)$, so T_p preserve $\mathrm{H}_1(X_0(N), \mathbf{Z})$. 2

7.4 Modular symbols and rational homology

In this section we sketch a beautiful proof, due to Manin, of a result that is crucial to our understanding of rationality properties of special values of L-functions. For example, Mazur and Swinnerton-Dyer write in [MSD74, §6], "The modular symbol is essential for our theory of p-adic Mellin transforms," right before discussing this rationality result. Also, as we will see in the next section, this result implies that if E is an elliptic curve over $\mathbf{Q}$, then $L(E,1)/\Omega_E \in \mathbf{Q}$, which confirms a consequence of the Birch and Swinnerton-Dyer conjecture.

Theorem 7.4.1 (Manin)**.** *For any $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, we have*

$$\{\alpha,\beta\} \in \mathrm{H}_1(X_0(N), \mathbf{Q}).$$

Proof (sketch). Since $\{\alpha,\beta\} = \{\alpha,\infty\} - \{\beta,\infty\}$, it suffices to show that $\{\alpha,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Q})$ for all $\alpha \in \mathbf{Q}$. We content ourselves with proving that $\{0,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$, since the proof for general $\{0,\alpha\}$ is almost the same.

[2] Say where, when I write this later.

We will use that the eigenvalues of T_p on $\mathrm{H}_1(X_0(N), \mathbf{R})$ have absolute value bounded by $2\sqrt{p}$, a fact that was proved by Weil (the Riemann hypothesis for curves over finite fields). Let $p \nmid N$ be a prime. Then

$$T_p(\{0,\infty\}) = \{0,\infty\} + \sum_{r=0}^{p-1} \left\{\frac{r}{p}, \infty\right\} = (1+p)\{0,\infty\} + \sum_{r=0}^{p-1} \left\{\frac{r}{p}, 0\right\},$$

so

$$(1 + p - T_p)(\{0,\infty\}) = \sum_{r=0}^{p-1} \left\{0, \frac{r}{p}\right\}.$$

Since $p \nmid N$, the cusps 0 and r/p are equivalent (use the Euclidean algorithm to find a matrix in $\mathrm{SL}_2(\mathbf{Z})$ of the form $\left(\begin{smallmatrix} r & * \\ p & * \end{smallmatrix}\right)$), so the modular symbols $\{0, r/p\}$, for $r = 0, 1, \ldots, p-1$ all lie in $\mathrm{H}_1(X_0(N), \mathbf{Z})$. Since the eigenvalues of T_p have absolute value at most $2\sqrt{p}$, the linear transformation $1 + p - T_p$ of $\mathrm{H}_1(X_0(N), \mathbf{Z})$ is invertible. It follows that some integer multiple of $\{0,\infty\}$ lies in $\mathrm{H}_1(X_0(N), \mathbf{Z})$, as claimed. □

There are general theorems about the denominator of $\{\alpha, \beta\}$ in some cases. Example 7.3.1 above demonstrated the following theorem in the case $N = 11$.

Theorem 7.4.2 (Ogg [Ogg71])**.** *Let N be a prime. Then the image*

$$[\{0,\infty\}] \in \mathrm{H}_1(X_0(N), \mathbf{Q}) / \mathrm{H}_1(X_0(N), \mathbf{Z})$$

has order equal to the numerator of $(N-1)/12$.

7.5 Special values of L-functions

This section is a preview of one of the central arithmetic results we will discuss in more generality later in this book.[3]

The celebrated modularity theorem of Wiles et al. asserts that there is a correspondence between isogeny classes of elliptic curves E of conductor N and normalized new modular eigenforms $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ with $a_n \in \mathbf{Z}$. This correspondence is characterized by the fact that for all primes $p \nmid N$, we have $a_p = p + 1 - \#E(\mathbf{F}_p)$.

Recall that a modular form for $\Gamma_0(N)$ of weight 2 is a holomorphic function $f : \mathfrak{h} \to \mathbf{C}$ that is "holomorphic at the cusps" and such that for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$,

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z).$$

Suppose E is an elliptic curve that corresponds to a modular form f. If $L(E, s)$ is the L-function attached to E, then

$$L(E, s) = L(f, s) = \sum \frac{a_n}{n^s},$$

so, by a theorem of Hecke which we will prove [later][4] , $L(f, s)$ is holomorphic on

[3] Where?

[4] where?

all $\mathbf{C}$. Note that $L(f,s)$ is the Mellin transform of the modular form f:

$$L(f,s) = (2\pi)^s \Gamma(s)^{-1} \int_0^{i\infty} (-iz)^s f(z) \frac{dz}{z}. \tag{7.5.1}$$

The Birch and Swinnerton-Dyer conjecture concerns the leading coefficient of the series expansion of $L(E,s)$ about $s=1$. A special case is that if $L(E,1) \neq 0$, then

$$\frac{L(E,1)}{\Omega_E} = \frac{\prod c_p \cdot \#Ш(E)}{\#E(\mathbf{Q})^2_{\mathrm{tor}}}.$$

Here $\Omega_E = |\int_{E(\mathbf{R})} \omega|$, where ω is a "Néron" differential 1-form on E, i.e., a generator for $\mathrm{H}^0(\mathcal{E}, \Omega^1_{\mathcal{E}/\mathbf{Z}})$, where $\mathcal{E}$ is the Néron model of E. (The Néron model of E is the unique, up to unique isomorphism, smooth group scheme $\mathcal{E}$ over $\mathbf{Z}$, with generic fiber E, such that for all smooth schemes S over $\mathbf{Z}$, the natural map $\mathrm{Hom}_{\mathbf{Z}}(S,\mathcal{E}) \to \mathrm{Hom}_{\mathbf{Q}}(S \times \mathrm{Spec}(\mathbf{Q}), E)$ is an isomorphism.) In particular, the conjecture asserts that for any elliptic curve E we have $L(E,1)/\Omega_E \in \mathbf{Q}$.

Theorem 7.5.1. *Let E be an elliptic curve over $\mathbf{Q}$. Then $L(E,1)/\Omega_E \in \mathbf{Q}$.*

Proof (sketch). By the modularity theorem of Wiles et al., E is modular, so there is a surjective morphism $\pi_E : X_0(N) \to E$, where N is the conductor of E. This implies that there is a newform f that corresponds to (the isogeny class of) E, with $L(f,s) = L(E,s)$. Also assume, without loss of generality, that E is "optimal" in its isogeny class, which means that if $X_0(N) \to E' \to E$ is a sequence of morphism whose composition is π_E and E' is an elliptic curve, then $E' = E$.

By Equation 7.5.1, we have

$$L(E,1) = 2\pi \int_0^{i\infty} -izf(z)dz/z. \tag{7.5.2}$$

If $q = e^{2\pi i z}$, then $dq = 2\pi i q dz$, so $2\pi i f(z)dz = dq/q$, and (7.5.2) becomes

$$L(E,1) = -\int_0^{i\infty} f(q)dq.$$

Recall that $\Omega_E = |\int_{E(\mathbf{R})} \omega|$, where ω is a Néron differential on E. The expression $f(q)dq$ defines a differential on the modular curve $X_0(N)$, and there is a rational number c, the *Manin constant*, such that $\pi_E^* \omega = c f(q)dq$. More is true: Edixhoven proved (as did Ofer Gabber) that $c \in \mathbf{Z}$; also Manin conjectured that $c = 1$ and Edixhoven proved (unpublished) that if $p \mid c$, then $p = 2, 3, 5, 7$.

A standard fact is that if

$$\mathcal{L} = \left\{ \int_\gamma \omega \,:\, \gamma \in \mathrm{H}_1(E, \mathbf{Z}) \right\}$$

is the period lattice of E associated to ω, then $E(\mathbf{C}) \cong \mathbf{C}/\mathcal{L}$. Note that Ω_E is either the least positive real element of $\mathcal{L}$ or twice this least positive element (if $E(\mathbf{R})$ has two real components).

The next crucial observation is that by Theorem 7.4.1, there is an integer n such that $n\{0,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$. This is relevant because if

$$\mathcal{L}' = \left\{ \int_\gamma f(q)dq \,:\, \gamma \in \mathrm{H}_1(X_0(N), \mathbf{Z}) \right\} \subset \mathbf{C}.$$

then $\mathcal{L} = \frac{1}{c}\mathcal{L}' \subset \mathcal{L}'$. This assertion follows from our hypothesis that E is optimal and standard facts about complex tori and Jacobians, which we will prove later [in this course/book].

One can show that $L(E,1) \in \mathbf{R}$, for example, by writing down an explicit real convergent series that converges to $L(E,1)$. This series is used in algorithms to compute $L(E,1)$, and the derivation of the series uses properties of modular forms that we have not yet developed. Another approach is to use complex conjugation to define an involution $*$ on $\mathrm{H}_1(X_0(N), \mathbf{R})$, then observe that $\{0,\infty\}$ is fixed by $*$. (The involution $*$ is given on modular symbols by $*\{\alpha,\beta\} = \{-\alpha,-\beta\}$.)

Since $L(E,1) \in \mathbf{R}$, the integral

$$\int_{n\{0,\infty\}} f(q)dq = n\int_0^{i\infty} f(q)dq = -nL(E,1) \in \mathcal{L}'$$

lies in the subgroup $(\mathcal{L}')^+$ of elements fixed by complex conjugation. If c is the Manin constant, we have $cnL(E,1) \in \mathcal{L}^+$. Since Ω_E is the least nonzero element of $\mathcal{L}^+$ (or twice it), it follows that $2cnL(E,1)/\Omega_E \in \mathbf{Z}$, which proves the proposition. □

8
Modular Forms of Higher Level

8.1 Modular Forms on $\Gamma_1(N)$

Fix integers $k \geq 0$ and $N \geq 1$. Recall that $\Gamma_1(N)$ is the subgroup of elements of $\mathrm{SL}_2(\mathbf{Z})$ that are of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ when reduced modulo N.

Definition 8.1.1 (Modular Forms)**.** The space of *modular forms* of level N and weight k is

$$M_k(\Gamma_1(N)) = \left\{ f : f(\gamma\tau) = (c\tau + d)^k f(\tau) \text{ all } \gamma \in \Gamma_1(N) \right\},$$

where the f are assumed holomorphic on $\mathfrak{h} \cup \{\text{cusps}\}$ (see below for the precise meaning of this). The space of *cusp forms* of level N and weight k is the subspace $S_k(\Gamma_1(N))$ of $M_k(\Gamma_1(N))$ of modular forms that vanish at all cusps.

Remark 8.1.2. In the beginning of this book (e.g., Section 1.3) we often wrote $S_k(N)$ for $S_k(\Gamma_1(N))$. Since there are many congruence subgroups, to avoid confusion we will write out $S_k(\Gamma_1(N))$ in the rest of this book.

Suppose $f \in M_k(\Gamma_1(N))$. The group $\Gamma_1(N)$ contains the matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, so

$$f(z+1) = f(z),$$

and for f to be holomorphic at infinity means that f has a Fourier expansion

$$f = \sum_{n=0}^{\infty} a_n q^n.$$

To explain what it means for f to be holomorphic at all cusps, we introduce some additional notation. For $\alpha \in \mathrm{GL}_2^+(\mathbf{R})$ and $f : \mathfrak{h} \to \mathbf{C}$ define another function $f_{|[\alpha]_k}$ as follows:

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

It is straightforward to check that $f_{|[\alpha\alpha']_k} = (f_{|[\alpha]_k})_{|[\alpha']_k}$. Note that we do not have to make sense of $f_{|[\alpha]_k}(\infty)$, since we only assume that f is a function on $\mathfrak{h}$ and not $\mathfrak{h}^*$.

Using our new notation, the transformation condition required for $f : \mathfrak{h} \to \mathbf{C}$ to be a modular form for $\Gamma_1(N)$ of weight k is simply that f be fixed by the $[\ ]_k$-action of $\Gamma_1(N)$. Suppose $x \in \mathbf{P}^1(\mathbf{Q})$ is a cusp, and choose $\alpha \in \mathrm{SL}_2(\mathbf{Z})$ such that $\alpha(\infty) = x$. Then $g = f_{|[\alpha]_k}$ is fixed by the $[\ ]_k$ action of $\alpha^{-1}\Gamma_1(N)\alpha$.

Lemma 8.1.3. *Let $\alpha \in \mathrm{SL}_2(\mathbf{Z})$. Then there exists a positive integer h such that $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \alpha^{-1}\Gamma_1(N)\alpha$.*

Proof. This follows from the general fact that the set of congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$ is closed under conjugation by elements $\alpha \in \mathrm{SL}_2(\mathbf{Z})$, and every congruence subgroup contains an element of the form $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$. If G is a congruence subgroup, then $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma(N) \subset G$ for some N, and $\alpha^{-1}\Gamma(N)\alpha = \Gamma(N)$, since $\Gamma(N)$ is normal, so $\Gamma(N) \subset \alpha^{-1}G\alpha$. □

Letting h be as in the lemma, we have $g(z + h) = g(z)$. Then the condition that f be holomorphic at the cusp x is that

$$g(z) = \sum_{n \geq 0} b_{n/h} q^{1/h}$$

on the upper half plane. We say that f vanishes at x if $b_0 = 0$, and a *cusp form* is a form that vanishes at every cusp.

8.2 Diamond bracket and Hecke operators

In this section we consider the spaces of modular forms $S_k(\Gamma_1(N), \varepsilon)$, for Dirichlet characters ε mod N, and explicitly describe the action of the Hecke operators on these spaces.

8.2.1 Diamond bracket operators

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 $(\mathbf{Z}/N\mathbf{Z})^*$. From this structure we obtain an action of $(\mathbf{Z}/N\mathbf{Z})^*$ on $S_k(\Gamma_1(N))$, and use it to decompose $S_k(\Gamma_1(N))$ as a direct sum of more manageable chunks $S_k(\Gamma_1(N), \varepsilon)$.

Definition 8.2.1 (Dirichlet character)**.** A *Dirichlet character* ε modulo N is a homomorphism

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

We extend ε to a map $\varepsilon : \mathbf{Z} \to \mathbf{C}$ by setting $\varepsilon(m) = 0$ if $(m, N) \neq 1$ and $\varepsilon(m) = \varepsilon(m \bmod N)$ otherwise. If $\varepsilon : \mathbf{Z} \to \mathbf{C}$ is a Dirichlet character, the *conductor* of ε is the smallest positive integer c such that ε arises from a homomorphism $(\mathbf{Z}/c\mathbf{Z})^* \to \mathbf{C}^*$.

Remarks 8.2.2.

1. If ε is a Dirichlet character modulo N and M is a multiple of N then ε induces a Dirichlet character mod M. If M is a divisor of N then ε is induced by a Dirichlet character modulo M if and only if M divides the conductor of ε.

2. The set of Dirichlet characters forms a group, which is non-canonically isomorphic to $(\mathbf{Z}/N\mathbf{Z})^*$ (it is the dual of this group).

3. The mod N Dirichlet characters all take values in $\mathbf{Q}(e^{2\pi i/e})$ where e is the exponent of $(\mathbf{Z}/N\mathbf{Z})^*$. When N is an odd prime power, the group $(\mathbf{Z}/N\mathbf{Z})^*$ is cyclic, so $e = \varphi(\varphi(N))$. This double-φ can sometimes cause confusion.

4. There are many ways to represent Dirichlet characters with a computer (see, e.g., [Ste07a, Ch. 4]). One way is to fix generators for $(\mathbf{Z}/N\mathbf{Z})^*$ in any way you like and represent ε by the images of each of these generators. Assume for the moment that N is odd. To make the representation more "canonical", reduce to the prime power case by writing $(\mathbf{Z}/N\mathbf{Z})^*$ as a product of cyclic groups corresponding to prime divisors of N. A "canonical" generator for $(\mathbf{Z}/p^r\mathbf{Z})^*$ is then the smallest positive integer s such that $s \bmod p^r$ generates $(\mathbf{Z}/p^r\mathbf{Z})^*$. Store the character that sends s to $e^{2\pi i n/\varphi(\varphi(p^r))}$ by storing the integer n. For general N, store the list of integers n_p, one p for each prime divisor of N (unless $p = 2$, in which case you store two integers n_2 and n_2', where $n_2 \in \{0, 1\}$).

Definition 8.2.3. Let $\overline{d} \in (\mathbf{Z}/N\mathbf{Z})^*$ and $f \in S_k(\Gamma_1(N))$. By Lemma 5.2.6, the map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective, so there exists a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ such that $d \equiv \overline{d} \pmod N$. The *diamond bracket d operator* is then

$$f(\tau)|\langle d\rangle = f_{|[\gamma]_k} = f(\gamma\tau)(c\tau + d)^{-k}.$$

The definition of $\langle d\rangle$ does not depend on the choice of lift matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, since any two lifts differ by an element of $\Gamma(N)$ and f is fixed by $\Gamma(N)$ since it is fixed by $\Gamma_1(N)$.

For each Dirichlet character ε mod N let

$$\begin{aligned} S_k(\Gamma_1(N), \varepsilon) &= \{f : f|\langle d\rangle = \varepsilon(d)f \text{ all } d \in (\mathbf{Z}/N\mathbf{Z})^*\} \\ &= \{f : f_{|[\gamma]_k} = \varepsilon(d_\gamma)f \text{ all } \gamma \in \Gamma_0(N)\}, \end{aligned}$$

where d_γ is the lower-right entry of γ.

When $f \in S_k(\Gamma_1(N), \varepsilon)$, we say that f has *Dirichlet character* ε. In the literature, sometimes f is said to be of "nebentypus" ε.

Lemma 8.2.4. *The operator $\langle d\rangle$ on the finite-dimensional vector space $S_k(\Gamma_1(N))$ is diagonalizable.*

Proof. There exists $n \geq 1$ such that $I = \langle 1\rangle = \langle d^n\rangle = \langle d\rangle^n$, so the characteristic polynomial of $\langle d\rangle$ divides the square-free polynomial $X^n - 1$. □

Note that $S_k(\Gamma_1(N), \varepsilon)$ is the $\varepsilon(d)$ eigenspace of $\langle d\rangle$. Thus we have a direct sum decomposition

$$S_k(\Gamma_1(N)) = \bigoplus_{\varepsilon:(\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*} S_k(\Gamma_1(N), \varepsilon).$$

We have $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \Gamma_0(N)$, so if $f \in S_k(\Gamma_1(N), \varepsilon)$, then

$$f(\tau)(-1)^{-k} = \varepsilon(-1) f(\tau).$$

Thus $S_k(\Gamma_1(N), \varepsilon) = 0$, unless $\varepsilon(-1) = (-1)^k$, so about half of the direct summands $S_k(\Gamma_1(N), \varepsilon)$ vanish.

8.2.2 Hecke operators on q-expansions

Suppose

$$f = \sum_{n=1}^{\infty} a_n q^n \in S_k(\Gamma_1(N), \varepsilon),$$

and let p be a prime. Then

$$f|T_p = \begin{cases} \displaystyle\sum_{n=1}^{\infty} a_{np} q^n + p^{k-1}\varepsilon(p) \sum_{n=1}^{\infty} a_n q^{pn}, & p \nmid N \\ \displaystyle\sum_{n=1}^{\infty} a_{np} q^n + 0. & p \mid N. \end{cases}$$

Note that $\varepsilon(p) = 0$ when $p \mid N$, so the second part of the formula is redundant.

When $p \mid N$, T_p is often denoted U_p in the literature, but we will not do so here. Also, the ring $\mathbf{T}$ generated by the Hecke operators is commutative, so it is harmless, though potentially confusing, to write $T_p(f)$ instead of $f|T_p$.

We record the relations

$$T_m T_n = T_{mn}, \quad (m, n) = 1,$$
$$T_{p^k} = \begin{cases} (T_p)^k, & p \mid N \\ T_{p^{k-1}} T_p - \varepsilon(p) p^{k-1} T_{p^{k-2}}, & p \nmid N. \end{cases}$$

WARNING: When $p \mid N$, the operator T_p on $S_k(\Gamma_1(N), \varepsilon)$ need not be diagonalizable.

8.3 Old and new subspaces

Let M and N be positive integers such that $M \mid N$ and let $t \mid \frac{N}{M}$. If $f(\tau) \in S_k(\Gamma_1(M))$ then $f(t\tau) \in S_k(\Gamma_1(N))$. We thus have maps

$$S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N))$$

for each divisor $t \mid \frac{N}{M}$. Combining these gives a map

$$\varphi_M : \bigoplus_{t \mid (N/M)} S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)).$$

Definition 8.3.1 (Old Subspace)**.** The *old subspace* of $S_k(\Gamma_1(N))$ is the subspace generated by the images of the φ_M for all $M \mid N$ with $M \neq N$.

Definition 8.3.2 (New Subspace)**.** The *new subspace* of $S_k(\Gamma_1(N))$ is the complement of the old subspace with respect to the Petersson inner product (see Section 4.6).

[[TODO: We only defined the Petersson inner product in Section 4.6 in the case when $N = 1$.]]

Definition 8.3.3 (Newform)**.** A *newform* is an element f of the new subspace of $S_k(\Gamma_1(N))$ that is an eigenvector for every Hecke operator, which is normalized so that the coefficient of q in f is 1.

If $f = \sum a_n q^n$ is a newform then the coefficients a_n are algebraic integers, which have deep arithmetic significance. For example, when f has weight 2, there is an associated abelian variety A_f over $\mathbf{Q}$ of dimension $[\mathbf{Q}(a_1, a_2, \ldots) : \mathbf{Q}]$ such that $\prod L(f^\sigma, s) = L(A_f, s)$, where the product is over the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of f. The abelian variety A_f was constructed by Shimura as follows. Let $J_1(N)$ be the Jacobian of the modular curve $X_1(N)$. As we will see, the ring $\mathbf{T}$ of Hecke operators acts naturally on $J_1(N)$. Let I_f be the kernel of the homomorphism $\mathbf{T} \to \mathbf{Z}[a_1, a_2, \ldots]$ that sends T_n to a_n. Then

$$A_f = J_1(N)/I_f J_1(N).$$

In the converse direction, it is a deep theorem of Breuil, Conrad, Diamond, Taylor, and Wiles that if E is any elliptic curve over $\mathbf{Q}$, then E is isogenous to A_f for some f of level equal to the conductor N of E. More generally, building on this work, it is now known that if A is a simple abelian variety over $\mathbf{Q}$ with $\mathrm{End}(A) \otimes \mathbf{Q}$ a number field of degree $\dim(A)$, then A is isogenous to A_f for some newform f (see [Rib92, KW08]).

When f has weight greater than 2, in [Sch90] Scholl constructs, in an analogous way, a Grothendieck motive $\mathcal{M}_f$ attached to f.

9
Newforms and Euler Products

In this chapter we discuss the work of Atkin, Lehner, and W. Li on newforms and their associated L-series and Euler products. Then we discuss explicitly how U_p, for $p \mid N$, acts on old forms, and how U_p can fail to be diagonalizable. Then we describe a canonical generator for $S_k(\Gamma_1(N))$ as a free module over $\mathbf{T}_{\mathbf{C}}$. Finally, we observe that the subalgebra of $\mathbf{T}_{\mathbf{Q}}$ generated by Hecke operators T_n with $(n, N) = 1$ is isomorphic to a product of number fields.

9.1 Atkin-Lehner-Li theory

The results of [Li75] about newforms are proved using many linear transformations that do not necessarily preserve $S_k(\Gamma_1(N), \varepsilon)$. Thus we introduce more general spaces of cusp forms, which these transformations preserve. These spaces are also useful because they make precise how the space of cusp forms for the principal congruence subgroup

$$\Gamma(N) = \ker(\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}))$$

can be understood in terms of spaces $S_k(\Gamma_1(M), \varepsilon)$ for various M and ε, which justifies our usual focus on these latter spaces. This section follows [Li75] closely.

Let M and N be positive integers and define

$$\Gamma_0(M, N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \ : \ M \mid c, N \mid b \right\},$$

and

$$\Gamma(M, N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(M, N) \ : \ a \equiv d \equiv 1 \pmod{MN} \right\}.$$

Note that $\Gamma_0(M, 1) = \Gamma_0(M)$ and $\Gamma(M, 1) = \Gamma_1(M)$. Let $S_k(M, N)$ denote the space of cusp forms for $\Gamma(M, N)$.

If ε is a Dirichlet character modulo MN such that $\varepsilon(-1) = (-1)^k$, let $S_k(M, N, \varepsilon)$ denote the space of all cups forms for $\Gamma(M, N)$ of weight k and character ε. This is the space of holomorphic functions $f : \mathfrak{h} \to \mathbf{C}$ that satisfy the usual vanishing conditions at the cusps and such that for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(M, N)$,

$$f| \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \varepsilon(d) f.$$

We have

$$S_k(M, N) = \oplus_\varepsilon S_k(M, N, \varepsilon).$$

We now introduce operators between various $S_k(M, N)$. Note that, except when otherwise noted, the notation we use for these operators below is as in [Li75], which conflicts with notation in various other books. When in doubt, check the definitions.

Let

$$f| \begin{pmatrix} a & b \\ c & d \end{pmatrix} (\tau) = (ad - bc)^{k/2} (c\tau + d)^{-k} f\left(\frac{a\tau + b}{c\tau + d}\right).$$

This is like before, but we omit the weight k from the bar notation, since k will be fixed for the whole discussion.

For any d and $f \in S_k(M, N, \varepsilon)$, define

$$f|U_d^N = d^{k/2-1} f\Big| \left(\sum_{u \bmod d} \begin{pmatrix} 1 & uN \\ 0 & d \end{pmatrix} \right),$$

where the sum is over *any* set u of representatives for the integers modulo d. Note that the N in the notation U_d^N is a superscript, not a power of N. Also, let

$$f|B_d = d^{-k/2} f| \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$f|C_d = d^{k/2} f| \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}.$$

In [Li75], C_d is denoted W_d, which would be confusing, since in the literature W_d is usually used to denote a completely different operator (the Atkin-Lehner operator, which is denoted V_d^M in [Li75]).

Since $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma(M, N)$, any $f \in S_k(M, N, \varepsilon)$ has a Fourier expansion in terms of powers of $q_N = q^{1/N}$. We have

$$\left(\sum_{n \geq 1} a_n q_N^n \right) |U_d^N = \sum_{n \geq 1} a_{nd} q_N^n,$$

$$\left(\sum_{n \geq 1} a_n q_N^n \right) |B_d = \sum_{n \geq 1} a_n q_N^{nd},$$

and

$$\left(\sum_{n \geq 1} a_n q_N^n \right) |C_d = \sum_{n \geq 1} a_n q_{Nd}^n.$$

The second two equalities are easy to see; for the first, write everything out and use that for $n \geq 1$, the sum $\sum_u e^{2\pi i u n/d}$ is 0 or d if $d \nmid n$, $d \mid n$, respectively, much as in Section 3.5.2.

The maps B_d and C_d define injective maps between various spaces $S_k(M, N, \varepsilon)$. To understand B_d, use the matrix relation

$$\begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x & dy \\ z/d & w \end{pmatrix} \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix},$$

and for C_d use

$$\begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x & y/d \\ zd & w \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}.$$

If $d \mid N$ then $B_d : S_k(M, N, \varepsilon) \to S_k(dM, N/d, \varepsilon)$ is an isomorphism, and if $d \mid M$, then $C_d : S_k(M, N) \to S_k(M/d, Nd, \varepsilon)$ is also an isomorphism. In particular, taking $d = N$, we obtain an isomorphism

$$B_N : S_k(M, N, \varepsilon) \to S_k(MN, 1, \varepsilon) = S_k(\Gamma_1(MN), \varepsilon). \tag{9.1.1}$$

Putting these maps together allows us to completely understand the cusp forms $S_k(\Gamma(N))$ in terms of spaces $S_k(\Gamma_1(N^2), \varepsilon)$, for all Dirichlet characters ε that arise from characters modulo N. This is because $S_k(\Gamma(N))$ is isomorphic to the direct sum of $S_k(N, N, \varepsilon)$, as ε various over all Dirichlet characters modulo N.

For any prime p, we define the pth *Hecke operator* on $S_k(M, N, \varepsilon)$ by

$$T_p = U_p^N + \varepsilon(p) p^{k-1} B_p.$$

Note that $T_p = U_p^N$ when $p \mid N$, since then $\varepsilon(p) = 0$. In terms of Fourier expansions, we have

$$\left(\sum_{n \geq 1} a_n q_N^n \right) | T_p = \sum_{n \geq 1} \left(a_{np} + \varepsilon(p) p^{k-1} a_{n/p} \right) q_N^n,$$

where $a_{n/p} = 0$ if $p \nmid n$.

The operators we have just defined satisfy several commutativity relations. Suppose p and q are prime. Then $T_p B_q = B_q T_p$, $T_p C_q = C_q T_p$, and $T_p U_q^N = U_q^N T_p$ if $(p, qMN) = 1$. Moreover $U_d^N B_{d'} = B_{d'} U_d^N$ if $(d, d') = 1$.

Remark 9.1.1. Because of these relations, (9.1.1) describes $S_k(\Gamma(N))$ as a module over the ring generated by T_p for $p \nmid N$.

Definition 9.1.2 (Old Subspace)**.** The *old subspace* $S_k(M, N, \varepsilon)_{\text{old}}$ is the subspace of $S_k(M, N, \varepsilon)$ generated by all $f|B_d$ and $g|C_e$ where $f \in S_k(M', N)$, $g \in S_k(M, N')$, and M', N' are proper factors of M, N, respectively, and $d \mid M/M'$, $e \mid N/N'$.

Since T_p commutes with B_d and C_e, the Hecke operators T_p preserve $S_k(M, N, \varepsilon)_{\text{old}}$, for $p \nmid MN$. Also, B_N defines an isomorphism

$$S_k(M, N, \varepsilon)_{\text{old}} \cong S_k(MN, 1, \varepsilon)_{\text{old}}.$$

Definition 9.1.3 (Petersson Inner Product)**.** If $f, g \in S_k(\Gamma(N))$, the *Petersson inner product* of f and g is

$$\langle f, g \rangle = \frac{1}{[\mathrm{SL}_2(\mathbf{Z}) : \Gamma(N)]} \int_D f(z) \overline{g(z)} y^{k-2} \, dx \, dy,$$

where D is a fundamental domain for $\Gamma(N)$ and $z = x + iy$.

This Petersson pairing is normalized so that if we consider f and g as elements of $\Gamma(N')$ for some multiple N' of N, then the resulting pairing is the same (since the volume of the fundamental domain scales by the index).

Theorem 9.1.4 (Petersson)**.** *If $p \nmid N$ and $f \in S_k(\Gamma_1(N), \varepsilon)$, then $\langle f|T_p, g\rangle = \varepsilon(p)\langle f, g|T_p\rangle$.*

See [Lan95, §VII.5, Thm. 5.1] for a proof of Theorem 9.1.4.

Remark 9.1.5. Theorem 9.1.4 implies that when $p \nmid N$ the adjoint of T_p is $\varepsilon(p)T_p$, so T_p commutes with its adjoint, hence T_p is *normal*, which implies that T_p is diagonalizable. Be careful, because the T_p, with $p \mid N$, need not be diagonalizable (see Section 9.2.3).

Definition 9.1.6 (New Subspace)**.** The *new subspace* $S_k(M, N, \varepsilon)_{\text{new}}$ is the orthogonal complement of $S_k(M, N, \varepsilon)_{old}$ in $S_k(M, N, \varepsilon)$ with respect to the Petersson inner product.

Both the old and new subspaces of $S_k(M, N, \varepsilon)$ are preserved by the Hecke operators T_p with $p \nmid NM$.

Remark 9.1.7. Li [Li75] also gives a purely algebraic definition of the new subspace as the intersection of the kernels of various trace maps from $S_k(M, N, \varepsilon)$ that are obtained by averaging over coset representatives.

Definition 9.1.8 (Newform)**.** A *newform* $f = \sum a_n q_N^n \in S_k(M, N, \varepsilon)$ is an element of $S_k(M, N, \varepsilon)_{\text{new}}$ that is an eigenform for all T_p, for $p \nmid NM$, and is normalized so that $a_1 = 1$.

Li introduces the crucial "Atkin-Lehner operator" W_q^M (denoted V_q^M in [Li75]), which plays a key roll in all the proofs, and is defined as follows. For a positive integer M and prime q, let $\alpha = \text{ord}_q(M)$ and use the extended Euclidean algorithm to find a choice of integers x, y, z such that $q^{2\alpha}x - yMz = q^\alpha$. Then W_q^M is the operator defined by slashing with the matrix $\begin{pmatrix} q^\alpha x & y \\ Mz & q^\alpha \end{pmatrix}$. Li shows that if $f \in S_k(M, 1, \varepsilon)$, then $f|W_q^M|W_q^M = \varepsilon(q^\alpha)f$, so W_q^M is invertible. Care must be taken, because the operator W_q^M need not commute with $T_p = U_p^N$, when $p \mid M$.

After proving many technical but elementary lemmas about the operators B_d, C_d, U_p^N, T_p, and W_q^M, Li uses the lemmas to deduce the following theorems, whose proofs are relatively elementary.

Theorem 9.1.9. *Suppose $f = \sum a_n q_N^n \in S_k(M, N, \varepsilon)$ and $a_n = 0$ for all n with $\gcd(n, r) = 1$, where r is a fixed positive integer. Then $f \in S_k(M, N, \varepsilon)_{\text{old}}$.*

From the theorem we see that if f and g are newforms in $S_k(M, N, \varepsilon)$, and if for all but finitely many primes p, the T_p eigenvalues of f and g are the same, then $f - g$ is an old form, so $f - g = 0$, hence $f = g$. Thus the eigenspaces in the new subspace corresponding to the systems of Hecke eigenvalues associated to the T_p, with $p \nmid MN$, each have dimension 1. This is known as *multiplicity one*.

Theorem 9.1.10. Let $f = \sum a_n q_N^n$ be a newform in $S_k(M, N, \varepsilon)$, p a prime with $p \nmid MN$, and $q \mid MN$ a prime. Then

1. $f|T_p = a_p f$, $f|U_q^N = a_q f$, and for all $n \geq 1$,
$$a_p a_n = a_{np} + \varepsilon(p)p^{k-1}a_{n/p},$$
$$a_q a_n = a_{nq}.$$
If $L(f,s) = \sum_{n\geq 1} a_n n^{-s}$ is the Dirichlet series associated to f, then $L(f,s)$ has an Euler product
$$L(f,s) = \prod_{\text{primes } p} (1 - a_p p^{-s} + \varepsilon(p)p^{k-1}p^{-2s})^{-1}.$$
Note that when $p \mid NM$, we have $1 - a_p p^{-s} + \varepsilon(p)p^{k-1}p^{-2s} = 1 - a_p p^{-s}$.
2. (a) If ε is not induced by a character mod MN/q, then $|a_q| = q^{(k-1)/2}$.
(b) If ε is induced by a character mod MN/q, then $a_q = 0$ if $q^2 \mid MN$, and $a_q^2 = \varepsilon(q)q^{k-2}$ if $q^2 \nmid MN$.

9.2 The U_p operator

Let N be a positive integer and M a divisor of N. For each divisor d of N/M we define a map
$$\alpha_d : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)) : \quad f(\tau) \mapsto f(d\tau).$$
Thus α_d is just the map B_d from Section 26.4.2. We verify that $f(d\tau) \in S_k(\Gamma_1(N))$ as follows. Recall that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we write
$$(f|[\gamma]_k)(\tau) = \det(\gamma)^{k-1}(cz+d)^{-k}f(\gamma(\tau)).$$
The transformation condition for f to be in $S_k(\Gamma_1(N))$ is that $f|[\gamma]_k(\tau) = f(\tau)$. Let $f(\tau) \in S_k(\Gamma_1(M))$ and let $\iota_d = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $f|[\iota_d]_k(\tau) = d^{k-1}f(d\tau)$ is a modular form on $\Gamma_1(N)$ since $\iota_d^{-1}\Gamma_1(M)\iota_d$ contains $\Gamma_1(N)$. Moreover, if f is a cusp form then so is $f|[\iota_d]_k$.

Proposition 9.2.1. *If $f \in S_k(\Gamma_1(M))$ is nonzero, then the images*
$$\left\{\alpha_d(f) \,:\, d \mid \frac{N}{M}\right\}$$
are linearly independent.

Proof. If the q-expansion of f is $\sum a_n q^n$, then the q-expansion of $\alpha_d(f)$ is $\sum a_n q^{dn}$. The matrix of coefficients of the q-expansions of $\alpha_d(f)$, for $d \mid (N/M)$, is upper triangular. Thus the q-expansions of the $\alpha_d(f)$ are linearly independent, hence the $\alpha_d(f)$ are linearly independent, since the map that sends a cusp form to its q-expansion is linear. □

When $p \mid N$, we denote by U_p the Hecke operator T_p acting on the space $S_k(\Gamma_1(N))$. For clarity, in this section only we will denote by T_p^M, the Hecke operator $T_p \in \mathrm{End}(S_k(\Gamma_1(M)))$. For $f = \sum a_n q^n \in S_k(\Gamma_1(N))$, we have
$$f|U_p = \sum a_{np} q^n.$$

Suppose $f = \sum a_n q^n \in S_k(\Gamma_1(M))$ is a normalized eigenform for all of the Hecke operators T_n and $\langle n \rangle$, and p is a prime that does not divide M. Then

$$f|T_p^M = a_p f \quad \text{and} \quad f|\langle p \rangle = \varepsilon(p) f.$$

Assume $N = p^r M$, where $r \geq 1$ is an integer. Let

$$f_i(\tau) = f(p^i \tau) = \alpha_{p^i}(f),$$

so $f_0, \ldots, f_r$ are the images of f under the maps $\alpha_{p^0}, \ldots, \alpha_{p^r}$, respectively, and $f = f_0$. We have

$$\begin{aligned} f|T_p^M &= \sum_{n \geq 1} a_{np} q^n + \varepsilon(p) p^{k-1} \sum a_n q^{pn} \\ &= f_0|U_p + \varepsilon(p) p^{k-1} f_1, \end{aligned}$$

so

$$f_0|U_p = f|T_p^M - \varepsilon(p) p^{k-1} f_1 = a_p f_0 - \varepsilon(p) p^{k-1} f_1. \tag{9.2.1}$$

Also

$$f_1|U_p = \left(\sum a_n q^{pn} \right) |U_p = \sum a_n q^n = f_0.$$

More generally, for any $i \geq 1$, we have $f_i|U_p = f_{i-1}$.

The operator U_p preserves the two dimensional vector space spanned by f_0 and f_1, and the matrix of U_p with respect to the basis f_0, f_1 is

$$A = \begin{pmatrix} a_p & 1 \\ -\varepsilon(p) p^{k-1} & 0 \end{pmatrix},$$

which has characteristic polynomial

$$X^2 - a_p X + p^{k-1} \varepsilon(p). \tag{9.2.2}$$

When $r \geq 3$, the operator U_p does not act diagonalizably on the space spanned by $f_0, f_1, \ldots, f_r$. See Section 9.2.3 below.

9.2.1 A Connection with Galois representations

Equation (9.2.2) leads to a striking connection with Galois representations. Let f be a newform and let $K = K_f$ be the field generated over $\mathbf{Q}$ by the Fourier coefficients of f. Let ℓ be a prime and λ a prime of the ring of integers of K lying over ℓ. Then, as we mentioned in Section 2.1, Deligne (and Serre, when $k = 1$, and Shimura when $k = 2$) constructed a representation

$$\rho_{f,\lambda} = \rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, K_\lambda).$$

If $p \nmid N\ell$, then ρ_λ is unramified at p, so if $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ if a Frobenius element, then $\rho_\lambda(\mathrm{Frob}_p)$ is well defined, up to conjugation. Moreover,

$$\begin{aligned} \det(\rho_\lambda(\mathrm{Frob}_p)) &= p^{k-1} \varepsilon(p), \quad \text{and} \\ \mathrm{tr}(\rho_\lambda(\mathrm{Frob}_p)) &= a_p. \end{aligned}$$

Thus the characteristic polynomial of $\rho_\lambda(\mathrm{Frob}_p) \in \mathrm{GL}_2(K_\lambda)$ is

$$X^2 - a_p X + p^{k-1} \varepsilon(p),$$

which is the same as (9.2.2).

9.2.2 When is U_p semisimple?

Question 9.2.2. Is U_p semisimple on the span of f_0 and f_1?

If the eigenvalues of U_p acting on the span of f_0 and f_1 are distinct, then the answer is yes. If the eigenvalues are the same, then $X^2 - a_p X + p^{k-1}\varepsilon(p)$ has discriminant 0, so $a_p^2 = 4p^{k-1}\varepsilon(p)$, hence

$$a_p = 2p^{\frac{k-1}{2}}\sqrt{\varepsilon(p)}.$$

Open Problem 9.2.3. Does there exist an eigenform $f = \sum a_n q^n \in S_k(\Gamma_1(N))$ such that $a_p = 2p^{\frac{k-1}{2}}\sqrt{\varepsilon(p)}$?

It is a curious fact that the Ramanujan conjectures, which were proved by Deligne in 1973, imply that $|a_p| \leq 2p^{(k-1)/2}$, so the above equality remains taunting. When $k = 2$, Coleman and Edixhoven proved in [CE98] that $|a_p| < 2p^{(k-1)/2}$.

9.2.3 An Example of non-semisimple U_p

Suppose $f = f_0$ is a normalized eigenform. Let W be the space spanned by f_0, f_1 and let V be the space spanned by f_0, f_1, f_2, f_3. Then U_p acts on V/W by $\overline{f}_2 \mapsto 0$ and $\overline{f}_3 \mapsto \overline{f}_2$. Thus the matrix of the action of U_p on V/W is $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, which is nonzero and nilpotent, hence not semisimple. Since W is invariant under U_p this shows that U_p is not semisimple on V, i.e., U_p is not diagonalizable.

9.3 The Cusp forms are free of rank 1 over $\mathbf{T}_\mathbf{C}$

9.3.1 Level 1

Suppose $N = 1$, so $\Gamma_1(N) = \mathrm{SL}_2(\mathbf{Z})$. Using the Petersson inner product, we see that all the T_n are diagonalizable, so $S_k = S_k(\Gamma_1(1))$ has a basis

$$f_1, \ldots, f_d$$

of normalized eigenforms where $d = \dim S_k$. This basis is canonical up to ordering. Let $\mathbf{T}_\mathbf{C} = \mathbf{T} \otimes \mathbf{C}$ be the ring generated over $\mathbf{C}$ by all Hecke operators T_n. Then, having fixed the basis above, there is a canonical map

$$\mathbf{T}_\mathbf{C} \hookrightarrow \mathbf{C}^d: \quad T \mapsto (\lambda_1, \ldots, \lambda_d),$$

where $f_i|T = \lambda_i f_i$. This map is injective and $\dim \mathbf{T}_\mathbf{C} = d$, so the map is an isomorphism of $\mathbf{C}$-vector spaces.

The form

$$v = f_1 + \cdots + f_n$$

generates S_k as a $\mathbf{T}$-module. Note that v is canonical since it does not depend on the ordering of the f_i. Since v corresponds to the vector $(1, \ldots, 1)$ and $\mathbf{T} \cong \mathbf{C}^d$ acts on $S_k \cong \mathbf{C}^d$ componentwise, this is just the statement that $\mathbf{C}^d$ is generated by $(1, \ldots, 1)$ as a $\mathbf{C}^d$-module.

Recall from Section 4.2 that there is a perfect bilinear pairing $S_k \times \mathbf{T}_\mathbf{C} \to \mathbf{C}$ given by

$$\langle f, T_n \rangle = a_1(f|T_n) = a_n(f),$$

where $a_n(f)$ denotes the nth Fourier coefficient of f. Thus we have simultaneously:

1. S_k is free of rank 1 over $\mathbf{T}_{\mathbf{C}}$, and
2. $S_k \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C})$ as $\mathbf{T}$-modules.

Combining these two facts yields an isomorphism

$$\mathbf{T}_{\mathbf{C}} \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C}). \tag{9.3.1}$$

This isomorphism sends an element $T \in \mathbf{T}$ to the homomorphism

$$X \mapsto \langle v|T, X\rangle = a_1(v|T|X).$$

Since the identification $S_k \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C})$ is canonical and since the vector v is canonical, we see that the isomorphism (9.3.1) is canonical.

Recall that M_k has as basis the set of products $E_4^a E_6^b$, where $4a + 6b = k$, and S_k is the subspace of forms where the constant coefficient of their q-expansion is 0. Thus there is a basis of S_k consisting of forms whose q-expansions have coefficients in $\mathbf{Q}$. Let $S_k(\mathbf{Z}) = S_k \cap \mathbf{Z}[[q]]$, be the submodule of S_k generated by cusp forms with Fourier coefficients in $\mathbf{Z}$, and note that $S_k(\mathbf{Z}) \otimes \mathbf{Q} \cong S_k(\mathbf{Q})$. Also, the explicit formula $(\sum a_n q^n)|T_p = \sum a_{np} q^n + p^{k-1} \sum a_n q^{np}$ implies that the Hecke algebra $\mathbf{T}$ preserves $S_k(\mathbf{Z})$.

Proposition 9.3.1. *We have $v \in S_k(\mathbf{Z})$.*

Proof. This is because $v = \sum \mathrm{Tr}(T_n) q^n$, and, as we observed above, there is a basis so that the matrices T_n have integer coefficients, so their traces are integers. □

Example 9.3.2. When $k = 36$, we have

$$\begin{aligned} v = {} & 3q + 139656q^2 - 104875308q^3 + 34841262144q^4 + 892652054010q^5 \\ & - 4786530564384q^6 + 878422149346056q^7 + \cdots . \end{aligned}$$

The normalized newforms f_1, f_2, f_3 are

$$\begin{aligned} f_i = {} & q + aq^2 + (-1/72a^2 + 2697a + 478011548)q^3 + (a^2 - 34359738368)q^4 \\ & (a^2 - 34359738368)q^4 + (-69/2a^2 + 14141780a + 1225308030462)q^5 + \cdots , \end{aligned}$$

for a each of the three roots of $X^3 - 139656X^2 - 59208339456X - 1467625047588864$.

9.3.2 General level

Now we consider the case for general level N. Recall that there are maps

$$S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)),$$

for all M dividing N and all divisor d of N/M.

Definition 9.3.3. The *old subspace* of $S_k(\Gamma_1(N))$ is the space generated by all images of these maps with $M|N$ but $M \neq N$. The *new subspace* is the orthogonal complement of the old subspace with respect to the Petersson inner product.

There is an algebraic definition of the new subspace. One defines trace maps

$$S_k(\Gamma_1(N)) \to S_k(\Gamma_1(M))$$

for all $M < N$, $M \mid N$ which are adjoint to the above maps (with respect to the Petersson inner product). Then f is in the new part of $S_k(\Gamma_1(N))$ if and only if f is in the kernels of all of the trace maps.

It follows from Atkin-Lehner-Li theory that the T_n acts semisimply on the new subspace $S_k(\Gamma_1(M))_{\text{new}}$ for all $M \geq 1$, since the common eigenspaces for all T_n each have dimension 1. Thus $S_k(\Gamma_1(M))_{\text{new}}$ has a basis of normalized eigenforms. We have a natural map

$$\bigoplus_{M|N} S_k(\Gamma_1(M))_{\text{new}} \hookrightarrow S_k(\Gamma_1(N)).$$

The image in $S_k(\Gamma_1(N))$ of an eigenform f for some $S_k(\Gamma_1(M))_{\text{new}}$ is called a *newform* of level $M_f = M$. Note that a newform of level less than N is often not an eigenform for all of the Hecke operators acting on $S_k(\Gamma_1(N))$; for example, if $N = p^r M$ with $p \nmid M$ a prime, then f is not eigenform for T_p (see Equation 9.2.1 above).

Let

$$v = \sum_f f(q^{\frac{N}{M_f}}) \in S_k(\Gamma_1(N)),$$

where the sum is taken over all newforms f of weight k and some level $M \mid N$. This generalizes the v constructed above when $N = 1$ and has many of the same good properties. For example, $S_k(\Gamma_1(N))$ is free of rank 1 over $\mathbf{T}_{\mathbf{C}}$ with basis element v. Moreover, the coefficients of v lie in $\mathbf{Z}$, but to show this we need to know that $S_k(\Gamma_1(N))$ has a basis whose q-expansions lie in $\mathbf{Q}[[q]]$. This is true, but we will not prove it here. One way to proceed is to use the Tate curve to construct a q-expansion map $\mathrm{H}^0(X_1(N), \Omega_{X_1(N)/\mathbf{Q}}) \to \mathbf{Q}[[q]]$ that is compatible with the usual Fourier expansion map.

Example 9.3.4. The space $S_2(\Gamma_1(22))$ has dimension 6. There is a single newform of level 11,

$$f = q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 + \cdots .$$

There are four newforms of level 22, the four $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of

$$g = q - \zeta q^2 + (-\zeta^3 + \zeta - 1)q^3 + \zeta^2 q^4 + (2\zeta^3 - 2)q^5 + (\zeta^3 - 2\zeta^2 + 2\,\zeta - 1)q^6 - 2\zeta^2 q^7 + \ldots$$

where ζ is a primitive 10th root of unity.

Warning 9.3.5. Let $S = S_2(\Gamma_0(88))$, and let $v = \sum \mathrm{Tr}(T_n)q^n$. Then S has dimension 9, but the Hecke span of v only has dimension 7. Thus the more "canonical looking" element $\sum \mathrm{Tr}(T_n)q^n$ is not a generator for S.

```
sage: S = CuspForms(88)
sage: B = S.sturm_bound(); B
25
sage: f = QQ[['q']]([0]+[S.T(n).trace() for n in [1..B]], B+1)
sage: f
```

```
9*q - 2*q^2 - 4*q^3 - 2*q^5 + 2*q^6 - 8*q^7 + ... + O(q^26)
sage: f = S(f)
sage: span([S.T(n)(f).element() for n in [1..B]]).dimension()
7
```

Remark 9.3.6. Recall Proposition 4.6.5 that the Fourier coefficients of each normalized eigenform in S_k are totally real algebraic integers. A *CM field* is a quadratic imaginary extension of a totally real field. For example, when $n > 2$, the field $\mathbf{Q}(\zeta_n)$ is a CM field, with totally real subfield $\mathbf{Q}(\zeta_n)^+ = \mathbf{Q}(\zeta_n + 1/\zeta_n)$. More generally, one shows that the eigenvalues of any newform $f \in S_k(\Gamma_1(N))$ generate a totally real or CM field.

9.4 Decomposing the anemic Hecke algebra

We first observe that it make no difference whether or not we include the Diamond bracket operators in the Hecke algebra. Then we note that the $\mathbf{Q}$-algebra generated by the Hecke operators of index coprime to the level is isomorphic to a product of fields corresponding to the Galois conjugacy classes of newforms. Here the Galois group $G_{\mathbf{Q}} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on newforms by acting on their coefficients, and a Galois conjugacy class means an orbit for the action of $G_{\mathbf{Q}}$.

Proposition 9.4.1. *The operators $\langle d \rangle$ on $S_k(\Gamma_1(N))$ lie in $\mathbf{Z}[\ldots, T_n, \ldots]$.*

Proof. Dirichlet's theorem on primes in arithmetic progression (see [Lan94, VIII.4]) asserts that every residue class modulo N coprime to N contains infinitely many primes. Thus it is enough to show $\langle p \rangle \in \mathbf{Z}[\ldots, T_n, \ldots]$ for primes $p \nmid N$, since each nonzero $\langle d \rangle$ is of the form $\langle p \rangle$ for some prime p. Since $p \nmid N$, we have (see Section 8.2.2) that

$$T_{p^2} = T_p^2 - \langle p \rangle p^{k-1},$$

so $\langle p \rangle p^{k-1} = T_p^2 - T_{p^2}$. Again, by Dirichlet's theorem on primes in arithmetic progression, there is another prime q 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

$$\langle p \rangle = \langle p \rangle (ap^{k-1} + bq^{k-1}) = a({T_p}^2 - T_{p^2}) + b({T_q}^2 - T_{q^2}) \in \mathbf{Z}[\ldots, T_n, \ldots].$$

□

Let S be a space of cusp forms, such as $S_k(\Gamma_1(N))$ or $S_k(\Gamma_1(N), \varepsilon)$. Let

$$f_1, \ldots, f_d \in S$$

be representatives for the Galois conjugacy classes of newforms in S of level N_{f_i} dividing N. For each i, let $K_i = \mathbf{Q}(\ldots, a_n(f_i), \ldots)$ be the field generated by the Fourier coefficients of f_i.

Definition 9.4.2 (Anemic Hecke Algebra)**.** The *anemic Hecke algebra* is the subalgebra

$$\mathbf{T}_0 = \mathbf{Z}[\ldots, T_n, \ldots : \gcd(n, N) = 1] \subset \mathbf{T}$$

of $\mathbf{T}$ obtained by adjoining to $\mathbf{Z}$ only those Hecke operators T_n with n relatively prime to N.

Proposition 9.4.3. *We have* $\mathbf{T}_0 \otimes \mathbf{Q} \cong \prod_{i=1}^{d} K_i$*, where* K_i *is the number field generated by the eigenvalues of* f_i*.*

The map sends T_n to $(a_n(f_1), \ldots, a_n(f_d))$. The proposition follows from the discussion above and Atkin-Lehner theory.

Example 9.4.4. When $S = S_2(\Gamma_1(22))$, then $\mathbf{T}_0 \otimes \mathbf{Q} \cong \mathbf{Q} \times \mathbf{Q}(\zeta_{10})$ (see Example 9.3.4). When $S = S_2(\Gamma_0(37))$, then $\mathbf{T}_0 \otimes \mathbf{Q} \cong \mathbf{Q} \times \mathbf{Q}$.

Remark 9.4.5. The index $[\mathbf{T} : \mathbf{T}_0]$ is usually not finite. As explained in Section 9.3.2, the space $S_k(\Gamma_1(N))$ is free of rank 1 over $\mathbf{T}_\mathbf{C}$, so $\dim \mathbf{T}_\mathbf{C} = \dim S_k(\Gamma_1(N))$. Proposition 9.4.3 implies that $\dim(\mathbf{T}_0)_\mathbf{C} = d$ is the number of newforms of level dividing N. If there is at least one oldform in $S_k(\Gamma_1(N))$, then $d < \dim S_k(\Gamma_1(N))$, because that oldform has at least two linearly independent images in $S_k(\Gamma_1(N))$

10
Some Explicit Genus Computations

This chapter is about computing the genus of certain modular curves, or equivalently, the dimensions of certain spaces of cusp forms. Section 10.1 explains the connection between genus and the dimension of a space of cusp form, and the general picture regarding ramification of certain covers of modular curves. Then in Section 10.2, we explain our strategy to compute the genus of various modular curves using an Euler characteristic argument. In Section 10.3 we apply this strategy in the case of $X(N)$, and Section 10.4 treats the case of $X_0(N)$ with N prime. Finally, in Section 10.5 we discuss a natural isomorphism between two spaces of mod p modular forms, which is suggested by dimension considerations.

We do not treat any other cases explicitly in this book. For the general case of $X_0(N)$, the reader might look at [Shi94, §1.6], for $X_1(N)$ at [DI95, §9.1], and at [DS05, Ch. 3]. See also Section 5.3 for a different approach in to computing the genus of $X(N)$.

10.1 Computing the dimension of $S_2(\Gamma)$

Let $k = 2$ unless otherwise noted, and let $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ be a congruence subgroup. Then $S_2(\Gamma) \cong H^0(X_\Gamma, \Omega^1)$ where $X_\Gamma = (\Gamma\backslash\mathcal{H}) \cup (\Gamma\backslash\mathbf{P}^1(\mathbf{Q}))$. By definition $\dim H^0(X_\Gamma, \Omega^1)$ is the genus of X_Γ.

Since $\Gamma \subset \Gamma(1)$ there is a covering $X_\Gamma \to X_{\Gamma(1)} \xrightarrow{j} \mathbf{P}^1(\mathbf{C})$, which can only possibly be ramified at points above $0, 1728, \infty \in \mathbf{P}^1(\mathbf{C})$. Here 0 corresponds to $\rho = e^{2\pi i/3}$ and 1728 corresponds to i under the j-invariant map. We illustrate this

as follows:

$$\begin{array}{ccc} \Gamma\backslash\mathcal{H} & \longrightarrow & X_\Gamma \\ \downarrow & & \downarrow \\ \Gamma(1)\backslash\mathcal{H} & \longrightarrow & X_{\Gamma(1)} \\ & & \downarrow j \\ & & \mathbf{P}^1(\mathbf{C}) \end{array}$$

Example 10.1.1. Suppose $\Gamma = \Gamma_0(N)$. The degree of the covering is the index $(\mathrm{SL}_2(\mathbf{Z})/\{\pm 1\} : \Gamma_0(N)/\{\pm 1\})$. As explained in Theorem 5.2.5, a point on $Y_{\Gamma(1)}$ corresponds to an isomorphism class of elliptic curves over $\mathbf{C}$, whereas a points on $Y_0(N)$ correspond to an isomorphism class of a pair consisting of an elliptic curve and a cyclic subgroup of order N. The map $Y_0(N) \to Y_{\Gamma(1)} = Y_0(1)$ forgets the extra structure of the cyclic subgroup.

10.2 Application of Riemann-Hurwitz

Now we compute the genus of X_Γ by applying the Riemann-Hurwitz formula. We recall some standard facts about the Euler characteristic of a topological space, and also the Riemann-Hurwitz formula.

The Euler characteristic χ is additive in the sense that if A and B are disjoint spaces then

$$\chi(A \cup B) = \chi(A) + \chi(B).$$

Also, if X is a compact Riemann surface of genus g, then $\chi(X) = 2 - 2g$, and $\chi(\{\text{point}\}) = 1$. Thus

$$\chi(X - \{p_1, \ldots, p_n\}) = \chi(X) - n\chi(\{\text{point}\}) = (2 - 2g) - n.$$

If $X \to Y$ is an unramified covering of degree d then a special case of the Riemann-Hurwitz formula is the assertion that

$$\chi(X) = d \cdot \chi(Y). \tag{10.2.1}$$

Identifying $X_{\Gamma(1)}$ and $\mathbf{P}^1(\mathbf{C})$ using the j map, consider the covering

$$\begin{array}{c} X_\Gamma - \{\text{points over } 0, 1728, \infty\} \\ \downarrow \\ X_{\Gamma(1)} - \{0, 1728, \infty\}. \end{array}$$

This covering is unramified, as was explained in Proposition 5.3.3.

Since $X_{\Gamma(1)}$ has genus 0, the space $X_{\Gamma(1)} - \{0, 1728, \infty\}$ has Euler characteristic $2 - 3 = -1$. If we let $g = \chi(X_\Gamma)$ then

$$\chi(X_\Gamma - \{\text{points over } 0, 1728, \infty\}) = 2 - 2g - n_0 - n_{1728} - n_\infty,$$

where n_p denotes the number of points lying over p. Thus $-d = 2 - 2g - n_0 - n_{1728} - n_\infty$ whence

$$2g - 2 = d - n_0 - n_{1728} - n_\infty.$$

Conclusion: Finding the genus of *any* modular curve X_Γ amounts to calculating:

1. the degree d of the j function,
2. the number $n_0 + n_{1728}$ of points on X_Γ with j-invariant 0 or 1728 and,
3. the number n_∞ of cusps for Γ.

10.3 The Genus of $X(N)$

Let $N > 3$ and consider the modular curve $X = X(N)$. There is a natural covering map $X \to X(1) \xrightarrow{j} \mathbf{C}$. Let d be the degree, then as we saw above

$$2g - 2 = d - n_0 - n_{1728} - n_\infty$$

where g is the genus of X and n_p is the number of points lying over p. Since $n_0 = d/3$ and $n_{1728} = d/2$,

$$2g - 2 = \frac{d}{6} - n_\infty.$$

Now we count the number n_∞ of cusps of $X(N)$, that is, the size of $\Gamma(N)\backslash\mathbf{P}^1(\mathbf{Q})$. There is a surjective map from $\mathrm{SL}_2(\mathbf{Z})$ to $\mathbf{P}^1(\mathbf{Q})$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The subgroup of elements of $\mathrm{SL}_2(\mathbf{Z})$ that stabilize $(1, 0)$ is

$$U = \left\{ \pm \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathbf{Z} \right\}.$$

The cusps of $X(N)$ are the elements of

$$\Gamma(N)\backslash(\mathrm{SL}_2(\mathbf{Z})/U) = (\Gamma(N)\backslash \mathrm{SL}_2(\mathbf{Z}))/U = \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/U$$

which has order

$$\frac{\#\,\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})}{2N} = \frac{d}{N}.$$

Substituting this into the above formula gives

$$2g - 2 = \frac{d}{6} - \frac{d}{N} = \frac{d}{6N}(N - 6),$$

so

$$g = 1 + \frac{d}{12N}(N - 6).$$

When N is prime,

$$d = \frac{1}{2} \cdot \#\,\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}) = \frac{1}{2} \cdot \frac{(N^2 - 1)(N^2 - N)}{N - 1}.$$

Thus when $N = 5$, $d = 60$ so $g = 0$, and when $N = 7$, $d = 168$ so $g = 3$. When $N = 2011$, we have $g = 337852526$.

10.4 The Genus of $X_0(N)$, for N prime

Suppose $N > 3$ and N is prime. The covering map $X_0(N) \to X(1)$ is of degree $N+1$ since a point of $X_0(N)$ corresponds to an elliptic curve along with a subgroup of order N and there are $N+1$ such subgroups because N is prime. Since N is prime, $X_0(N)$ has two cusps; they are the orbit of ∞ which is unramified and 0 which is ramified of order N. Thus

$$2g - 2 = N + 1 - 2 - n_{1728} - n_0.$$

Here n_0 is the number of pairs (E, C) (modulo isomorphism) such that E has j-invariant 0. So we consider $E = \mathbf{C}/\mathbf{Z}[\frac{-1+i\sqrt{3}}{2}]$ which has endomorphism ring $\mathrm{End}(E) = \mathbf{Z}[\mu_6]$. Now $\mu_6/\pm 1$ acts on the cyclic subgroups C so, letting ω be a primitive cube root of unity, we have

$$(E, C) \cong (E, \omega C) \cong (E, \omega^2 C).$$

This might lead one to think that n_0 is $(N+1)/3$, but it may be bigger if, for example, $C = \omega C$. Thus we must count those C so that $\omega C = C$ or $\omega^2 C = C$, that is, those C which are stable under $\mathcal{O} = \mathbf{Z}[\frac{-1+i\sqrt{3}}{2}]$. So we must compute the number of stable $\mathcal{O}/N\mathcal{O}$-submodules of order N. This depends on the structure of $\mathcal{O}/N\mathcal{O}$:

$$\mathcal{O}/N\mathcal{O} = \begin{cases} \mathbf{F}_N \oplus \mathbf{F}_N & \text{if } (\frac{-3}{N}) = 1 \ (N \text{ splits}) \\ \mathbf{F}_{N^2} & \text{if } (\frac{-3}{N}) = -1 \ (N \text{ stays inert}) \end{cases}$$

Since $\mathcal{O}/N\mathcal{O} = \mathbf{F}_{N^2}$ is a field it has no submodules of order N, whereas $\mathbf{F}_N \oplus \mathbf{F}_N$ has two $\mathcal{O}/N\mathcal{O}$-submodules of order N, namely $\mathbf{F}_N \oplus 0$ and $0 \oplus \mathbf{F}_N$. Thus

$$n_0 = \begin{cases} \frac{N+1}{3} & \text{if } N \equiv 2 \pmod 3 \\ \frac{N-1}{3} + 2 & \text{if } N \equiv 1 \pmod 3 \end{cases}$$

Exercise 10.4.1. It is an exercise in elegance to write this as a single formula involving the quadratic symbol.

By similar reasoning one shows that

$$n_{1728} = \begin{cases} \frac{N+1}{2} & \text{if } N \equiv 3 \pmod 4 \\ \frac{N-1}{2} + 2 & \text{if } N \equiv 1 \pmod 4 \end{cases}$$

We can now compute the genus of $X_0(N)$ for any prime N. For example, if $N = 37$ then $2g - 2 = 36 - (2 + 18) - (14) = 2$ so $g = 2$. Similarly, $X_0(13)$ has genus 0 and $X_0(11)$ has genus 1. In general, $X_0(N)$ has genus approximately $N/12$.

Serre constructed a nice formula for the above genus. Suppose $N > 3$ is a prime and write $N = 12a + b$ with $0 \le b \le 11$. Then Serre's formula is

b	1	5	7	11
g	$a-1$	a	a	$a+1$

10.5 Modular forms mod p

Let N be a positive integer, let p be a prime and assume Γ is either $\Gamma_0(N)$ or $\Gamma_1(N)$.

Definition 10.5.1. Let $M_k(\Gamma, \mathbf{Z}) = M_k(\Gamma, \mathbf{C}) \cap \mathbf{Z}[[q]]$. Then

$$M_k(\Gamma, \mathbf{F}_p) = M_k(\Gamma, \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{F}_p$$

is the space of *modular forms mod p* of weight k.

Suppose $p = N$. Then one has *Serre's Equality*:

$$M_{p+1}(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_p) = M_2(\Gamma_0(p), \mathbf{F}_p)$$

[[TODO: Insert a dimension formula calculation to give evidence for this equality, since that's the whole reason this section is in this chapter.]]

The map from the right hand side to the left hand side is accomplished via a certain normalized Eisenstein series. Recall from Section 4.1 that for $\mathrm{SL}_2(\mathbf{Z})$ we have Eisenstein series

$$G_k = -\frac{B_k}{2k} + \sum_{n=1}^{\infty} \left(\sum_{d|n} d^{k-1} \right) q^n$$

and

$$E_k = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} (\sum_{d|n} d^{k-1}) q^n.$$

One finds $\mathrm{ord}_p(-\frac{B_k}{2k})$ using Kummer congruences [[TODO: elaborate]]. In particular, $\mathrm{ord}_p(B_{p-1}) = -1$, so $E_{p-1} \equiv 1 \pmod{p}$. Thus multiplication by E_{p-1} increases the weight by $p-1$ but does not change the q-expansion mod p. We thus get a map

$$M_2(\Gamma_0(p), \mathbf{F}_p) \to M_{p+1}(\Gamma_0(p), \mathbf{F}_p).$$

The map

$$M_{p+1}(\Gamma_0(p), \mathbf{F}_p) \to M_{p+1}(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_p)$$

is the trace map (which is dual to the natural inclusion going the other way) and is accomplished by averaging in order to get a form invariant under $\mathrm{SL}_2(\mathbf{Z})$.

11
The Field of Moduli

In this chapter we will study the field of definition of the modular curve $X(N)$. We assume the reader if familiar with the correspondence between function fields and nonsingular projective algebraic curves, as explained in [Har77, §I.6]. For example, the function field of $\mathbf{P}^1_{\mathbf{Q}}$ is $\mathbf{Q}(t)$. We will also freely use basic facts about the j-invariant of an elliptic curve, including that two curves are isomorphic over an algebraically closed field if and only if they have the same j-invariant.

11.1 Algebraic definition of $X(N)$

If E is an elliptic curve given by a Weierstrass equation $y^2 = 4x^3 - g_2x - g_3$, then

$$j(E) = j(g_2, g_3) = \frac{1728 g_2^3}{g_2^3 - 27 g_3^2}.$$

The j-invariant determines the isomorphism class of E over $\mathbf{C}$, since $j : X(1) \to \mathbf{P}^1_{\mathbf{C}}$ has degree 1 (see [Ser73, VII.3.3]). There is an elliptic curve $E/\mathbf{Q}(t)$ such that $j(E) = t$, for example, the elliptic curve with Weierstrass equation

$$y^2 = 4x^3 - \frac{27t}{t - 1728}x - \frac{27t}{t - 1728}. \tag{11.1.1}$$

For the moment, let k be any field, E/k an arbitrary elliptic curve and N a positive integer prime to $\operatorname{char} k$. Upon fixing a choice of basis for $E[N]$, we have $E[N](\overline{k}) \cong (\mathbf{Z}/N\mathbf{Z})^2$. Let $k(E[N])$ be the field obtained by adjoining the coordinates of the N-torsion points of E to k, so we have a tower of fields $\overline{k} \supset k(E[N]) \supset k$. There is a Galois representation on the N torsion of E:

$$\operatorname{Gal}(\overline{k}/k) \xrightarrow{\rho_E} \operatorname{Aut}(E[N]) \cong \operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})$$

and $\operatorname{Gal}(\overline{k}/k(E[N])) = \ker(\rho_E)$. Also $\operatorname{Gal}(k(E[N])/k) \hookrightarrow \operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})$.

Applying the above observations with $k = \mathbf{Q}(t)$ and E the curve from (11.1.1) with j-invariant t, shows that the Galois group of the extension $\mathbf{Q}(t)(E[N])$ of $\mathbf{Q}(t)$ is contained in $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$. Let $X(N)$ be the algebraic curve corresponding to the function field $\mathbf{Q}(t)(E[N])$. As we will see in Proposition 11.4.2 below,

$$\overline{\mathbf{Q}} \cap (\mathbf{Q}(t)(E[N])) \subset \mathbf{Q}(\mu_N),$$

so the curve $X(N)$ is defined over $\mathbf{Q}(\mu_N)$.

The main input we need is that the representation ρ_E is surjective. Our strategy to prove this starts by composing ρ_E with the natural map $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}) \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ to define a map

$$\overline{\rho}_E : \mathrm{Gal}(\overline{k}/k) \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}.$$

The following proposition shows that we loose little in passing to $\overline{\rho}_E$.

Proposition 11.1.1. *$\overline{\rho}_E$ is surjective if and only if ρ_E is surjective.*

Proof. If $\overline{\rho}_E$ is surjective then either $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ or its negative lies in the image of ρ. Thus $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ lies in the image of ρ_E. Since $\overline{\rho}_E$ is surjective this implies that ρ_E is surjective. The converse is trivial. □

11.2 Digression on moduli

Recall from Theorem 5.2.5 that the non-cuspidal points in $X_0(N)(\mathbf{C})$ are the set of $\mathbf{C}$-isomorphism classes of pairs (E, C) where $E/\mathbf{C}$ is an elliptic curve and C is a cyclic subgroup of order N. Assume for the moment that $X_0(N)$ has a some sort of canonical model of algebraic curve over $\mathbf{Q}$, and let $Y_0(N) = X_0(N) - \{ \text{ cusps } \}$. It is reasonable to assume that $Y_0(N)(\overline{\mathbf{Q}})$ is the set of $\overline{\mathbf{Q}}$-isomorphism classes of pairs (E, C) where $E/\overline{\mathbf{Q}}$ is an elliptic curve and C is a cyclic subgroup of order N. Each $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ would then act on the set of these points by sending the class of (E, C) to the class of $({}^{\sigma}E, {}^{\sigma}C)$, where ${}^{\sigma}E$ is the elliptic curve got by applying σ to the coefficients of an equation for E, and ${}^{\sigma}C \subset {}^{\sigma}E(\overline{\mathbf{Q}})[N]$ is the image of C under σ.

Let K be a number field. Since $\overline{\mathbf{Q}}$ is a separable extension of K, we have

$$X_0(N)(K) = X_0(N)(\overline{\mathbf{Q}})^{\mathrm{Gal}(\overline{\mathbf{Q}}/K)}.$$

Thus $Y_0(N)(K)$ is the set of isomorphism classes of pairs $(E, C) \in Y_0(N)(\overline{\mathbf{Q}})$ such that for all $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/K)$, there exists some isomorphism $({}^{\sigma}E, {}^{\sigma}C) \approx (E, C)$ defined over $\overline{\mathbf{Q}}$. There is a map

$$\{K\text{-isomorphism classes of pairs } (E, C)/K\} \to Y_0(N)(K)$$

that is "notoriously" non-injective. For example, when $N = 1$, the map $X_0(1) \xrightarrow{j} \mathbf{P}^1$ identifies $X_0(1)$ with the j-line. With this identification, the above map then sends an elliptic curve E/K to its j-invariant. Since all quadratic twists of E are also defined over K and have the same j-invariant, there are infinitely many different elements of the left hand side that all map to the same point $j = j(E)$ in the right hand side.

The paper [DR73] contains a proof that the map is surjective, i.e., that any isomorphism class $[(E, C)]$, with E and C defined over $\overline{\mathbf{Q}}$ that is fixed by all $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/K)$ has a representative element (E, C) with E and C defined over K. When $N = 1$, [DR73] observe that the map is surjective directly, i.e., they answer this question:

Question 11.2.1. If $E/\overline{K}$ is isomorphic to all its Galois conjugates, is there a curve E'/K that is isomorphic to E over $\overline{K}$?

For $N > 1$ they show that certain obstructions vanish. [[TODO: precise ref?]]

11.3 When is ρ_E surjective?

Let K be a field of characteristic 0.

Proposition 11.3.1. *Let E_1 and E_2 be elliptic curves defined over K with equal j-invariants, so $E_1 \approx E_2$ over $\overline{K}$. Assume E_1 and E_2 do not have complex multiplication over $\overline{K}$. Then ρ_{E_1} is surjective if and only if ρ_{E_2} is surjective.*

Proof. Assume ρ_{E_1} is surjective. Since E_1 does not have complex multiplication over $\overline{K}$, we have $\mathrm{Aut}\, E_1 = \{\pm 1\}$. Choose an isomorphism $\varphi : E_1 \xrightarrow{\sim} E_2$ over $\overline{K}$. Then for any $\sigma \in \mathrm{Gal}(\overline{K}/K)$ we have the diagram

$$\begin{array}{ccc} E_1 & \xrightarrow{\varphi} & E_2 \\ {\scriptstyle =}\downarrow & & {\scriptstyle =}\downarrow \\ {}^{\sigma}E_1 & \xrightarrow{\sigma\varphi} & {}^{\sigma}E_2 \end{array},$$

where we note that ${}^{\sigma}E_1 = E_1$ and ${}^{\sigma}E_1 = E_2$, since both curves are defined over K. Thus $\sigma\varphi = \pm\varphi$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$, so $\varphi : E_1[N] \to E_2[N]$ defines an equivalence $\overline{\rho}_{E_1} \cong \overline{\rho}_{E_2}$. Since ρ_{E_1} is surjective this implies that $\overline{\rho}_{E_2}$ is surjective; Proposition 11.1.1 then implies that ρ_{E_2} is surjective. □

Let $K = \mathbf{C}(j)$, with j transcendental over $\mathbf{C}$. Let E/K be an elliptic curve such as (11.1.1) with j-invariant j. Fix a positive integer N and let

$$\rho_E : \mathrm{Gal}(\overline{K}/K) \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$$

be the associated Galois representation on the N-torsion of E. Then one can prove using an algebraic definition of the Weil pairing that $\det \rho_E$ is the cyclotomic character, which is trivial since $\mathbf{C} \subset K$ and $\mathbf{C}$ contains the Nth roots of unity. Thus the image of ρ_E lands inside of $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. Our next theorem states that a "generic elliptic curve", i.e., a curve with j-invariant j, has maximal possible Galois action on its division points.

Theorem 11.3.2. $\rho_E : \mathrm{Gal}(\overline{K}/K) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ *is surjective.*

Igusa [Igu56] [[TODO: I could not find anything about this in [Igu56], but I only looked for a few minutes. In [DR73, §2] they say that Igusa proves this, though, and sketch his proof.]] found an algebraic proof of this theorem, but we content ourselves with making some comments on how an analytic proof goes.

Proof. The field $\mathbf{C}(j) = K = \mathcal{F}_1$ is the field of modular functions for $\mathrm{SL}_2(\mathbf{Z})$. Suppose $N \geq 3$ and let $\mathcal{F}_N$ be the field of mereomorphic functions for $\Gamma(N)$, i.e., meromorphic functions on $\mathfrak{h}^*$ that are invariant under $\Gamma(N)$. One can prove [] that the extension $\mathcal{F}_N/\mathcal{F}_1$ is a Galois extension with Galois group $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\} \cong \overline{\Gamma(1)}/\overline{\Gamma(N)}$, where $\overline{\Gamma(N)}$ denotes the image of $\Gamma(N)$ in $\mathrm{PSL}_2(\mathbf{Z})$.

Let E be an elliptic curve over K with j-invariant j. We will show that $\mathrm{Gal}(\mathcal{F}_N/\mathcal{F}_1)$ acts transitively on the x-coordinates of the N-torsion points of E. This will show that $\overline{\rho}_E$ maps surjectively onto $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ [[TODO: Why?]]. Then by Proposition 11.1.1, ρ_E maps surjectively onto $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$, as claimed. [[TODO: Many details need to be added based on Rohrlich's proof!]]

We will now view the x-coordinates of the points in $E[N]$ as functions on $\mathcal{H}$ that are invariant under $\Gamma(N)$. (Thus $K(E[N]/\{\pm 1\}) \subset \mathcal{F}_N$.) Let $\tau \in \mathcal{H}$ and let $\Lambda_\tau = \mathbf{Z}\tau + \mathbf{Z}$. Consider $\wp(z, \Lambda_\tau)$, which gives the x-coordinate of $\mathbf{C}/\Lambda_\tau$ in its standard form $y^2 = 4x^3 - g_2 x - g_3$. Define, for each nonzero $(r, s) \in (\mathbf{Z}/N\mathbf{Z})^2/\{\pm 1\}$, a function

$$f_{(r,s)} : \mathcal{H} \to \mathbf{C} : \quad \tau \mapsto \frac{g_2(\tau)}{g_3(\tau)} \wp\left(\frac{r\tau + s}{N}, \Lambda_\tau\right).$$

We first prove that for any $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$,

$$f_{(r,s)}(\alpha\tau) = f_{(r,s)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(\tau),$$

as follows. Indeed, $\wp$ is homogeneous of degree -2, g_2 is modular of weight 4 and g_3 is modular of weight 6, so [[TODO: Weierstrass $\wp$ notation below is missing second argument.]]

$$\begin{aligned} f_{(r,s)}(\alpha\tau) &= \frac{g_2(\alpha\tau)}{g_3(\alpha\tau)} \wp\left(\frac{r\alpha\tau + s}{N}\right) \\ &= (c\tau + d)^{-2} \frac{g_2(\tau)}{g_3(\tau)} \wp\left(\frac{ra\tau + rb + cs\tau + sd}{N(c\tau + d)}\right) \\ &= \frac{g_2(\tau)}{g_3(\tau)} \wp\left(\frac{(ra + sc)\tau + rb + sd}{N}\right) = f_{(r,s)\alpha}(\tau) \end{aligned}$$

Let $E_{j(\tau)}$ denote an elliptic curve over $\mathbf{C}$ with j-invariant $j(\tau)$. If $\tau \in \mathcal{H}$ with $g_2(\tau), g_3(\tau) \neq 0$ then the $f_{(r,s)}(\tau)$ are the x-coordinates of the nonzero N-division points of $E_{j(\tau)}$. The various $f_{(r,s)}(\tau)$ are distinct. Thus $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ acts transitively on the $f_{(r,s)}$. The consequence is that the $N^2 - 1$ nonzero division points of our generic curve E have x-coordinates in $\overline{\mathcal{F}}_N$ equal to the $f_{(r,s)} \in \mathcal{F}_N$. □

11.4 Observations

Proposition 11.4.1. *If $E/\mathbf{Q}(\mu_N)(t)$ is an elliptic curve with $j(E) = t$, then ρ_E has image $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$.*

Proof. Since $\mathbf{Q}(\mu_N)$ contains the Nth roots of unity, the Nth cyclotomic character is trivial, hence the determinant of ρ_E is trivial. Thus the image of ρ_E lies in $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. In the other direction, there is a natural inclusion

$$\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}) = \mathrm{Gal}(\mathbf{C}(t)(E[N])/\mathbf{C}(t)) \hookrightarrow \mathrm{Gal}(\mathbf{Q}(\mu_N)(t)(E[N])/\mathbf{Q}(\mu_N)(t)).$$

□

Proposition 11.4.2. *If $E/\mathbf{Q}(t)$ is an elliptic curve with $j(E) = t$, then ρ_E has image $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ and $\overline{\mathbf{Q}} \cap (\mathbf{Q}(t)(E[N])) = \mathbf{Q}(\mu_N)$.*

Proof. Since $\mathbf{Q}(t)$ contains no Nth roots of unity (recall $N \geq 3$), the mod N cyclotomic character, and hence $\det \rho_E$, is surjective onto $(\mathbf{Z}/N\mathbf{Z})^*$. Since the image of ρ_E already contains $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ it must equal $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$. For the second assertion consider the diagram

$$\begin{array}{ccc}
\overline{\mathbf{Q}} & & \mathbf{Q}(t)(E[N]) \\
| & & | \,\mathrm{SL}_2 \\
\mathbf{Q}(\mu_N) & \text{———} & \mathbf{Q}(\mu_N)(t) \\
(\mathbf{Z}/N\mathbf{Z})^* \,| & & | \,\mathrm{GL}_2 \,/\, \mathrm{SL}_2 = (\mathbf{Z}/N\mathbf{Z})^* \\
\mathbf{Q} & \text{———} & \mathbf{Q}(t)
\end{array}$$

□

This gives a way to view $X_0(N)$ as a projective algebraic curve over $\mathbf{Q}$. Let $K = \mathbf{Q}(t)$ and let $L = K(E[N]) \supset \mathbf{Q}(\mu_N)(t)$. Then

$$H = \{\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)\} \subset \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}) = \mathrm{Gal}(L/K).$$

The fixed field L^H is an extension of $\mathbf{Q}(t)$ of transcendence degree 1 with field of constants $\overline{\mathbf{Q}} \cap L^H = \mathbf{Q}$, i.e., a projective algebraic curve defined over $\mathbf{Q}$.

11.5 A descent problem

Consider the following exercise, which may be approached in an honest or dishonest way.

Exercise 11.5.1. Suppose L/K is a finite Galois extension and $G = \mathrm{Gal}(L/K)$. Let E/L be an elliptic curve, assume $\mathrm{Aut}_L\, E = \{\pm 1\}$, and suppose that for all $g \in G$, there is an isomorphism ${}^gE \xrightarrow{\sim} E$ over L. Show that there exists E_0/K such that $E_0 \cong E$ over L.

Caution! The exercise is *false* as stated. Both the dishonest and honest approaches below work only if L is a separable closure of K. Now: can one construct a counterexample?

Discussion. First the hard, but "honest" way to look at this problem. For notions on descent see [Ser88, V.20]. By descent theory, to give E_0 is the same as to give a family $(\lambda_g)_{g\in G}$ of maps $\lambda_g : {}^gE \xrightarrow{\sim} E$ such that $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ where ${}^g\lambda_h = g \circ \lambda_h \circ g^{-1}$. Note that $\lambda_g \circ {}^g\lambda_h$ maps ${}^{gh}E \to E$. This is the natural condition to impose, because if $f : E_0 \xrightarrow{\sim} E$ and we let $\lambda_g = f \circ {}^g(f^{-1})$ then $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$.

Using our hypothesis choose, for each $g \in G$, an isomorphism

$$\lambda_g : {}^gE \xrightarrow{\sim} E.$$

Define a map c by

$$c(g, h) = \lambda_g \circ {}^g\lambda_h \circ \lambda_{gh}^{-1}.$$

Note that $c(g,h) \in \operatorname{Aut} E = \{\pm 1\}$ so c defines an element of

$$H^2(G, \{\pm 1\}) \subset H^2(\operatorname{Gal}(\overline{L}/K), \{\pm 1\}) = \operatorname{Br}(K)[2].$$

Here $\operatorname{Br}(K)[2]$ denotes the 2-torsion of the Brauer group

$$\operatorname{Br}(K) = H^2(\operatorname{Gal}(\overline{L}/K), \overline{L}^*).$$

This probably leads to an honest proof.

The dishonest approach is to note that $g(j(E)) = j(E)$ for all $g \in G$, since all conjugates of E are isomorphic and $j({}^gE) = g(j(E))$. Thus $j(E) \in K$ (assuming K is perfect), so we can define E_0/K by substituting $j(E)$ into (11.1.1). This gives an elliptic curve E_0 defined over K but isomorphic to E over $\overline{K}$.

11.6 Second look at the descent exercise

We have been discussing the following problem. Suppose L/K is a Galois extension with $\operatorname{char} K = 0$, and let E/L be an elliptic curve. Suppose that for all $\sigma \in G = \operatorname{Gal}(L/K)$, ${}^\sigma E \cong E$ over L. Conclude that there is an elliptic curve E_0/K such that $E_0 \cong E$ over L. The conclusion may fail to hold if L is a finite extension of K, but the exercise is true when $L = \overline{K}$. First we give a descent argument which holds when $L = \overline{K}$ and then give a counterexample to the more general statement.

For $g, h \in G = \operatorname{Gal}(L/K)$ we define an automorphism $c(g,h) \in \operatorname{Aut} E = \{\pm 1\}$. Choose for every $g \in \operatorname{Gal}(L/K)$ some isomorphism

$$\lambda_g : {}^gE \xrightarrow{\sim} E.$$

If the λ_g were to all satisfy the compatibility criterion $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ then by descent theory we could find a K-structure on E, that is a model for E defined over K and isomorphic to E over L. Define $c(g,h)$ by $c(g,h)\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ so $c(g,h)$ measures how much the λ_g fail to satisfy the compatibility criterion. Since $c(g,h)$ is a cocycle it defines an elements of $H^2(G, \{\pm 1\})$. We want to know that this element is trivial. When $L = \overline{K}$, the map $H^2(G, \{\pm 1\}) \to H^2(G, L^*)$ is injective. To see this first consider the exact sequence

$$0 \to \{\pm 1\} \to \overline{K}^* \xrightarrow{2} \overline{K}^* \to 0$$

where $2 : \overline{K}^* \to \overline{K}^*$ is the squaring map. Taking cohomology yields an exact sequence

$$H^1(G, \overline{K}^*) \to H^2(G, \{\pm 1\}) \to H^2(G, \overline{K}^*).$$

By Hilbert's theorem 90 ([Ser79] Ch. X, Prop. 2), $H^1(G, \{\pm 1\}) = 0$. Thus we have an exact sequence

$$0 \to H^2(G, \{\pm 1\}) \to H^2(G, \overline{K}^*)[2] \to 0.$$

Thus $H^2(G, \{\pm 1\})$ naturally sits inside $H^2(G, L^*)$.

[[TODO: To finish Ribet does something with differentials and $H^0({}^gE, \Omega^1)$ which I don't understand.]]

The counterexample in the case when L/K is finite was provided by Kevin Buzzard [[TODO: (who said Coates gave it to him)]]. Let $L = \mathbf{Q}(i)$, $K = \mathbf{Q}$ and E be

the elliptic curve with Weirstrass equation $iy^2 = x^3 + x + 1$. Then E is isomorphic to its conjugate over L but one can show directly that E has no model over $\mathbf{Q}$. [[TODO: BUT: Isn't $y^2 = x^3 + x + 1$ a model for this curve over $\mathbf{Q}$. It is isomorphic to that curve over $\overline{\mathbf{Q}}$.]]

11.7 Action of GL_2

Let $N > 3$ be an integer and $E/\mathbf{Q}(j)$ an elliptic curve with j-invariant $j(E) = j$. Then there is a Galois extension

$$\begin{array}{c} \mathcal{F}_N = \mathbf{Q}(j)(E[N]/\{\pm 1\}) \\ \Big| \\ \mathcal{F}_1 = \mathbf{Q}(j) \end{array}$$

with Galois group $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$. Think of $\mathbf{Q}(j)(E[N]/\{\pm 1\})$ as the field obtained from $\mathbf{Q}(j)$ by adjoining the x-coordinates of the N-torsion points of E. Note that this situation differs from the previous situation in Section 11.3 in that the base field $\mathbf{C}$ has been replaced by $\mathbf{Q}$.

Consider

$$\mathcal{F} = \bigcup_N \mathcal{F}_N$$

which corresponds to a projective system of modular curves . Let $\mathcal{A}_f$ be the ring of finite adèles; thus

$$\mathcal{A}_f = \hat{\mathbf{Q}} = \hat{\mathbf{Z}} \otimes \mathbf{Q} \subset \prod_p \mathbf{Q}_p.$$

We think of $\mathcal{A}_f$ as

$$\{(x_p) \in \prod \mathbf{Q}_p \,:\, x_p \in \mathbf{Z}_p \text{ for almost all } p\}.$$

The group $\mathrm{GL}_2(\mathcal{A}_f)$ acts on $\mathcal{F}$. To understand what this action is we first consider the subgroup $\mathrm{GL}_2(\hat{\mathbf{Z}})$ of $\mathrm{GL}_2(\mathcal{A}_f)$.

It can be shown that

$$\mathcal{F} = \mathbf{Q}(f_{N,(r,s)} : (r,s) \in (\mathbf{Z}/N\mathbf{Z})^2 - \{(0,0)\}, N \geq 1)$$

where $f_{N,(r,s)}$ is as defined in Section 11.3. We define an action of $\mathrm{GL}_2(\hat{\mathbf{Z}})$ on $\mathcal{F}$ as follows. To describe how $g \in \mathrm{GL}_2(\hat{\mathbf{Z}})$ acts on $f_{N,(r,s)}$, first map g into $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ via the natural reduction map, then note that $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ acts on $f_{N,(r,s)}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot f_{N,(r,s)} = f_{N,(r,s)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} = f_{N,(ra+sc,rb+sd)}.$$

Let E be an elliptic curve. Then the universal Tate module is

$$T(E) = \varprojlim_{N \geq 1} E[N] = \prod_p T_p(E).$$

There is some isomorphism $\alpha : \hat{\mathbf{Z}}^2 \xrightarrow{\sim} T(E)$. Via right composition, $\mathrm{GL}_2(\hat{\mathbf{Z}})$ acts on the collection of all such isomorphisms α. So $\mathrm{GL}_2(\hat{\mathbf{Z}})$ acts naturally on pairs (E, α) with the action doing nothing to E. An important point to be grasped when constructing objects like Shimura varieties is that we must "free ourselves" and allow $\mathrm{GL}_2(\hat{\mathbf{Z}})$ to act on the E's as well.

Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbf{Q}).$$

Let $\tau \in \mathcal{H}$ and let $E = E_\tau$ be the elliptic curve determined by the lattice $\Lambda_\tau = \mathbf{Z}\tau + \mathbf{Z}$. Let

$$\alpha_\tau : \Lambda_\tau = \mathbf{Z}\tau + \mathbf{Z} \xrightarrow{\sim} \mathbf{Z}^2$$

be the isomorphism defined by $\tau \mapsto (1, 0)$ and $1 \mapsto (0, 1)$. Now view $\alpha = \alpha_\tau$ as a map

$$\alpha : \mathbf{Z}^2 \xrightarrow{\sim} H_1(E(\mathbf{C}), \mathbf{Z}).$$

Tensoring with $\mathbf{Q}$ then gives another map (also denoted α)

$$\alpha : \mathbf{Q}^2 \xrightarrow{\sim} H_1(E, \mathbf{Q}).$$

Then $\alpha \circ g$ is another isomorphism

$$\mathbf{Q}^2 \xrightarrow{\alpha \circ g} H_1(E, \mathbf{Q}),$$

which induces an isomorphism $\mathbf{Z}^2 \xrightarrow{\sim} L' \subset H_1(E, \mathbf{Q})$ where L' is a lattice. There exists an elliptic curve $E'/\mathbf{C}$ and a map $\lambda \in \mathrm{Hom}(E', E) \otimes \mathbf{Q}$, which induces a map (also denoted λ)

$$\lambda : H_1(E', \mathbf{Z}) \xrightarrow{\sim} L' \subset H_1(E, \mathbf{Q})$$

on homology groups.

Now we can define an action on pairs (E, α) by sending (E, α) to (E', α'). Here α' is the map $\alpha' : \mathbf{Z}^2 \to H_1(E', \mathbf{Z})$ given by the composition

$$\mathbf{Z}^2 \xrightarrow{\alpha g} L' \xrightarrow{\lambda^{-1}} H_1(E', \mathbf{Z}).$$

In more concrete terms the action is

$$g : (E_\tau, \alpha_\tau) \mapsto (E'_\tau, \alpha'_\tau)$$

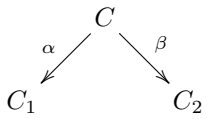
where $\tau' = g\tau = \frac{a\tau + b}{c\tau + d}$.

12
Hecke Operators as Correspondences

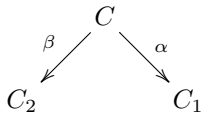
Our goal is to view the Hecke operators T_n and $\langle d \rangle$ as objects defined over $\mathbf{Q}$ that act in a compatible way on modular forms, modular Jacobians, and homology. In order to do this, we will define the Hecke operators as correspondences.

12.1 The Definition

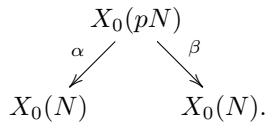
Definition 12.1.1 (Correspondence)**.** Let C_1 and C_2 be curves. A *correspondence* $C_1 \rightsquigarrow C_2$ is a curve C together with nonconstant morphisms $\alpha : C \to C_1$ and $\beta : C \to C_2$. We represent a correspondence by a diagram



Given a correspondence $C_1 \rightsquigarrow C_2$ the *dual correspondence* $C_2 \rightsquigarrow C_1$ is obtained by looking at the diagram in a mirror



In defining Hecke operators, we will focus on the simple case when the modular curve is $X_0(N)$ and Hecke operator is T_p, where $p \nmid N$. We will view T_p as a correspondence $X_0(N) \rightsquigarrow X_0(N)$, so there is a curve $C = X_0(pN)$ and maps α and β fitting into a diagram



The maps α and β are degeneracy maps which forget data. To define them, we view $X_0(N)$ as classifying isomorphism classes of pairs (E, C), where E is an elliptic curve and C is a cyclic subgroup of order N (we will not worry about what happens at the cusps, since any rational map of nonsingular curves extends uniquely to a morphism [Har77, Ch. I, Prop. 6.8]). Similarly, $X_0(pN)$ classifies isomorphism classes of pairs (E, G) where $G = C \oplus D$, C is cyclic of order N and D is cyclic of order p. Note that since $(p, N) = 1$, the group G is cyclic of order pN and the subgroups C and D are uniquely determined by G. The map α forgets the subgroup D of order p, and β quotients out by D:

$$\alpha : (E, G) \mapsto (E, C) \tag{12.1.1}$$

$$\beta : (E, G) \mapsto (E/D, (C + D)/D) \tag{12.1.2}$$

We translate this into the language of complex analysis by thinking of $X_0(N)$ and $X_0(pN)$ as quotients of the upper half plane. The first map α corresponds to the map

$$\Gamma_0(pN)\backslash\mathfrak{h} \to \Gamma_0(N)\backslash\mathfrak{h}$$

induced by the inclusion $\Gamma_0(pN) \hookrightarrow \Gamma_0(N)$. The second map β is constructed by composing the isomorphism

$$\Gamma_0(pN)\backslash\mathfrak{h} \xrightarrow{\sim} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1} \backslash\mathfrak{h} \tag{12.1.3}$$

with the map to $\Gamma_0(N)\backslash\mathfrak{h}$ induced by the inclusion

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1} \subset \Gamma_0(N).$$

The isomorphism (12.1.3) is induced by $z \mapsto \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) z = pz$; explicitly, it is

$$\Gamma_0(pN)z \mapsto \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) \Gamma_0(pN) \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) z.$$

(Note that this is well-defined.)

The maps α and β induce pullback maps on differentials

$$\alpha^*, \beta^* : \mathrm{H}^0(X_0(N), \Omega^1) \to \mathrm{H}^0(X_0(pN), \Omega^1).$$

We can identify $S_2(\Gamma_0(N))$ with $H^0(X_0(N), \Omega^1)$ by sending the cusp form $f(z)$ to the holomorphic differential $f(z)dz$. Doing so, we obtain two maps

$$\alpha^*, \beta^* : S_2(\Gamma_0(N)) \to S_2(\Gamma_0(pN)).$$

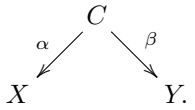
Since α is induced by the identity map on the upper half plane, we have $\alpha^*(f) = f$, where we view $f = \sum a_n q^n$ as a cusp form with respect to the smaller group $\Gamma_0(pN)$. Also, since β^* is induced by $z \mapsto pz$, we have

$$\beta^*(f) = p \sum_{n=1}^{\infty} a_n q^{pn}.$$

The factor of p is because

$$\beta^*(f(z)dz) = f(pz)d(pz) = pf(pz)dz.$$

Let X, Y, and C be curves, and α and β be nonconstant holomorphic maps, so we have a correspondence



By first pulling back, then pushing forward, we obtain induced maps on differentials

$$H^0(X,\Omega^1) \xrightarrow{\alpha^*} H^0(C,\Omega^1) \xrightarrow{\beta_*} H^0(Y,\Omega^1).$$

The composition $\beta_* \circ \alpha^*$ is a map $H^0(X,\Omega^1) \to H^0(Y,\Omega^1)$. If we consider the dual correspondence, which is obtained by switching the roles of X and Y, we obtain a map $H^0(Y,\Omega^1) \to H^0(X,\Omega^1)$.

Now let α and β be as in (12.1.1). Then we can recover the action of T_p on modular forms of weight 2 by considering the induced map

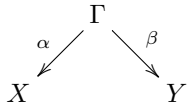
$$\beta_* \circ \alpha^* : H^0(X_0(N),\Omega^1) \to H^0(X_0(N),\Omega^1)$$

and using that $S_2(\Gamma_0(N)) \cong H^0(X_0(N),\Omega^1)$.

12.2 Maps induced by correspondences

In this section we will see how correspondences induce maps on divisor groups, which in turn induce maps on Jacobians.

Suppose $\varphi : X \to Y$ is a morphism of curves. Let $\Gamma \subset X \times Y$ be the graph of φ. This gives a correspondence



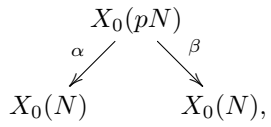
We can reconstruct φ from the correspondence by using that $\varphi(x) = \beta(\alpha^{-1}(x))$.

More generally, suppose Γ is a curve and that $\alpha : \Gamma \to X$ has degree $d \geq 1$. View $\alpha^{-1}(x)$ as a divisor on Γ (it is the formal sum of the points lying over x, counted with appropriate multiplicities). Then $\beta(\alpha^{-1}(x))$ is a divisor on Y. We thus obtain a map

$$\mathrm{Div}^n(X) \xrightarrow{\beta\circ\alpha^{-1}} \mathrm{Div}^{dn}(Y),$$

where $\mathrm{Div}^n(X)$ is the group of divisors of degree n on X. In particular, setting $d = 0$, we obtain a map $\mathrm{Div}^0(X) \to \mathrm{Div}^0(Y)$.

We now apply the above construction to T_p. Recall that T_p is the correspondence



where α and β are as in Section 12.1 and the induced map is

$$(E,C) \overset{\alpha^*}{\mapsto} \sum_{D\in E[p]} (E, C\oplus D) \overset{\beta_*}{\mapsto} \sum_{D\in E[p]} (E/D, (C+D)/D).$$

Thus we have a map $\mathrm{Div}(X_0(N)) \to \mathrm{Div}(X_0(N))$. This strongly resembles the first definition we gave of T_p on level 1 forms, where T_p was a correspondence of lattices.

12.3 Induced maps on Jacobians of curves

Let X be a curve of genus g over a field k. Recall that there is an important association

$$\left\{ \text{ curves } X/k \right\} \longrightarrow \left\{ \text{ Jacobians } \mathrm{Jac}(X) = J(X) \text{ of curves } \right\}$$

between curves and their Jacobians.

Definition 12.3.1 (Jacobian)**.** Let X be a curve of genus g over a field k. Then the *Jacobian* of X is an abelian variety of dimension g over k whose underlying group is functorially isomorphic to the group of divisors of degree 0 on X modulo linear equivalence. (For a more precise definition, see Section **??** (Jacobians section)[1] .) 1

There are many constructions of the Jacobian of a curve. We first consider the Albanese construction. Recall that over $\mathbf{C}$, any abelian variety is isomorphic to $\mathbf{C}^g/L$, where L is a lattice (and hence a free $\mathbf{Z}$-module of rank $2g$). There is an embedding

$$\begin{aligned} \iota : \mathrm{H}_1(X, \mathbf{Z}) &\hookrightarrow \mathrm{H}^0(X, \Omega^1)^* \\ \gamma &\mapsto \int_\gamma \bullet \end{aligned}$$

Then we realize $\mathrm{Jac}(X)$ as a quotient

$$\mathrm{Jac}(X) = \mathrm{H}^0(X, \Omega^1)^* / \iota(\mathrm{H}_1(X, \mathbf{Z})).$$

In this construction, $\mathrm{Jac}(X)$ is most naturally viewed as covariantly associated to X, in the sense that if $X \to Y$ is a morphism of curves, then the resulting map $\mathrm{H}^0(X, \Omega^1)^* \to \mathrm{H}^0(Y, \Omega^1)^*$ on tangent spaces induces a map $\mathrm{Jac}(X) \to \mathrm{Jac}(Y)$.

There are other constructions in which $\mathrm{Jac}(X)$ is contravariantly associated to X. For example, if we view $\mathrm{Jac}(X)$ as $\mathrm{Pic}^0(X)$, and $X \to Y$ is a morphism, then pullback of divisor classes induces a map $\mathrm{Jac}(Y) = \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X) = \mathrm{Jac}(X)$.

If $F : X \rightsquigarrow Y$ is a correspondence, then F induces an a map $\mathrm{Jac}(X) \to \mathrm{Jac}(Y)$ and also a map $\mathrm{Jac}(Y) \to \mathrm{Jac}(X)$. If $X = Y$, so that X and Y are the same, it can often be confusing to decide which duality to use. Fortunately, for T_p, with p prime to N, it does not matter which choice we make. But it matters a lot if $p \mid N$ since then we have non-commuting confusable operators and this has resulted in mistakes in the literature.

12.4 More on Hecke operators

Our goal is to move things down to $\mathbf{Q}$ from $\mathbf{C}$ or $\overline{\mathbf{Q}}$. In doing this we want to understand T_n (or T_p), that is, how they act on the associated Jacobians and

[1] insert this

how they can be viewed as correspondences. In characteristic p the formulas of Eichler-Shimura will play an important role.

We consider T_p as a correspondence on $X_1(N)$ or $X_0(N)$. To avoid confusion we will mainly consider T_p on $X_0(N)$ with $p \nmid N$. Thus assume, unless otherwise stated, that $p \nmid N$. Remember that T_p was defined to be the correspondence

$$\begin{array}{ccc} & X_0(pN) & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\beta} \\ X_0(N) & & X_0(N) \end{array}$$

Think of $X_0(pN)$ as consisting of pairs $(\underline{E}, D)$ where D is a cyclic subgroup of E of order p and $\underline{E}$ is the *enhanced* elliptic curve consisting of an elliptic curve E along with a cyclic subgroup of order N. The degeneracy map α forgets the subgroup D and the degeneracy map β divides by it. By contravariant functoriality we have a commutative diagram

$$\begin{array}{ccc} H^0(X_0(N),\Omega^1) & \xrightarrow{T_p^*=\alpha_*\circ\beta^*} & H^0(X_0(N),\Omega^1) \\ \| & & \| \\ S_2(\Gamma_0(N)) & \xrightarrow{T_p} & S_2(\Gamma_0(N)) \end{array}$$

Our convention to define T_p^* as $\alpha_* \circ \beta^*$ instead of $\beta_* \circ \alpha^*$ was completely psychological because there is a canonical duality relating the two. We chose the way we did because of the analogy with the case of a morphism $\varphi : Y \to X$ with graph Γ which induces a correspondence

$$\begin{array}{ccc} & \Gamma & \\ {\scriptstyle\pi_1}\swarrow & & \searrow{\scriptstyle\pi_2} \\ Y & & X \end{array}$$

Since the morphism φ induces a map on global sections in the other direction

$$H^0(X,\Omega^1) = \Gamma(X) \xrightarrow{\varphi^*} \Gamma(Y) = H^0(Y,\Omega^1)$$

it is psychologically natural for more general correspondence such as T_p to map from the right to the left.

The morphisms α and β in the definition of T_p are defined over $\mathbf{Q}$. This can be seen using the general theory of representable functors. Thus since T_p is defined over $\mathbf{Q}$ most of the algebraic geometric objects we will construct related to T_p will be defined over $\mathbf{Q}$.

12.5 Hecke operators acting on Jacobians

The Jacobian $J(X_0(N)) = J_0(N)$ is an abelian variety defined over $\mathbf{Q}$. There are both covariant and contravariant ways to construct $J_0(N)$. Thus a map α :

$X_0(pN) \to X_0(N)$ induces maps

$$\begin{array}{ccc} J_0(pN) & = & J_0(pN) \\ {\scriptstyle \alpha^*}\uparrow & & \downarrow{\scriptstyle \alpha_*} \\ J_0(N) & \xrightarrow[p+1]{} & J_0(N) \end{array}$$

Note that $\alpha_* \circ \alpha^* : J_0(N) \to J_0(N)$ is just multiplication by $\deg(\alpha) = p+1$, since there are $p+1$ subgroups of order p in $\underline{E}$. (At least when $p \nmid N$, when $p|N$ there are only p subgroups.)

There are two possible ways to define T_p as an endomorphism of $J_0(N)$. We could either define T_p as $\beta_* \circ \alpha^*$ or equivalently as $\alpha_* \circ \beta^*$ (assuming still that $p \nmid N$).

12.5.1 The Albanese Map

There is a way to map the curve $X_0(N)$ into its Jacobian since the underlying group structure of $J_0(N)$ is

$$J_0(N) = \frac{\left\{ \text{ divisors of degree 0 on } X_0(N) \right\}}{\left\{ \text{ principal divisors} \right\}}$$

Once we have chosen a rational point, say ∞, on $X_0(N)$ we obtain the Albanese map

$$\theta : X_0(N) \to J_0(N) : x \mapsto x - \infty$$

which sends a point x to the divisor $x - \infty$. The map θ gives us a way to pullback differentials on $J_0(N)$. Let $\operatorname{Cot} J_0(N)$ denote the cotangent space of $J_0(N)$ (or the space of regular differentials). The diagram

$$\begin{array}{ccc} \operatorname{Cot} J_0(N) & \xleftarrow{\xi_p^*} & \operatorname{Cot} J_0(N) \\ {\scriptstyle \theta^*}\downarrow{\scriptstyle \wr} & & {\scriptstyle \wr}\downarrow{\scriptstyle \theta^*} \\ H^0(X_0(N), \Omega^1) & \xleftarrow{T_p^*} & H^0(X_0(N), \Omega^1) \end{array}$$

may be taken to give a definition of ξ_p since there is a unique endomorphism $\xi_p : J_0(N) \to J_0(N)$ inducing a map ξ_p^* which makes the diagram commute.

Now suppose Γ is a correspondence $X \rightsquigarrow Y$ so we have a diagram

$$\begin{array}{ccccc} & & \Gamma & & \\ & {\scriptstyle \alpha}\swarrow & & \searrow{\scriptstyle \beta} & \\ X & & & & Y \end{array}$$

For example, think of Γ as the graph of a morphism $\varphi : X \to Y$. Then Γ should induce a natural map

$$H^0(Y, \Omega^1) \longrightarrow H^0(X, \Omega^1).$$

Taking Jacobians we see that the composition

$$J(X) \xrightarrow{\alpha^*} J(\Gamma) \xrightarrow{\beta_*} J(Y)$$

gives a map $\beta_* \circ \alpha^* : J(X) \to J(Y)$. On cotangent spaces this induces a map

$$\alpha^* \circ \beta_* : H^0(Y, \Omega^1) \to H^0(X, \Omega^1).$$

Now, after choice of a rational point, the map $X \to J(X)$ induces a map $\operatorname{Cot} J(X) \to H^0(X, \Omega^1)$. This is in fact independent of the choice of rational point since differentials on $J(X)$ are invariant under translation.

The map $J(X) \to J(Y)$ is preferred in the literature. It is said to be induced by the Albanese functoriality of the Jacobian. We could have just as easily defined a map from $J(Y) \to J(X)$. To see this let

$$\psi = \beta_* \circ \alpha^* : J(X) \to J(Y).$$

Dualizing induces a map $\psi^\vee = \alpha_* \circ \beta^*$:

$$\begin{array}{ccc} J(X)^\vee & \xleftarrow{\quad\psi^\vee\quad} & J(Y)^\vee \\ \downarrow \cong & & \cong \uparrow \\ J(X) & & J(Y) \end{array}$$

Here we have used autoduality of Jacobians. This canonical duality is discussed in [MFK94] and [Mum70] and in Milne's article in [Sch65].

12.5.2 The Hecke algebra

We now have $\xi_p = T_p \in \operatorname{End} J_0(N)$ for every prime p. If $p|N$, then we must decide between $\alpha_* \circ \beta^*$ and $\beta_* \circ \alpha^*$. The usual choice is the one which induces the usual T_p on cusp forms. If you don't like your choice you can get out of it with Atkin-Lehner operators.

Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_p, \ldots] \subset \operatorname{End} J_0(N)$$

then $\mathbf{T}$ is the same as $\mathbf{T}_\mathbf{Z} \subset \operatorname{End}(S_2(\Gamma_0(N)))$. To see this first note that there is a map $\mathbf{T} \to \mathbf{T}_\mathbf{Z}$ which is not a prior injective, but which is injective because elements of $\operatorname{End} J_0(N)$ are completely determined by their action on $\operatorname{Cot} J_0(N)$.

12.6 The Eichler-Shimura relation

Suppose $p \nmid N$ is a prime. The Hecke operator T_p and the Frobenius automorphism Frob_p induce, by functoriality, elements of $\operatorname{End}(J_0(N)_{\mathbf{F}_p})$, which we also denote T_p and Frob_p. The Eichler-Shimura relation asserts that the relation

$$T_p = \operatorname{Frob}_p + \operatorname{Frob}_p^\vee \tag{12.6.1}$$

holds in $\operatorname{End}(J_0(N)_{\mathbf{F}_p})$. In this section we sketch the main idea behind why (12.6.1) holds. For more details and a proof of the analogous statement for $J_1(N)$, see [Con01].[2]

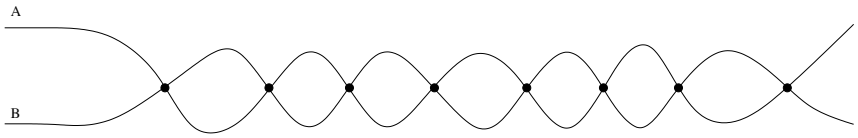


FIGURE 12.6.1. The reduction mod p of the Deligne-Rapoport model of $X_0(Np)$

Since $J_0(N)$ is an abelian variety defined over $\mathbf{Q}$, it is natural to ask for the primes p such that $J_0(N)$ have good reduction. In the 1950s Igusa showed[3] that $J_0(N)$ has good reduction for all $p \nmid N$. He viewed $J_0(N)$ as a scheme over $\mathrm{Spec}(\mathbf{Q})$, then "spread things out" to make an abelian scheme over $\mathrm{Spec}(\mathbf{Z}[1/N])$. He did this by taking the Jacobian of the normalization of $X_0(N)$ (which is defined over $\mathbf{Z}[1/N]$) in $\mathbf{P}^n_{\mathbf{Z}[1/N]}$. 3

The Eichler-Shimura relation is a formula for T_p in characteristic p, or more precisely, for the corresponding endomorphisms in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$ for all p for which $J_0(N)$ has good reduction at p. If $p \nmid N$, then $X_0(N)_{\mathbf{F}_p}$ has many of the same properties as $X_0(N)_{\mathbf{Q}}$. In particular, the noncuspidal points on $X_0(N)_{\mathbf{F}_p}$ classify isomorphism classes of enhanced elliptic curves $\underline{E} = (E, C)$, where E is an elliptic curve over $\mathbf{F}_p$ and C is a cyclic subgroup of E of order N. (Note that two pairs are considered *isomorphic* if they are isomorphic over $\overline{\mathbf{F}}_p$.)

Next we ask what happens to the map $T_p : J_0(N) \to J_0(N)$ under reduction modulo p. To this end, consider the correspondence

$$\begin{array}{ccc} & X_0(Np) & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\beta} \\ X_0(N) & & X_0(N) \end{array}$$

that defines T_p. The curve $X_0(N)$ has good reduction at p, but $X_0(Np)$ typically does not. Deligne and Rapaport [DR73] showed that $X_0(Np)$ has relatively benign reduction at p. Over $\mathbf{F}_p$, the reduction $X_0(Np)_{\mathbf{F}_p}$ can be viewed as two copies of $X_0(N)$ glued at the supersingular points, as illustrated in Figure 12.6.1.

The set of supersingular points

$$\Sigma \subset X_0(N)(\overline{\mathbf{F}}_p)$$

is the set of points in $X_0(N)$ represented by pairs $\underline{E} = (E, C)$, where E is a supersingular elliptic curve (so $E(\overline{\mathbf{F}}_p)[p] = 0$). There are exactly $g+1$ supersingular points, where g is the genus of $X_0(N)$.[4] 4

Consider the correspondence $T_p : X_0(N) \rightsquigarrow X_0(N)$ which takes an enhanced elliptic curve $\underline{E}$ to the sum $\sum \underline{E}/D$ of all quotients of $\underline{E}$ by subgroups D of order p. This is the correspondence

$$\begin{array}{ccc} & X_0(pN) & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\beta} \\ X_0(N) & & X_0(N), \end{array} \tag{12.6.2}$$

[2]Add more references to original source materials...
[3]Ken, what's a reference for this?
[4]Reference.

where the map α forgets the subgroup of order p, and β quotients out by it. From this one gets $T_p : J_0(N) \to J_0(N)$ by functoriality.

Remark 12.6.1. There are many ways to think of $J_0(N)$. The cotangent space $\mathrm{Cot}\, J_0(N)$ of $J_0(N)$ is the space of holomorphic (or translation invariant) differentials on $J_0(N)$, which is isomorphic to $S_2(\Gamma_0(N))$. This gives a connection between our geometric definition of T_p and the definition, presented earlier,[5] of T_p as an operator on a space of cusp forms. 5

The Eichler-Shimura relation takes place in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$. Since $X_0(N)$ reduces "nicely" in characteristic p, we can apply the Jacobian construction to $X_0(N)_{\mathbf{F}_p}$.

Lemma 12.6.2. *The natural reduction map*

$$\mathrm{End}(J_0(N)) \hookrightarrow \mathrm{End}(J_0(N)_{\mathbf{F}_p})$$

is injective.

Proof. Let $\ell \nmid Np$ be a prime. By [ST68, Thm. 1, Lem. 2], the reduction to characteristic p map induces an isomorphism

$$J_0(N)(\overline{\mathbf{Q}})[\ell^\infty] \cong J_0(N)(\overline{\mathbf{F}}_p)[\ell^\infty].$$

If $\varphi \in \mathrm{End}(J_0(N))$ reduces to the 0 map in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$, then $J_0(N)(\overline{\mathbf{Q}})[\ell^\infty]$ must be contained in $\ker(\varphi)$. Thus φ induces the 0 map on $\mathrm{Tate}_\ell(J_0(N))$, so $\varphi = 0$. □

Let $F : X_0(N)_{\mathbf{F}_p} \to X_0(N)_{\mathbf{F}_p}$ be the Frobenius map in characteristic p. Thus, if $K = K(X_0(N))$ is the function field of the nonsingular curve $X_0(N)$, then $F : K \to K$ is induced by the pth power map $a \mapsto a^p$.

Remark 12.6.3. The Frobenius map corresponds to the pth powering map on points. For example, if $X = \mathrm{Spec}(\mathbf{F}_p[t])$, and $z = (\mathrm{Spec}(\overline{\mathbf{F}}_p) \to X)$ is a point defined by a homomorphism $\alpha : \mathbf{F}_p[t] \mapsto \overline{\mathbf{F}}_p$, then $F(z)$ is the composite

$$\mathbf{F}_p[t] \xrightarrow{x \mapsto x^p} \mathbf{F}_p[t] \xrightarrow{\alpha} \overline{\mathbf{F}}_p.$$

If $\alpha(t) = \xi$, then $F(z)(t) = \alpha(t^p) = \xi^p$.

By both functorialities, F induces maps on the Jacobian of $X_0(N)_{\mathbf{F}_p}$:

$$\mathrm{Frob}_p = F_* \quad \text{and} \quad \mathrm{Ver}_p = \mathrm{Frob}_p^\vee = F^*,$$

which we illustrate as follows:

$$J_0(N)_{\mathbf{F}_p} \underset{\mathrm{Frob}_p}{\overset{\mathrm{Ver}_p}{\rightleftarrows}} J_0(N)_{\mathbf{F}_p}$$

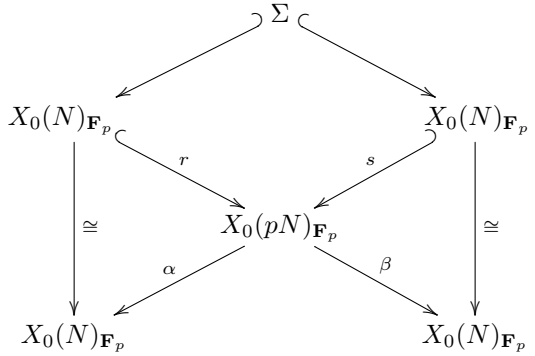
Note that $\mathrm{Ver}_p \circ \mathrm{Frob}_p = \mathrm{Frob}_p \circ \mathrm{Ver}_p = [p]$ since p is the degree of F (for example, if $K = \mathbf{F}_p(t)$, then $F(K) = \mathbf{F}_p(t^p)$ is a subfield of degree p, so the map induced by F has degree p).

[5] more precise

Theorem 12.6.4 (Eichler-Shimura Relation)**.** *Let N be a positive integer and $p \nmid N$ be a prime. Then the following relation holds:*

$$T_p = \mathrm{Frob}_p + \mathrm{Ver}_p \in \mathrm{End}(J_0(N)_{\mathbf{F}_p}).$$

Sketch of Proof. We view $X_0(pN)_{\mathbf{F}_p}$ as two copies of $X_0(N)_{\mathbf{F}_p}$ glued along corresponding supersingular points Σ, as in Figure 12.6.1. This diagram and the correspondence (12.6.2) that defines T_p translate into the following diagram of schemes over $\mathbf{F}_p$:



The maps r and s are defined as follows. Recall that a point of $X_0(N)_{\mathbf{F}_p}$ is an enhanced elliptic curve $\underline{E} = (E, C)$ consisting of an elliptic curve E (not necessarily defined over $\mathbf{F}_p$) along with a cyclic subgroup C of order N. We view a point on $X_0(Np)$ as a triple $(E, C, E \to E')$, where (E, C) is as above and $E \to E'$ is an isogeny of degree p. We use an isogeny instead of a cyclic subgroup of order p because $E(\overline{\mathbf{F}}_p)[p]$ has order either 1 or p, so the data of a cyclic subgroup of order p holds very little information.

The map r sends $\underline{E}$ to $(\underline{E}, \varphi)$, where φ is the isogeny of degree p,

$$\varphi : E \xrightarrow{\mathrm{Frob}} E^{(p)}.$$

Here $E^{(p)}$ is the curve obtained from E by hitting all defining equations by Frobenious, that is, by pth powering the coefficients of the defining equations for E. We introduce $E^{(p)}$ since if E is not defined over $\mathbf{F}_p$, then Frobenious does not define an endomorphism of E. Thus r is the map

$$r: \quad \underline{E} \mapsto (\underline{E}, E \xrightarrow{\mathrm{Frob}_p} E^{(p)}),$$

and similarly we define s to be the map

$$s: \quad \underline{E} \mapsto (E^{(p)}, C, E \xleftarrow{\mathrm{Ver}_p} E^{(p)})$$

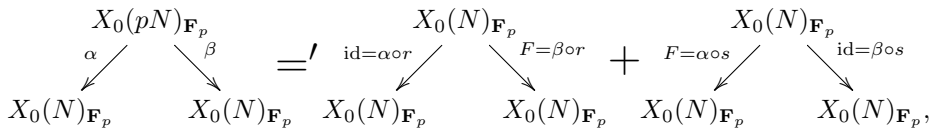
where Ver_p is the dual of Frob_p (so $\mathrm{Ver}_p \circ \mathrm{Frob}_p = \mathrm{Frob}_p \circ \mathrm{Ver}_p = [p]$).

We view α as the map sending $(\underline{E}, E \to E')$ to $\underline{E}$, and similarly we view β as the map sending $(\underline{E}, E \to E')$ to the pair (E', C'), where C' is the image of C in E' via $E \to E'$. Thus

$$\begin{aligned} \alpha: &\quad (E \to E') \mapsto E \\ \beta: &\quad (E' \to E) \mapsto E' \end{aligned}$$

It now follows immediately that $\alpha \circ r = \mathrm{id}$ and $\beta \circ s = \mathrm{id}$. Note also that $\alpha \circ s = \beta \circ r = F$ is the map $E \mapsto E^{(p)}$.

Away from the finitely many supersingular points, we may view $X_0(pN)_{\mathbf{F}_p}$ as the disjoint union of two copies of $X_0(N)_{\mathbf{F}_p}$. Thus away from the supersingular points, we have the following equality of correspondences:



where $F = \mathrm{Frob}_p$, and the $='$ means equality away from the supersingular points. Note that we are simply "pulling back" the correspondence; in the first summand we use the inclusion r, and in the second we use the inclusion s.

This equality of correspondences implies that the equality

$$T_p = \mathrm{Frob}_p + \mathrm{Ver}_p$$

of endomorphisms holds on a dense subset of $J_0(N)_{\mathbf{F}_p}$, hence on all $J_0(N)_{\mathbf{F}_p}$. □

12.7 Applications of the Eichler-Shimura relation

12.7.1 The Characteristic polynomial of Frobenius

How can we apply the relation $T_p = \mathrm{Frob} + \mathrm{Ver}$ in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$? Let $\ell \nmid pN$ be a prime and consider the ℓ-adic Tate module

$$\mathrm{Tate}_\ell(J_0(N)) = \left(\varprojlim J_0(N)[\ell^\nu]\right) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$$

which is a vector space of dimension $2g$ over $\mathbf{Q}_\ell$, where g is the genus of $X_0(N)$ or the dimension of $J_0(N)$. Reduction modulo p induces an isomorphism

$$\mathrm{Tate}_\ell(J_0(N)) \to \mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$$

(see the proof of Lemma 12.6.2). On $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$ we have linear operators Frob_p, Ver_p and T_p which, as we saw in Section 12.6, satisfy

$$\begin{aligned} \mathrm{Frob}_p + \mathrm{Ver}_p &= T_p, \qquad \text{and} \\ \mathrm{Frob}_p \circ \mathrm{Ver}_p &= \mathrm{Ver}_p \circ \mathrm{Frob}_p = [p]. \end{aligned}$$

The endomorphism $[p]$ is invertible on $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$, since p is prime to ℓ, so Ver_p and Frob_p are also invertible and

$$T_p = \mathrm{Frob}_p + [p]\,\mathrm{Frob}_p^{-1}.$$

Multiplying both sides by Frob_p and rearranging, we see that

$$\mathrm{Frob}_p^2 - T_p\,\mathrm{Frob}_p + [p] = 0 \in \mathrm{End}(\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})).$$

This is a beautiful quadratic relation, so we should be able to get something out of it. We will come back to this shortly, but first we consider the various objects acting on the ℓ-adic Tate module.

The module $\mathrm{Tate}_\ell(J_0(N))$ is acted upon in a natural way by

1. The Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of $\mathbf{Q}$, and
2. $\mathrm{End}_{\mathbf{Q}}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ (which acts by functoriality).

These actions commute with each other since endomorphisms defined over $\mathbf{Q}$ are not affected by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Reducing modulo p, we also have the following commuting actions:

3. The Galois group $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ of $\mathbf{F}_p$, and
4. $\mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$.

Note that a decomposition group group $D_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts, after quotienting out by the corresponding inertia group, in the same way as $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and the action is unramified, so action 3 is a special case of action 1.

The Frobenius elements $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and $\mathrm{Frob}_\in \mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ induce the same operator on $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$. Note that while φ_p naturally lives in a quotient of a decomposition group, one often takes a lift to get an element in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

On $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$ we have a quadratic relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0.$$

This relation plays a role when one separates out pieces of $J_0(N)$ in order to construct Galois representations attached to newforms of weight 2. Let

$$R = \mathbf{Z}[\ldots, T_p, \ldots] \subset \mathrm{End}\, J_0(N),$$

where we only adjoin those T_p with $p \nmid N$. Think of R as a reduced Hecke algebra; in particular, R is a subring of $\mathbf{T}$. Then

$$R \otimes \mathbf{Q} = \prod_{i=1}^{r} E_i,$$

where the E_i are totally real number fields. The factors E_i are in bijection with the Galois conjugacy classes of weight 2 newforms f on $\Gamma_0(M)$ (for some $M|N$). The bijection is the map

$$f \mapsto \mathbf{Q}(\text{coefficients of } f) = E_i$$

Observe that the map is the same if we replace f by one of its conjugates. This decomposition is a decomposition of a subring

$$R \otimes \mathbf{Q} \subset \mathrm{End}(J_0(N)) \otimes \mathbf{Q} \stackrel{\text{def}}{=} \mathrm{End}(J_0(N) \otimes \mathbf{Q}).$$

Thus it induces a direct product decomposition of $J_0(N)$, so $J_0(N)$ gets divided up into subvarieties which correspond to conjugacy classes of newforms.

The relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0 \tag{12.7.1}$$

suggests that

$$\mathrm{tr}(\varphi_p) = T_p \qquad \text{and} \quad \det \varphi_p = p. \tag{12.7.2}$$

This is true, but (12.7.2) does not follow formally just from the given quadratic relation. It can be proved by combining (12.7.1) with the Weil pairing.

12.7.2 The Cardinality of $J_0(N)(\mathbf{F}_p)$

Proposition 12.7.1. *Let $p \nmid N$ be a prime, and let f be the characteristic polynomial of T_p acting on $S_2(\Gamma_0(N))$. Then*

$$\#J_0(N)(\mathbf{F}_p) = f(p+1).$$

[6] 6

[6] Add details later, along with various generalizations.

13
Abelian Varieties

This chapter provides foundational background about abelian varieties and Jacobians, with an aim toward what we will need later when we construct abelian varieties attached to modular forms. We will not give complete proofs of very much, but will try to give precise references whenever possible, and many examples.

We will follow the articles by Rosen [Ros86] and Milne [Mil86] on abelian varieties. We will try primarily to explain the statements of the main results about abelian varieties, and prove results when the proofs are not too technical and enhance understanding of the statements.

13.1 Abelian varieties

Definition 13.1.1 (Variety)**.** A *variety* X over a field k is a finite-type separated scheme over k that is geometrically integral.

The condition that X be geometrically integral means that $X_{\overline{k}}$ is reduced (no nilpotents in the structure sheaf) and irreducible.

Definition 13.1.2 (Group variety)**.** A *group variety* is a group object in the category of varieties. More precisely, a group variety X over a field k is a variety equipped with morphisms

$$m : X \times X \to X \quad \text{and} \quad i : X \to X$$

and a point $1_X \in A(k)$ such that m, i, and 1_X satisfy the axioms of a group; in particular, for every k-algebra R they give $X(R)$ a group structure that depends in a functorial way on R.

Definition 13.1.3 (Abelian Variety)**.** An *abelian variety* A over a field k is a complete group variety.

Theorem 13.1.4. *Suppose A is an abelian variety. Then*

1. *The group law on A is commutative.*

2. *A is projective, i.e., there is an embedding from A into $\mathbf{P}^n$ for some n.*

3. *If $k = \mathbf{C}$, then $A(k)$ is analytically isomorphic to V/L, where V is a finite-dimensional complex vector space and L is a lattice in V. (A lattice is a free $\mathbf{Z}$-module of rank equal to $2\dim V$ such that $\mathbf{R}L = V$.)*

Proof. Part 1 is not too difficult, and can be proved by showing that every morphism of abelian varieties is the composition of a homomorphism with a translation, then applying this result to the inversion map (see [Mil86, Cor. 2.4]). Part 2 is proved with some effort in [Mil86, §7]. Part 3 is proved in [Mum70, §I.1] using the exponential map from Lie theory from the tangent space at 0 to A. □

13.2 Complex tori

Let A be an abelian variety over $\mathbf{C}$. By Theorem 13.1.4, there is a complex vector space V and a lattice L in V such that $A(\mathbf{C}) = V/L$, that is to say, $A(\mathbf{C})$ is a complex torus.

More generally, if V is any complex vector space and L is a lattice in V, we call the quotient $T = V/L$ a *complex torus*. In this section, we prove some results about complex tori that will help us to understand the structure of abelian varieties, and will also be useful in designing algorithms for computing with abelian varieties.

The differential 1-forms and first homology of a complex torus are easy to understand in terms of T. If $T = V/L$ is a complex torus, the tangent space to $0 \in T$ is canonically isomorphic to V. The $\mathbf{C}$-linear dual $V^* = \mathrm{Hom}_{\mathbf{C}}(V, \mathbf{C})$ is isomorphic to the $\mathbf{C}$-vector space $\Omega(T)$ of holomorphic differential 1-forms on T. Since $V \to T$ is the universal covering of T, the first homology $\mathrm{H}_1(T, \mathbf{Z})$ of T is canonically isomorphic to L.

13.2.1 Homomorphisms

Suppose $T_1 = V_1/L_1$ and $T_2 = V_2/L_2$ are two complex tori. If $\varphi : T_1 \to T_2$ is a (holomorphic) homomorphism, then φ induces a $\mathbf{C}$-linear map from the tangent space of T_1 at 0 to the tangent space of T_2 at 0. The tangent space of T_i at 0 is canonically isomorphic to V_i, so φ induces a $\mathbf{C}$-linear map $V_1 \to V_2$. This maps

sends L_1 into L_2, since $L_i = \mathrm{H}_1(T_i, \mathbf{Z})$. We thus have the following diagram:

$$\begin{array}{ccc} 0 & & 0 \\ \downarrow & & \downarrow \\ L_1 & \xrightarrow{\rho_{\mathbf{Z}}(\varphi)} & L_2 \\ \downarrow & & \downarrow \\ V_1 & \xrightarrow{\rho_{\mathbf{C}}(\varphi)} & L_2 \\ \downarrow & & \downarrow \\ T_1 & \xrightarrow{\varphi} & T_2 \\ \downarrow & & \downarrow \\ 0 & & 0 \end{array} \tag{13.2.1}$$

We obtain two faithful representations of $\mathrm{Hom}(T_1, T_2)$,

$$\rho_{\mathbf{C}} : \mathrm{Hom}(T_1, T_2) \to \mathrm{Hom}_{\mathbf{C}}(V_1, V_2)$$

$$\rho_{\mathbf{Z}} : \mathrm{Hom}(T_1, T_2) \to \mathrm{Hom}_{\mathbf{Z}}(L_1, L_2).$$

Suppose $\psi \in \mathrm{Hom}_{\mathbf{Z}}(L_1, L_2)$. Then $\psi = \rho_{\mathbf{Z}}(\varphi)$ for some $\varphi \in \mathrm{Hom}(T_1, T_2)$ if and only if there is a complex linear homomorphism $f : V_1 \to V_2$ whose restriction to L_1 is ψ. Note that $f = \psi \otimes \mathbf{R}$ is uniquely determined by ψ, so ψ arises from some φ precisely when f is $\mathbf{C}$-linear. This is the case if and only if $fJ_1 = J_2 f$, where $J_n : V_n \to V_n$ is the $\mathbf{R}$-linear map induced by multiplication by $i = \sqrt{-1} \in \mathbf{C}$.

Example 13.2.1.

1. Suppose $L_1 = \mathbf{Z} + \mathbf{Z}i \subset V_1 = \mathbf{C}$. Then with respect to the basis $1, i$, we have $J_1 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. One finds that $\mathrm{Hom}(T_1, T_1)$ is the free $\mathbf{Z}$-module of rank 2 whose image via $\rho_{\mathbf{Z}}$ is generated by J_1 and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. As a ring $\mathrm{Hom}(T_1, T_1)$ is isomorphic to $\mathbf{Z}[i]$.

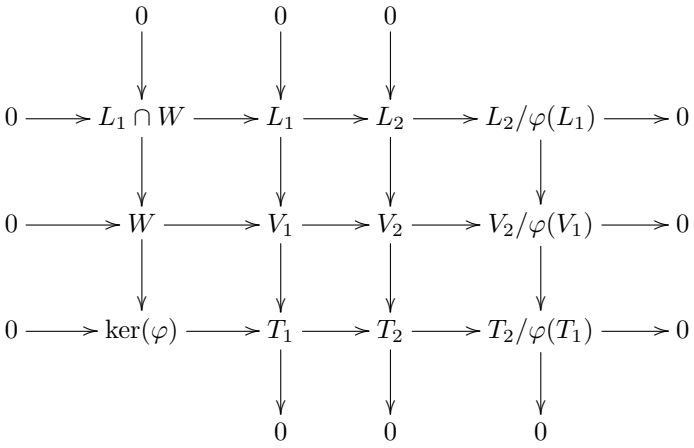
2. Suppose $L_1 = \mathbf{Z} + \mathbf{Z}\alpha i \subset V_1 = \mathbf{C}$, with $\alpha^3 = 2$. Then with respect to the basis $1, \alpha i$, we have $J_1 = \left(\begin{smallmatrix} 0 & -\alpha \\ 1/\alpha & 0 \end{smallmatrix}\right)$. Only the scalar integer matrices commute with J_1.

Proposition 13.2.2. *Let T_1 and T_2 be complex tori. Then* $\mathrm{Hom}(T_1, T_2)$ *is a free* $\mathbf{Z}$*-module of rank at most* $4 \dim T_1 \cdot \dim T_2$.

Proof. The representation $\rho_{\mathbf{Z}}$ is faithful (injective) because φ is determined by its action on L_1, since L_1 spans V_1. Thus $\mathrm{Hom}(T_1, T_2)$ is isomorphic to a subgroup of $\mathrm{Hom}_{\mathbf{Z}}(L_1, L_2) \cong \mathbf{Z}^d$, where $d = 2 \dim V_1 \cdot 2 \dim V_2$. □

Lemma 13.2.3. *Suppose $\varphi : T_1 \to T_2$ is a homomorphism of complex tori. Then the image of φ is a subtorus of T_2 and the connected component of* $\ker(\varphi)$ *is a subtorus of T_1 that has finite index in* $\ker(\varphi)$.

Proof. Let $W = \ker(\rho_{\mathbf{C}}(\varphi))$. Then the following diagram, which is induced by φ, has exact rows and columns:



Using the snake lemma, we obtain an exact sequence

$$0 \to L_1 \cap W \to W \to \ker(\varphi) \to L_2/\varphi(L_1) \to V_2/\varphi(V_1) \to T_2/\varphi(T_1) \to 0.$$

Note that $T_2/\varphi(T_1)$ is compact because it is the continuous image of a compact set, so the cokernel of φ is a torus (it is given as a quotient of a complex vector space by a lattice).

The kernel $\ker(\varphi) \subset T_1$ is a closed subset of the compact set T_1, so is compact. Thus $L_1 \cap W$ is a lattice in W. The map $L_2/\varphi(L_1) \to V_2/\varphi(V_1)$ has kernel generated by the saturation of $\varphi(L_1)$ in L_2, so it is finite, so the torus $W/(L_1 \cap W)$ has finite index in $\ker(\varphi)$. □

Remark 13.2.4. The category of complex tori is not an abelian category because kernels need not be in the category.

13.2.2 Isogenies

Definition 13.2.5 (Isogeny)**.** An *isogeny* $\varphi : T_1 \to T_2$ of complex tori is a surjective morphism with finite kernel. The *degree* $\deg(\varphi)$ of φ is the order of the kernel of φ.

Note that $\deg(\varphi \circ \varphi') = \deg(\varphi)\deg(\varphi')$.

Lemma 13.2.6. *Suppose that φ is an isogeny. Then the kernel of φ is isomorphic to the cokernel of $\rho_{\mathbf{Z}}(\varphi)$.*

Proof. (This is essentially a special case of Lemma 13.2.3.) Apply the snake lemma to the morphism (13.2.1) of short exact sequences, to obtain a six-term exact sequence

$$0 \to K_L \to K_V \to K_T \to C_L \to C_V \to C_T \to 0,$$

where K_X and C_X are the kernel and cokernel of $X_1 \to X_2$, for $X = L, V, T$, respectively. Since φ is an isogeny, the induced map $V_1 \to V_2$ must be an isomorphism, since otherwise the kernel would contain a nonzero subspace (modulo a lattice), which would be infinite. Thus $K_V = C_V = 0$. It follows that $K_T \cong C_L$, as claimed. □

One consequence of the lemma is that if φ is an isogeny, then

$$\deg(\varphi) = [L_1 : \rho_{\mathbf{Z}}(\varphi)(L_1)] = |\det(\rho_{\mathbf{Z}}(\varphi))|.$$

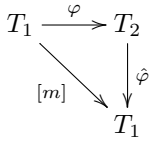
Proposition 13.2.7. *Let T be a complex torus of dimension d, and let n be a positive integer. Then multiplication by n, denoted $[n]$, is an isogeny $T \to T$ with kernel $T[n] \cong (\mathbf{Z}/n\mathbf{Z})^{2d}$ and degree n^{2d}.*

Proof. By Lemma 13.2.6, $T[n]$ is isomorphic to L/nL, where $T = V/L$. Since $L \approx \mathbf{Z}^{2d}$, the proposition follows. □

We can now prove that isogeny is an equivalence relation.

Proposition 13.2.8. *Suppose $\varphi : T_1 \to T_2$ is a degree m isogeny of complex tori of dimension d. Then there is a unique isogeny $\hat{\varphi} : T_2 \to T_1$ of degree m^{2d-1} such that $\hat{\varphi} \circ \varphi = \varphi \circ \hat{\varphi} = [m]$.*

Proof. Since $\ker(\varphi) \subset \ker([m])$, the map $[m]$ factors through φ, so there is a morphism $\hat{\varphi}$ such that $\hat{\varphi} \circ \varphi = [m]$:



We have

$$(\varphi \circ \hat{\varphi} - [m]) \circ \varphi = \varphi \circ \hat{\varphi} \circ \phi - [m] \circ \varphi = \varphi \circ \hat{\varphi} \circ \phi - \varphi \circ [m] = \varphi \circ (\hat{\varphi} \circ \phi - [m]) = 0.$$

This implies that $\varphi \circ \hat{\varphi} = [m]$, since φ is surjective. Uniqueness is clear since the difference of two such morphisms would vanish on the image of φ. To see that $\hat{\varphi}$ has degree m^{2d-1}, we take degrees on both sides of the equation $\hat{\varphi} \circ \varphi = [m]$. □

13.2.3 Endomorphisms

The ring $\mathrm{End}(T) = \mathrm{Hom}(T, T)$ is called the *endomorphism ring* of the complex torus T. The *endomorphism algebra* of T is $\mathrm{End}_0(T) = \mathrm{End}(T) \otimes_{\mathbf{Z}} \mathbf{Q}$.

Definition 13.2.9 (Characteristic polynomial)**.** The *characteristic polynomial* of $\varphi \in \mathrm{End}(T)$ is the characteristic polynomial of the $\rho_{\mathbf{Z}}(\varphi)$. Thus the characteristic polynomial is a monic polynomial of degree $2 \dim T$.

13.3 Abelian varieties as complex tori

In this section we introduce extra structure on a complex torus $T = V/L$ that will enable us to understand whether or not T is isomorphic to $A(\mathbf{C})$, for some abelian variety A over $\mathbf{C}$. When $\dim T = 1$, the theory of the Weierstrass $\wp$ function implies that T is always $E(\mathbf{C})$ for some elliptic curve. In contrast, the generic torus of dimension > 1 does not arise from an abelian variety.

In this section we introduce the basic structures on complex tori that are needed to understand which tori arise from abelian varieties, to construct the dual of an

abelian variety, to see that $\mathrm{End}_0(A)$ is a semisimple $\mathbf{Q}$-algebra, and to understand the polarizations on an abelian variety. For proofs, including extensive motivation from the one-dimensional case, read the beautifully written book [SD74] by Swinnerton-Dyer, and for another survey that strongly influenced the discussion below, see Rosen's [Ros86].

13.3.1 Hermitian and Riemann forms

Let V be a finite-dimensional complex vector space.

Definition 13.3.1 (Hermitian form)**.** A *Hermitian form* is a conjugate-symmetric pairing

$$H : V \times V \to \mathbf{C}$$

that is $\mathbf{C}$-linear in the first variable and $\mathbf{C}$-antilinear in the second. Thus H is $\mathbf{R}$-bilinear, $H(iu, v) = iH(u, v) = H(u, \bar{i}v)$, and $H(u, v) = \overline{H(v, u)}$.

Write $H = S + iE$, where $S, E : V \times V \to \mathbf{R}$ are real bilinear pairings.

Proposition 13.3.2. *Let H, S, and E be as above.*

1. *We have that S is symmetric, E is antisymmetric, and*

$$S(u, v) = E(iu, v), \quad S(iu, iv) = S(u, v), \quad E(iu, iv) = E(u, v).$$

2. *Conversely, if E is a real-valued antisymmetric bilinear pairing on V such that $E(iu, iv) = E(u, v)$, then $H(u, v) = E(iu, v) + iE(u, v)$ is a Hermitian form on V. Thus there is a bijection between the Hermitian forms on V and the real, antisymmetric bilinear forms E on V such that $E(iu, iv) = E(u, v)$.*

Proof. To see that S is symmetric, note that $2S = H + \overline{H}$ and $H + \overline{H}$ is symmetric because H is conjugate symmetric. Likewise, $E = (H - \overline{H})/(2i)$, so

$$E(v, u) = \frac{1}{2i}\left(H(v, u) - \overline{H(v, u)}\right) = \frac{1}{2i}\left(\overline{H(u, v)} - H(u, v)\right) = -E(u, v),$$

which implies that E is antisymmetric. To see that $S(u, v) = E(iu, v)$, rewrite both $S(u, v)$ and $E(iu, v)$ in terms of H and simplify to get an identity. The other two identities follow since

$$H(iu, iv) = iH(u, iv) = i\bar{i}H(u, v) = H(u, v).$$

Suppose $E : V \times V \to \mathbf{R}$ is as in the second part of the proposition. Then

$$H(iu, v) = E(i^2u, v) + iE(iu, v) = -E(u, v) + iE(iu, v) = iH(u, v),$$

and the other verifications of linearity and antilinearity are similar. For conjugate symmetry, note that

$$\begin{aligned} H(v, u) &= E(iv, u) + iE(v, u) = -E(u, iv) - iE(u, v) \\ &= -E(iu, -v) - iE(u, v) = H(u, v). \end{aligned}$$

□

Note that the set of Hermitian forms is a group under addition.

Definition 13.3.3 (Riemann form)**.** A *Riemann form* on a complex torus $T = V/L$ is a Hermitian form H on V such that the restriction of $E = \mathrm{Im}(H)$ to L is integer valued. If $H(u,u) \geq 0$ for all $u \in V$ then H is *positive semi-definite* and if H is positive and $H(u,u) = 0$ if and only if $u = 0$, then H is *nondegenerate*.

Theorem 13.3.4. *Let T be a complex torus. Then T is isomorphic to $A(\mathbf{C})$, for some abelian variety A, if and only if there is a nondegenerate Riemann form on T.*

This is a nontrivial theorem, which we will not prove here. It is proved in [SD74, Ch.2] by defining an injective map from positive divisors on $T = V/L$ to positive semi-definite Riemann forms, then constructing positive divisors associated to theta functions on V. If H is a nondegenerate Riemann form on T, one computes the dimension of a space of theta functions that corresponds to H in terms of the determinant of $E = \mathrm{Im}(H)$. Since H is nondegenerate, this space of theta functions is nonzero, so there is a corresponding nondegenerate positive divisor D. Then a basis for

$$L(3D) = \{f \,:\, (f) + 3D \text{ is positive } \} \cup \{0\}$$

determines an embedding of T in a projective space.

Why the divisor $3D$ instead of D above? For an elliptic curve $y^2 = x^3 + ax + b$, we could take D to be the point at infinity. Then $L(3D)$ consists of the functions with a pole of order at most 3 at infinity, which contains 1, x, and y, which have poles of order 0, 2, and 3, respectively.

Remark 13.3.5. (Copied from page 39 of [SD74].) When $n = \dim V > 1$, however, a general lattice L will admit no nonzero Riemann forms. For if $\lambda_1, \ldots, \lambda_{2n}$ is a base for L then E as an $\mathbf{R}$-bilinear alternating form is uniquely determined by the $E(\lambda_i, \lambda_j)$, which are integers; and the condition $E(z,w) = E(iz, iw)$ induces linear relations with real coefficients between $E(\lambda_i, \lambda_j)$, which for general L have no nontrivial integer solutions.

13.3.2 Complements, quotients, and semisimplicity of the endomorphism algebra

Lemma 13.3.6. *If T possesses a nondegenerate Riemann form and $T' \subset T$ is a subtorus, then T' also possesses a nondegenerate Riemann form.*

Proof. If H is a nondegenerate Riemann form on a torus T and T' is a subtorus of T, then the restriction of H to T' is a nondegenerate Riemann form on T' (the restriction is still nondegenerate because H is positive definite). □

Lemma 13.3.6 and Lemma 13.2.3 together imply that the kernel of a homomorphism of abelian varieties is an extension of an abelian variety by a finite group.

Lemma 13.3.7. *If T possesses a nondegenerate Riemann form and $T \to T'$ is an isogeny, then T' also possesses a nondegenerate Riemann form.*

Proof. Suppose $T = V/L$ and $T' = V'/L'$. Since the isogeny is induced by an isomorphism $V \to V'$ that sends L into L', we may assume for simplicity that $V = V'$ and $L \subset L'$. If H is a nondegenerate Riemann form on V/L, then $E = \mathrm{Re}(H)$ need not be integer valued on L'. However, since L has finite index in L', there

is some integer d so that dE is integer valued on L'. Then dH is a nondegenerate Riemann form on V/L'. □

Note that Lemma 13.3.7 implies that the quotient of an abelian variety by a finite subgroup is again an abelian variety.

Theorem 13.3.8 (Poincare Reducibility)**.** *Let A be an abelian variety and suppose $A' \subset A$ is an abelian subvariety. Then there is an abelian variety $A'' \subset A$ such that $A = A' + A''$ and $A' \cap A''$ is finite. (Thus A is isogenous to $A' \times A''$.)*

Proof. We have $A(\mathbf{C}) \approx V/L$ and there is a nondegenerate Riemann form H on V/L. The subvariety A' is isomorphic to V'/L', where V' is a subspace of V and $L' = V' \cap L$. Let V'' be the orthogonal complement of V' with respect to H, and let $L'' = L \cap V''$. To see that L'' is a lattice in V'', it suffices to show that L'' is the orthogonal complement of L' in L with respect to $E = \mathrm{Im}(H)$, which, because E is integer valued, will imply that L'' has the correct rank. First, suppose that $v \in L''$; then, by definition, v is in the orthogonal complement of L' with respect to H, so for any $u \in L'$, we have $0 = H(u, v) = S(u, v) + iE(u, v)$, so $E(u, v) = 0$. Next, suppose that $v \in L$ satisfies $E(u, v) = 0$ for all $u \in L'$. Since $V' = \mathbf{R}L'$ and E is $\mathbf{R}$-bilinear, this implies $E(u, v) = 0$ for any $u \in V'$. In particular, since V' is a complex vector space, if $u \in L'$, then $S(u, v) = E(iu, v) = 0$, so $H(u, v) = 0$.

We have shown that L'' is a lattice in V'', so $A'' = V''/L''$ is an abelian subvariety of A. Also $L' + L''$ has finite index in L, so there is an isogeny $V'/L' \oplus V''/L'' \to V/L$ induced by the natural inclusions. □

Proposition 13.3.9. *Suppose $A' \subset A$ is an inclusion of abelian varieties. Then the quotient A/A' is an abelian variety.*

Proof. Suppose $A = V/L$ and $A' = V'/L'$, where V' is a subspace of V. Let $W = V/V'$ and $M = L/(L \cap V')$. Then, W/M is isogenous to the complex torus V''/L'' of Theorem 13.3.8 via the natural map $V'' \to W$. Applying Lemma 13.3.7 completes the proof. □

Definition 13.3.10. An abelian variety A is *simple* if it has no nonzero proper abelian subvarieties.

Proposition 13.3.11. *The algebra* $\mathrm{End}_0(A)$ *is semisimple.*

Proof. Using Theorem 13.3.8 and induction, we can find an isogeny

$$A \simeq A_1^{n_1} \times A_2^{n_2} \times \cdots \times A_r^{n_r}$$

with each A_i simple. Since $\mathrm{End}_0(A) = \mathrm{End}(A) \otimes \mathbf{Q}$ is unchanged by isogeny, and $\mathrm{Hom}(A_i, A_j) = 0$ when $i \neq j$, we have

$$\mathrm{End}_0(A) = \mathrm{End}_0(A_1^{n_1}) \times \mathrm{End}_0(A_2^{n_2}) \times \cdots \times \mathrm{End}_0(A_r^{n_r})$$

Each of $\mathrm{End}_0(A_i^{n_i})$ is isomorphic to $M_{n_i}(D_i)$, where $D_i = \mathrm{End}_0(A_i)$. By Schur's Lemma, $D_i = \mathrm{End}_0(A_i)$ is a division algebra over $\mathbf{Q}$ (proof: any nonzero endomorphism has trivial kernel, and any injective linear transformation of a $\mathbf{Q}$-vector space is invertible), so $\mathrm{End}_0(A)$ is a product of matrix algebras over division algebras over $\mathbf{Q}$, which proves the proposition. □

13.3.3 Theta functions

Suppose $T = V/L$ is a complex torus.

Definition 13.3.12 (Theta function)**.** Let $M : V \times L \to C$ and $J : L \to \mathbf{C}$ be set-theoretic maps such that for each $\lambda \in L$ the map $z \mapsto M(z, \lambda)$ is $\mathbf{C}$-linear. A *theta function* of type (M, J) is a function $\theta : V \to \mathbf{C}$ such that for all $z \in V$ and $\lambda \in L$, we have

$$\theta(z + \lambda) = \theta(z) \cdot \exp(2\pi i(M(z, \lambda) + J(\lambda))).$$

Suppose that $\theta(z)$ is a nonzero holomorphic theta function of type (M, J). The $M(z, \lambda)$, for various λ, cannot be unconnected. Let $F(z, \lambda) = 2\pi i(M(z, \lambda) + J(\lambda))$.

Lemma 13.3.13. *For any $\lambda, \lambda' \in L$, we have*

$$F(z, \lambda + \lambda') = F(z + \lambda, \lambda') + F(z, \lambda) \pmod{2\pi i}.$$

Thus

$$M(z, \lambda + \lambda') = M(z, \lambda) + M(z, \lambda'), \tag{13.3.1}$$

and

$$J(\lambda + \lambda') - J(\lambda) - J(\lambda') \equiv M(\lambda, \lambda') \pmod{\mathbf{Z}}.$$

Proof. Page 37 of [SD74]. □

Using (13.3.1) we see that M extends uniquely to a function $\tilde{M} : V \times V \to \mathbf{C}$ which is $\mathbf{C}$-linear in the first argument and $\mathbf{R}$-linear in the second. Let

$$E(z, w) = \tilde{M}(z, w) - M(w, z),$$

$$H(z, w) = E(iz, w) + iE(z, w).$$

Proposition 13.3.14. *The pairing H is Riemann form on T with real part E.*

We call H the Riemann form associated to θ.

13.4 A Summary of duality and polarizations

Suppose A is an abelian variety over an arbitrary field k. In this section we summarize the most important properties of the dual abelian variety $A^\vee$ of A. First we review the language of sheaves on a scheme X, and define the Picard group of X as the group of invertible sheaves on X. The dual of A is then a variety whose points correspond to elements of the Picard group that are algebraically equivalent to 0. Next, when the ground field is $\mathbf{C}$, we describe how to view $A^\vee$ as a complex torus in terms of a description of A as a complex torus. We then define the Néron-Severi group of A and relate it to polarizations of A, which are certain homomorphisms $A \to A^\vee$. Finally we observe that the dual is functorial.

13.4.1 Sheaves

We will use the language of sheaves, as in [Har77], which we now quickly recall. A *pre-sheaf of abelian groups* $\mathcal{F}$ on a scheme X is a contravariant functor from the category of open sets on X (morphisms are inclusions) to the category of abelian groups. Thus for every open set $U \subset X$ there is an abelian group $\mathcal{F}(U)$, and if $U \subset V$, then there is a restriction map $\mathcal{F}(V) \to \mathcal{F}(U)$. (We also require that $\mathcal{F}(\emptyset) = 0$, and the map $\mathcal{F}(U) \to \mathcal{F}(U)$ is the identity map.) A *sheaf* is a pre-sheaf whose sections are determined locally (for details, see [Har77, §II.1]).

Every scheme X is equipped with its structure sheaf $\mathcal{O}_X$, which has the property that if $U = \mathrm{Spec}(R)$ is an affine open subset of X, then $\mathcal{O}_X(U) = R$. A *sheaf of $\mathcal{O}_X$-modules* is a sheaf $\mathcal{M}$ of abelian groups on X such that each abelian group has the structure of $\mathcal{O}_X$-module, such that the restriction maps are module morphisms. A *locally-free sheaf* of $\mathcal{O}_X$-modules is a sheaf $\mathcal{M}$ of $\mathcal{O}_X$-modules, such that X can be covered by open sets U so that $\mathcal{M}|_U$ is a free $\mathcal{O}_X$-module, for each U.

13.4.2 The Picard group

An *invertible sheaf* is a sheaf $\mathcal{L}$ of $\mathcal{O}_X$-modules that is locally free of rank 1. If $\mathcal{L}$ is an invertible sheaf, then the sheaf-theoretic Hom, $\mathcal{L}^\vee = \mathcal{H}\mathrm{om}(\mathcal{L}, \mathcal{O}_X)$ has the property that $\mathcal{L}^\vee \otimes \mathcal{L} = \mathcal{O}_X$. The group $\mathrm{Pic}(X)$ of invertible sheaves on a scheme X is called *the Picard group* of X. See [Har77, §II.6] for more details.

Let A be an abelian variety over a field k. An invertible sheaf $\mathcal{L}$ on A is *algebraically equivalent to* 0 if there is a connected variety T over k, an invertible sheaf $\mathcal{M}$ on $A \times_k T$, and $t_0, t_1 \in T(k)$ such that $\mathcal{M}_{t_0} \cong \mathcal{L}$ and $\mathcal{M}_{t_1} \cong \mathcal{O}_A$. Let $\mathrm{Pic}^0(A)$ be the subgroup of elements of $\mathrm{Pic}(A)$ that are algebraically equivalent to 0.

The *dual* $A^\vee$ of A is a (unique up to isomorphism) abelian variety such that for every field F that contains the base field k, we have $A^\vee(F) = \mathrm{Pic}^0(A_F)$. For the precise definition of $A^\vee$ and a proof that $A^\vee$ exists, see [Mil86, §9–10].

13.4.3 The Dual as a complex torus

When A is defined over the complex numbers, so $A(\mathbf{C}) = V/L$ for some vector space V and some lattice L, [Ros86, §4] describes a construction of $A^\vee$ as a complex torus, which we now describe. Let

$$V^* = \{f \in \mathrm{Hom}_{\mathbf{R}}(V, \mathbf{C}) \,:\, f(\alpha t) = \overline{\alpha} f(t), \ \text{ all } \ \alpha \in \mathbf{C},\, t \in V\}.$$

Then V^* is a complex vector space of the same dimension as V and the map $\langle f, v\rangle = \mathrm{Im}\, f(t)$ is an $\mathbf{R}$-linear pairing $V^* \times V \to \mathbf{R}$. Let

$$L^* = \{f \in V^* \,:\, \langle f, \lambda\rangle \in \mathbf{Z}, \ \text{all } \lambda \in L\}.$$

Since A is an abelian variety, there is a nondegenerate Riemann form H on A. The map $\lambda : V \to V^*$ defined by $\lambda(v) = H(v, \cdot)$ is an isomorphism of complex vector spaces. If $v \in L$, then $\lambda(v) = H(v, \cdot)$ is integer valued on L, so $\lambda(L) \subset L^*$. Thus λ induces an isogeny of complex tori $V/L \to V^*/L^*$, so by Lemma 13.3.7 the torus V^*/L^* possesses a nondegenerate Riemann form (it's a multiple of H). In [Ros86, §4], Rosen describes an explicit isomorphism between V^*/L^* and $A^\vee(\mathbf{C})$.

13.4.4 The Néron-Severi group and polarizations

Let A be an abelian variety over a field k. Recall that $\mathrm{Pic}(A)$ is the group of invertible sheaves on A, and $\mathrm{Pic}^0(A)$ is the subgroup of invertible sheaves that are algebraically equivalent to 0. The *Néron-Severi group* of A is the quotient $\mathrm{Pic}(A)/\mathrm{Pic}^0(A)$, so by definition we have an exact sequence

$$0 \to \mathrm{Pic}^0(A) \to \mathrm{Pic}(A) \to \mathrm{NS}(A) \to 0.$$

Suppose $\mathcal{L}$ is an invertible sheaf on A. One can show that the map $A(k) \to \mathrm{Pic}^0(A)$ defined by $a \mapsto t_a^*\mathcal{L} \otimes \mathcal{L}^{-1}$ is induced by homomorphism $\varphi_{\mathcal{L}} : A \to A^\vee$. (Here $t_a^*\mathcal{L}$ is the pullback of the sheaf $\mathcal{L}$ by translation by a.) Moreover, the map $\mathcal{L} \mapsto \varphi_{\mathcal{L}}$ induces a homomorphism from $\mathrm{Pic}(A) \to \mathrm{Hom}(A, A^\vee)$ with kernel $\mathrm{Pic}^0(A)$. The group $\mathrm{Hom}(A, A^\vee)$ is free of finite rank, so $\mathrm{NS}(A)$ is a free abelian group of finite rank. Thus $\mathrm{Pic}^0(A)$ is saturated in $\mathrm{Pic}(A)$ (i.e., the cokernel of the inclusion $\mathrm{Pic}^0(A) \to \mathrm{Pic}(A)$ is torsion free).

Definition 13.4.1 (Polarization)**.** A *polarization* on A is a homomorphism $\lambda : A \to A^\vee$ such that $\lambda_{\overline{k}} = \varphi_{\mathcal{L}}$ for some $\mathcal{L} \in \mathrm{Pic}(A_{\overline{k}})$. A polarization is *principal* if it is an isomorphism.

When the base field k is algebraically closed, the polarizations are in bijection with the elements of $\mathrm{NS}(A)$. For example, when $\dim A = 1$, we have $\mathrm{NS}(A) = \mathbf{Z}$, and the polarizations on A are multiplication by n, for each integer n.

13.4.5 The Dual is functorial

The association $A \mapsto A^\vee$ extends to a contravariant functor on the category of abelian varieties. Thus if $\varphi : A \to B$ is a homomorphism, there is a natural choice of homomorphism $\varphi^\vee : B^\vee \to A^\vee$. Also, $(A^\vee)^\vee = A$ and $(\varphi^\vee)^\vee = \varphi$.

Theorem 13.4.2 below describes the kernel of $\varphi^\vee$ in terms of the kernel of φ. If G is a finite group scheme, the *Cartier dual* of G is $\mathrm{Hom}(G, \mathbf{G}_m)$. For example, the Cartier dual of $\mathbf{Z}/m\mathbf{Z}$ is μ_m and the Cartier dual of μ_m is $\mathbf{Z}/m\mathbf{Z}$. (If k is algebraically closed, then the Cartier dual of G is just G again.)

Theorem 13.4.2. *If $\varphi : A \to B$ is a surjective homomorphism of abelian varieties with kernel G, so we have an exact sequence $0 \to G \to A \to B \to 0$, then the kernel of $\varphi^\vee$ is the Cartier dual of G, so we have an exact sequence $0 \to G^\vee \to B^\vee \to A^\vee \to 0$.*

13.5 Jacobians of curves

We begin this lecture about Jacobians with an inspiring quote of David Mumford:

> "The Jacobian has always been a corner-stone in the analysis of algebraic curves and compact Riemann surfaces. [...] Weil's construction [of the Jacobian] was the basis of his epoch-making proof of the Riemann Hypothesis for curves over finite fields, which really put characteristic p algebraic geometry on its feet." – Mumford, *Curves and Their Jacobians*, page 49.

13.5.1 Divisors on curves and linear equivalence

Let X be a projective nonsingular algebraic curve over an algebraically field k. A *divisor* on X is a formal finite $\mathbf{Z}$-linear combination $\sum_{i=1}^m n_i P_i$ of closed points in X. Let $\mathrm{Div}(X)$ be the group of all divisors on X. The *degree* of a divisor $\sum_{i=1}^m n_i P_i$ is the integer $\sum_{i=1}^m n_i$. Let $\mathrm{Div}^0(X)$ denote the subgroup of divisors of degree 0.

Suppose k is a perfect field (for example, k has characteristic 0 or k is finite), but do not require that k be algebraically closed. Let the group of divisors on X over k be the subgroup

$$\mathrm{Div}(X) = \mathrm{Div}(X/k) = \mathrm{H}^0(\mathrm{Gal}(\overline{k}/k), \mathrm{Div}(X/\overline{k}))$$

of elements of $\mathrm{Div}(X/\overline{k})$ that are fixed by every automorphism of $\overline{k}/k$. Likewise, let $\mathrm{Div}^0(X/k)$ be the elements of $\mathrm{Div}(X/k)$ of degree 0.

A *rational function* on an algebraic curve X is a function $X \to \mathbf{P}^1$, defined by polynomials, which has only a finite number of poles. For example, if X is the elliptic curve over k defined by $y^2 = x^3 + ax + b$, then the field of rational functions on X is the fraction field of the integral domain $k[x, y]/(y^2 - (x^3 + ax + b))$. Let $K(X)$ denote the field of all rational functions on X defined over k.

There is a natural homomorphism $K(X)^* \to \mathrm{Div}(X)$ that associates to a rational function f its divisor

$$(f) = \sum \mathrm{ord}_P(f) \cdot P$$

where $\mathrm{ord}_P(f)$ is the order of vanishing of f at P. Since X is nonsingular, the local ring of X at a point P is isomorphic to $k[[t]]$. Thus we can write $f = t^r g(t)$ for some unit $g(t) \in k[[t]]$. Then $R = \mathrm{ord}_P(f)$.

Example 13.5.1. If $X = \mathbf{P}^1$, then the function $f = x$ has divisor $(0) - (\infty)$. If X is the elliptic curve defined by $y^2 = x^3 + ax + b$, then

$$(x) = (0, \sqrt{b}) + (0, -\sqrt{b}) - 2\infty,$$

and

$$(y) = (x_1, 0) + (x_2, 0) + (x_3, 0) - 3\infty,$$

where x_1, x_2, and x_3 are the roots of $x^3 + ax + b = 0$. A uniformizing parameter t at the point ∞ is x/y. An equation for the elliptic curve in an affine neighborhood of ∞ is $Z = X^3 + aXZ^2 + bZ^3$ (where $\infty = (0, 0)$ with respect to these coordinates) and $x/y = X$ in these new coordinates. By repeatedly substituting Z into this equation we see that Z can be written in terms of X.

It is a standard fact in the theory of algebraic curves that if f is a nonzero rational function, then $(f) \in \mathrm{Div}^0(X)$, i.e., the number of poles of f equals the number of zeros of f. For example, if X is the Riemann sphere and f is a polynomial, then the number of zeros of f (counted with multiplicity) equals the degree of f, which equals the order of the pole of f at infinity.

The *Picard group* $\mathrm{Pic}(X)$ of X is the group of divisors on X modulo linear equivalence. Since divisors of functions have degree 0, the subgroup $\mathrm{Pic}^0(X)$ of divisors on X of degree 0, modulo linear equivalence, is well defined. Moreover, we have an exact sequence of abelian groups

$$0 \to K(X)^* \to \mathrm{Div}^0(X) \to \mathrm{Pic}^0(X) \to 0.$$

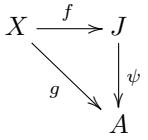
Thus for any algebraic curve X we have associated to it an abelian group $\mathrm{Pic}^0(X)$. Suppose $\pi : X \to Y$ is a morphism of algebraic curves. If D is a divisor on Y, the pullback $\pi^*(D)$ is a divisor on X, which is defined as follows. If $P \in \mathrm{Div}(Y/\overline{k})$ is a point, let $\pi^*(P)$ be the sum $\sum e_{Q/P} Q$ where $\pi(Q) = P$ and $e_{Q/P}$ is the ramification degree of Q/P. (Remark: If t is a uniformizer at P then $e_{Q/P} = \mathrm{ord}_Q(\pi^* t_P)$.) One can show that $\pi^* : \mathrm{Div}(Y) \to \mathrm{Div}(X)$ induces a homomorphism $\mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X)$. Furthermore, we obtain the contravariant *Picard functor* from the category of algebraic curves over a fixed base field to the category of abelian groups, which sends X to $\mathrm{Pic}^0(X)$ and $\pi : X \to Y$ to $\pi^* : \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X)$.

Alternatively, instead of defining morphisms by pullback of divisors, we could consider the push forward. Suppose $\pi : X \to Y$ is a morphism of algebraic curves and D is a divisor on X. If $P \in \mathrm{Div}(X/\overline{k})$ is a point, let $\pi_*(P) = \pi(P)$. Then π_* induces a morphism $\mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y)$. We again obtain a functor, called the covariant *Albanese functor* from the category of algebraic curves to the category of abelian groups, which sends X to $\mathrm{Pic}^0(X)$ and $\pi : X \to Y$ to $\pi_* : \mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y)$.

13.5.2 Algebraic definition of the Jacobian

First we describe some universal properties of the Jacobian under the hypothesis that $X(k) \neq \emptyset$. Thus suppose X is an algebraic curve over a field k and that $X(k) \neq \emptyset$. The Jacobian variety of X is an abelian variety J such that for an extension k'/k, there is a (functorial) isomorphism $J(k') \to \mathrm{Pic}^0(X/k')$. (I don't know whether this condition uniquely characterizes the Jacobian.)

Fix a point $P \in X(k)$. Then we obtain a map $f : X(k) \to \mathrm{Pic}^0(X/k)$ by sending $Q \in X(k)$ to the divisor class of $Q - P$. One can show that this map is induced by an injective morphism of algebraic varieties $X \hookrightarrow J$. This morphism has the following universal property: if A is an abelian variety and $g : X \to A$ is a morphism that sends P to $0 \in A$, then there is a unique homomorphism $\psi : J \to A$ of abelian varieties such that $g = \psi \circ f$:



This condition uniquely characterizes J, since if $f' : X \to J'$ and J' has the universal property, then there are unique maps $J \to J'$ and $J' \to J$ whose composition in both directions must be the identity (use the universal property with $A = J$ and $f = g$).

If X is an arbitrary curve over an arbitrary field, the Jacobian is an abelian variety that represents the "sheafification" of the "relative Picard functor". Look in Milne's article or Bosch-Lüktebohmert-Raynaud *Neron Models* for more details. Knowing this totally general definition won't be important for this course, since we will only consider Jacobians of modular curves, and these curves always have a rational point, so the above properties will be sufficient.

A useful property of Jacobians is that they are canonically principally polarized, by a polarization that arises from the "θ divisor" on J. In particular, there is always an isomorphism $J \to J^\vee = \mathrm{Pic}^0(J)$.

13.5.3 The Abel-Jacobi theorem

Over the complex numbers, the construction of the Jacobian is classical. It was first considered in the 19th century in order to obtain relations between integrals of rational functions over algebraic curves (see Mumford's book, *Curves and Their Jacobians*, Ch. III, for a nice discussion).

Let X be a Riemann surface, so X is a one-dimensional complex manifold. Thus there is a system of coordinate charts (U_α, t_α), where $t_\alpha : U_\alpha \to \mathbf{C}$ is a homeomorphism of U_α onto an open subset of $\mathbf{C}$, such that the change of coordinate maps are analytic isomorphisms. A *differential* 1*-form* on X is a choice of two continuous functions f and g to each local coordinate $z = x + iy$ on $U_\alpha \subset X$ such that $f\,dx + g\,dy$ is invariant under change of coordinates (i.e., if another local coordinate patch U'_α intersects U_α, then the differential is unchanged by the change of coordinate map on the overlap). If $\gamma : [0,1] \to X$ is a path and $\omega = f\,dx + g\,dy$ is a 1-form, then

$$\int_\gamma \omega := \int_0^1 \left(f(x(t), y(t)) \frac{dx}{dt} + g(x(t), y(t)) \frac{dy}{dt} \right) dt \in \mathbf{C}.$$

From complex analysis one sees that if γ is homologous to γ', then $\int_\gamma \omega = \int_{\gamma'} \omega$. In fact, there is a nondegenerate pairing

$$\mathrm{H}^0(X, \Omega^1_X) \times \mathrm{H}_1(X, \mathbf{Z}) \to \mathbf{C}$$

If X has genus g, then it is a standard fact that the complex vector space $\mathrm{H}^0(X, \Omega^1_X)$ of holomorphic differentials on X is of dimension g. The integration pairing defined above induces a homomorphism from integral homology to the dual V of the differentials:

$$\Phi : \mathrm{H}_1(X, \mathbf{Z}) \to V = \mathrm{Hom}(H^0(X, \Omega^1_X), \mathbf{C}).$$

This homomorphism is called the *period mapping*.

Theorem 13.5.2 (Abel-Jacobi)**.** *The image of* Φ *is a lattice in* V*.*

The proof involves repeated clever application of the residue theorem.

The intersection pairing

$$\mathrm{H}_1(X, \mathbf{Z}) \times \mathrm{H}_1(X, \mathbf{Z}) \to \mathbf{Z}$$

defines a nondegenerate alternating pairing on $L = \Phi(\mathrm{H}_1(X, \mathbf{Z}))$. This pairing satisfies the conditions to induce a nondegenerate Riemann form on V, which gives $J = V/L$ to structure of abelian variety. The abelian variety J is the Jacobian of X, and if $P \in X$, then the functional $\omega \mapsto \int_P^Q \omega$ defines an embedding of X into J. Also, since the intersection pairing is perfect, it induces an isomorphism from J to $J^\vee$.

Example 13.5.3. For example, suppose $X = X_0(23)$ is the modular curve attached to the subgroup $\Gamma_0(23)$ of matrices in $\mathrm{SL}_2(\mathbf{Z})$ that are upper triangular modulo 24. Then $g = 2$, and a basis for $\mathrm{H}_1(X_0(23), \mathbf{Z})$ in terms of modular symbols is

$$\{-1/19, 0\}, \quad \{-1/17, 0\}, \quad \{-1/15, 0\}, \quad \{-1/11, 0\}.$$

The matrix for the intersection pairing on this basis is

$$\begin{pmatrix} 0 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

With respect to a reduced integral basis for

$$\mathrm{H}^0(X, \Omega^1_X) \cong S_2(\Gamma_0(23)),$$

the lattice $\Phi(\mathrm{H}_1(X, \mathbf{Z}))$ of periods is (approximately) spanned by

```
[
    (0.59153223605591049412844857432 - 1.68745927346801253993135357636*i
        0.76280632445804716868108032384657147872 7 - 0.60368764497868211035115379488*i),
    (-0.59153223605591049412844857432 - 1.68745927346801253993135357636*i
        -0.76280632445804716868108032384657147872 7 - 0.60368764497868211035115379488*i),
    (-1.3543385605139576626809528899804 - 1.08377162848933042958019978085687487140 97*i
        -0.59153223605591049412844857401 + 0.48008398351064831922904598746 7*i),
    (-1.525612648916094337362160650 99 0.342548176804273349105263499648)
]
```

13.5.4 Every abelian variety is a quotient of a Jacobian

Over an infinite field, *every* abelin variety can be obtained as a quotient of a Jacobian variety. The modular abelian varieties that we will encounter later are, by definition, exactly the quotients of the Jacobian $J_1(N)$ of $X_1(N)$ for some N. In this section we see that merely being a quotient of a Jacobian does not endow an abelian variety with any special properties.

Theorem 13.5.4 (Matsusaka)**.** *Let A be an abelian variety over an algebraically closed field. Then there is a Jacobian J and a surjective map $J \to A$.*

This was originally proved in *On a generating curve of an abelian variety*, Nat. Sc. Rep. Ochanomizu Univ. **3** (1952), 1–4. Here is the Math Review by P. Samuel:

> An abelian variety A is said to be generated by a variety V (and a mapping f of V into A) if A is the group generated by $f(V)$. It is proved that every abelian variety A may be generated by a curve defined over the algebraic closure of def(A). A first lemma shows that, if a variety V is the carrier of an algebraic system $(X(M))_{M \in U}$ of curves ($X(M)$ being defined, non-singular and disjoint from the singular bunch of V for almost all M in the parametrizing variety U) if this system has a simple base point on V, and if a mapping f of V into an abelian variety is constant on some $X(M_0)$, then f is a constant; this is proved by specializing on M_0 a generic point M of U and by using specializations of cycles [Matsusaka, Mem. Coll. Sci. Kyoto Univ. Ser. A. Math. 26, 167–173 (1951); these Rev. 13, 379]. Another lemma notices that, for a normal projective variety V, a suitable linear family of plane sections of V may be taken as a family $(X(M))$. Then the main result follows from the complete reducibility theorem. This result is said to be the

basic tool for generalizing Chow's theorem ("the Jacobian variety of a curve defined over k is an abelian projective variety defined over k").

Milne [Mil86, §10] proves the theorem under the weaker hypothesis that the base field is infinite. We briefly sketch his proof now. If $\dim A = 1$, then A is the Jacobian of itself, so we may assume $\dim A > 1$. Embed A into $\mathbf{P}^n$, then, using the Bertini theorem, cut $A \subset \mathbf{P}^n$ by hyperplane sections $\dim(A) - 1$ times to obtain a nonsingular curve C on A of the form $A \cap V$, where V is a linear subspace of $\mathbf{P}^n$. Using standard arguments from Hartshorne [Har77], Milne shows (Lemma 10.3) that if W is a nonsingular variety and $\pi : W \to A$ is a finite morphism, then $\pi^{-1}(C)$ is geometrically connected (the main point is that the pullback of an ample invertible sheaf by a finite morphism is ample). (A morphism $f : X \to Y$ is *finite* if for every open affine subset $U = \mathrm{Spec}(R) \subset Y$, the inverse image $f^{-1}(U) \subset X$ is an affine open subset $\mathrm{Spec}(B)$ with B a finitely generated R-**module**. Finite morphisms have finite fibers, but not conversely.) We assume this lemma and deduce the theorem.

Let J be the Jacobian of C; by the universal property of Jacobians there is a unique homomorphism $f : J \to A$ coming from the inclusion $C \hookrightarrow A$. The image $A_1 = f(J)$ is an abelian subvariety since images of homomorphisms of abelian varieties are abelian varieties. By the Poincare reducibility theorem (we only proved this over $\mathbf{C}$, but it is true in general), there is an abelian subvariety $A_2 \subset A$ such that $A_1 + A_2 = A$ and $A_1 \cap A_2$ is finite. The isogeny $g : A_1 \times A_2 \to A$ given by $g(x, y) = x + y \in A$ is a finite morphism (any isogeny of abelian varieties is finite, flat, and surjective by Section 8 of [Mil86]). The inverse image $g^{-1}(A_1)$ is a union of $\#(A_1 \cap A_2)$ irreducible components; if this intersection is nontrivial, then likewise $g^{-1}(C)$ is reducible, which is a contradiction. This does not complete the proof, since it is possible that g is an isomorphism, so we use one additional trick. Suppose n is a positive integer coprime to the residue characteristic, and let

$$h = 1 \times [n] : A_1 \times A_2 \to A_1 \times A_2$$

be the identity map on the first factor and multiplication by n on the second. Then h is finite and $(h \circ g)^{-1}(A_1)$ is a union of $n^{2 \dim A_2} = \deg(h)$ irreducible components, hence $(h \circ g)^{-1}(C)$ is reducible, a contradiction.

Question 13.5.5. Is Theorem 13.5.4 false for some abelian variety A over some finite field k?

Question 13.5.6 (Milne)**.** Using the theorem we can obtain a sequence of Jacobian varieties $J_1, J_2, \ldots$ that form a complex

$$\cdots \to J_2 \to J_1 \to A \to 0.$$

(In each case the image of J_{i+1} is the connected component of the kernel of $J_i \to J_{i-1}$.) Is it possible to make this construction in such a way that the sequence terminates in 0?

Question 13.5.7 (Yau)**.** Let A be an abelian variety. What can be said about the minimum of the dimensions of all Jacobians J such that there is a surjective morphism $J \to A$?

Remark 13.5.8. Brian Conrad has explained to the author that if A is an abelian variety over an infinite field, then A can be embedded in a Jacobian J. This does

not follow directly from Theorem 13.5.4 above, since if $J \twoheadrightarrow A^\vee$, then the dual map $A \to J$ need not be injective.

13.6 Néron models

The main references for Néron models are as follows:

1. [AEC2]: Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*. Chapter IV of this book contains an extremely well written and motivated discussion of Néron models of elliptic curves over Dedekind domains with perfect residue field. In particular, Silverman gives an almost complete construction of Néron models of elliptic curves. Silverman very clearly really wants his reader to understand the construction. Highly recommended.

2. [BLR]: Bosch, Lütkebohmert, Raynaud, *Néron Models*. This is an excellent and accessible book that contains a complete construction of Néron models and some of their generalizations, a discussion of their functorial properties, and a sketch of the construction of Jacobians of families of curves. The goal of this book was to redo in scheme-theoretic language Néron original paper, which is written in a language that was ill-adapted to the subtleties of Néron models.

3. Artin, *Néron Models*, in Cornell-Silverman. This is the first-ever exposition of Néron's original paper in the language of schemes.

13.6.1 What are Néron models?

Suppose E is an elliptic curve over $\mathbf{Q}$. If Δ is the minimal discriminant of E, then E has good reduction at p for all $p \nmid \Delta$, in the sense that E extends to an abelian scheme $\mathcal{E}$ over $\mathbf{Z}_p$ (i.e., a "smooth" and "proper" group scheme). One can not ask for E to extend to an abelian scheme over $\mathbf{Z}_p$ for all $p \mid \Delta$. One can, however, ask whether there is a notion of "good" model for E at these bad primes. To quote [BLR, page 1], "It came as a surprise for arithmeticians and algebraic geometers when A. Néron, relaxing the condition of properness and concentrating on the group structure and the smoothness, discovered in the years 1961–1963 that such models exist in a canonical way."

Before formally defining Néron models, we describe what it means for a morphism $f : X \to Y$ of schemes to be smooth. A morphism $f : X \to Y$ is finite type if for every open affine $U = \mathrm{Spec}(R) \subset Y$ there is a finite covering of $f^{-1}(U)$ by open affines $\mathrm{Spec}(S)$, such that each S is a finitely generated R-algebra.

Definition 13.6.1. A morphism $f : X \to Y$ is *smooth at* $x \in X$ if it is of finite type and there are open affine neighborhoods $\mathrm{Spec}(A) \subset X$ of x and $\mathrm{Spec}(R) \subset Y$ of $f(x)$ such that

$$A \cong R[t_1, \ldots, t_{n+r}]/(f_1, \ldots, f_n)$$

for elements $f_1, \ldots, f_n \in R[t_1, \ldots, t_{n+r}]$ and all $n \times n$ minors of the Jacobian matrix $(\partial f_i / \partial t_j)$ generate the unit ideal of A. The morphism f is *étale* at x if, in addition, $r = 0$. A morphism is *smooth of relative dimension* d if it is smooth at x for every $x \in X$ and $r = d$ in the isomorphism above.

Smooth morphisms behave well. For example, if f and g are smooth and $f \circ g$ is defined, then $f \circ g$ is automatically smooth. Also, smooth morphisms are closed under base extension: if $f : X \to Y$ is a smooth morphism over S, and S' is a scheme over S, then the induced map $X \times_S S' \to Y \times_S S'$ is smooth. (If you've never seen products of schemes, it might be helpful to know that $\mathrm{Spec}(A) \times \mathrm{Spec}(B) = \mathrm{Spec}(A \otimes B)$. Read [Har77, §II.3] for more information about fiber products, which provide a geometric way to think about tensor products. Also, we often write $X_{S'}$ as shorthand for $X \times_S S'$.)

We are now ready for the definition. Suppose R is a Dedekind domain with field of fractions K (e.g., $R = \mathbf{Z}$ and $K = \mathbf{Q}$).

Definition 13.6.2 (Néron model)**.** Let A be an abelian variety over K. The *Néron model* $\mathcal{A}$ of A is a smooth commutative group scheme over R such that for any smooth morphism $S \to R$ the natural map of abelian groups

$$\mathrm{Hom}_R(S, \mathcal{A}) \to \mathrm{Hom}_K(S \times_R K, A)$$

is a bijection. This is called the Néron mapping property: In more compact notation, it says that there is an isomorphism $\mathcal{A}(S) \cong A(S_K)$.

Taking $S = \mathcal{A}$ in the definition we see that $\mathcal{A}$ is unique, up to a unique isomorphism.

It is a deep theorem that Néron models exist. Fortunately, Bosch, Lütkebohmert, and Raynaud devoted much time to create a carefully written book [BLR90] that explains the construction in modern language. Also, in the case of elliptic curves, Silverman's second book [Sil94] is extremely helpful.

The basic idea of the construction is to first observe that if we can construct a Néron model at each localization $R_{\mathfrak{p}}$ at a nonzero prime ideal of R, then each of these local models can be glued to obtain a global Néron model (this uses that there are only finitely many primes of bad reduction). Thus we may assume that R is a discrete valuation ring.

The next step is to pass to the "strict henselization" R' of R. A local ring R with maximal ideal $\wp$ is henselian if "every simple root lifts uniquely"; more precisely, if whenever $f(x) \in R[x]$ and $\alpha \in R$ is such that $f(\alpha) \equiv 0 \pmod{\wp}$ and $f'(\alpha) \not\equiv 0 \pmod{\wp}$, there is a unique element $\tilde{\alpha} \in R$ such that $\tilde{\alpha} \equiv \alpha \pmod{\wp}$ and $f(\tilde{\alpha}) = 0$. The strict henselization of a discrete valuation ring R is an extension of R that is henselian and for which the residue field of R' is the separable closure of the residue field of R (when the residue field is finite, the separable close is just the algebraic closure). The strict henselization is not too much bigger than R, though it is typically not finitely generated over R. It is, however, much smaller than the completion of R (e.g., $\mathbf{Z}_p$ is uncountable). The main geometric property of a strictly henselian ring R with residue field k is that if X is a smooth scheme over R, then the reduction map $X(R) \to X(k)$ is surjective.

Working over the strict henselization, we first resolve singularities. Then we use a generalization of the theorem that Weil used to construct Jacobians to pass from a birational group law to an actual group law. We thus obtain the Néron model over the strict henselization of R. Finally, we use Grothendieck's faithfully flat descent to obtain a Néron model over R.

When A is the Jacobian of a curve X, there is an alternative approach that involves the "minimal proper regular model" of X. For example, when A is an elliptic curve, it is the Jacobian of itself, and the Néron model can be constructed in

terms of the minimal proper regular model $\mathcal{X}$ of A as follows. In general, the model $\mathcal{X} \to R$ is not also smooth. Let $\mathcal{X}'$ be the smooth locus of $\mathcal{X} \to R$, which is obtained by removing from each closed fiber $\mathcal{X}_{\mathbf{F}_p} = \sum n_i C_i$ all irreducible components with multiplicity $n_i \geq 2$ and all singular points on each C_i, and all points where at least two C_i intersect each other. Then the group structure on A extends to a group structure on $\mathcal{X}'$, and $\mathcal{X}'$ equipped with this group structure is the Néron model of A.

Explicit determination of the possibilities for the minimal proper regular model of an elliptic curve was carried out by Kodaira, then Néron, and finally in a very explicit form by Tate. Tate codified a way to find the model in what's called "Tate's Algorithm" (see Antwerp IV, which is available on my web page: `http://modular.fas.harvard.edu/scans/antwerp/`, and look at Silverman, chapter IV, which also has important implementation advice).

13.6.2 The Birch and Swinnerton-Dyer conjecture and Néron models

Throughout this section, let A be an abelian variety over $\mathbf{Q}$ and let $\mathcal{A}$ be the corresponding Néron model over $\mathbf{Z}$. We work over $\mathbf{Q}$ for simplicity, but could work over any number field.

Let $L(A, s)$ be the Hasse-Weil L-function of A (see Section [to be written][1]). Let $r = \mathrm{ord}_{s=1} L(A, s)$ be the analytic rank of A. The Birch and Swinnerton-Dyer Conjecture asserts that $A(\mathbf{Q}) \approx \mathbf{Z}^r \oplus A(\mathbf{Q})_{\mathrm{tor}}$ and

$$\frac{L^{(r)}(A,1)}{r!} = \frac{(\prod c_p) \cdot \Omega_A \cdot \mathrm{Reg}_A \cdot \#Ш(A)}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^\vee(\mathbf{Q})_{\mathrm{tor}}}.$$

We have not defined most of the quantities appearing in this formula. In this section, we will define the Tamagawa numbers c_p, the real volume Ω_A, and the Shafarevich-Tate group $Ш(A)$ in terms of the Néron model $\mathcal{A}$ of A.

We first define the Tamagawa numbers c_p, which are the orders groups of connected components. Let p be a prime and consider the closed fiber $\mathcal{A}_{\mathbf{F}_p}$, which is a smooth commutative group scheme over $\mathbf{F}_p$. Then $\mathcal{A}_{\mathbf{F}_p}$ is a disjoint union of one or more connected components. The connected component $\mathcal{A}^0_{\mathbf{F}_p}$ that contains the identity element is a subgroup of $\mathcal{A}_{\mathbf{F}_p}$ (Intuition: the group law is continuous and the continuous image of a connected set is connected, so the group structure restricts to $\mathcal{A}^0_{\mathbf{F}_p}$).

Definition 13.6.3 (Component Group)**.** The *component group* of A at p is

$$\Phi_{A,p} = \mathcal{A}_{\mathbf{F}_p} / \mathcal{A}^0_{\mathbf{F}_p}.$$

Fact: The component group $\Phi_{A,p}$ is a finite flat group scheme over $\mathbf{F}_p$, and[2] for all but finitely many primes p, we have $\Phi_{A,p} = 0$.

Definition 13.6.4 (Tamagawa Numbers)**.** The *Tamagawa number* of A at a prime p is

$$c_p = \#\Phi_{A,p}(\mathbf{F}_p).$$

[1] Add reference.
[2] Reference?

Next we define the real volume Ω_A. Choose a basis

$$\omega_1, \ldots, \omega_d \in \mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}})$$

for the global differential 1-forms on $\mathcal{A}$, where $d = \dim A$. The wedge product $w = \omega_1 \wedge \omega_2 \wedge \cdots \wedge \omega_d$ is a global d-form on $\mathcal{A}$. Then w induces a differential d-form on the real Lie group $A(\mathbf{R})$.

Definition 13.6.5 (Real Volume)**.** The *real volume* of A is

$$\Omega_A = \left| \int_{A(\mathbf{R})} w \right| \in \mathbf{R}_{>0}.$$

Finally, we give a definition of the Shafarevich-Tate group in terms of the Néron model. Let $\mathcal{A}_0$ be the scheme obtained from the Néron model $\mathcal{A}$ over A by removing from each closed fiber all nonidentity components. Then $\mathcal{A}_0$ is again a smooth commutative group scheme, but it need not have the Néron mapping property.

Recall that an étale morphism is a morphism that is smooth of relative dimension 0. A sheaf of abelian groups on the étale site $\mathbf{Z}_{\text{ét}}$ is a functor (satisfying certain axioms) from the category of étale morphism $X \to \mathbf{Z}$ to the category of abelian groups. There are enough sheaves on $\mathbf{Z}_{\text{ét}}$ so that there is a cohomology theory for such sheaves, which is called étale cohomology. In particular if $\mathcal{F}$ is a sheaf on $\mathbf{Z}_{\text{ét}}$, then for every integer q there is an abelian group $\mathrm{H}^q(\mathbf{Z}_{\text{ét}}, \mathcal{F})$ associated to $\mathcal{F}$ that has the standard properties of a cohomology functor.

The group schemes $\mathcal{A}_0$ and $\mathcal{A}$ both determine sheaves on the étale site, which we will also denote by $\mathcal{A}_0$ and $\mathcal{A}$.

Definition 13.6.6 (Shafarevich-Tate Group)**.** Suppose $A(\mathbf{R})$ is connected, i.e., that $\mathcal{A}_0 = \mathcal{A}$. Then the *Shafarevich-Tate* group of A is $\mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A})$. More generally, suppose only that $A(\mathbf{R})$ is connected. Then the Shafarevich-Tate group is the image of the natural map

$$f : \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}_0) \to \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}).$$

Even more generally, if $A(\mathbf{R})$ is not connected, then there is a natural map $r : \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}) \to \mathrm{H}^1(\mathrm{Gal}(\mathbf{C}/\mathbf{R}), A(\mathbf{C}))$ and $Ш(A) = \mathrm{im}(f) \cap \ker(r)$.

Mazur proves in the appendix to [Maz72] that this definition is equivalent to the usual Galois cohomology definition. To do this, he considers the exact sequence $0 \to \mathcal{A}_0 \to \mathcal{A} \to \Phi_A \to 0$, where Φ_A is a sheaf version of $\oplus_p \Phi_{A,p}$. The main input is Lang's Theorem, which implies that over a local field, unramified Galois cohomology is the same as the cohomology of the corresponding component group.[3]

3

Conjecture 13.6.7 (Shafarevich-Tate)**.** *The group* $\mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A})$ *is finite.*

When A has rank 0, all component groups $\Phi_{A,p}$ are trivial, $A(\mathbf{R})$ is connected, and $A(\mathbf{Q})_{\text{tor}}$ and $A^\vee(\mathbf{Q})_{\text{tor}}$ are trivial, the Birch and Swinnerton-Dyer conjecture takes the simple form

$$\frac{L(A,1)}{\Omega_A} = \#\,\mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}).$$

Later[4] , when A is modular, we will (almost) interpret $L(A,1)/\Omega_A$ as the order of 4

[3]Reference for Lang's Lemma, etc.

[4]Where?

a certain group that involves modular symbols. Thus the BSD conjecture asserts that two groups have the same order; however, they are not isomorphic, since, e.g., when $\dim A = 1$ the modular symbols group is always cyclic, but the Shafarevich-Tate group is never cyclic (unless it is trivial).

13.6.3 Functorial properties of Neron models

The definition of Néron model is functorial, so one might expect the formation of Néron models to have good functorial properties. Unfortunately, it doesn't.

Proposition 13.6.8. *Let A and B be abelian varieties. If $\mathcal{A}$ and $\mathcal{B}$ are the Néron models of A and B, respectively, then the Néron model of $A \times B$ is $\mathcal{A} \times \mathcal{B}$.*

Suppose $R \subset R'$ is a finite extension of discrete valuation rings with fields of fractions $K \subset K'$. Sometimes, given an abelian variety A over a field K, it is easier to understand properties of the abelian variety, such as reduction, over K'. For example, you might have extra information that implies that $A_{K'}$ decomposes as a product of well-understood abelian varieties. It would thus be useful if the Néron model of $A_{K'}$ were simply the base extension $\mathcal{A}_{R'}$ of the Néron model of A over R. This is, however, frequently not the case.

Distinguishing various types of ramification will be useful in explaining how Néron models behave with respect to base change, so we now recall the notions of tame and wild ramification. If π generates the maximal ideal of R and v' is the valuation on R', then the extension is *unramified* if $v'(\pi) = 1$. It is *tamely ramified* if $v'(\pi)$ is not divisible by the residue characteristic of R, and it is *wildly ramified* if $v'(\pi)$ is divisible by the residue characteristic of R. For example, the extension $\mathbf{Q}_p(p^{1/p})$ of $\mathbf{Q}_p$ is wildly ramified.

Example 13.6.9. If R is the ring of integers of a p-adic field, then for every integer n there is a unique unramified extension of R of degree n. See [Cp86, §I.7], where Fröhlich uses Hensel's lemma to show that the unramified extensions of $K = \mathrm{Frac}(R)$ are in bijection with the finite (separable) extensions of the residue class field.

The Néron model does not behave well with respect to base change, except in some special cases. For example, suppose A is an abelian variety over the field of fractions K of a discrete valuation ring R. If K' is the field of fractions of a finite unramified extension R' of R, then the Néron model of $A_{K'}$ is $\mathcal{A}_{R'}$, where $\mathcal{A}$ is the Néron model of A over R. Thus the Néron model over an unramified extension is obtained by base extending the Néron model over the base. This is not too surprising because in the construction of Néron model we first passed to the strict henselization of R, which is a limit of unramified extensions.

Continuing with the above notation, if K' is tamely ramified over K, then in general $\mathcal{A}_{R'}$ need *not* be the Néron model of $A_{K'}$. Assume that K' is Galois over K. In [Edi92a], Bas Edixhoven describes the Néron model of A_K in terms of $\mathcal{A}_{R'}$. To describe his main theorem, we introduce the restriction of scalars of a scheme.

Definition 13.6.10 (Restriction of Scalars)**.** Let $S' \to S$ be a morphism of schemes and let X' be a scheme over S'. Consider the functor

$$\mathcal{R}(T) = \mathrm{Hom}_{S'}(T \times_S S', X')$$

on the category of all schemes T over S. If this functor is representable, the representing object $X = \mathrm{Res}_{S'/S}(X')$ is called the *restriction of scalars* of X' to S.

Edixhoven's main theorem is that if G is the Galois group of K' over K and $X = \mathrm{Res}_{R'/R}(\mathcal{A}_{R'})$ is the restriction of scalars of $\mathcal{A}_{R'}$ down to R, then there is a natural map $\mathcal{A} \to X$ whose image is the closed subscheme X^G of fixed elements.

We finish this section with some cautious remarks about exactness properties of Néron models. If $0 \to A \to B \to C \to 0$ is an exact sequence of abelian varieties, then the functorial definition of Néron models produces a complex of Néron models

$$0 \to \mathcal{A} \to \mathcal{B} \to \mathcal{C} \to 0,$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are the Néron models of A, B, and C, respectively. This complex can fail to be exact at every point. For an in-depth discussion of conditions when we have exactness, along with examples that violate exactness, see [BLR90, Ch. 7], which says: "we will see that, except for quite special cases, there will be a defect of exactness, the defect of right exactness being much more serious than the one of left exactness."

To give examples in which right exactness fails, it suffices to give an optimal quotient $B \to C$ such that for some p the induced map $\Phi_{B,p} \to \Phi_{C,p}$ on component groups is not surjective (recall that optimal means $A = \ker(B \to C)$ is an abelian variety). Such quotients, with B and C modular, arise naturally in the context of Ribet's level optimization. For example, the elliptic curve E given by $y^2 + xy = x^3 + x^2 - 11x$ is the optimal new quotient of the Jacobian $J_0(33)$ of $X_0(33)$. The component group of E at 3 has order 6, since E has semistable reduction at 3 (since $3 \mid\mid 33$) and $\mathrm{ord}_3(j(E)) = -6$. The image of the component group of $J_0(33)$ in the component group of E has order 2:

```
> OrderOfImageOfComponentGroupOfJ0N(ModularSymbols("33A"),3);
2
```

Note that the modular form associated to E is congruent modulo 3 to the form corresponding to $J_0(11)$, which illustrates the connection with level optimization.

14

Abelian Varieties Attached to Modular Forms

In this chapter we describe how to decompose $J_1(N)$, up to isogeny, as a product of abelian subvarieties A_f corresponding to Galois conjugacy classes of cusp forms f of weight 2. This was first accomplished by Shimura (see [Shi94, Theorem 7.14]). We also discuss properties of the Galois representation attached to f. [1] 1

In this chapter we will work almost exclusively with $J_1(N)$. However, everything goes through exactly as below with $J_1(N)$ replaced by $J_0(N)$ and $S_2(\Gamma_1(N))$ replaced by $S_2(\Gamma_0(N))$. Since, $J_1(N)$ has dimension much larger than $J_0(N)$, so for computational investigations it is frequently better to work with $J_0(N)$.

See Brian Conrad's appendix to [ribet-stein: Lectures on Serre's Conjectures] for a much more extensive exposition of the construction discussed below, which is geared toward preparing the reader for Deligne's more general construction of Galois representations associated to newforms of weight $k \geq 2$ (for that, see Conrad's book ...).

14.1 Decomposition of the Hecke algebra

Let N be a positive integer and let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(J_1(N))$$

be the algebra of all Hecke operators acting on $J_1(N)$. Recall from Section 9.4 that the anemic Hecke algebra is the subalgebra

$$\mathbf{T}_0 = \mathbf{Z}[\ldots, T_n, \ldots : (n, N) = 1] \subset \mathbf{T}$$

of $\mathbf{T}$ obtained by adjoining to $\mathbf{Z}$ only those Hecke operators T_n with n relatively prime to N.

---

[1] Rewrite intro after chapter is done, and point to where each thing is done.

Remark 14.1.1. Viewed as $\mathbf{Z}$-modules, $\mathbf{T}_0$ need not be saturated in $\mathbf{T}$, i.e., $\mathbf{T}/\mathbf{T}_0$ need not be torsion free. For example, if $\mathbf{T}$ is the Hecke algebra associated to $S_2(\Gamma_1(24))$ then $\mathbf{T}/\mathbf{T}_0 \cong \mathbf{Z}/2\mathbf{Z}$. Also, if $\mathbf{T}$ is the Hecke algebra associated to $S_2(\Gamma_0(54))$, then $\mathbf{T}/\mathbf{T}_0 \cong \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}$. [2] 2

If $f = \sum a_n q^n$ is a newform, then the field $K_f = \mathbf{Q}(a_1, a_2, \ldots)$ has finite degree over $\mathbf{Q}$, since the a_n are the eigenvalues of a family of commuting operators with integral characteristic polynomials. The *Galois conjugates* of f are the newforms $\sigma(f) = \sum \sigma(a_n)q^n$, for $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. There are $[K_f : \mathbf{Q}]$ Galois conjugates of f.

As in Section 9.4, we have a canonical decomposition

$$\mathbf{T}_0 \otimes \mathbf{Q} \cong \prod_f K_f, \tag{14.1.1}$$

where f varies over a set of representatives for the Galois conjugacy classes of newforms in $S_2(\Gamma_1(N))$ of level dividing N. For each f, let

$$\pi_f = (0, \ldots, 0, 1, 0, \ldots, 0) \in \prod K_f$$

be projection onto the factor K_f of the product (14.1.1). Since $\mathbf{T}_0 \subset \mathbf{T}$, and $\mathbf{T}$ has no additive torsion, we have $\mathbf{T}_0 \otimes \mathbf{Q} \subset \mathbf{T} \otimes \mathbf{Q}$, so these projectors π_f lie in $\mathbf{T}_\mathbf{Q} = \mathbf{T} \otimes \mathbf{Q}$. Since $\mathbf{T}_\mathbf{Q}$ is commutative and the π_f are mutually orthogonal idempotents whose sum is $(1, 1, \ldots, 1)$, we see that $\mathbf{T}_\mathbf{Q}$ breaks up as a product of algebras

$$\mathbf{T}_\mathbf{Q} \cong \prod_f L_f, \qquad t \mapsto \sum_f t\pi_f.$$

14.1.1 The Dimension of the algebras L_f

Proposition 14.1.2. *If f, L_f and K_f are as above, then $\dim_{K_f} L_f$ is the number of divisors of N/N_f where N_f is the level of the newform f.*

Proof. Let V_f be the complex vector space spanned by all images of Galois conjugates of f via all maps α_d with $d \mid N/N_f$. It follows from [Atkin-Lehner-Li theory – multiplicity one][3] that the images via α_d of the Galois conjugates of f are linearly independent. 3 (Details: More generally, if f and g are newforms of level M, then by Proposition 9.2.1, $B(f) = \{\alpha_d(f) : d \mid N/N_f\}$ is a linearly independent set and likewise for $B(g)$. Suppose some nonzero element f' of the span of $B(f)$ equals some element g' of the span of $B(g)$. Since T_p, for $p \nmid N$, commutes with α_d, we have $T_p(f') = a_p(f)f'$ and $T_p(g') = a_p(g)g'$, so $0 = T_p(0) = T_p(f' - g') = a_p(f)f' - a_p(g)g'$. Since $f' = g'$, this implies that $a_p(f) = a_p(g)$. Because a newform is determined by the eigenvalues of T_p for $p \nmid N$, it follows that $f = g$.) Thus the $\mathbf{C}$-dimension of V_f is the number of divisors of N/N_f times $\dim_\mathbf{Q} K_f$.

The factor L_f is isomorphic to the image of $\mathbf{T}_\mathbf{Q} \subset \mathrm{End}(S_k(\Gamma_1(N)))$ in $\mathrm{End}(V_f)$. As in Section **??**, there is a single element $v \in V_f$ so that $V_f = \mathbf{T}_\mathbf{C} \cdot v$. Thus the image of $\mathbf{T}_\mathbf{Q}$ in $\mathrm{End}(V_f)$ has dimension $\dim_\mathbf{C} V_f$, and the result follows. □

---

[2] I'm including the MAGMA scripts I used to check this as comments in the latex source, until I find the right way to justify these computational remarks. Maybe the remarks should point to an appendix where they are all justified?

[3] fill in

Let's examine a particular case of this proposition. Suppose p is a prime and $f = \sum a_n q^n$ is a newform of level N_f coprime to p, and let $N = p \cdot N_f$. We will show that

$$L_f = K_f[U]/(U^2 - a_p U + p), \tag{14.1.2}$$

hence $\dim_{K_f} L_f = 2$ which, as expected, is the number of divisors of $N/N_f = p$. The first step is to view L_f as the space of operators generated by the Hecke operators T_n acting on the span V of the images $f(dz) = f(q^d)$ for $d \mid (N/N_f) = p$. If $n \neq p$, then T_n acts on V as the scalar a_n, and when $n = p$, the Hecke operator T_p acts on $S_k(\Gamma_1(p \cdot N_f))$ as the operator also denoted U_p. By Section 9.2, we know that U_p corresponds to the matrix $\left(\begin{smallmatrix} a_p & 1 \\ -p & 0 \end{smallmatrix}\right)$ with respect to the basis $f(q), f(q^p)$ of V. Thus U_p satisfies the relation $U_p^2 - a_p U + p$. Since U_p is not a scalar matrix, this minimal polynomial of U_p is quadratic, which proves (14.1.2).

More generally, see [DDT94, Lem. 4.4] (Diamond-Darmon-Taylor)[4] for an explicit presentation of L_f as a quotient 4

$$L_f \cong K_f[\ldots, U_p, \ldots]/I$$

where I is an ideal and the U_p correspond to the prime divisors of N/N_f.

14.2 Decomposition of $J_1(N)$

Let f be a newform in $S_2(\Gamma_1(N))$ of level a divisor M of N, so $f \in S_2(\Gamma_1(M))_{\text{new}}$ is a normalized eigenform for all the Hecke operators of level M. We associate to f an abelian subvariety A_f of $J_1(N)$, of dimension $[L_f : \mathbf{Q}]$, as follows. Recall that π_f is the fth projector in $\mathbf{T}_0 \otimes \mathbf{Q} = \prod_g K_g$. We can not define A_f to be the image of $J_1(N)$ under π_f, since π_f is only, a priori, an element of $\text{End}(J_1(N)) \otimes \mathbf{Q}$. Fortunately, there exists a positive integer n such that $n\pi_f \in \text{End}(J_1(N))$, and we let

$$A_f = n\pi_f(J_1(N)).$$

This is independent of the choice of n, since the choices for n are all multiples of the "denominator" n_0 of π_f, and if A is any abelian variety and n is a positive integer, then $nA = A$.

The natural map $\prod_f A_f \to J_1(N)$, which is induced by summing the inclusion maps, is an isogeny. Also A_f is simple if f is of level N, and otherwise A_f is isogenous to a power of $A'_f \subset J_1(N_f)$. Thus we obtain an isogeny decomposition of $J_1(N)$ as a product of $\mathbf{Q}$-simple abelian varieties.

Remark 14.2.1. The abelian varieties A_f frequently decompose further over $\overline{\mathbf{Q}}$, i.e., they are not absolutely simple, and it is an interesting problem to determine an isogeny decomposition of $J_1(N)_{\overline{\mathbf{Q}}}$ as a product of simple abelian varieties. It is still not known precisely how to do this computationally for any particular N. [5] 5

This decomposition can be viewed in another way over the complex numbers. As a complex torus, $J_1(N)(\mathbf{C})$ has the following model:

$$J_1(N)(\mathbf{C}) = \text{Hom}(S_2(\Gamma_1(N)), \mathbf{C})/H_1(X_1(N), \mathbf{Z}).$$

[4] Remove parens.

[5] Add more/pointers/etc.

The action of the Hecke algebra $\mathbf{T}$ on $J_1(N)(\mathbf{C})$ is compatible with its action on the cotangent space $S_2(\Gamma_1(N))$. This construction presents $J_1(N)(\mathbf{C})$ naturally as $V/\mathcal{L}$ with V a complex vector space and $\mathcal{L}$ a lattice in V. The anemic Hecke algebra $\mathbf{T}_0$ then decomposes V as a direct sum $V = \bigoplus_f V_f$. The Hecke operators act on V_f and $\mathcal{L}$ in a compatible way, so $\mathbf{T}_0$ decomposes $\mathcal{L} \otimes \mathbf{Q}$ in a compatible way. Thus $\mathcal{L}_f = V_f \cap \mathcal{L}$ is a lattice in V_f, so we may $A_f(\mathbf{C})$ view as the complex torus $V_f/\mathcal{L}_f$.

Lemma 14.2.2. *Let $f \in S_2(\Gamma_1(N))$ be a newform of level dividing N and $A_f = n\pi_f(J_1(N))$ be the corresponding abelian subvariety of $J_1(N)$. Then the Hecke algebra $\mathbf{T} \subset \mathrm{End}(J_1(N))$ leaves A_f invariant.*

Proof. The Hecke algebra $\mathbf{T}$ is commutative, so if $t \in \mathbf{T}$, then

$$tA_f = tn\pi_f(J_1(N)) = n\pi_f(tJ_1(N)) \subset n\pi_f(J_1(N)) = A_f.$$

□

Remark 14.2.3. Viewing $A_f(\mathbf{C})$ as $V_f/\mathcal{L}_f$ is extremely useful computationally, since $\mathcal{L}$ can be computed using modular symbols, and $\mathcal{L}_f$ can be cut out using the Hecke operators. For example, if f and g are nonconjugate newforms of level dividing N, we can explicitly compute the group structure of $A_f \cap A_g \subset J_1(N)$ by doing a computation with modular symbols in $\mathcal{L}$. More precisely, we have

$$A_f \cap A_g \cong (\mathcal{L}/(\mathcal{L}_f + \mathcal{L}_g))_{\mathrm{tor}}.$$

Note that A_f depends on viewing f as an element of $S_2(\Gamma_1(N))$ for some N. Thus it would be more accurate to denote A_f by $A_{f,N}$, where N is any multiple of the level of f, and to reserve the notation A_f for the case $N = 1$. Then $\dim A_{f,N}$ is $\dim A_f$ times the number of divisors of N/N_f.

14.2.1 Aside: intersections and congruences

Suppose f and g are not Galois conjugate. Then the intersection $\Psi = A_f \cap A_g$ is finite, since $V_f \cap V_g = 0$, and the integer $\#\Psi$ is of interest. This cardinality is related to congruence between f and g, but the exact relation is unclear. For example, one might expect that $p \mid \#\Psi$ if and only if there is a prime $\wp$ of the compositum $K_f.K_g$ of residue characteristic p such that $a_q(f) \equiv a_q(g) \pmod{\wp}$ for all $q \nmid N$. If $p \mid \#\Psi$, then such a prime $\wp$ exists (take $\wp$ to be induced by a maximal ideal in the support of the nonzero $\mathbf{T}$-module $\Psi[p]$). The converse is frequently true, but is sometimes false. For example, if N is the prime 431 and

$$f = q - q^2 + q^3 - q^4 + q^5 - q^6 - 2q^7 + \cdots$$
$$g = q - q^2 + 3q^3 - q^4 - 3q^5 - 3q^6 + 2q^7 + \cdots,$$

then $f \equiv g \pmod 2$, but $A_f \cap A_g = 0$. This example implies that "multiplicity one fails" for level 431 and $p = 2$, so the Hecke algebra associated to $J_0(431)$ is not Gorenstein (see [Lloyd Kilford paper][6] for more details). 6

[6] fix reference

14.3 Galois representations attached to A_f

It is important to emphasize the case when f is a newform of level N, since then A_f is $\mathbf{Q}$-simple[7] and there is a compatible family of 2-dimensional ℓ-adic representations attached to f, which arise from torsion points on A_f. 7

Proposition 14.1.2 implies that $L_f = K_f$. Fix such an f, let $A = A_f$, let $K = K_f$, and let

$$d = \dim A = \dim_{\mathbf{Q}} K = [K : \mathbf{Q}].$$

Let ℓ be a prime and consider the $\mathbf{Q}_\ell$-adic Tate module $\mathrm{Tate}_\ell(A)$ of A:

$$\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes \varprojlim_{\nu>0} A[\ell^\nu].$$

Note that as a $\mathbf{Q}_\ell$-vector space $\mathrm{Tate}_\ell(A) \cong \mathbf{Q}_\ell^{2d}$, since $A[n] \cong (\mathbf{Z}/n\mathbf{Z})^{2d}$, as groups.

There is a natural action of the ring $K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$ on $\mathrm{Tate}_\ell(A)$. By algebraic number theory

$$K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell = \prod_{\lambda|\ell} K_\lambda,$$

where λ runs through the primes of the ring $\mathcal{O}_K$ of integers of K lying over ℓ and K_λ denotes the completion of K with respect to the absolute value induced by λ. Thus $\mathrm{Tate}_\ell(A)$ decomposes as a product

$$\mathrm{Tate}_\ell(A) = \prod_{\lambda|\ell} \mathrm{Tate}_\lambda(A)$$

where $\mathrm{Tate}_\lambda(A)$ is a K_λ vector space.

Lemma 14.3.1. *Let the notation be as above. Then for all λ lying over ℓ,*

$$\dim_{K_\lambda} \mathrm{Tate}_\lambda(A) = 2.$$

Proof. Write $A = V/\mathcal{L}$, with $V = V_f$ a complex vector space and $\mathcal{L}$ a lattice. Then $\mathrm{Tate}_\lambda(A) \cong \mathcal{L} \otimes \mathbf{Q}_\ell$ as K_λ-modules (not as $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-modules!), since $A[\ell^n] \cong \mathcal{L}/\ell^n\mathcal{L}$, and $\varprojlim_n \mathcal{L}/\ell^n\mathcal{L} \cong \mathbf{Z}_\ell \otimes \mathcal{L}$. Also, $\mathcal{L} \otimes \mathbf{Q}$ is a vector space over K, which must have dimension 2, since $\mathcal{L} \otimes \mathbf{Q}$ has dimension $2d = 2\dim A$ and K has degree d. Thus

$$\mathrm{Tate}_\lambda(A) \cong \mathcal{L} \otimes K_\lambda \approx (K \oplus K) \otimes_K K_\lambda \cong K_\lambda \oplus K_\lambda$$

has dimension 2 over K_λ. □

Now consider $\mathrm{Tate}_\lambda(A)$, which is a K_λ-vector space of dimension 2. The Hecke operators are defined over $\mathbf{Q}$, so $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on $\mathrm{Tate}_\ell(A)$ in a way compatible with the action of $K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$. We thus obtain a homomorphism

$$\rho_\ell = \rho_{f,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{K\otimes\mathbf{Q}_\ell} \mathrm{Tate}_\ell(A) \approx \mathrm{GL}_2(K \otimes \mathbf{Q}_\ell) \cong \prod_\lambda \mathrm{GL}_2(K_\lambda).$$

Thus ρ_ℓ is the direct sum of ℓ-adic Galois representations ρ_λ where

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{End}_{K_\lambda}(\mathrm{Tate}_\lambda(A))$$

[7] add pointer to where this is proved.

gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathrm{Tate}_\lambda(A)$.

If $p \nmid \ell N$, then ρ_λ is unramified at p (see [ST68, Thm. 1]). In this case it makes sense to consider $\rho_\lambda(\varphi_p)$, where $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a Frobenius element at p. Then $\rho_\lambda(\varphi_p)$ has a well-defined trace and determinant, or equivalently, a well-defined characteristic polynomial $\Phi(X) \in K_\lambda[X]$.

Theorem 14.3.2. *Let $f \in S_2(\Gamma_1(N), \varepsilon)$ be a newform of level N with Dirichlet character ε. Suppose $p \nmid \ell N$, and let $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be a Frobenius element at p. Let $\Phi(X)$ be the characteristic polynomial of $\rho_\lambda(\varphi_p)$. Then*

$$\Phi(X) = X^2 - a_p X + p \cdot \varepsilon(p),$$

where a_p is the pth coefficient of the modular form f (thus a_p is the image of T_p in E_f and $\varepsilon(p)$ is the image of $\langle p \rangle$).

Let $\varphi = \varphi_p$. By the Cayley-Hamilton theorem

$$\rho_\lambda(\varphi)^2 - \mathrm{tr}(\rho_\lambda(\varphi))\rho_\lambda(\varphi) + \det(\rho_\lambda(\varphi)) = 0.$$

Using the Eichler-Shimura congruence relation (see [8]) we will show that $\mathrm{tr}(\rho_\lambda(\varphi)) = a_p$, but we defer the proof of this until ...[9] .

We will prove that $\det(\rho_\lambda(\varphi)) = p$ in the special case when $\varepsilon = 1$. This will follow from the equality

$$\det(\rho_\lambda) = \chi_\ell, \tag{14.3.1}$$

where χ_ℓ is the ℓth cyclotomic character

$$\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^* \subset K_\lambda^*,$$

which gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on μ_{ℓ^∞}. We have $\chi_\ell(\varphi) = p$ because φ induces induces pth powering map on μ_{ℓ^∞}.

It remains to establish (14.3.1). The simplest case is when A is an elliptic curve. In [Sil92,][10] , Silverman shows that $\det(\rho_\ell) = \chi_\ell$ using the Weil pairing. We will consider the Weil pairing in more generality in the next section, and use it to establish (14.3.1).

14.3.1 The Weil pairing

Let $T_\ell(A) = \varprojlim_{n \geq 1} A[\ell^n]$, so $\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes T_\ell(A)$. The Weil pairing is a non-degenerate perfect pairing

$$e_\ell : T_\ell(A) \times T_\ell(A^\vee) \to \mathbf{Z}_\ell(1).$$

(See e.g., [Mil86, §16] for a summary of some of its main properties.)

Remark 14.3.3. Identify $\mathbf{Z}/\ell^n\mathbf{Z}$ with μ_{ℓ^n} by $1 \mapsto e^{-2\pi i/\ell^n}$, and extend to a map $\mathbf{Z}_\ell \to \mathbf{Z}_\ell(1)$. If $J = \mathrm{Jac}(X)$ is a Jacobian, then the Weil pairing on J is induced by the canonical isomorphism

$$T_\ell(J) \cong \mathrm{H}^1(X, \mathbf{Z}_\ell) = \mathrm{H}^1(X, \mathbf{Z}) \otimes \mathbf{Z}_\ell,$$

[8] Next week!
[9] when?
[10] get ref

and the cup product pairing

$$\mathrm{H}^1(X, \mathbf{Z}_\ell) \otimes_{\mathbf{Z}_\ell} \mathrm{H}^1(X, \mathbf{Z}_\ell) \xrightarrow{\cup} \mathbf{Z}_\ell.$$

For more details see the discussion on pages 210–211 of Conrad's appendix to [RS01], and the references therein. In particular, note that $\mathrm{H}^1(X, \mathbf{Z}_\ell)$ is isomorphic to $\mathrm{H}_1(X, \mathbf{Z}_\ell)$, because $\mathrm{H}_1(X, \mathbf{Z}_\ell)$ is self-dual because of the intersection pairing. It is easy to see that $\mathrm{H}_1(X, \mathbf{Z}_\ell) \cong T_\ell(J)$ since by Abel-Jacobi $J \cong T_0(J)/\mathrm{H}_1(X, \mathbf{Z})$, where $T_0(J)$ is the tangent space at J at 0 (see Lemma 14.3.1). [11] 11

Here $\mathbf{Z}_\ell(1) \cong \varprojlim \mu_{\ell^n}$ is isomorphic to $\mathbf{Z}_\ell$ as a ring, but has the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ induced by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\varprojlim \mu_{\ell^n}$. Given $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, there is an element $\chi_\ell(\sigma) \in \mathbf{Z}_\ell^*$ such that $\sigma(\zeta) = \zeta^{\chi_\ell(\sigma)}$, for every ℓ^nth root of unity ζ. If we view $\mathbf{Z}_\ell(1)$ as just $\mathbf{Z}_\ell$ with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then the action of $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathbf{Z}_\ell(1)$ is left multiplication by $\chi_\ell(\sigma) \in \mathbf{Z}_\ell^*$.

Definition 14.3.4 (Cyclotomic Character)**.** The homomorphism

$$\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^*$$

is called the *ℓ-adic cyclotomic character*.

If $\varphi : A \to A^\vee$ is a polarization (so it is an isogeny defined by translation of an ample invertible sheaf), we define a pairing

$$e_\ell^\varphi : T_\ell(A) \times T_\ell(A) \to \mathbf{Z}_\ell(1) \tag{14.3.2}$$

by $e_\ell^\varphi(a, b) = e_\ell(a, \varphi(b))$. The pairing (14.3.2) is a skew-symmetric, nondegenerate, bilinear pairing that is $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-equivariant, in the sense that if $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then

$$e_\ell^\varphi(\sigma(a), \sigma(b)) = \sigma \cdot e_\ell^\varphi(a, b) = \chi_\ell(\sigma) e_\ell^\varphi(a, b).$$

We now apply the Weil pairing in the special case $A = A_f \subset J_1(N)$. Abelian varieties attached to modular forms are equipped with a canonical polarization called the *modular polarization*. The canonical principal polarization of $J_1(N)$ is an isomorphism $J_1(N) \xrightarrow{\sim} J_1(N)^\vee$, so we obtain the modular polarization $\varphi = \varphi_A : A \to A^\vee$ of A, as illustrated in the following diagram:

$$\begin{array}{ccc} J_1(N) & \xrightarrow{\text{autoduality}\,\cong} & J_1(N)^\vee \\ \uparrow & & \downarrow \\ A & \xrightarrow{\text{polarization}\,\varphi_A} & A^\vee \end{array}$$

Consider (14.3.2) with $\varphi = \varphi_A$ the modular polarization. Tensoring over $\mathbf{Q}$ and restricting to $\mathrm{Tate}_\lambda(A)$, we obtain a nondegenerate skew-symmetric bilinear pairing

$$e : \mathrm{Tate}_\lambda(A) \times \mathrm{Tate}_\lambda(A) \to \mathbf{Q}_\ell(1). \tag{14.3.3}$$

The nondegeneracy follows from the nondegeneracy of e_ℓ^φ and the observation that

$$e_\ell^\varphi(\mathrm{Tate}_\lambda(A), \mathrm{Tate}_{\lambda'}(A)) = 0$$

[11] Remove or expand?

when $\lambda \neq \lambda'$. This uses the Galois equivariance of e_ℓ^ϕ carries over to Galois equivariance of e, in the following sense. If $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and $x, y \in \mathrm{Tate}_\lambda(A)$, then

$$e(\sigma x, \sigma y) = \sigma e(x, y) = \chi_\ell(\sigma) e(x, y).$$

Note that σ acts on $\mathbf{Q}_\ell(1)$ as multiplication by $\chi_\ell(\sigma)$.

14.3.2 The Determinant

There are two proofs of the theorem, a fancy proof and a concrete proof. We first present the fancy proof. The pairing e of (14.3.3) is a skew-symmetric and bilinear form so it determines a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-equivarient homomorphism

$$\bigwedge_{K_\lambda}^{2} \mathrm{Tate}_\lambda(A) \to \mathbf{Q}_\ell(1). \tag{14.3.4}$$

It is not *a priori* true that we can take the wedge product over K_λ instead of $\mathbf{Q}_\ell$, but we can because $e(tx, y) = e(x, ty)$ for any $t \in K_\lambda$. This is where we use that A is attached to a newform with trivial character, since when the character is nontrivial, the relation between $e(T_p x, y)$ and $e(x, T_p y)$ will involve $\langle p \rangle$. Let $D = \bigwedge^2 \mathrm{Tate}_\lambda(A)$ and note that $\dim_{K_\lambda} D = 1$, since $\mathrm{Tate}_\lambda(A)$ has dimension 2 over K_λ.

There is a canonical isomorphism

$$\mathrm{Hom}_{\mathbf{Q}_\ell}(D, \mathbf{Q}_\ell(1)) \cong \mathrm{Hom}_{K_\lambda}(D, K_\lambda(1)),$$

and the map of (14.3.4) maps to an isomorphism $D \cong K_\lambda(1)$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-modules. Since the representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on D is the determinant, and the representation on $K_\lambda(1)$ is the cyclotomic character χ_ℓ, it follows that $\det \rho_\lambda = \chi_\ell$.

Next we consider a concrete proof. If $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then we must show that $\det(\sigma) = \chi_\ell(\sigma)$. Choose a basis $x, y \in \mathrm{Tate}_\lambda(A)$ of $\mathrm{Tate}_\lambda(A)$ as a 2 dimensional K_λ vector space. We have $\sigma(x) = ax + cy$ and $\sigma(y) = bx + dy$, for $a, b, c, d \in K_\lambda$. Then

$$\begin{aligned}
\chi_\ell(\sigma) e(x, y) &= \langle \sigma x, \sigma y \rangle \\
&= e(ax + cy, bx + dy) \\
&= e(ax, bx) + e(ax, dy) + e(cy, bx) + e(cy, dy) \\
&= e(ax, dy) + e(cy, bx) \\
&= e(adx, y) - e(bcx, y) \\
&= e((ad - bc)x, y) \\
&= (ad - bc) e(x, y)
\end{aligned}$$

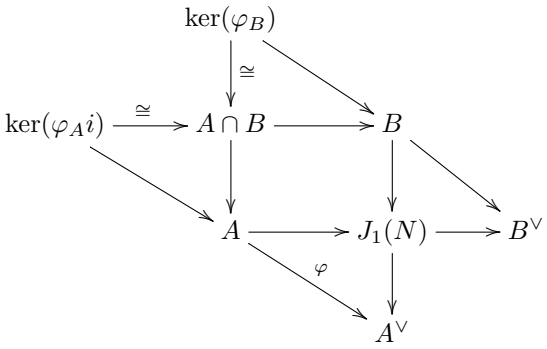
To see that $e(ax, bx) = 0$, note that

$$e(ax, bx) = e(abx, x) = -e(x, abx) = -e(ax, bx).$$

Finally, since e is nondegenerate, there exists x, y such that $e(x, y) \neq 0$, so $\chi_\ell(\sigma) = ad - bc = \det(\sigma)$.

14.4 Remarks about the modular polarization

Let A and φ be as in Section 14.3.1. The degree $\deg(\varphi)$ of the modular polarization of A is an interesting arithmetic invariant of A. If $B \subset J_1(N)$ is the sum of all modular abelian varieties A_g attached to newforms $g \in S_2(\Gamma_1(N))$, with g not a Galois conjugate of f and of level dividing N, then $\ker(\varphi) \cong A \cap B$, as illustrated in the following diagram:



Note that $\ker(\varphi_B)$ is also isomorphic to $A \cap B$, as indicated in the diagram.

In connection with Section **??**, the quantity $\ker(\varphi_A) = A \cap B$ is closely related to congruences between f and eigenforms orthogonal to the Galois conjugates of f.

When A has dimension 1, we may alternatively view A as a quotient of $X_1(N)$ via the map

$$X_1(N) \to J_1(N) \to A^\vee \cong A.$$

Then $\varphi_A : A \to A$ is pullback of divisors to $X_1(N)$ followed by push forward, which is multiplication by the degree. Thus $\varphi_A = [n]$, where n is the degree of the morphism $X_1(N) \to A$ of algebraic curves. The *modular degree* is

$$\deg(X_1(N) \to A) = \sqrt{\deg(\varphi_A)}.$$

More generally, if A has dimension greater than 1, then $\deg(\varphi_A)$ has order a perfect square (for references, see [Mil86, Thm. 13.3]), and we define the *modular degree* to be $\sqrt{\deg(\varphi_A)}$.

Let f be a newform of level N. In the spirit of Section 14.2.1 we use congruences to define a number related to the modular degree, called the congruence number. For a subspace $V \subset S_2(\Gamma_1(N))$, let $V(\mathbf{Z}) = V \cap \mathbf{Z}[[q]]$ be the elements with integral q-expansion at ∞ and $V^\perp$ denotes the orthogonal complement of V with respect to the Petersson inner product. The *congruence number* of f is

$$r_f = \#\frac{S_2(\Gamma_1(N))(\mathbf{Z})}{V_f(\mathbf{Z}) + V_f^\perp(\mathbf{Z})},$$

where V_f is the complex vector space spanned by the Galois conjugates of f. We thus have two positive associated to f, the congruence number r_f and the modular degree m_f of of A_f.

Theorem 14.4.1. $m_f \mid r_f$

Ribet mentions this in the case of elliptic curves in [ZAGIER, 1985] [Zag85a], but the statement is given incorrectly in that paper (the paper says that $r_f \mid m_f$, which is wrong). The proof for dimension greater than one is in [AGASHE-STEIN, Manin constant...]. Ribet also subsequently proved that if $p^2 \nmid N$, then $\mathrm{ord}_p(m_f) = \mathrm{ord}_p(r_f)$.

We can make the same definitions with $J_1(N)$ replaced by $J_0(N)$, so if $f \in S_2(\Gamma_0(N))$ is a newform, $A_f \subset J_0(N)$, and the congruence number measures congruences between f and other forms in $S_2(\Gamma_0(N))$. In [FM99, Ques. 4.4], they ask whether it is always the case that $m_f = r_f$ when A_f is an elliptic curve, and m_f and r_f are defined relative to $\Gamma_0(N)$. I[12] implemented an algorithm in MAGMA to compute r_f, and found the first few counterexamples, which occur when 12

$$N = 54, 64, 72, 80, 88, 92, 96, 99, 108, 120, 124, 126, 128, 135, 144.$$

For example, the elliptic curve A labeled 54B1 in [Cre97] has $r_A = 6$ and $m_A = 2$. To see directly that $3 \mid r_A$, observe that if f is the newform corresponding to E and g is the newform corresponding to $X_0(27)$, then $g(q) + g(q^2)$ is congruent to f modulo 3. This is consistent with Ribet's theorem that if $p \mid r_A/m_A$ then $p^2 \mid N$. There seems to be no absolute bound on the p that occur.

It would be interesting to determine the answer to the analogue of the question of Frey-Mueller for $\Gamma_1(N)$. For example, if $A \subset J_1(54)$ is the curve isogeneous to 54B1, then $m_A = 18$ is divisible by 3. However, I do not know[13] r_A in this case, because I haven't written a program to compute it for $\Gamma_1(N)$. 13

[12] change...
[13] fix

15
Modularity of Abelian Varieties

15.1 Modularity over $\mathbf{Q}$

Definition 15.1.1 (Modular Abelian Variety)**.** Let A be an abelian variety over $\mathbf{Q}$. Then A is *modular* if there exists a positive integer N and a surjective map $J_1(N) \to A$ defined over $\mathbf{Q}$.

The following theorem is the culmination of a huge amount of work, which started with Wiles's successful attack [Wil95] on Fermat's Last Theorem, and finished with [BCDT01].

Theorem 15.1.2 (Breuil, Conrad, Diamond, Taylor, Wiles)**.** *Let E be an elliptic curve over $\mathbf{Q}$. Then E is modular.*

We will say nothing about the proof here.[1] 1

If A is an abelian variety over $\mathbf{Q}$, let $\mathrm{End}_{\mathbf{Q}}(A)$ denote the ring of endomorphisms of A that are defined over $\mathbf{Q}$.

Definition 15.1.3 (GL_2-type)**.** An abelian variety A over $\mathbf{Q}$ is of GL_2*-type* if the endomorphism algebra $\mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$ contains a number field of degree equal to the dimension of A.

For example, every elliptic curve E over $\mathbf{Q}$ is trivially of GL_2-type, since $\mathbf{Q} \subset \mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(E)$.

Proposition 15.1.4. *If A is an abelian variety over $\mathbf{Q}$, and $K \subset \mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$ is a field, then $[K : \mathbf{Q}]$ divides $\dim A$.*

Proof. As discussed in [Rib92, §2], K acts faithfully on the tangent space $\mathrm{Tan}_0(A/\mathbf{Q})$ over $\mathbf{Q}$ to A at 0, which is a $\mathbf{Q}$ vector space of dimension $\dim(A)$. Thus $\mathrm{Tan}_0(A/\mathbf{Q})$ is a vector space over K, hence has $\mathbf{Q}$-dimension a multiple of $[K : \mathbf{Q}]$. □

[1] Also pointer to later in book.

Proposition 15.1.4 implies, in particular, that if E is an elliptic curve over $\mathbf{Q}$, then $\mathrm{End}_{\mathbf{Q}}(E) = \mathbf{Q}$. Recall[2] that *$E$ has CM* or is a *complex multiplication* elliptic curve if $\mathrm{End}_{\overline{\mathbf{Q}}}(E) \neq \mathbf{Z}$). Proposition 15.1.4 implies that if E is a CM elliptic curve, the extra endomorphisms are *never* defined over $\mathbf{Q}$. 2

Proposition 15.1.5. *Suppose $A = A_f \subset J_1(N)$ is an abelian variety attached to a newform of level N. Then A is of GL_2-type.*

Proof. The endomorphism ring of A_f contains $\mathcal{O}_f = \mathbf{Z}[\ldots, a_n(f), \ldots]$, hence the field $K_f = \mathbf{Q}(\ldots, a_n(f), \ldots)$ is contained in $\mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$. Since $A_f = n\pi J_1(N)$, where π is a projector onto the factor K_f of the anemic Hecke algebra $\mathbf{T}_0 \otimes_{\mathbf{Z}} \mathbf{Q}$, we have $\dim A_f = [K_f : \mathbf{Q}]$. (One way to see this is to recall that the tangent space $T = \mathrm{Hom}(S_2(\Gamma_1(N)), \mathbf{C})$ to $J_1(N)$ at 0 is free of rank 1 over $\mathbf{T}_0 \otimes_{\mathbf{Z}} \mathbf{C}$.) $\square$

Conjecture 15.1.6 (Ribet)**.** *Every abelian variety over $\mathbf{Q}$ of GL_2-type is modular.*

Suppose

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

is an odd irreducible continuous Galois representation, where odd means that

$$\det(\rho(c)) = -1,$$

where c is complex conjugation. We say that ρ is *modular* if there is a newform $f \in S_k(\Gamma_1(N))$, and a prime ideal $\wp \subset \mathcal{O}_f$ such that for all $\ell \nmid Np$, we have

$$\begin{aligned} \mathrm{Tr}(\rho(\mathrm{Frob}_\ell)) &\equiv a_\ell \pmod{\wp}, \\ \mathrm{Det}(\rho(\mathrm{Frob}_\ell)) &\equiv \ell^{k-1} \cdot \varepsilon(\ell) \pmod{\wp}. \end{aligned}$$

Here χ_p is the p-adic cyclotomic character, and ε is the (Nebentypus) character of the newform f.

Conjecture 15.1.7 (Serre)**.** *Every odd irreducible continuous representation*

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

is modular. Moreover, there is a formula for the "optimal" weight $k(\rho)$ and level $N(\rho)$ of a newform that gives rise to ρ.

In [Ser87], Serre describes the formula for the weight and level. Also, it is now known due to work of Ribet, Edixhoven, Coleman, Voloch, Gross, and others that if ρ is modular, then ρ arises from a form of the conjectured weight and level, except in some cases when $p = 2$. (For more details see the survey paper [RS01].) However, the full Conjecture 15.1.7 is known in very few cases.

Remark 15.1.8. There is interesting recent work of Richard Taylor which connects Conjecture 15.1.7 with the open question of whether every variety of a certain type has a point over a solvable extension of $\mathbf{Q}$. The question of the existence of solvable points ("solvability of varieties in radicals") seems very difficult. For example, we don't even know the answer for genus one curves, or have a good reason to make a conjecture either way (as far as I know[3]). There's a book of Mike Fried that discusses this solvability question.[4] 3 4

[2] from where

[3] fix

[4] Find the exact reference in that book and the exact book.

Serre's conjecture is very strong. For example, it would imply modularity of all abelian varieties over $\mathbf{Q}$ that could possibly be modular, and the proof of this implication does not rely on Theorem 15.1.2.

Theorem 15.1.9 (Ribet)**.** *Serre's conjectures on modularity of all odd irreducible mod p Galois representations implies Conjecture 15.1.6.*

To give the reader a sense of the connection between Serre's conjecture and modularity, we sketch some of the key ideas of the proof of Theorem 15.1.9; for more details the reader may consult Sections 1–4 of [Rib92].

Without loss, we may assume that A is $\mathbf{Q}$-simple. As explained in the not trivial [Rib92, Thm. 2.1], this hypothesis implies that

$$K = \mathbf{Q} \otimes_{\mathbf{Z}} \mathrm{End}_{\mathbf{Q}}(A)$$

is a number field of degree $\dim(A)$. The Tate modules

$$\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes \varprojlim_{n \geq 1} A[\ell^n]$$

are free of rank two over $K \otimes \mathbf{Q}_\ell$, so the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathrm{Tate}_\ell(A)$ defines a representation

$$\rho_{A,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(K \otimes \mathbf{Q}_\ell).$$

Remarks 15.1.10. That these representations take values in GL_2 is why such A are said to be "of GL_2-type". Also, note that the above applies to $A = A_f \subset J_1(N)$, and the ℓ-adic representations attached to f are just the factors of $\rho_{A,\ell}$ coming from the fact that $K \otimes \mathbf{Q}_\ell \cong \prod_{\lambda|\ell} K_\lambda$.

The deepest input to Ribet's proof is Faltings's isogeny theorem, which Faltings proved in order to prove Mordell's conjecture (there are only a finite number of L-rational points on any curve over L of genus at least 2).

If B is an abelian variety over $\mathbf{Q}$, let

$$L(B,s) = \prod_{\text{all primes } p} \frac{1}{\det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(A)\right)} = \prod_p L_p(B,s),$$

where ℓ is a prime of good reduction (it makes no difference which one).

Theorem 15.1.11 (Faltings)**.** *Let A and B be abelian varieties. Then A is isogenous to B if and only if $L_p(A,s) = L_p(B,s)$ for almost all p.*

Using an analysis of Galois representations and properties of conductors and applying results of Faltings, Ribet finds an infinite set Λ of primes of K such that all $\rho_{A,\lambda}$ are irredudible and there only finitely many Serre invariants $N(\rho_{A,\lambda})$ and $k(\rho_{A,\lambda})$. For each of these λ, by Conjecture 15.1.7 there is a newform f_λ of level $N(\rho_{A,\lambda}))$ and weight $k(\rho_{A,\lambda})$ that gives rise to the mod ℓ representation $\rho_{A,\lambda}$. Since Λ is infinite, but there are only finitely many Serre invariants $N(\rho_{A,\lambda}))$, $k(\rho_{A,\lambda})$, there must be a single newform f and an infinite subset Λ' of Λ so that for every $\lambda \in \Lambda'$ the newform f gives rise to $\rho_{A,\lambda}$.

Let $B = A_f \subset J_1(N)$ be the abelian variety attached to f. Fix any prime p of good reduction. There are infinitely many primes $\lambda \in \Lambda'$ such that $\rho_{A,\lambda} \cong \rho_{B,\tilde{\lambda}}$ for some $\tilde{\lambda}$, and for these λ,

$$\det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|A[\lambda]\right) = \det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|B[\tilde{\lambda}]\right).$$

This means that the degree two polynomials in p^{-s} (over the appropriate fields, e.g., $K \otimes \mathbf{Q}_\ell$ for A)

$$\det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(A)\right)$$

and

$$\det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(B)\right)$$

are congruent modulo infinitely many primes. Therefore they are equal. By Theorem 15.1.11, it follows that A is isogenous to $B = A_f$, so A is modular.

15.2 Modularity of elliptic curves over $\overline{\mathbf{Q}}$

Definition 15.2.1 (Modular Elliptic Curve)**.** An elliptic curve E over $\overline{\mathbf{Q}}$ is *modular* if there is a surjective morphism $X_1(N) \to E$ for some N.

Definition 15.2.2 ($\mathbf{Q}$-curve)**.** An elliptic curve E over $\mathbf{Q}$-bar is a $\mathbf{Q}$-*curve* if for every $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ there is an isogeny $E^\sigma \to E$ (over $\overline{\mathbf{Q}}$).

Theorem 15.2.3 (Ribet)**.** *Let E be an elliptic curve over $\overline{\mathbf{Q}}$. If E is modular, then E is a $\mathbf{Q}$-curve, or E has CM.*

This theorem is proved in [Rib92, §5].

Conjecture 15.2.4 (Ribet)**.** *Let E be an elliptic curve over $\overline{\mathbf{Q}}$. If E is a $\mathbf{Q}$-curve, then E is modular.*

In [Rib92, §6], Ribet proves that Conjecture 15.1.7 implies Conjecture 15.2.4. He does this by showing that if a $\mathbf{Q}$-curve E does not have CM then there is a $\mathbf{Q}$-simple abelian variety A over $\mathbf{Q}$ of GL_2-type such that E is a simple factor of A over $\overline{\mathbf{Q}}$. This is accomplished finding a model for E over a Galois extension K of $\mathbf{Q}$, restricting scalars down to $\mathbf{Q}$ to obtain an abelian variety $B = \mathrm{Res}_{K/\mathbf{Q}}(E)$, and using Galois cohomology computations (mainly in H^2's) to find the required A of GL_2-type inside B. Then Theorem 15.1.9 and our assumption that Conjecture 15.1.7 is true together immediately imply that A is modular.

Ellenberg and Skinner [ES00] have recently used methods similar to those used by Wiles to prove strong theorems toward Conjecture 15.2.4. See also Ellenberg's survey [Ell02], which discusses earlier modularity results of Hasegawa, Hashimoto, Hida, Momose, and Shimura, and gives an example to show that there are infinitely many $\mathbf{Q}$-curves whose modularity is not known.

Theorem 15.2.5 (Ellenberg, Skinner)**.** *Let E be a $\mathbf{Q}$-curve over a number field K with semistable reduction at all primes of K lying over 3, and suppose that K is unramified at 3. Then E is modular.*

15.3 Modularity of abelian varieties over $\overline{\mathbf{Q}}$

Hida discusses modularity of abelian varieties over $\overline{\mathbf{Q}}$ in [Hid00]. Let A be an abelian variety over $\overline{\mathbf{Q}}$. For any subalgebra $E \subset \mathbf{Q}\otimes\mathrm{End}(A/\overline{\mathbf{Q}})$, let $\mathbf{Q}\otimes\mathrm{End}_E(A/\overline{\mathbf{Q}})$ be the subalgebra of endomorphism that commute with E.

Definition 15.3.1 (Real Multiplication Abelian Variety)**.** An abelian variety A over $\overline{\mathbf{Q}}$ is a *real multiplication abelian variety* if there is a totally real field K with $[K : \mathbf{Q}] = \dim(A)$ such that

$$K \subset \mathbf{Q} \otimes \mathrm{End}(A/\overline{\mathbf{Q}}) \qquad \text{and} \qquad K = \mathbf{Q} \otimes \mathrm{End}_K(A/\overline{\mathbf{Q}}).$$

If E is a CM elliptic curve, then E is *not* a real multiplication abelian variety, because the extra CM endomorphisms in $\mathrm{End}(E/\overline{\mathbf{Q}})$ commute with $K = \mathbf{Q}$, so $K \neq \mathbf{Q} \otimes \mathrm{End}_K(A/\overline{\mathbf{Q}})$.

In analogy with $\mathbf{Q}$-curves, Hida makes the following definition.

Definition 15.3.2 ($\mathbf{Q}$-RMAV)**.** Let A be an abelian variety over $\overline{\mathbf{Q}}$ with real multiplication by a field K. Then A is a $\mathbf{Q}$-*real multiplication abelian variety* (abbreviated $\mathbf{Q}$-RMAV) if for every $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ there is a K-linear isogeny $A^\sigma \to A$. Here K acts on A^σ via the canonical isomorphism $\mathrm{End}(A) \to \mathrm{End}(A^\sigma)$, which exists since applying σ is nothing but relabeling everything.

I haven't had sufficient time to absorb Hida's paper to see just what he proves about such abelian varieties, so I'm not quite sure how to formulate the correct modularity conjectures and theorems in this generality.

Elizabeth Pyle's Ph.D. thesis [Pyl] under Ribet about "building blocks" is also very relevant to this section.

16
L-functions

16.1 L-functions attached to modular forms

Let $f = \sum_{n \geq 1} a_n q^n \in S_k(\Gamma_1(N))$ be a cusp form.

Definition 16.1.1 (L-series)**.** The *L-series* of f is

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

Definition 16.1.2 (Λ-function)**.** The *completed Λ function* of f is

$$\Lambda(f, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f, s),$$

where

$$\Gamma(s) = \int_0^\infty e^{-t}t^s \frac{dt}{t}$$

is the Γ function (so $\Gamma(n) = (n-1)!$ for positive integers n).

We can view $\Lambda(f, s)$ as a (Mellin) transform of f, in the following sense:

Proposition 16.1.3. *We have*

$$\Lambda(f, s) = N^{s/2} \int_0^\infty f(iy)y^s \frac{dy}{y},$$

and this integral converges for $\mathrm{Re}(s) > \frac{k}{2} + 1$.

Proof. We have

$$\begin{aligned}\int_0^\infty f(iy)y^s\frac{dy}{y} &= \int_0^\infty \sum_{n=1}^\infty a_n e^{-2\pi ny}y^s\frac{dy}{y}\\ &= \sum_{n=1}^\infty a_n\int_0^\infty e^{-t}(2\pi n)^{-s}t^s\frac{dt}{t} \qquad (t=2\pi ny)\\ &= (2\pi)^{-s}\Gamma(s)\sum_{n=1}^\infty \frac{a_n}{n^s}.\end{aligned}$$

To go from the first line to the second line, we reverse the summation and integration and perform the change of variables $t = 2\pi ny$. (We omit discussion of convergence.[1]) □

16.1.1 Analytic continuation and functional equations

We define the *Atkin-Lehner operator* W_N on $S_k(\Gamma_1(N))$ as follows. If $w_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0\end{smallmatrix}\right)$, then $[w_N^2]_k$ acts as $(-N)^{k-2}$, so if

$$W_N(f) = N^{1-\frac{k}{2}}\cdot f|[w_N]_k,$$

then $W_N^2 = (-1)^k$. (Note that W_N is an involution when k is even.) It is easy to check directly that if $\gamma \in \Gamma_1(N)$, then $w_N\gamma w_N^{-1} \in \Gamma_1(N)$, so W_N preserves $S_k(\Gamma_1(N))$. Note that in general W_N does *not* commute with the Hecke operators T_p, for $p \mid N$.

The following theorem is mainly due to Hecke (and maybe other people, at least in this generality). For a very general version of this theorem, see [Li75].[2]

Theorem 16.1.4. *Suppose $f \in S_k(\Gamma_1(N), \chi)$ is a cusp form with character χ. Then $\Lambda(f, s)$ extends to an entire (holomorphic on all of $\mathbf{C}$) function which satisfies the functional equation*

$$\Lambda(f, s) = i^k\Lambda(W_N(f), k - s).$$

Since $N^{s/2}(2\pi)^{-s}\Gamma(s)$ is everywhere nonzero, Theorem 16.1.4 implies that $L(f, s)$ also extends to an entire function.

It follows from Definition 16.1.2 that $\Lambda(cf, s) = c\Lambda(f, s)$ for any $c \in \mathbf{C}$. Thus if f is a W_N-eigenform, so that $W_N(f) = wf$ for some $w \in \mathbf{C}$, then the functional equation becomes

$$\Lambda(f, s) = i^k w\Lambda(f, k - s).$$

If $k = 2$, then W_N is an involution, so $w = \pm 1$, and the sign in the functional equation is $\varepsilon(f) = i^k w = -w$, which is the negative of the sign of the Atkin-Lehner involution W_N on f. It is straightforward to show that $\varepsilon(f) = 1$ if and only if $\operatorname{ord}_{s=1} L(f, s)$ is even. Parity observations such as this are extremely useful when trying to understand the Birch and Swinnerton-Dyer conjecture.

---

[1] change for book

[2] Add refs. – see page 58 of Diamond-Im.

Sketch of proof of Theorem 16.1.4 when $N = 1$. We follow [Kna92, §VIII.5] closely.

Note that since $w_1 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$, the condition $W_1(f) = f$ is satisfied for any $f \in S_k(1)$. This translates into the equality

$$f\left(-\frac{1}{z}\right) = z^k f(z). \tag{16.1.1}$$

Write $z = x + iy$ with x and y real. Then (16.1.1) along the positive imaginary axis (so $z = iy$ with y positive real) is

$$f\left(\frac{i}{y}\right) = i^k y^k f(iy). \tag{16.1.2}$$

From Proposition 16.1.3 we have

$$\Lambda(f, s) = \int_0^\infty f(iy) y^{s-1} dy, \tag{16.1.3}$$

and this integral converges for $\mathrm{Re}(s) > \frac{k}{2} + 1$.

Again using growth estimates, one shows that

$$\int_1^\infty f(iy) y^{s-1} dy$$

converges for all $s \in \mathbf{C}$, and defines an entire function. Breaking the path in (16.1.3) at 1, we have for $\mathrm{Re}(s) > \frac{k}{2} + 1$ that

$$\Lambda(f, s) = \int_0^1 f(iy) y^{s-1} dy + \int_1^\infty f(iy) y^{s-1} dy.$$

Apply the change of variables $t = 1/y$ to the first term and use (16.1.2) to get

$$\begin{aligned} \int_0^1 f(iy) y^{s-1} dy &= \int_\infty^1 -f(i/t) t^{1-s} \frac{1}{t^2} dt \\ &= \int_1^\infty f(i/t) t^{-1-s} dt \\ &= \int_1^\infty i^k t^k f(it) t^{-1-s} dt \\ &= i^k \int_1^\infty f(it) t^{k-1-s} dt. \end{aligned}$$

Thus

$$\Lambda(f, s) = i^k \int_1^\infty f(it) t^{k-s-1} dt + \int_1^\infty f(iy) y^{s-1} dy.$$

The first term is just a translation of the second, so the first term extends to an entire function as well. Thus $\Lambda(f, s)$ extends to an entire function.

The proof of the general case for $\Gamma_0(N)$ is almost the same, except the path is broken at $1/\sqrt{N}$, since $i/\sqrt{N}$ is a fixed point for w_N. □

16.1.2 A Conjecture about nonvanishing of $L(f, k/2)$

Suppose $f \in S_k(1)$ is an eigenform. If $k \equiv 2 \pmod 4$, then $L(f, k/2) = 0$ for reasons related to the discussion after the statement of Theorem 16.1.4. On the other hand, if $k \equiv 0 \pmod 4$, then $\operatorname{ord}_{s=k/2} L(f, k/2)$ is even, so $L(f, k/2)$ may or may not vanish.

Conjecture 16.1.5. *Suppose $k \equiv 0 \pmod 4$. Then $L(f, k/2) \neq 0$.*

According to [CF99], Conjecture 16.1.5 is true for weight k if there is some n such that the characteristic polynomial of T_n on $S_k(1)$ is irreducible. Thus Maeda's conjecture implies Conjecture 16.1.5. Put another way, if you find an f of level 1 and weight $k \equiv 0 \pmod 4$ such that $L(f, k/2) = 0$, then Maeda's conjecture is false for weight k.

16.1.3 Euler products

Euler products make very clear how L-functions of eigenforms encode deep arithmetic information about representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Given a "compatible family" of ℓ-adic representations ρ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, one can define an Euler product $L(\rho, s)$, but in general it is very hard to say anything about the analytic properties of $L(\rho, s)$. However, as we saw above, when ρ is attached to a modular form, we know that $L(\rho, s)$ is entire.

Theorem 16.1.6. *Let $f = \sum a_n q^n$ be a newform in $S_k(\Gamma_1(N), \varepsilon)$, and let $L(f, s) = \sum_{n \geq 1} a_n n^{-s}$ be the associated Dirichlet series. Then $L(f, s)$ has an Euler product*

$$L(f, s) = \prod_{p \mid N} \frac{1}{1 - a_p p^{-s}} \cdot \prod_{p \nmid N} \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{k-1} p^{-2s}}.$$

Note that it is not really necessary to separate out the factors with $p \mid N$ as we have done, since $\varepsilon(p) = 0$ whenever $p \mid N$. Also, note that the denominators are of the form $F(p^{-s})$, where

$$F(X) = 1 - a_p X + \varepsilon(p) p^{k-1} X^2$$

is the reverse of the characteristic polynomial of Frob_p acting on any of the ℓ-adic representations attached to f, with $p \neq \ell$.

Recall that if p is a prime, then for every $r \geq 2$ the Hecke operators satisfy the relationship

$$T_{p^r} = T_{p^{r-1}} T_p - p^{k-1} \varepsilon(p) T_{p^{r-2}}. \tag{16.1.4}$$

Lemma 16.1.7. *For every prime p we have the formal equality*

$$\sum_{r \geq 0} T_{p^r} X^r = \frac{1}{1 - T_p X + \varepsilon(p) p^{k-1} X^2}. \tag{16.1.5}$$

Proof. Multiply both sides of (16.1.5) by $1 - T_p X + \varepsilon(p) p^{k-1} X^2$ to obtain the equation

$$\sum_{r \geq 0} T_{p^r} X^r - \sum_{r \geq 0} (T_{p^r} T_p) X^{r+1} + \sum_{r \geq 0} (\varepsilon(p) p^{k-1} T_{p^r}) X^{r+2} = 1.$$

[[THERE IS A COMMENTED OUT POSTSCRIPT PICTURE – look at source and redo it in Sage]]

FIGURE 16.1.1. Graph of $L(E, s)$ for s real, for curves of ranks 0 to 3.

This equation is true if and only if the lemma is true. Equality follows by checking the first few terms and shifting the index down by 1 for the second sum and down by 2 for the third sum, then using (16.1.4). □

Note that $\varepsilon(p) = 0$ when $p \mid N$, so when $p \mid N$

$$\sum_{r \geq 0} T_{p^r} X^r = \frac{1}{1 - T_p X}.$$

Since the eigenvalues a_n of f also satisfy (16.1.4), we obtain each factor of the Euler product of Theorem 16.1.6 by substituting the a_n for the T_n and p^{-s} for X into (16.1.4). For $(n, m) = 1$, we have $a_{nm} = a_n a_m$, so

$$\sum_{n \geq 1} \frac{a_n}{n^s} = \prod_p \left(\sum_{r \geq 0} \frac{a_{p^r}}{p^{rs}} \right),$$

which gives the full Euler product for $L(f, s) = \sum a_n n^{-s}$.

16.1.4 Visualizing L-function

A. Shwayder did his Harvard junior project with me[3] on visualizing L-functions of elliptic curves (or equivalently, of newforms $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ with $a_n \in \mathbf{Z}$ for all n. The graphs in Figures 16.1.1–**??** of $L(E, s)$, for s real, and $|L(E, s)|$, for s complex, are from his paper. 3

[3] refine

17

The Birch and Swinnerton-Dyer Conjecture

This chapter is about the conjecture of Birch and Swinnerton-Dyer on the arithmetic of abelian varieties. We focus primarily on abelian varieties attached to modular forms.

In the 1960s, Sir Peter Swinnerton-Dyer worked with the EDSAC computer lab at Cambridge University, and developed an operating system that ran on that computer (so he told me once). He and Bryan Birch programmed EDSAC to compute various quantities associated to elliptic curves. They then formulated the conjectures in this chapter in the case of dimension 1 (see [Bir65, Bir71, SD67]). Tate formulated the conjectures in a functorial way for abelian varieties of arbitrary dimension over global fields in [Tat66], and proved that if the conjecture is true for an abelian variety A, then it is also true for each abelian variety isogenous to A.

Suitably interpreted, the conjectures may by viewed as generalizing the analytic class number formula, and Bloch and Kato generalized the conjectures to Grothendieck motives in [BK90].

17.1 The Rank conjecture

Let A be an abelian variety over a number field K.

Definition 17.1.1 (Mordell-Weil Group)**.** The *Mordell-Weil group* of A is the abelian group AK) of all K-rational points on A.

Theorem 17.1.2 (Mordell-Weil)**.** *The Mordell-Weil group $A(K)$ of A is finitely generated.*

The proof is nontrivial and combines two ideas. First, one proves the "weak Mordell-Weil theorem": for any integer m the quotient $A(K)/mA(K)$ is finite. This is proved by combining Galois cohomology techniques with standard finiteness theorems from algebraic number theory. The second idea is to introduce the Néron-

Tate canonical height $h : A(K) \to \mathbf{R}_{\geq 0}$ and use properties of h to deduce, from finiteness of $A(K)/mA(K)$, that $A(K)$ itself is finitely generated.

Definition 17.1.3 (Rank)**.** By the structure theorem $A(K) \cong \mathbf{Z}^r \oplus G_{\mathrm{tor}}$, where r is a nonnegative integer and G_{tor} is the torsion subgroup of G. The *rank* of A is r.

Let $f \in S_2(\Gamma_1(N))$ be a newform of level N, and let $A = A_f \subset J_1(N)$ be the corresponding abelian variety. Let $f_1, \ldots, f_d$ denote the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of f, so if $f = \sum a_n q^n$, then $f_i = \sum \sigma(a_n) q^n$, for some $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

Definition 17.1.4 (L-function of A)**.** We define the L-function of $A = A_f$ (or any abelian variety isogenous to A) to be

$$L(A, s) = \prod_{i=1}^{d} L(f_i, s).$$

By Theorem 16.1.4, each $L(f_i, s)$ is an entire function on $\mathbf{C}$, so $L(A, s)$ is entire. In Section 17.4 we will discuss an intrinsic way to define $L(A, s)$ that does not require that A be attached to a modular form. However, in general we do not know that $L(A, s)$ is entire.

Conjecture 17.1.5 (Birch and Swinnerton-Dyer)**.** *The rank of* $A(\mathbf{Q})$ *is equal to* $\mathrm{ord}_{s=1} L(A, s)$.

One motivation for Conjecture 17.1.5 is the following *formal* observation. Assume for simplicity of notation that $\dim A = 1$. By Theorem 16.1.6, the L-function $L(A, s) = L(f, s)$ has an Euler product representation

$$L(A, s) = \prod_{p \mid N} \frac{1}{1 - a_p p^{-s}} \cdot \prod_{p \nmid N} \frac{1}{1 - a_p p^{-s} + p \cdot p^{-2s}},$$

which is valid for $\mathrm{Re}(s)$ sufficiently large. (Note that $\varepsilon = 1$, since A is a modular elliptic curve, hence a quotient of $X_0(N)$.) There is no loss in considering the product $L^*(A, s)$ over only the good primes $p \nmid N$, since $\mathrm{ord}_{s=1} L(A, s) = \mathrm{ord}_{s=1} L^*(A, s)$ (because $\prod_{p \mid N} \frac{1}{1 - a_p p^{-s}}$ is nonzero at $s = 1$). We then have *formally* that

$$\begin{aligned} L^*(A, 1) &= \prod_{p \nmid N} \frac{1}{1 - a_p p^{-1} + p^{-1}} \\ &= \prod_{p \nmid N} \frac{p}{p - a_p + 1} \\ &= \prod_{p \nmid N} \frac{p}{\#A(\mathbf{F}_p)} \end{aligned}$$

The intuition is that if the rank of A is large, i.e., $A(\mathbf{Q})$ is large, then each group $A(\mathbf{F}_p)$ will also be large since it has many points coming from reducing the elements of $A(\mathbf{Q})$ modulo p. It seems likely that if the groups $\#A(\mathbf{F}_p)$ are unusually large, then $L^*(A, 1) = 0$, and computational evidence suggests the more precise Conjecture 17.1.5.

Example 17.1.6. Let A_0 be the elliptic curve $y^2 + y = x^3 - x^2$, which has rank 0 and conductor 11, let A_1 be the elliptic curve $y^2 + y = x^3 - x$, which has rank 1 and

conductor 37, let A_2 be the elliptic curve $y^2 + y = x^3 + x^2 - 2x$, which has rank 2 and conductor 389, and finally let A_3 be the elliptic curve $y^2 + y = x^3 - 7x + 6$, which has rank 3 and conductor 5077. By an exhaustive search, these are known to be the smallest-conductor elliptic curves of each rank. Conjecture 17.1.5 is known to be true for them, the most difficult being A_3, which relies on the results of [GZ86].

The following diagram illustrates $|\#A_i(\mathbf{F}_p)|$ for $p < 100$, for each of these curves. The height of the red line (first) above the prime p is $|\#A_0(\mathbf{F}_p)|$, the green line (second) gives the value for A_1, the blue line (third) for A_2, and the black line (fourth) for A_3. The intuition described above suggests that the clumps should look like triangles, with the first line shorter than the second, the second shorter than the third, and the third shorter than the fourth—however, this is visibly not the case. The large Mordell-Weil group over $\mathbf{Q}$ does not increase the size of every $E(\mathbf{F}_p)$ as much as we might at first suspect. Nonetheless, the first line is no longer than the last line for every p except $p = 41, 79, 83, 97$.

[[THERE IS A COMMENTED OUT POSTSCRIPT PICTURE – look at source and redo it in Sage]]

Remark 17.1.7. Suppose that $L(A, 1) \neq 0$. Then assuming the Riemann hypothesis for $L(A, s)$ (i.e., that $L(A, s) \neq 0$ for $\mathrm{Re}(s) > 1$), Goldfeld [Gol82] proved that the Euler product for $L(A, s)$, formally evaluated at 1, converges but *does not* converge to $L(A, 1)$. Instead, it converges (very slowly) to $L(A, 1)/\sqrt{2}$. For further details and insight into this strange behavior, see [Con03].

Remark 17.1.8. The Clay Math Institute has offered a one million dollar prize for a proof of Conjecture 17.1.5 for elliptic curves over $\mathbf{Q}$. See [Wil00].

Theorem 17.1.9 (Kolyvagin-Logachev)**.** *Suppose $f \in S_2(\Gamma_0(N))$ is a newform such that $\mathrm{ord}_{s=1} L(f, s) \leq 1$. Then Conjecture 17.1.5 is true for A_f.*

Theorem 17.1.10 (Kato)**.** *Suppose $f \in S_2(\Gamma_1(N))$ and $L(f, 1) \neq 0$. Then Conjecture 17.1.5 is true for A_f.*

17.2 Refined rank zero conjecture

Let $f \in S_2(\Gamma_1(N))$ be a newform of level N, and let $A_f \subset J_1(N)$ be the corresponding abelian variety.

The following conjecture refines Conjecture 17.1.5 in the case $L(A, 1) \neq 0$. We recall some of the notation below, where we give a formula for $L(A, 1)/\Omega_A$, which can be computed up to an vinteger, which we call the Manin index. Note that the definitions, results, and proofs in this section are all true exactly as stated with $X_1(N)$ replaced by $X_0(N)$, which is relevant if one wants to do computations.

Conjecture 17.2.1 (Birch and Swinnerton-Dyer)**.** *Suppose $L(A, 1) \neq 0$. Then*

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

By Theorem 17.1.10, the group $Ш(A)$ is finite, so the right hand side makes sense. The right hand side is a rational number, so if Conjecture 17.2.1 is true, then the quotient $L(A, 1)/\Omega_A$ should also be a rational number. In fact, this is

[[THERE IS A COMMENTED OUT POSTSCRIPT PICTURE – look at source and redo it in Sage]]

FIGURE 17.2.1. Graphs of real solutions to $y^2z = x^3 - xz^2$ on three affine patches

true, as we will prove below (see Theorem 17.2.11). Below we will discuss aspects of the proof of rationality in the case that A is an elliptic curve, and at the end of this section we give a proof of the general case.

In to more easily understanding $L(A,1)/\Omega_A$, it will be easiest to work with $A = A_f^\vee$, where $A_f^\vee$ is the dual of A_f. We view A naturally as a quotient of $J_1(N)$ as follows. Dualizing the map $A_f \hookrightarrow J_1(N)$ we obtain a surjective map $J_1(N) \to A_f^\vee$. Passing to the dual doesn't affect whether or not $L(A,1)/\Omega_A$ is rational, since changing A by an isogeny does not change $L(A,1)$, and only changes Ω_A by multiplication by a nonzero rational number.

17.2.1 The Number of real components

Definition 17.2.2 (Real Components)**.** Let c_∞ be the number of connected components of $A(\mathbf{R})$.

If A is an elliptic curve, then $c_\infty = 1$ or 2, depending on whether the graph of the affine part of $A(\mathbf{R})$ in the plane $\mathbf{R}^2$ is connected. For example, Figure 17.2.1 shows the real points of the elliptic curve defined by $y^2 = x^3 - x$ in the three affine patches that cover $\mathbf{P}^2$. The completed curve has two real components.

In general, there is a simple formula for c_∞ in terms of the action of complex conjugation on $\mathrm{H}_1(A(\mathbf{R}), \mathbf{Z})$, which can be computed using modular symbols. The formula is

$$\log_2(c_\infty) = \dim_{\mathbf{F}_2} A(\mathbf{R})[2] - \dim(A).$$

17.2.2 The Manin index

The map $J_1(N) \to A$ induces a map $\mathcal{J} \to \mathcal{A}$ on Néron models. Pullback of differentials defines a map

$$\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}) \to \mathrm{H}^0(\mathcal{J}, \Omega^1_{\mathcal{J}/\mathbf{Z}}). \tag{17.2.1}$$

One can show[1] that there is a q-expansion map 1

$$\mathrm{H}^0(\mathcal{J}, \Omega^1_{\mathcal{J}/\mathbf{Z}}) \to \mathbf{Z}[[q]] \tag{17.2.2}$$

which agrees with the usual q-expansion map after tensoring with $\mathbf{C}$. (For us $X_1(N)$ is the curve that parameterizes pairs $(E, \mu_N \hookrightarrow E)$, so that there is a q-expansion map with values in $\mathbf{Z}[[q]]$.)

Let φ_A be the composition of (17.2.1) with (17.2.2).

Definition 17.2.3 (Manin Index)**.** The *Manin index* c_A of A is the index of $\varphi_A(\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}))$ in its saturation. I.e., it is the order of the quotient group

$$\left(\frac{\mathbf{Z}[[q]]}{\varphi_A(\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}))} \right)_{\mathrm{tor}}.$$

[1] reference or further discussion

Open Problem 17.2.4. Find an algorithm to compute c_A.

Manin conjectured that $c_A = 1$ when $\dim A = 1$, and I think $c_A = 1$ in general.

Conjecture 17.2.5 (Agashe, Stein)**.** $c_A = 1$.

This conjecture is false if A is not required to be attached to a newform, even if $A_f \subset J_1(N)^{\text{new}}$. For example, Adam Joyce, a student of Kevin Buzzard, found an $A \subset J_1(431)$ (and also $A' \subset J_0(431)$) whose Manin constant is 2. Here A is isogenous over $\mathbf{Q}$ to a product of two elliptic curves. [2] Also, the Manin index for $J_0(33)$ (viewed as a quotient of $J_0(33)$) is divisible by 3, because there is a cusp form in $S_2(\Gamma_0(33))$ that has integer Fourier expansion at ∞, but not at one of the other cusps.

Theorem 17.2.6. *If $f \in S_2(\Gamma_0(N))$ then the Manin index c of $A_f^\vee$ can only divisible by 2 or primes whose square divides N. Moreover, if $4 \nmid N$, then* $\text{ord}_2(c) \leq \dim(A_f)$.

The proof involves applying nontrivial theorems of Raynaud about exactness of sequences of differentials, then using a trick with the Atkin-Lehner involution, which was introduced by Mazur in [Maz78], and finally one applies the "q-expansion principle" in characteristic p to deduce the result (see [**?**]). Also, Edixhoven claims he can prove that if A_f is an elliptic curve then c_A is only divisible by 2, 3, 5, or 7. His argument use his semistable models for $X_0(p^2)$, but my understanding is that the details are not all written up.[3]

17.2.3 The Real volume Ω_A

Definition 17.2.7 (Real Volume)**.** The *real volume* Ω_A of $A(\mathbf{R})$ is the volume of $A(\mathbf{R})$ with respect to a measure obtained by wedging together a basis for $\mathrm{H}^0(\mathcal{A}, \Omega^1)$.

If A is an elliptic curve with *minimal* Weierstrass equation

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

then one can show that

$$\omega = \frac{dx}{2y + a_1x + a_3} \tag{17.2.3}$$

is a basis for $\mathrm{H}^0(\mathcal{A}, \Omega^1)$. Thus

$$\Omega_A = \int_{A(\mathbf{R})} \frac{dx}{2y + a_1x + a_3}.$$

There is a fast algorithm for computing Ω_A, for A an elliptic curve, which relies on the quickly-convergent Gauss arithmetic-geometric mean (see [Cre97, §3.7]). For example, if A is the curve defined by $y^2 = x^3 - x$ (this is a minimal model), then

$$\Omega_A \sim 2 \times 2.622057554292119810464839589.$$

For a general abelian variety A, it is an open problem to compute Ω_A. However, we can compute Ω_A/c_A, where c_A is the Manin index of A, by explicitly computing A as a complex torus using the period mapping Φ, which we define in the next section.

---

[2] Add better reference.

[3] Refine.

17.2.4 The Period mapping

Let

$$\Phi : \mathrm{H}_1(X_1(N), \mathbf{Z}) \to \mathrm{Hom}_{\mathbf{C}}(\mathbf{C}f_1 + \cdots + \mathbf{C}f_d, \mathbf{C})$$

be the *period mapping* on integral 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 f_i of f. Extend Φ to $\mathrm{H}_1(X_1(N), \mathbf{Q})$ by $\mathbf{Q}$-linearity. We normalize Φ so that $\Phi(\{0, \infty\})(f) = L(f, 1)$. More explicitly, for $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, we have

$$\Phi(\{\alpha, \beta\})(f) = -2\pi i \int_\alpha^\beta f(z)dz.$$

The motivation for this normalization is that

$$L(f, 1) = -2\pi i \int_0^{i\infty} f(z)dz, \tag{17.2.4}$$

which we see immediately from the Mellin transform definition of $L(f, s)$:

$$L(f, s) = (2\pi)^s \Gamma(s)^{-1} \int_0^{i\infty} (-iz)^s f(z) \frac{dz}{z}.$$

17.2.5 The Manin-Drinfeld theorem

Recall the Manin-Drinfeld theorem, which we proved long ago, asserts that $\{0, \infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Q})$. We proved this by explicitly computing $(p + 1 - T_p)(\{0, \infty\})$, for $p \nmid N$, noting that the result is in $\mathrm{H}_1(X_0(N), \mathbf{Z})$, and inverting $p + 1 - T_p$. Thus there is an integer n such that $n\{0, \infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$.

Suppose that $A = A_f^\vee$ is an elliptic curve quotient of $J_0(N)$. Rewriting (17.2.4) in terms of Φ, we have $\Phi(\{0, \infty\}) = L(f, 1)$. Let ω be a minimal differential on A, as in (17.2.3), so $\omega = -c_A \cdot 2\pi i f(z)dz$, where c_A is the Manin index of A, and the equality is after pulling ω back to $\mathrm{H}^0(X_0(N), \Omega) \cong S_2(\Gamma_0(N))$. Note that when we defined c_A, there was no factor of $2\pi i$, since we compared ω with $f(q)\frac{dq}{q}$, and $q = e^{2\pi i z}$, so $dq/q = 2\pi i dz$.

17.2.6 The Period lattice

The *period lattice* of A with respect to a nonzero differential g on A is

$$\mathcal{L}_g = \left\{ \int_\gamma g : \gamma \in \mathrm{H}_1(A, \mathbf{Z}) \right\},$$

and we have $A(\mathbf{C}) \cong \mathbf{C}/\mathcal{L}_g$. This is the Abel-Jacobi theorem, and the significance of g is that we are choosing a basis for the one-dimensional $\mathbf{C}$-vector space $\mathrm{Hom}(\mathrm{H}^0(A, \Omega), \mathbf{C})$, in order to embed the image of $\mathrm{H}_1(A, \mathbf{Z})$ in $\mathbf{C}$.

The integral $\int_{A(\mathbf{R})} g$ is "visible" in terms of the complex torus representation of $A(\mathbf{C}) = \mathbf{C}/\mathcal{L}_g$. More precisely, if $\mathcal{L}_g$ is not rectangular, then $A(\mathbf{R})$ may be identified with the part of the real line in a fundamental domain for $\mathcal{L}_g$, and $\int_{A(\mathbf{R})} g$ is the length of this segment of the real line. If $\mathcal{L}_g$ is rectangular, then it is that line along with another line above it that is midway to the top of the fundamental domain.

The real volume, which appears in Conjecture 17.2.1, is

$$\Omega_A = \int_{A(\mathbf{R})} \omega = -c_A \cdot 2\pi i \int_{A(\mathbf{R})} f.$$

Thus Ω_A is the least positive real number in $\mathcal{L}_\omega = -c_A \cdot 2\pi i \mathcal{L}_f$, when the period lattice is not rectangular, and twice the least positive real number when it is.

17.2.7 The Special value $L(A, 1)$

Proposition 17.2.8. *We have $L(f, 1) \in \mathbf{R}$.*

Proof. With the right setup, this would follow immediately from the fact that $z \mapsto -\overline{z}$ fixes the homology class $\{0, \infty\}$. However, we don't have such a setup, so we give a direct proof.

Just as in the proof of the functional equation for $\Lambda(f, s)$, use that f is an eigenvector for the Atkin-Lehner operator W_N and (17.2.4) to write $L(f, 1)$ as the sum of two integrals from $i/\sqrt{N}$ to $i\infty$. Then use the calculation

$$\begin{aligned}
\overline{2\pi i \int_{i/\sqrt{N}}^{i\infty} \sum_{n=1}^{\infty} a_n e^{2\pi i n z} dz} &= -2\pi i \sum_{n=1}^{\infty} a_n \overline{\int_{i/\sqrt{N}}^{i\infty} e^{2\pi i n z} dz} \\
&= -2\pi i \sum_{n=1}^{\infty} a_n \overline{\frac{1}{2\pi i n} e^{-2\pi n/\sqrt{N}}} \\
&= 2\pi i \sum_{n=1}^{\infty} a_n \frac{1}{2\pi i n} e^{2\pi n/\sqrt{N}}
\end{aligned}$$

to see that $\overline{L(f, 1)} = L(f, 1)$. □

Remark 17.2.9. The BSD conjecture implies that $L(f, 1) \geq 0$, but this is unknown (it follows from GRH for $L(f, s)$).

17.2.8 Rationality of $L(A, 1)/\Omega_A$

Proposition 17.2.10. *Suppose $A = A_f$ is an elliptic curve. Then $L(A, 1)/\Omega_A \in \mathbf{Q}$. More precisely, if n is the smallest multiple of $\{0, \infty\}$ that lies in $\mathrm{H}_1(X_0(N), \mathbf{Z})$ and c_A is the Manin constant of A, then $2n \cdot c_A \cdot L(A, 1)/\Omega_A \in \mathbf{Z}$.*

Proof. By the Manin-Drinfeld theorem $n\{0, \infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$, so

$$n \cdot L(f, 1) = -n \cdot 2\pi i \cdot \int_0^{i\infty} f(z) dz \in -2\pi i \cdot \mathcal{L}_f = \frac{1}{c_A} \mathcal{L}_\omega.$$

Combining this with Proposition 17.2.8, we see that

$$n \cdot c_A \cdot L(f, 1) \in \mathcal{L}_\omega^+,$$

where $\mathcal{L}_\omega^+$ is the submodule fixed by complex conjugation (i.e., $\mathcal{L}_\omega^+ = \mathcal{L} \cap \mathbf{R}$). When the period lattice is not rectangular, Ω_A generates $\mathcal{L}_\omega^+$, and when it is rectangular, $\frac{1}{2}\Omega_A$ generates. Thus $n \cdot c_A \cdot L(f, 1)$ is an integer multiple of $\frac{1}{2}\Omega_A$, which proves the proposition. □

Proposition 17.2.10 can be more precise and generalized to abelian varieties $A = A_f^\vee$ attached to newforms. One can also replace n by the order of the image of $(0) - (\infty)$ in $A(\mathbf{Q})$.

Theorem 17.2.11 (Agashe, Stein)**.** *Suppose $f \in S_2(\Gamma_1(N))$ is a newform and let $A = A_f^\vee$ be the abelian variety attached to f. Then we have the following equality of rational numbers:*

$$\frac{|L(A,1)|}{\Omega_A} = \frac{1}{c_\infty \cdot c_A} \cdot [\Phi(H_1(X_1(N), \mathbf{Z}))^+ : \Phi(\mathbf{T}\{0,\infty\})].$$

Note that $L(A,1) \in \mathbf{R}$, so $|L(A,1)| = \pm L(A,1)$, and one expects, of course, that $L(A,1) \geq 0$.

For V and W lattices in an $\mathbf{R}$-vector space M, the *lattice index* $[V : W]$ is by definition the absolute value of the determinant of a change of basis taking a basis for V to a basis for W, or 0 if W has rank smaller than the dimension of M.

Proof. Let $\tilde{\Omega}_A$ be the measure of $A(\mathbf{R})$ with respect to a basis for $S_2(\Gamma_1(N), \mathbf{Z})[I_f]$, where I_f is the annihilator in $\mathbf{T}$ of f. Note that $\tilde{\Omega}_A \cdot c_A = \Omega_A$, where c_A is the Manin index. Unwinding the definitions, we find that

$$\tilde{\Omega}_A = c_\infty \cdot [\mathrm{Hom}(S_2(\Gamma_1(N), \mathbf{Z})[I_f], \mathbf{Z}) : \Phi(H_1(X_0(N), \mathbf{Z}))^+].$$

For any ring R the pairing[4] 4

$$\mathbf{T}_R \times S_2(\Gamma_1(N), R) \to R$$

given by $\langle T_n, f\rangle = a_1(T_n f)$ is perfect, so $(\mathbf{T}/I_f) \otimes R \cong \mathrm{Hom}(S_2(\Gamma_1(N), R)[I_f], R)$. Using this pairing, we may view Φ as a map

$$\Phi : H_1(X_1(N), \mathbf{Q}) \to (\mathbf{T}/I_f) \otimes \mathbf{C},$$

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 $(\mathbf{T}/I_f) \otimes \mathbf{C}$, if we view $\Phi(\{0,\infty\})$ as the operator "left-multiplication by $\Phi(\{0,\infty\})$", then

$$\det(\Phi(\{0,\infty\})) = \prod_i L(f_i, 1) = L(A, 1),$$

Letting $H = H_1(X_0(N), \mathbf{Z})$, we have

$$\begin{aligned}
[\Phi(H)^+ : \Phi(\mathbf{T}\{0,\infty\})] &= [\Phi(H)^+ : (\mathbf{T}/I_f) \cdot \Phi(\{0,\infty\})] \\
&= [\Phi(H)^+ : \mathbf{T}/I_f] \cdot [\mathbf{T}/I_f : \mathbf{T}/I_f \cdot \Phi(\{0,\infty\})] \\
&= \frac{c_\infty}{\tilde{\Omega}_A} \cdot |\det(\Phi(\{0,\infty\}))| \\
&= \frac{c_\infty c_A}{\Omega_A} \cdot |L(A,1)|,
\end{aligned}$$

[4] reference!

which proves the theorem.

□

Remark 17.2.12. Theorem 17.2.11 is false, in general, when A is a quotient of $J_1(N)$ not attached to a single $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-orbit of newforms. It could be modified to handle this more general case, but the generalization seems not to has been written down.

17.3 General refined conjecture

Conjecture 17.3.1 (Birch and Swinnerton-Dyer)**.** *Let* $r = \mathrm{ord}_{s=1} L(A, s)$*. Then* r *is the rank of* $A(\mathbf{Q})$*, the group* Ш(A) *is finite, and*

$$\frac{L^{(r)}(A,1)}{r!} = \frac{\#Ш(A) \cdot \Omega_A \cdot \mathrm{Reg}_A \cdot \prod_{p|N} c_p}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^\vee(\mathbf{Q})_{\mathrm{tor}}}.$$

17.4 The Conjecture for non-modular abelian varieties

Conjecture 17.3.1 can be extended to general abelian varieties over global fields. Here we discuss only the case of a general abelian variety A over $\mathbf{Q}$. We follow the discussion in [Lan91, 95-94] (Lang, Number Theory III)[5] , which describes Gross's formulation of the conjecture for abelian varieties over number fields, and to which we refer the reader for more details. 5

For each prime number ℓ, the ℓ-adic *Tate module* associated to A is

$$\mathrm{Ta}_\ell(A) = \varprojlim_n A(\overline{\mathbf{Q}})[\ell^n].$$

Since $A(\overline{\mathbf{Q}})[\ell^n] \cong (\mathbf{Z}/\ell^n\mathbf{Z})^{2\dim(A)}$, we see that $\mathrm{Ta}_\ell(A)$ is free of rank $2\dim(A)$ as a $\mathbf{Z}_\ell$-module. Also, since the group structure on A is defined over $\mathbf{Q}$, $\mathrm{Ta}_\ell(A)$ comes equipped with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$:

$$\rho_{A,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(\mathrm{Ta}_\ell(A)) \approx \mathrm{GL}_{2d}(\mathbf{Z}_\ell).$$

Suppose p is a prime and let $\ell \neq p$ be another prime. Fix any embedding $\overline{\mathbf{Q}} \hookrightarrow \overline{\mathbf{Q}}_p$, and notice that restriction defines a homorphism $r : \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let $G_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be the image of r. The inertia group $I_p \subset G_p$ is the kernel of the natural surjective reduction map, and we have an exact sequence

$$0 \to I_p \to \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \to 0.$$

The Galois group $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ is isomorphic to $\widehat{\mathbf{Z}}$ with canonical generator $x \mapsto x^p$. Lifting this generator, we obtain an element $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, which is well-defined up to an element of I_p. Viewed as an element of $G_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, the element Frob_p is well-defined up I_p and our choice of embedding $\overline{\mathbf{Q}} \hookrightarrow \overline{\mathbf{Q}}_p$. One

[5] Remove paren.

can show that this implies that $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is well-defined up to I_p and conjugation by an element of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

For a G_p-module M, let

$$M^{I_p} = \{x \in M : \sigma(x) = x \text{ all } \sigma \in I_p\}.$$

Because I_p acts trivially on M^{I_p}, the action of the element $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on M^{I_p} is well-defined up to conjugation (I_p acts trivially, so the "up to I_p" obstruction vanishes). Thus the characteristic polynomial of Frob_p on M^{I_p} is well-defined, which is why $L_p(A,s)$ is well-defined. The *local L-factor* of $L(A,s)$ at p is

$$L_p(A,s) = \frac{1}{\det\left(I - p^{-s}\,\mathrm{Frob}_p^{-1} \mid \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Ta}_\ell(A), \mathbf{Z}_\ell)^{I_p}\right)}.$$

Definition 17.4.1. $L(A,s) = \prod_{\text{all } p} L_p(A,s)$

For all but finitely many primes $\mathrm{Ta}_\ell(A)^{I_p} = \mathrm{Ta}_\ell(A)$. For example, if $A = A_f$ is attached to a newform $f = \sum a_n q^n$ of level N and $p \nmid \ell \cdot N$, then $\mathrm{Ta}_\ell(A)^{I_p} = \mathrm{Ta}_\ell(A)$. In this case, the Eichler-Shimura relation implies that $L_p(A,s)$ equals $\prod L_p(f_i,s)$, where the $f_i = \sum a_{n,i} q^n$ are the Galois conjugates of f and $L_p(f_i,s) = (1 - a_{p,i} \cdot p^{-s} + p^{1-2s})^{-1}$. The point is that Eichler-Shimura can be used to show that the characteristic polynomial of Frob_p is $\prod_{i=1}^{\dim(A)}(X^2 - a_{p,i}X + p^{1-2s})$.

Theorem 17.4.2. $L(A_f,s) = \prod_{i=1}^{d} L(f_i,s)$.

17.5 Visibility of Shafarevich-Tate groups

Let K be a number field. Suppose

$$0 \to A \to B \to C \to 0$$

is an exact sequence of abelian varieties over K. (Thus each of A, B, and C is a complete group variety over K, whose group is automatically abelian.) Then there is a corresponding long exact sequence of cohomology for the group $\mathrm{Gal}(\overline{\mathbf{Q}}/K)$:

$$0 \to A(K) \to B(K) \to C(K) \xrightarrow{\delta} \mathrm{H}^1(K,A) \to \mathrm{H}^1(K,B) \to \mathrm{H}^1(K,C) \to \cdots$$

The study of the Mordell-Weil group $C(K) = \mathrm{H}^0(K,C)$ is popular in arithmetic geometry. For example, the Birch and Swinnerton-Dyer conjecture (BSD conjecture), which is one of the million dollar Clay Math Problems, asserts that the dimension of $C(K) \otimes \mathbf{Q}$ equals the ordering vanishing of $L(C,s)$ at $s = 1$.

The group $\mathrm{H}^1(K,A)$ is also of interest in connection with the BSD conjecture, because it contains the Shafarevich-Tate group

$$Ш(A) = Ш(A/K) = \mathrm{Ker}\left(\mathrm{H}^1(K,A) \to \bigoplus_v \mathrm{H}^1(K_v,A)\right) \subset \mathrm{H}^1(K,A),$$

where the sum is over all places v of K (e.g., when $K = \mathbf{Q}$, the fields K_v are $\mathbf{Q}_p$ for all prime numbers p and $\mathbf{Q}_\infty = \mathbf{R}$).

The group $A(K)$ is fundamentally different than $\mathrm{H}^1(K, C)$. The Mordell-Weil group $A(K)$ is finitely generated, whereas the first Galois cohomology $\mathrm{H}^1(K, C)$ is far from being finitely generated—in fact, every element has finite order and there are infinitely many elements of any given order.

This talk is about "dimension shifting", i.e., relating information about $\mathrm{H}^0(K, C)$ to information about $\mathrm{H}^1(K, A)$.

17.5.1 Definitions

Elements of $\mathrm{H}^0(K, C)$ are simply points, i.e., elements of $C(K)$, so they are relatively easy to "visualize". In contrast, elements of $\mathrm{H}^1(K, A)$ are Galois cohomology classes, i.e., equivalence classes of set-theoretic (continuous) maps $f : \mathrm{Gal}(\overline{\mathbf{Q}}/K) \to A(\overline{\mathbf{Q}})$ such that $f(\sigma\tau) = f(\sigma) + \sigma f(\tau)$. Two maps are equivalent if their difference is a map of the form $\sigma \mapsto \sigma(P) - P$ for some fixed $P \in A(\overline{\mathbf{Q}})$. From this point of view H^1 is more mysterious than H^0.

There is an alternative way to view elements of $\mathrm{H}^1(K, A)$. The WC group of A is the group of isomorphism classes of principal homogeneous spaces for A, where a principal homogeneous space is a variety X and a map $A \times X \to X$ that satisfies the same axioms as those for a simply transitive group action. Thus X is a twist as variety of A, but $X(K) = \emptyset$, unless $X \approx A$. Also, the nontrivial elements of $Ш(A)$ correspond to the classes in WC that have a K_v-rational point for all places v, but no K-rational point.

Mazur introduced the following definition in order to help unify diverse constructions of principal homogeneous spaces:

Definition 17.5.1 (Visible)**.** The *visible subgroup* of $\mathrm{H}^1(K, A)$ in B is

$$\begin{aligned} \mathrm{Vis}_B\, \mathrm{H}^1(K, A) &= \mathrm{Ker}(\mathrm{H}^1(K, A) \to \mathrm{H}^1(K, B)) \\ &= \mathrm{Coker}(B(K) \to C(K)). \end{aligned}$$

Remark 17.5.2. Note that $\mathrm{Vis}_B\, \mathrm{H}^1(K, A)$ *does* depend on the embedding of A into B. For example, suppose $B = B_1 \times A$. Then there could be nonzero visible elements if A is embedding into the first factor, but there will be no nonzero visible elements if A is embedded into the second factor. Here we are using that $\mathrm{H}^1(K, B_1 \times A) = \mathrm{H}^1(K, B_1) \oplus \mathrm{H}^1(K, A)$.

The connection with the WC group of A is as follows. Suppose

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

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 one can show that the class of X in the WC group of A corresponds to c.

Lemma 17.5.3. *The group* $\mathrm{Vis}_B\, \mathrm{H}^1(K, A)$ *is finite.*

Proof. Since $\mathrm{Vis}_B\, \mathrm{H}^1(K, A)$ is a homomorphic image of the finitely generated group $C(K)$, it is also finitely generated. On the other hand, it is a subgroup of $\mathrm{H}^1(K, A)$, so it is a torsion group. The lemma follows since a finitely generated torsion abelian group is finite. □

17.5.2 Every element of $\mathrm{H}^1(K, A)$ is visible somewhere

Proposition 17.5.4. *Let $c \in \mathrm{H}^1(K, A)$. Then there exists an abelian variety $B = B_c$ and an embedding $A \hookrightarrow B$ such that c is visible in B.*

Proof. By definition of Galois cohomology, there is a finite extension L of K such that $\mathrm{res}_L(c) = 0$. Thus c maps to 0 in $\mathrm{H}^1(L, A_L)$. By a slight generalization of the Shapiro Lemma from group cohomology (which can be proved by dimension shifting; see, e.g.,[6] Atiyah-Wall in Cassels-Frohlich), there is a canonical isomorphism

$$\mathrm{H}^1(L, A_L) \cong \mathrm{H}^1(K, \mathrm{Res}_{L/K}(A_L)) = \mathrm{H}^1(K, B),$$

where $B = \mathrm{Res}_{L/K}(A_L)$ is the Weil restriction of scalars of A_L back down to K. The restriction of scalars B is an abelian variety of dimension $[L : K] \cdot \dim A$ that is characterized by the existence of functorial isomorphisms

$$\mathrm{Mor}_K(S, B) \cong \mathrm{Mor}_L(S_L, A_L),$$

for any K-scheme S, i.e., $B(S) = A_L(S_L)$. In particular, setting $S = A$ we find that the identity map $A_L \to A_L$ corresponds to an injection $A \hookrightarrow B$. Moreover, $c \mapsto \mathrm{res}_L(c) = 0 \in \mathrm{H}^1(K, B)$. □

Remark 17.5.5. The abelian variety B in Proposition 17.5.4 is a twist of a power of A.

17.5.3 Visibility in the context of modularity

Usually we focus on visibility of elements in $Ш(A)$. There are a number of other results about visibility in various special cases, and large tables of examples in the context of elliptic curves and modular abelian varieties. There are also interesting modularity questions and conjectures in this context.

Motivated by the desire to understand the Birch and Swinnerton-Dyer conjecture more explicitly, I developed[7] (with significant input from Agashe, Cremona, Mazur, and Merel) computational techniques for unconditionally constructing Shafarevich-Tate groups of modular abelian varieties $A \subset J_0(N)$ (or $J_1(N)$). For example, if $A \subset J_0(389)$ is the 20-dimensional simple factor, then

$$\mathbf{Z}/5\mathbf{Z} \times \mathbf{Z}/5\mathbf{Z} \subset Ш(A),$$

as predicted by the Birch and Swinnerton-Dyer conjecture. See [CM00] for examples when $\dim A = 1$. We will spend the rest of this section discussing the examples of [AS05, AS02] in more detail.

Tables 17.5.1–17.5.4 illustrate the main computational results of [AS05]. These tables were made by gathering data about certain arithmetic invariants of the 19608 abelian varieties A_f of level ≤ 2333. Of these, exactly 10360 have satisfy $L(A_f, 1) \neq 0$, and for these with $L(A_f, 1) \neq 0$, we compute a divisor and multiple of 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

[6] Fix

[7] change.

odd prime, and we prove for 37 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.

The meaning of the tables is as follows. The first column lists a level N and an isogeny class, which uniquely specifies an abelian variety $A = A_f \subset J_0(N)$. The nth isogeny class is given by the nth letter of the alphabet. We will not discuss the ordering further, except to note that usually, the dimension of A, which is given in the second column, is enough to determine A. When $L(A,1) \neq 0$, Conjecture 17.2.1 predicts that

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

We view the quotient $L(A,1)/\Omega_A$, which is a rational number, as a single quantity. We can compute multiples and divisors of every quantity appearing in the right hand side of this equation, and this yields columns three and four, which are a divisor S_ℓ and a multiple S_u of the conjectural order of $Ш(A)$ (when $S_u = S_\ell$, we put an equals sign in the S_u column). Column five, which is labeled odd $\deg(\varphi_A)$, contains the odd part of the degree of the polarization

$$\varphi_A : (A \hookrightarrow J_0(N) \cong J_0(N)^\vee \to A^\vee). \tag{17.5.1}$$

The second set of columns, columns six and seven, contain an abelian variety $B = B_g \subset J_0(N)$ such that $\#(A \cap B)$ is divisible by an odd prime divisor of S_ℓ and $L(B,1) = 0$. When $\dim(B) = 1$, we have verified that B is an elliptic curve of rank 2. The eighth column $A \cap B$ contains the group structure of $A \cap B$, where e.g., $[2^2302^2]$ is shorthand notation for $(\mathbf{Z}/2\mathbf{Z})^2 \oplus (\mathbf{Z}/302\mathbf{Z})^2$. The final column, labeled Vis, contains a divisor of the order of $\mathrm{Vis}_{A+B}(Ш(A))$.

The following proposition explains the significance of the odd $\deg(\varphi_A)$ column.

Proposition 17.5.6. *If $p \nmid \deg(\varphi_A)$, then $p \nmid \mathrm{Vis}_{J_0(N)}(\mathrm{H}^1(\mathbf{Q}, A))$.*

Proof. There exists a complementary morphism $\hat{\varphi}_A$, such that $\varphi_A \circ \hat{\varphi}_A = \hat{\varphi}_A \circ \varphi_A = [n]$, where n is the degree of φ_A. If $c \in \mathrm{H}^1(\mathbf{Q}, A)$ maps to 0 in $\mathrm{H}^1(\mathbf{Q}, J_0(N))$, then it also maps to 0 under the following composition

$$\mathrm{H}^1(\mathbf{Q}, A) \to \mathrm{H}^1(\mathbf{Q}, J_0(N)) \to \mathrm{H}^1(\mathbf{Q}, A^\vee) \xrightarrow{\hat{\varphi}_A} \mathrm{H}^1(\mathbf{Q}, A).$$

Since this composition is $[n]$, it follows that $c \in \mathrm{H}^1(\mathbf{Q}, A)[n]$, which proves the proposition. □

Remark 17.5.7. Since the degree of φ_A does not change if we extend scalars to a number field K, the subgroup of $\mathrm{H}^1(K, A)$ visible in $J_0(N)_K$, still has order divisible only by primes that divide $\deg(\varphi_A)$.

The following theorem explains the significance of the B column, and how it was used to deduce the Vis column.

Theorem 17.5.8. *Suppose A and B are abelian subvarieties of an abelian variety C over $\mathbf{Q}$ and that $A(\overline{\mathbf{Q}}) \cap B(\overline{\mathbf{Q}})$ is finite. Assume also that $A(\mathbf{Q})$ is finite. Let N be an integer divisible by the residue characteristics of primes of bad reduction for C (e.g., N could be the conductor of C). Suppose p is a prime such that*

$$p \nmid 2 \cdot N \cdot \#((A+B)/B)(\mathbf{Q})_{\mathrm{tor}} \cdot \#B(\mathbf{Q})_{\mathrm{tor}} \cdot \prod_\ell c_{A,\ell} \cdot c_{B,\ell},$$

where $c_{A,\ell} = \#\Phi_{A,\ell}(\mathbf{F}_\ell)$ is the Tamagawa number of A at ℓ (and similarly for B). Suppose furthermore that $B(\overline{\mathbf{Q}})[p] \subset A(\overline{\mathbf{Q}})$ as subgroups of $C(\overline{\mathbf{Q}})$. Then there is a natural injection

$$B(\mathbf{Q})/pB(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_C(Ш(A)).$$

A complete proof of a generalization of this theorem can be found in [AS02].

Sketch of Proof. Without loss of generality, we may assume $C = A + B$. Our hypotheses yield a diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & B[p] & \longrightarrow & B & \xrightarrow{p} & B & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \vdots & & \\ 0 & \longrightarrow & A & \longrightarrow & C & \longrightarrow & B' & \longrightarrow & 0, \end{array}$$

where $B' = C/A$. Taking $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-cohomology, we obtain the following diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & B(\mathbf{Q}) & \xrightarrow{p} & B(\mathbf{Q}) & \longrightarrow & B(\mathbf{Q})/pB(\mathbf{Q}) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \vdots & & \\ 0 & \longrightarrow & C(\mathbf{Q})/A(\mathbf{Q}) & \longrightarrow & B'(\mathbf{Q}) & \longrightarrow & \mathrm{Vis}_C(\mathrm{H}^1(\mathbf{Q}, A)) & \longrightarrow & 0. \end{array}$$

The snake lemma and our hypothesis that $p \nmid \#(C/B)(\mathbf{Q})_{\mathrm{tor}}$ imply that the rightmost vertical map is an injection

$$i : B(\mathbf{Q})/pB(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_C(\mathrm{H}^1(\mathbf{Q}, A)), \tag{17.5.2}$$

since $C(A)/(A(\mathbf{Q}) + B(\mathbf{Q}))$ is a sub-quotient of $(C'/B)(\mathbf{Q})$.

We show that the image of (17.5.2) lies in $Ш(A)$ using a local analysis at each prime, which we now sketch. At the archimedian prime, no work is needed since $p \neq 2$. At non-archimedian primes ℓ, one uses facts about Néron models (when $\ell = p$) and our hypothesis that p does not divide the Tamagawa numbers of B (when $\ell \neq p$) to show that if $x \in B(\mathbf{Q})/pB(\mathbf{Q})$, then the corresponding cohomology class $\mathrm{res}_\ell(i(x)) \in \mathrm{H}^1(\mathbf{Q}_\ell, A)$ splits over the maximal unramified extension. However,

$$\mathrm{H}^1(\mathbf{Q}_\ell^{\mathrm{ur}}/\mathbf{Q}_\ell, A) \cong \mathrm{H}^1(\overline{\mathbf{F}}_\ell/\mathbf{F}_\ell, \Phi_{A,\ell}(\overline{\mathbf{F}}_\ell)),$$

and the right hand cohomology group has order $c_{A,\ell}$, which is coprime to p. Thus $\mathrm{res}_\ell(i(x)) = 0$, which completes the sketch of the proof. □

17.5.4 Future directions

The data in Tables 17.5.1-17.5.4 could be investigated further.

It should be possible to replace the hypothesis that $B[p] \subset A$, with the weaker hypothesis that $B[\mathfrak{m}] \subset A$, where $\mathfrak{m}$ is a maximal ideal of the Hecke algebra $\mathbf{T}$. For example, this improvement would help one to show that 5^2 divides the order of the Shafarevich-Tate group of **1041E**. Note that for this example, we only know that $L(B, 1) = 0$, not that $B(\mathbf{Q})$ has positive rank (as predicted by Conjecture 17.1.5), which is another obstruction.

One can consider visibility at a higher level. For example, there are elements of order 3 in the Shafarevich-Tate group of **551H** that are not visible in $J_0(551)$, but these elements are visible in $J_0(2 \cdot 551)$, according to the computations in [Ste04] (Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using MAGMA).[8] 8

Conjecture 17.5.9 (Stein)**.** *Suppose $c \in Ш(A_f)$, where $A_f \subset J_0(N)$. Then there exists M such that c is visible in $J_0(NM)$. In other words, every element of $Ш(A_f)$ is "modular".*

17.6 Kolyvagin's Euler system of Heegner points

In this section we will briefly sketch some of the key ideas behind Kolyvagin's proof of Theorem 17.1.9. We will follow [Rub89] very closely. Two other excellent references are [Gro91] and [McC91]. Kolyvagin's original papers[9] on this theorem 9
are not so easy to read because they are all translations from Russian, but none of the three papers cited above give a complete proof of his theorem.

We only sketch a proof of the following special case of Kolyvagin's theorem.

Theorem 17.6.1. *Suppose E is an elliptic curve over $\mathbf{Q}$ such that $L(E,1) \neq 0$. Then $E(\mathbf{Q})$ is finite and $Ш(E)[p] = 0$ for almost all primes p.*

The strategy of the proof is as follows. Applying Galois cohomology to the multiplication by p sequence

$$0 \to E[p] \to E \xrightarrow{p} E \to 0,$$

we obtain a short exact sequence

$$0 \to E(\mathbf{Q})/pE(\mathbf{Q}) \to \mathrm{H}^1(\mathbf{Q}, E[p]) \to \mathrm{H}^1(\mathbf{Q}, E)[p] \to 0.$$

The inverse image of $Ш(E)[p]$ in $\mathrm{H}^1(\mathbf{Q}, E[p])$ is called the *Selmer group of E at p*, and we will denote it by $\mathrm{Sel}^{(p)}(E)$. We have an exact sequence

$$0 \to E(\mathbf{Q})/pE(\mathbf{Q}) \to \mathrm{Sel}^{(p)}(E) \to Ш(E)[p] \to 0.$$

Because $E(\mathbf{Q})$ is finitely generated, to prove Theorem 17.6.1 it suffices to prove that $\mathrm{Sel}^{(p)}(E) = 0$ for all but finitely many primes p. We do this by using complex multiplication elliptic curves to construct Galois cohomology classes with precise local behavior.

It is usually easier to say something about the Galois cohomology of a module over a local field than over $\mathbf{Q}$, so we introduce some more notation to help formalize this. For any prime p and place v we have a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E(\mathbf{Q})/pE(\mathbf{Q}) & \longrightarrow & \mathrm{H}^1(\mathbf{Q}, E[p]) & \longrightarrow & \mathrm{H}^1(\mathbf{Q}, E)[p] & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow{\scriptstyle \mathrm{res}_v} & & \downarrow{\scriptstyle \mathrm{res}_v} & & \\
0 & \longrightarrow & E(\mathbf{Q}_v)/pE(\mathbf{Q}_v) & \longrightarrow & \mathrm{H}^1(\mathbf{Q}_v, E[p]) & \longrightarrow & \mathrm{H}^1(\mathbf{Q}_v, E)[p] & \longrightarrow & 0
\end{array}$$

[8] remove

[9] Add references.

TABLE 17.5.1. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	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	$[2^2302^2]$	–
1070M	7	$3\cdot5^2$	$3^2\cdot5^2$	$3\cdot5\cdot_{1720261}$	**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	$[2^262^2]$	–
1429B	64	5^2	=	$_{1}$	NONE			
1443G	5	7^2	=	$7^2\cdot_{18525}$	**1443C**	1	$[7^114^1]$	–
1446N	7	3^2	=	$3\cdot_{17459029}$	**1446A**	1	$[12^2]$	–

TABLE 17.5.2. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	Vis
1466H$*$	23	13^2	$=$	$13 \cdot {\scriptstyle 25631993723}$	**1466B**	1	$[26^2]$	13^2
1477C$*$	24	13^2	$=$	$13 \cdot {\scriptstyle 57037637}$	**1477A**	1	$[13^2]$	13^2
1481C	71	13^2	$=$	${\scriptstyle 70825}$	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	$3 \cdot {\scriptstyle 759709}$	NONE			
1529D	36	5^2	$=$	${\scriptstyle 535641763}$	NONE			
1531D	73	3	3^2	3	**1531A**	1	$[48^2]$	–
1534J	6	3	3^2	$3^2 \cdot {\scriptstyle 635931}$	**1534B**	1	$[6^2]$	–
1551G	13	3^2	$=$	$3 \cdot {\scriptstyle 110659885}$	**141A**	1	$[15^2]$	–
1559B	90	11^2	$=$	${\scriptstyle 1}$	NONE			
1567D	69	$7^2 \cdot 41^2$	$=$	$7 \cdot 41$	**1567B**	3	$[4^4 1148^2]$	–
1570J$*$	6	11^2	$=$	$11 \cdot {\scriptstyle 228651397}$	**1570B**	1	$[11^2]$	11^2
1577E	36	3	3^2	$3^2 \cdot {\scriptstyle 15}$	**83A**	1	$[6^2]$	–
1589D	35	3^2	$=$	${\scriptstyle 6005292627343}$	NONE			
1591F$*$	35	31^2	$=$	$31 \cdot {\scriptstyle 2401}$	**1591A**	1	$[31^2]$	31^2
1594J	17	3^2	$=$	$3 \cdot {\scriptstyle 259338050025131}$	**1594A**	1	$[12^2]$	–
1613D$*$	75	5^2	$=$	$5 \cdot {\scriptstyle 19}$	**1613A**	1	$[20^2]$	5^2
1615J	13	3^4	$=$	$3^2 \cdot {\scriptstyle 13317421}$	**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	$=$	${\scriptstyle 6354841131}$	NONE			
1633D	27	$3^6 \cdot 7^2$	$=$	$3^5 \cdot 7 \cdot {\scriptstyle 31375}$	**1633A**	3	$[6^4 42^2]$	–
1634K	12	3^2	$=$	$3 \cdot {\scriptstyle 3311565989}$	**817A**	1	$[3^2]$	–
1639G$*$	34	17^2	$=$	$17 \cdot {\scriptstyle 82355}$	**1639B**	1	$[34^2]$	17^2
1641J$*$	24	23^2	$=$	$23 \cdot {\scriptstyle 1491344147471}$	**1641B**	1	$[23^2]$	23^2
1642D$*$	14	7^2	$=$	$7 \cdot {\scriptstyle 123398360851}$	**1642A**	1	$[7^2]$	7^2
1662K	7	11^2	$=$	$11 \cdot {\scriptstyle 16610917393}$	**1662A**	1	$[11^2]$	–
1664K	1	5^2	$=$	$5 \cdot {\scriptstyle 7}$	**1664N**	1	$[5^2]$	–
1679C	45	11^2	$=$	${\scriptstyle 6489}$	NONE			
1689E	28	3^2	$=$	$3 \cdot {\scriptstyle 172707180029157365}$	**563A**	1	$[3^2]$	–
1693C	72	1301^2	$=$	1301	**1693A**	3	$[2^4 2602^2]$	–
1717H$*$	34	13^2	$=$	$13 \cdot {\scriptstyle 345}$	**1717B**	1	$[26^2]$	13^2
1727E	39	3^2	$=$	${\scriptstyle 118242943}$	NONE			
1739F	43	659^2	$=$	$659 \cdot {\scriptstyle 151291281}$	**1739C**	2	$[2^2 1318^2]$	–
1745K	33	5^2	$=$	$5 \cdot {\scriptstyle 1971380677489}$	**1745D**	1	$[20^2]$	–
1751C	45	5^2	$=$	$5 \cdot {\scriptstyle 707}$	**103A**	2	$[505^2]$	–
1781D	44	3^2	$=$	${\scriptstyle 61541}$	NONE			
1793G$*$	36	23^2	$=$	$23 \cdot {\scriptstyle 8846589}$	**1793B**	1	$[23^2]$	23^2
1799D	44	5^2	$=$	${\scriptstyle 201449}$	NONE			
1811D	98	31^2	$=$	${\scriptstyle 1}$	NONE			
1829E	44	13^2	$=$	${\scriptstyle 3595}$	NONE			
1843F	40	3^2	$=$	${\scriptstyle 8389}$	NONE			
1847B	98	3^6	$=$	${\scriptstyle 1}$	NONE			
1871C	98	19^2	$=$	${\scriptstyle 14699}$	NONE			

TABLE 17.5.3. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	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	$[4^2196^2]$	–
1907D*	90	7^2	=	$7\cdot_{165}$	**1907A**	1	$[56^2]$	7^2
1909D*	38	3^4	=	$3^2\cdot_{9317}$	**1909A**	1	$[18^2]$	3^4
1913B*	1	3^2	=	$3\cdot_{103}$	**1913A**	1	$[3^2]$	3^2
1913E	84	$5^4\cdot61^2$	=	$5^2\cdot61\cdot_{103}$	**1913A**	1	$[10^2]$	–
					1913C	2	$[2^2610^2]$	–
1919D	52	23^2	=	$_{675}$	NONE			
1927E	45	3^2	3^4	$_{52667}$	NONE			
1933C	83	$3^2\cdot7$	$3^2\cdot7^2$	$3\cdot7$	**1933A**	1	$[42^2]$	3^2
1943E	46	13^2	=	$_{62931125}$	NONE			
1945E*	34	3^2	=	$3\cdot_{571255479184807}$	**389A**	1	$[3^2]$	3^2
1957E*	37	$7^2\cdot11^2$	=	$7\cdot11\cdot_{3481}$	**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*	94	29^2	=	29	**2027A**	1	$[58^2]$	29^2
2029C	90	$5^2\cdot269^2$	=	$5\cdot269$	**2029A**	2	$[2^22690^2]$	–
2031H*	36	11^2	=	$11\cdot_{1014875952355}$	**2031C**	1	$[44^2]$	11^2
2035K	16	11^2	=	$11\cdot_{218702421}$	**2035C**	1	$[11^122^1]$	–
2038F	25	5	5^2	$5^2\cdot_{92198576587}$	**2038A**	1	$[20^2]$	–
					1019B	1	$[5^2]$	–
2039F	99	$3^4\cdot5^2$	=	$_{13741381043009}$	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	$5\cdot7^2$	$5^2\cdot7^2$	$5^2\cdot7\cdot_{167359757}$	**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	$3^2\cdot11^2$	=	$3^2\cdot11\cdot_{874412923071571792611}$	**2103B**	1	$[33^2]$	11^2
2111B	112	211^2	=	$_{1}$	NONE			
2113B	91	7^2	=	$_{1}$	NONE			
2117E*	45	19^2	=	$19\cdot_{1078389}$	**2117A**	1	$[38^2]$	19^2

TABLE 17.5.4. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	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	$[2^2174^2]$	–
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	$[2^2302^2]$	–
2287B	109	71^2	=	$_1$	NONE			
2291C	52	3^2	=	$_{427943}$	NONE			
2293C	96	479^2	=	479	**2293A**	2	$[2^2958^2]$	–
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	$[2^6166682^2]$	–

Observe that

$$\mathrm{Sel}^{(p)}(E) = \bigcap_v \mathrm{res}_v^{-1}(\text{ image } E(\mathbf{Q}_v)).$$

For $s \in \mathrm{Sel}^{(p)}(E)$, let s_v denote the inverse image of $\mathrm{res}_v(s)$ in $E(\mathbf{Q}_v)/pE(\mathbf{Q}_v)$.

Our first proposition asserts that if we can construct a cohomology class with certain properties, then that cohomology class forces $\mathrm{Sel}^{(p)}(E)$ to be locally trivial.

Proposition 17.6.2. *Suppose ℓ is a prime such that $E(\mathbf{Q}_\ell)[p] \cong \mathbf{Z}/p\mathbf{Z}$ and suppose there is a cohomology class $c_\ell \in \mathrm{H}^1(\mathbf{Q}, E)[p]$ such that $\mathrm{res}_v(c_\ell) \neq 0$ if and only if $v = \ell$. Then $\mathrm{res}_\ell(\mathrm{Sel}^{(p)}(E)) = 0 \subset \mathrm{H}^1(\mathbf{Q}_\ell, E[p])$.*

Sketch of proof. Suppose $s \in \mathrm{Sel}^{(p)}(E)$. Using the local Tate pairing we obtain, for each v, a nondegenerate pairing

$$\langle\ ,\ \rangle_v : E(\mathbf{Q}_v)/pE(\mathbf{Q}_v) \times \mathrm{H}^1(\mathbf{Q}_v, E)[p] \to \mathbf{Z}/p\mathbf{Z}.$$

Unwinding the definitions, and using the fact that the sum of the local invariants of an element of $\mathrm{Br}(\mathbf{Q})$ is trivial, we see that our hypothesis that all but one restriction of c_ℓ is trivial implies that $\langle s_\ell, \mathrm{res}_\ell(c_\ell)\rangle = 0$. Since $\mathrm{res}_\ell(c_\ell) \neq 0$, this implies that $s_\ell = 0$. □

Let $\tau \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ denote a fixed choice of complex conjugation, which is defined by fixing an embedding of $\overline{\mathbf{Q}}$ into $\mathbf{C}$. For any module M on which τ acts, let

$$M^+ = \{x \in M : \tau(x) = x\} \qquad \text{and} \qquad M^- = \{x \in M : \tau(x) = -x\}.$$

If E is defined by $y^2 = x^3 + ax + b$ and $K = \mathbf{Q}(\sqrt{d})$, then the twist of E by K is the elliptic curve $dy^2 = x^3 + ax + b$.

Theorem 17.6.3. *There exists infinitely many quadratic imaginary fields K in which all primes dividing the conductor N of E split, and such that $L'(E^K, 1) \neq 0$, where E^K is the twist of E over K.*

Fix a K as in the theorem such that $\mathcal{O}_K^* = \{\pm 1\}$, i.e., so that the discriminant D of K is not -3 or -4.

Lemma 17.6.4. *Suppose $\ell \nmid pDN$ and $\mathrm{Frob}_\ell(K(E[p])/\mathbf{Q}) = [\tau]$. Then ℓ is inert in K, we have $a_\ell \equiv \ell + 1 \equiv 0 \pmod{p}$, and the four groups $E(\mathbf{Q}_\ell)[p]$, $E(\mathbf{F}_\ell)[p]$, $E(K_\ell)[p]^-$, and $E(\mathbf{F}_{\ell^2})[p]^-$ are all cyclic of order p.*

Proof. The first assertion is true because $\tau|_K$ has order 2. The characteristic polynomial of $\mathrm{Frob}_\ell(K(E[p])/\mathbf{Q})$ on $E[p]$ is $x^2 - a_\ell x + \ell$, and the characteristic polynomial of τ on $E[p]$ is $x^2 - 1$, which proves the second assertion. For the third, we have

$$(\mathbf{Z}/p\mathbf{Z})^2 \cong E(\overline{\mathbf{Q}})[p] \cong E(K_\ell)[p] \cong E(\mathbf{Q}_\ell)[p] \oplus E(K_\ell)[p]^-,$$

and each summand on the right must be nonzero since $\tau \neq 1$ on $E[p]$. □

For the rest of the argument, we assume that p is odd and does not divide the order of the class group of K. We will now use the theory of complex multiplication elliptic curves and class field theory to construct cohomology classes c_ℓ which we will use in Proposition 17.6.2.

Since every prime dividing the conductor N of E splits in K, there is an ideal $\mathcal{N}$ of K such that $\mathcal{O}_K/\mathcal{N} \cong \mathbf{Z}/N\mathbf{Z}$. Fix such an ideal for the rest of the argument. Let

$$\mathcal{N}^{-1} = \{x \in K : x\mathcal{N} \subset \mathcal{O}_K\}$$

be the inverse of $\mathcal{N}$ in the group of fractional ideals of $\mathcal{O}_K$. The pair

$$(\mathbf{C}/\mathcal{O}_K, \quad \mathcal{N}^{-1}/\mathcal{O}_K)$$

then defines a point on $x \in X_0(N)$. By complex multiplication theory, the point x is defined over the Hilbert Class Field H of K, which is the maximal unramified abelian extension of K. Also, class field theory implies that $\mathrm{Gal}(H/K)$ is isomorphic to the ideal class group of K.

Let $\pi : X_0(N) \to E$ be a (minimal) modular parametrization. Then $y_K = \mathrm{Tr}_{H/K}\, \pi(x_H) \in E(K)$, and we call

$$y = y_K - \tau(y_K) \in E(K)^-$$

the *Heegner point* associated to K. Our hypotheses on K imply, by the theorem of Gross and Zagier, that y has infinite order, a fact which will be crucial in our construction of cohomology classes c_ℓ. If we were to choose a K such that y were torsion (of order coprime to p), then our classes c_ℓ would be locally trivial, hence give candidate elements of $Ш(E)$ (see [Gro91, §11, pg. 254] and [McC91] for more on this connection).

Suppose $\ell \nmid N$ is inert in K (i.e., does not split). Let

$$\mathcal{O}_\ell = \mathbf{Z} + \ell\mathcal{O}_K,$$

which is the *order of conductor* ℓ in $\mathcal{O}_K$. Notice that $\mathcal{O}_K/\mathcal{O}_\ell \cong \mathbf{Z}/\ell\mathbf{Z}$ as *groups*. Since $K \subset \mathbf{C}$, it make sense to view $\mathcal{O}_\ell$ as a lattice in $\mathbf{C}$. The pair

$$(\mathbf{C}/\mathcal{O}_\ell, \quad (\mathcal{N} \cap \mathcal{O}_\ell)^{-1}/\mathcal{O}_\ell)$$

then defines a CM point $x_\ell \in X_0(N)$. This point won't be defined over the Hilbert class field H, though, but instead over an abelian extension $K[\ell]$ of K that is a cyclic extension of H of degree $\ell + 1$ totally ramified over H at ℓ and unramified everywhere else. Let

$$y_\ell = \pi(x_\ell) \in E(K[\ell]).$$

Proposition 17.6.5. *We have* $\mathrm{Tr}_{K[\ell]/H}(y_\ell) = a_\ell y_H$, *where* $y_H = \pi(x_H)$. *Also, for any prime* λ *of* $K[\ell]$ *lying above* ℓ, *we have*

$$\tilde{y}_\ell = \mathrm{Frob}_\ell(\tilde{y}_H),$$

as elements of $E(\mathbf{F}_{\ell^2})$.

We will not prove this proposition, except to note that it follows from the explicit descriptions of the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on CM points on $X_0(N)$ and of the Hecke correspondences.

We are almost ready to construct the cohomology classes c_ℓ. Suppose $\ell \nmid pDN$ is any prime such that $\mathrm{Frob}_\ell(K(E[p])/\mathbf{Q}) = [\tau]$. By Lemma 17.6.4, $[K[\ell] : H] = \ell + 1$ is divisible by p, so there is a unique extension H' of H of degree p contained in $K[\ell]$. Let φ be a lift of $\mathrm{Frob}_\ell(H/\mathbf{Q})$ to $\mathrm{Gal}(H'/\mathbf{Q})$, and let

$$z' = \mathrm{Tr}_{K[\ell]/H'}(y_\ell - \varphi(y_\ell)) - \frac{a_\ell}{p}(y_H - \varphi(y_H)) \in E(H').$$

Lemma 17.6.6. $\mathrm{Tr}_{H'/H}(z') = 0$.

Proof. Using properties of the trace and Proposition 17.6.5, we have

$$\begin{aligned}\mathrm{Tr}_{H'/H}(z') &= \mathrm{Tr}_{K[\ell]/H}(y_\ell) - \mathrm{Tr}_{K[\ell]/H}(\varphi(y_\ell)) - a_\ell(y_H) - a_\ell\varphi(y_H)\\ &= a_\ell y_H - a_\ell\varphi(y_H) - a_\ell(y_H) - a_\ell\varphi(y_H)\\ &= 0.\end{aligned}$$

□

For the rest of the proof we only consider primes p such that

$$\mathrm{H}^1(\mathbf{Q}_v^{\mathrm{unr}}/\mathbf{Q}_v, E(\mathbf{Q}_v^{\mathrm{unr}}))[p] = 0$$

for all v, where $\mathbf{Q}_v^{\mathrm{unr}}$ is the maximal unramified extension of $\mathbf{Q}_v$. (This only excludes finitely many primes, by [Milne, Arithmetic Duality Theorems, Prop. I.3.8].)

Proposition 17.6.7. *There is an element $c_\ell \in \mathrm{H}^1(\mathbf{Q}, E)[p]$ such that $\mathrm{res}_v(c_\ell) = 0$ for all $v \neq \ell$, and $\mathrm{res}_\ell(c_\ell) \neq 0$ if and only if $y \notin pE(K_\ell)$.*

Proof. Recall that we assumed that $p \nmid [H : K]$. Thus there is a unique extension K' of K of degree p in $K[\ell]$, which is totally ramified at ℓ and unramified everywhere else, such that $H' = HK'$. Let

$$z = \mathrm{Tr}_{H'/K'}(z') \in E(K').$$

By Lemma 17.6.6, $\mathrm{Tr}_{K'/K}(z) = 0$.

The extension $\mathrm{Gal}(K'/K)$ is cyclic (since it has degree p), so it is generated by a single element, say σ. The cohomology of a cyclic group is easy to understand (see Atiyah-Wall[10]). In particular, we have a canonical isomorphism 10

$$\mathrm{H}^1(K'/K, E(K')) \cong \frac{\ker(\mathrm{Tr}_{K'/K} : E(K') \to E(K))}{(\sigma - 1)E(K')}.$$

Define c'_ℓ to be the element of $\mathrm{H}^1(K'/K, E(K'))$ that corresponds to z under this isomorphism. A more careful analysis shows that c'_ℓ is the restriction of an element c_ℓ in $\mathrm{H}^1(\mathbf{Q}, E)[p]$.

If $v \neq \ell$ then

$$\mathrm{res}_v(c_\ell) \in \mathrm{H}^1(\mathbf{Q}_v^{\mathrm{unr}}/\mathbf{Q}_v, E(\mathbf{Q}_v^{\mathrm{unr}}))[p] = 0,$$

as asserted. The assertion about $\mathrm{res}_\ell(c_\ell)$ is obtained with some work by reducing modulo ℓ and applying Lemma 17.6.4 and Proposition 17.6.5. □

Remark 17.6.8. McCallum has observed that c_ℓ is represented by the 1-cocycle

$$\sigma \mapsto -\frac{(\sigma - 1)y_\ell}{\ell},$$

where $((\sigma-1)y_\ell)/\ell$ is the unique point $P \in E(K[\ell])$ such that $\ell P = (\sigma-1)y_\ell$. (The point P is unique because no nontrivial p-torsion on E is defined over $E(K[\ell])$.)

[10] expand

Corollary 17.6.9. *If $y \notin pE(K_\ell)$, then* $\mathrm{res}_\ell(\mathrm{Sel}^{(p)}(E)) = 0$.

Proof. Combine Propositions 17.6.2 and 17.6.7. □

If $t \in \mathrm{H}^1(K, E[p])$, denote by $\hat{t}$ the image of t under restriction:

$$\mathrm{H}^1(K, E[p]) \to \mathrm{Hom}(\mathrm{Gal}(\overline{\mathbf{Q}}/K(E[p])), E[p]). \tag{17.6.1}$$

It is useful to consider $\hat{t}$, since homomorphisms satisfy a local-to-global principle. A homomorphism $\varphi : \mathrm{Gal}(\overline{\mathbf{Q}}/F) \to M$ is 0 if and only if it is 0 when restricted to $\mathrm{Gal}(\overline{F}_\lambda/F_\lambda)$ for all primes λ of F (this fact follows from the Chebotarev density theorem).

For the rest of the proof we now assume that p is large enough that there are no $\mathbf{Q}$-rational cyclic subgrups of E of order p, and $\mathrm{H}^1(K(E[p])/K, E[p]) = 0$. (That we can do this follows from a theorem of Serre when E does not have CM, or CM theory when E does have CM.)

Lemma 17.6.10. *Suppose $t \in \mathrm{H}^1(K, E[p])^{\pm}$ and the image of $\hat{t}$ is cyclic. Then $t = 0$.*

Proof. Since τ acts on $\hat{t}$ by $\pm$, the image of $\hat{t}$ is rational over $\mathbf{Q}$. Thus the image of τ is trivial. The kernel of the restriction map (17.6.1) is also trivial by assumption, so $t = 0$. □

We are now ready to prove Theorem 17.6.1. Choose p large enough so that $y \notin pE(K)$, in addition to all other "sufficiently large" constraints that we put on p above.

Fix $s \in \mathrm{Sel}^{(p)}(E)$ and write $\hat{s}$ for the restriction of s to a homomorphism $\mathrm{Gal}(\overline{\mathbf{Q}}/K(E[p])) \to E[p]$. Our goal is to prove that $s = 0$. Write $\hat{y}$ for the restriction to $\mathrm{Gal}(\overline{\mathbf{Q}}/K(E[p]))$ of the image of y under the injection

$$E(K)^-/pE(K)^- \to \mathrm{H}^1(K, E[p])^-. \tag{17.6.2}$$

Fix a finite extension F of $K(E[p])$ that is Galois over $\mathbf{Q}$, so that both homomorphism $\hat{s}$ and $\hat{y}$ factor through $G = \mathrm{Gal}(F/K(E[p]))$.

Suppose $\gamma \in G$, and use the Chebotarev Density Theorem to find a prime $\ell \nmid pDN$ such that $\mathrm{Frob}_\ell(F/\mathbf{Q}) = [\gamma\tau]$. Then

$$\mathrm{Frob}_\ell(K(E[p])/\mathbf{Q}) = [\tau],$$

and

$$\mathrm{Frob}_\ell(F/K(E[p])) = [\gamma\tau]^{\mathrm{order}(\tau)} = [(\gamma\tau)^2].$$

By Lemma 17.6.10, $s_\ell = 0$ implies that $\hat{s}((\gamma\tau)^2) = 0$ for all $\gamma \in G$. Likewise, $y \in pE(K_\ell)$ implies that $\hat{y}((\gamma\tau)^2) = 0$ for all $\gamma \in G$. Since $s \in \mathrm{H}^1(\mathbf{Q}, E)$ we have $\tau(\hat{s}) = \hat{s}$ and by (17.6.2) we have $\tau(\hat{y}) = -\hat{y}$, so

$$\hat{s}((\gamma\tau)^2) = \hat{s}(\gamma) + \hat{s}(\tau\gamma\tau) = (1 + \tau)\hat{s}(\gamma) \tag{17.6.3}$$

$$\hat{y}((\gamma\tau)^2) = \hat{y}(\gamma) + \hat{y}(\tau\gamma\tau) = (1 - \tau)\hat{y}(\gamma). \tag{17.6.4}$$

By Corollary 17.6.9, for every $\gamma \in G$, one of $\hat{s}((\gamma\tau)^2)$ or $\hat{y}((\gamma\tau)^2)$ is 0, so by (17.6.3) and (17.6.4) at least one of the following holds:

$$\hat{s}(\gamma) \in E[p]^-$$

or

$$\hat{y}(\gamma) \in E[p]^+.$$

Thus

$$G = \hat{s}^{-1}(E[p]^-) \bigcup \hat{y}^{-1}(E[p]^+).$$

Since a group cannot be the union of two proper subgroups, either $\hat{s}(G) \subset E[p]^-$ or $\hat{y}(G) \subset E[p]^+$. By Lemma 17.6.10, either $s = 0$ or $y \in pE(K)$. But we assumed that $y \notin pE(K)$, so $s = 0$, as claimed.

17.6.1 A Heegner point when $N = 11$

Let E be the elliptic curve $X_0(11)$, which has conductor 11 and Weierstrass equation $y^2 + y = x^3 - x^2 - 10x - 20$. The following uses a nonstandard MAGMA package...

```
> E := EllipticCurve(CremonaDatabase(),"11A");
> E;
Elliptic Curve defined by y^2 + y = x^3 - x^2 - 10*x - 20 over Rational Field
>
[]
> A := [-D : D in [1..20] | HeegnerHypothesis(E,-D)]; A;
[ -7, -8, -19 ]
> P := [* *]; for D in A do time p,q := HeegnerPoint(E,D); Append(~P,<p,q>); end for;
> P;   // pairs the (heegner point, its trace to Q)
[* <(0.50000000000000000023668707008012648357 + 1.32287565553229529292727422603610*i :
    -2.00000000000000000434731999629536227 02 + 5.29150262212918118626810364338 7*i : 1),
    (11 : -61 : 1)>,
<(-2.99999999999999999717372417577 + 1.41421356237309505112821543066 8*i :
    2.99999999999999999372415320819 + 4.24264068711928515852918481721 2*i : 1),
     (-3/4 : 121/8 : 1)>,
<(-6.99999999999999999753082167926 + 8.71779788708134709829181547345*i :
   37.00000000000000004486713945213 6100922 + 8.71779788708134710182445372185*i : 1),
   (1161/16 : -33549/64 : 1)> *]
> P[1];
<(0.50000000000000000023668707008012648357 + 1.32287565553229529292727422603610*i :
   -2.00000000000000000434731999629536227 02 + 5.29150262212918118626810364338 7*i : 1),
  (11 : -61 : 1)>
> Sqrt(-7)/2;
2.64575131106459059050161575359*i
> Sqrt(-7)/2;
1.32287565553229529525080787679 5*i
> 2*Sqrt(-7);
5.29150262212918118100323150718 3*i
> // Thus  ((1+sqrt(-7))/2 , -2+2*sqrt(-7)) is the Heegner point
> // corresponding to Q(sqrt(-7)).
```

17.6.2 Kolyvagin's Euler system for curves of rank at least 2

[11]

11

[11]Summary of my work to show nontriviality in some cases. Make sure to also mention work of Rubin, Mazur, Howard, Weinstein, etc., in this section. One interesting idea might be to add my whole *book* on BSD into this book...

18

The Gorenstein Property for Hecke Algebras

18.1 Mod ℓ representations associated to modular forms

Suppose $f = \sum a_n q^n$ is a newform of exact level N and weight 2 for the congruence subgroup $\Gamma_0(N)$. Let $E = \mathbf{Q}(\ldots, a_n, \ldots)$ and let λ be a place of E lying over the prime ℓ of $\mathbf{Q}$. The action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the ℓ-adic Tate module of the associated abelian variety A_f gives rise to a representation

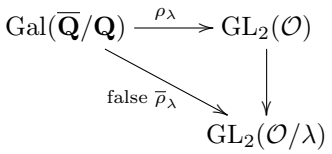
$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(E_\lambda)$$

that satisfies $\det(\rho_\lambda) = \chi_\ell$ and $\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_p)) = a_p$ for $p \nmid \ell N$. Using the following lemma, it is possible to reduce ρ_λ module λ.

Lemma 18.1.1. *Let $\mathcal{O}$ be the ring of integers of E_λ. Then ρ_λ is equivalent to a representation that takes values in $\mathrm{GL}_2(\mathcal{O})$.*

Proof. View $\mathrm{GL}_2(E_\lambda)$ as the group of automorphisms of a 2-dimensional E_λ-vector space V. A *lattice* $L \subset V$ is a free $\mathcal{O}$-module of rank 2 such that $L \otimes E_\lambda \cong V$. It suffices to find an $\mathcal{O}$-lattice L in V that is invariant under the action via ρ_λ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. For then the matrices of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ with respect to a basis of V consisting of vectors from L will have coefficients in $\mathcal{O}$. Choose any lattice $L_0 \subset V$. Since L_0 is discrete and $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact, the set of lattices $\rho_\lambda(g)L_0$ with $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is finite. Let $L = \sum \rho_\lambda(g)L_0$ be the sum of the finitely many conjugates of L_0; then L is Galois invariant. The sum is a lattice because it is finitely generated and torsion free, and $\mathcal{O}$ is a principal ideal domain. □

Choose an $\mathcal{O}$ as in the lemma, and tentatively write $\overline{\rho}_\lambda = \rho_\lambda \mod \lambda$:



The drawback to this definition is that $\overline{\rho}_\lambda$ is not intrinsic; the definition depends on making a choice of $\mathcal{O}$. Instead we define $\overline{\rho}_\lambda$ to be the semisimplification of the reduction of ρ_λ modulo λ. The semisimplification of a representation is the direct sum of the Jordan-Hölder factors in a filtration of a vector space affording the representation. We have $\det(\overline{\rho}_\lambda) \equiv \chi_\ell \pmod{\ell}$, where χ_ℓ is the mod ℓ cyclotomic character, and $\operatorname{tr}(\overline{\rho}_\lambda(\mathrm{Frob}_p)) = a_p \mod \lambda$. Thus the characteristic polynomials in the semisimplification $\overline{\rho}_\lambda$ are independent of our choice of reduction of ρ_λ. The following theorem implies that $\overline{\rho}_\lambda$ depends only on f and λ, and not on the choice of the reduction.

Theorem 18.1.2 (Brauer-Nesbitt)**.** *Suppose $\rho_1, \rho_2 : G \to \mathrm{GL}(V)$ are two finite dimensional semisimple representations of a group G over a finite field k. Assume furthermore that for every $g \in G$ the characteristic polynomial of $\rho_1(g)$ is the same as the characteristic polynomial of $\rho_2(g)$. Then ρ_1 and ρ_2 are equivalent.*

Proof. For a proof, see [CR62, §30, p. 215]. □

The Hecke operators T_n act as endomorphisms of $S_2(\Gamma_0(N))$. Let

$$\mathbf{T} := \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N))),$$

and recall that $\mathbf{T}$ is a commutative ring; as a $\mathbf{Z}$-module $\mathbf{T}$ has rank equal to $\dim_{\mathbf{C}} S_2(\Gamma_0(N))$. Let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ and set $k := \mathbf{T}/\mathfrak{m} \approx \mathbf{F}_{\ell^\nu}$.

Proposition 18.1.3. *There is a unique semisimple representation*

$$\rho_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(k)$$

such that $\rho_{\mathfrak{m}}$ is unramified outside ℓN and

$$\begin{aligned} \operatorname{tr}(\rho_{\mathfrak{m}}(\mathrm{Frob}_p)) &= T_p \mod \mathfrak{m} \\ \det(\rho_{\mathfrak{m}}(\mathrm{Frob}_p)) &= p \mod \mathfrak{m}. \end{aligned}$$

Proof. It is enough to prove the assertion with $\mathbf{T}$ replaced by the subalgebra

$$\mathbf{T}_0 = \mathbf{Z}[\{\ldots, T_n, \ldots : (n, N) = 1\}].$$

Indeed, the maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ pulls back to a maximal ideal $\mathfrak{m}_0$ of $\mathbf{T}_0$, and $k_0 = \mathbf{T}_0/\mathfrak{m}_0 \subset k$. Now

$$\mathbf{T}_0 \subset \mathbf{T}_0 \otimes \mathbf{Q} = \prod_{i=1}^{t} E_i$$

with the E_i number fields. Let $\mathcal{O}_{E_i}$ be the ring of integers of E_i and let $\mathcal{O} = \prod \mathcal{O}_{E_i}$. By the going up theorem there is a maximal ideal $\lambda \subset \prod \mathcal{O}_{E_i}$ lying over $\mathfrak{m}_0$:

$$\begin{array}{ccccc} \lambda & \hookrightarrow & \prod \mathcal{O}_{E_i} & \twoheadrightarrow & \mathcal{O}/\lambda \\ | & & | & & | \\ \mathfrak{m}_0 & \hookrightarrow & \mathbf{T}_0 & \twoheadrightarrow & k_0 \end{array}$$

Using the above construction, we make a representation

$$\overline{\rho}_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O}/\lambda)$$

that satisfies $\det(\overline{\rho}_\lambda) \equiv \chi_\ell \pmod{\lambda}$ and $\mathrm{tr}(\overline{\rho}_\lambda) = T_p \pmod{\lambda}$. Because of how we have set things up, T_p plays the role of a_p. Thus this representation has the required properties, but it takes values in $\mathrm{GL}_2(\mathcal{O}/\lambda)$ instead of k_0.

Since the characteristic polynomial of every $\overline{\rho}_\lambda(g)$ for $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ has coefficients in the subfield $k_0 \subset \mathcal{O}/\lambda$ there is a model for ρ_λ over k_0. This is a classical result of I. Schur. Brauer groups of finite fields are trivial (see e.g., [Ser79, X.7, Ex. a]), so the argument of [Ser77, §12.2] completes the proof of the proposition.

Alternatively, when the residue characteristic ℓ of k_0 is odd, the following more direct proof can be used. Complex conjugation acts through $\overline{\rho}_\lambda$ as a matrix with distinct $\mathbf{F}_\ell$-rational eigenvalues; another well known theorem of Schur [Sch06, IX a] (cf. [Wal85, Lemme I.1]) then implies that $\overline{\rho}_\lambda$ can be conjugated into a representation with values in $\mathrm{GL}(2, k_0)$. □

Let us look at this construction in another way. Write

$$\mathbf{T}_0 \otimes \mathbf{Q} = E_1 \times \cdots \times E_t$$

and recall that each number field E_i corresponds to a newform of level $M \mid N$; one can obtain E_i by adjoining the coefficients of some newform of level $M \mid N$ to $\mathbf{Q}$. Likewise, the Jacobian $J_0(N)$ is isogenous to a product $A_1 \times \cdots \times A_t$. Consider one of the factors, say E_1, and to fix ideas suppose that it corresponds to a newform of exact level N. Since $\mathrm{Tate}_\ell A_1$ is free of rank 2 over $E_1 \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$, we obtain a 2-dimensional representation ρ_λ. Reducing mod λ and semisimplifying gives the representation constructed in the above proposition. But it is also possible that one of the fields E_i corresponds to a newform f of level M properly dividing N. In this case, we repeat the whole construction with $J_0(M)$ to get a 2-dimensional representation.

Consider one of the abelian varieties A_1 as above, which we view as an abelian subvariety of $J_0(N)$. Then $\mathbf{T}$ acts on A_1; let $\overline{\mathbf{T}}$ denote the image of $\mathbf{T}$ in $\mathrm{End}\, A_1$. Although $\overline{\mathbf{T}}$ sits naturally in $\mathcal{O}_1$, which is the ring of integers of a field, $\overline{\mathbf{T}}$ might not be integrally closed. Consider the usual $\mathbf{Z}_\ell$-adic Tate module $\mathrm{Tate}_\ell(A_1) \approx \mathbf{Z}_\ell^{2\dim A_1}$, where $\dim A_1 = [E_1 : \mathbf{Q}]$. In the 1940s Weil proved that $\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell$ acts faithfully on the Tate module. By the theory of semilocal rings (see, e.g., [Eis95, Cor. 7.6]), we have

$$\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell = \prod_{\mathfrak{m} \mid \ell} \overline{\mathbf{T}}_{\mathfrak{m}},$$

where the product is over all maximal ideals of $\overline{\mathbf{T}}$ of residue characteristic ℓ. The idempotents $e_{\mathfrak{m}}$, in this decomposition, decompose Tate_ℓ as a product $\prod_{\mathfrak{m}} \mathrm{Tate}_{\mathfrak{m}}(A_1)$. It would be nice if Tate_ℓ were free of rank 2 over $\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell$ but this is not known to be true in general, although it has been verified in many special cases. For this to be true we must have that, for all $\mathfrak{m} \mid \ell$, that $\mathrm{Tate}_{\mathfrak{m}}(A_1)$ is free of rank 2 over $\overline{\mathbf{T}}_{\mathfrak{m}}$.

Next we put things in a finite context instead of a projective limit context. Let $J = J_0(N)$, then by Albanese or Picard functoriality $\mathbf{T} \subset \mathrm{End}\, J$. Let $\mathfrak{m} \subset \mathbf{T}$ be a maximal ideal. Let

$$J[\mathfrak{m}] = \{t \in J(\overline{\mathbf{Q}}) : xt = 0 \text{ for all } x \in \mathfrak{m}\}.$$

Note that $J[\mathfrak{m}] \subset J[\ell]$ where ℓ is the rational prime lying in $\mathfrak{m}$. Now $J[\ell]$ is an $\mathbf{F}_\ell$ vector space of rank $2g$ where g is the genus of $X_0(N)$. Although it is true that $J[\ell]$ is a $\mathbf{T}/\ell$-module, it is not convenient to work with $\mathbf{T}/\ell$ since it might not

be a product of fields because of unpleasant ramification. It is more convenient to work with $J[\mathfrak{m}]$ since $\mathbf{T}/\mathfrak{m}$ is a field. Thus, via this optic, $J[\mathfrak{m}]$ is a $k[G]$-module where $k = \mathbf{T}/\mathfrak{m}$ and $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The naive hope is that $J[\mathfrak{m}]$ is a model for $\rho_{\mathfrak{m}}$, at least when $\rho_{\mathfrak{m}}$ is irreducible. This does not quite work, but we do have the following theorem.

Theorem 18.1.4. *If $\ell \nmid 2N$ then $J[\mathfrak{m}]$ is a model for $\rho_{\mathfrak{m}}$.*

18.2 The Gorenstein property

Consider the Hecke algebra

$$\mathbf{T} := \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}\, J_0(N),$$

and let $\mathfrak{m} \subset \mathbf{T}$ be a maximal ideal of residue characteristic ℓ. We have constructed a semisimple representation

$$\rho_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{T}/\mathfrak{m}).$$

It is unramified outside ℓN, and for any prime $p \nmid \ell N$ we have

$$\begin{aligned} \mathrm{tr}(\rho_{\mathfrak{m}}(\mathrm{Frob}_p)) &= T_p \mod \mathfrak{m} \\ \det(\rho_{\mathfrak{m}}(\mathrm{Frob}_p)) &= p \mod \mathfrak{m}. \end{aligned}$$

We will usually be interested in the case when $\rho_{\mathfrak{m}}$ is irreducible. Let $\mathbf{T}_{\mathfrak{m}} = \varprojlim \mathbf{T}/\mathfrak{m}^i\mathbf{T}$ denote the completion of $\mathbf{T}$ at $\mathfrak{m}$. Note that $\mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Z}_\ell = \prod_{\mathfrak{m}|\ell} \mathbf{T}_{\mathfrak{m}}$. Our goal is to prove that if $\mathfrak{m} \nmid 2N$ and $\rho_{\mathfrak{m}}$ is irreducible, then $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein.

Definition 18.2.1. Let $\mathcal{O}$ be a complete discrete valuation ring. Let T be a local $\mathcal{O}$-algebra which as a module is finite and free over $\mathcal{O}$. Then T is a *Gorenstein $\mathcal{O}$-algebra* if there is an isomorphism of T-modules $T \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(T, \mathcal{O})$.

Thus $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein if there is an isomorphism $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_{\mathfrak{m}}, \mathbf{Z}_\ell) \approx \mathbf{T}_{\mathfrak{m}}$ of $\mathbf{T}_{\mathfrak{m}}$-modules. Intuitively, this means that $\mathbf{T}_{\mathfrak{m}}$ is "autodual".

Theorem 18.2.2. *Let $J = J_0(N)$ and let $\mathfrak{m}$ be a maximal ideal of the Hecke algebra. Assume that $\rho_{\mathfrak{m}}$ is irreducible and that $\mathfrak{m} \nmid 2N$. Then $\dim_{\mathbf{T}/\mathfrak{m}} J[\mathfrak{m}] = 2$ and $J[\mathfrak{m}]$, as a Galois module, is a 2-dimensional representation giving rise to $\rho_{\mathfrak{m}}$.*

An easy argument shows that the theorem implies $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein.

We first consider the structure of $W = J[\mathfrak{m}]$. Suppose the two dimensional representation

$$\rho_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{T}/\mathfrak{m}} V$$

constructed before is irreducible. Consider the semisimplification $W^{\mathrm{ß}}$ of W, thus $W^{\mathrm{ß}}$ is the direct sum of its Jordan-Hölder factors as a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. Mazur proved the following theorem.

Theorem 18.2.3. *There is some integer $t \geq 0$ so that*

$$W^{\mathrm{ß}} \cong V \times \cdots \times V = V^t.$$

If $\rho_{\mathfrak{m}}$ is in fact *absolutely irreducible* then it is a result of Boston, Lenstra, and Ribet [BpR91] that $W \cong V \times \cdots \times V$. A representation is absolutely irreducible if it is irreducible over the algebraic closure. It can be shown that if $\ell \neq 2$ and $\rho_{\mathfrak{m}}$ is irreducible then $\rho_{\mathfrak{m}}$ must be absolutely irreducible.

The construction of W is nice and gentle whereas the construction of V is accomplished via brute force.

Proof. (Mazur) We want to compare V with W. Let $d = \dim W$. Let

$$W^* = \mathrm{Hom}_{\mathbf{T}/\mathfrak{m}}(W, \mathbf{T}/\mathfrak{m}(1))$$

where $\mathbf{T}/\mathfrak{m}(1) = \mathbf{T}/\mathfrak{m} \otimes_{\mathbf{Z}} \mu_\ell$. We need to show that

$$W^{ß} \oplus W^{*ß} \cong V \times \cdots \times V = V^d$$

as representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Note that each side is a semisimple module of dimension $2d$. To obtain the isomorphism we show that the two representations have the same characteristic polynomials so they are isomorphic.

We want to show that the characteristic polynomial of Frob_p is the same for both $W^{ß} \oplus W^{ß}$ and $V \times \cdots \times V$. The characteristic polynomial of Frob_p on V is $X^2 - T_p X + p = (X - r)(X - pr^{-1})$ where r lies in a suitable algebraic closure. It follows that the characteristic polynomial of Frob_p on V^d is $(x - r)^d (x - pr^{-1})^d$. On W the characteristic polynomial of Frob_p is $(X - \alpha_1) \cdots (X - \alpha_d)$ where α_i is either r or pr^{-1}. This is because Eichler-Shimura implies Frob_p must satisfy $\mathrm{Frob}_p^2 - T_p \,\mathrm{Frob}_p + p = 0$. [[I don't see this implication.]] On W^* the characteristic polynomial of Frob_p is $(X - p\alpha_1^{-1}) \cdots (X - p\alpha_d^{-1})$. [[This is somehow tied up with the definition of W^* and I can't quite understand it.]] Thus on $W \oplus W^*$, the characteristic polynomial of Frob_p is

$$\prod_{i=1}^{d} (X - \alpha_i)(X - p\alpha_i^{-1}) = \prod_{i=1}^{d} (X - r)(X - pr^{-1}) = (X - r)^d (X - pr^{-1})^d.$$

Therefore the characteristic polynomial of Frob_p on $W \oplus W^*$ is the same as the characteristic polynomial of Frob_p on $V \times \cdots \times V$. The point is that although the α_i could all *a priori* be r or pr^{-1}, by adding in W^* everything pairs off correctly. [[I don't understand why we only have to check that the two representations agree on Frob_p. There are lots of other elements in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, right?]] □

We next show that $J[\mathfrak{m}] \neq 0$. This does not follow from the theorem proved above because it does not rule out the possibility that $t = 0$ and hence $W^{ß} \cong 0$. Suppose $J[\mathfrak{m}] = 0$, then we will show that $J[\mathfrak{m}^i] = 0$ for all $i \geq 1$. We consider the ℓ-divisible group $J_{\mathfrak{m}} = \cup_i J[\mathfrak{m}^i]$. To get a better feel for what is going on, temporarily forget about $\mathfrak{m}$ and just consider the Tate module corresponding to ℓ.

It is standard to consider the Tate module

$$\mathrm{Tate}_\ell J = \varprojlim J[\ell^i] \cong \mathbf{Z}_\ell^{2 \dim J}.$$

It is completely equivalent to consider

$$J_\ell := \cup_{i=1}^{\infty} J[\ell^i] \xrightarrow{\sim} (\mathbf{Q}_\ell/\mathbf{Z}_\ell)^{2 \dim J}.$$

Note that since $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ is not a ring the last isomorphism must be viewed as an isomorphism of abelian groups. In [Maz77] Mazur called $\mathrm{Tate}_\ell J \cong \mathrm{Hom}(\mathbf{Q}_\ell/\mathbf{Z}_\ell, J_\ell)$ the covariant Tate module. Call

$$\mathrm{Tate}_\ell^* J := \mathrm{Hom}(J_\ell, \mathbf{Q}_\ell/\mathbf{Z}_\ell) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell)$$

the contravariant Tate module. [[Why are the last two isomorphic?]] The covariant and contravariant Tate modules are related by a Weil pairing $J[\ell^i] \times J[\ell^i] \to \mu_{\ell^i}$. Taking projective limits we obtain a pairing

$$\langle \cdot, \cdot \rangle : \mathrm{Tate}_\ell J \times \mathrm{Tate}_\ell J \to \mathbf{Z}_\ell(1) = \varprojlim \mu_{\ell^i}.$$

This gives a map

$$\mathrm{Tate}_\ell J \to \mathrm{Hom}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell(1)) = (\mathrm{Tate}_\ell^* J)(1)$$

where $(\mathrm{Tate}_\ell^* J)(1) = (\mathrm{Tate}_\ell^* J) \otimes \mathbf{Z}_\ell(1)$. $\mathbf{Z}_\ell(1)$ is a $\mathbf{Z}_\ell$-module where

$$\sum a_i \ell^i \cdot \zeta = \zeta^{\sum a_i \ell^i}$$

[[This should probably be said long ago.]] This pairing is not a pairing of $\mathbf{T}$-modules, since if $t \in \mathbf{T}$ then $\langle tx, y \rangle = \langle x, t^\vee y \rangle$. It is more convenient to use an adapted pairing defined as follows. Let $w = w_V \in \mathrm{End}\, J_0(N)$ be the Atkin-Lehner involution so that $t^\vee = wtw$. Define a new $\mathbf{T}$-compatible pairing by $[x, y] := \langle x, wy \rangle$. Then

$$[tx, y] = \langle tx, wy \rangle = \langle x, t^\vee wy \rangle = \langle x, wty \rangle = [x, ty].$$

The pairing $[\cdot, \cdot]$ defines an isomorphism of $\mathbf{T} \otimes \mathbf{Z}_\ell$-modules

$$\mathrm{Tate}_\ell J \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell(1)).$$

Since $\mathbf{Z}_\ell(1)$ is a free module of rank 1 over $\mathbf{Z}_\ell$ a suitable choice of basis gives an isomorphism of $\mathbf{T} \otimes \mathbf{Z}_\ell$-modules

$$(\mathrm{Tate}_\ell^* J)(1) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell(1)) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell) = \mathrm{Tate}_\ell^* J.$$

Thus $\mathrm{Tate}_\ell J \xrightarrow{\sim} \mathrm{Tate}_\ell^* J$.

Proof. (That $J[\mathfrak{m}] \neq 0$.) The point is that the contravariant Tate module $\mathrm{Hom}(J_\ell, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$ is the Pontrjagin dual of T_ℓ. How does this relate to $\mathrm{Tate}_\mathfrak{m} J$? Since $\mathbf{T} \otimes \mathbf{Z}_\ell \cong \prod_{\mathfrak{m}|\ell} \mathbf{T}_\mathfrak{m}$, $\mathrm{Tate}_\ell J \cong \prod_{\mathfrak{m}|\ell} \mathrm{Tate}_\mathfrak{m} J$ so we can define $\mathrm{Tate}_\mathfrak{m}^* J := \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\mathfrak{m} J, \mathbf{Z}_\ell)$. Weil proved that $\mathrm{Tate}_\mathfrak{m} J \cong Tate_\mathfrak{m}^* J$ is nonzero. View $\mathrm{Tate}_\mathfrak{m}^* J$ as being dual to $J_\mathfrak{m}$ in the sense of Pontrjagin duality and so $(\mathrm{Tate}_\mathfrak{m}^* J)/(\mathfrak{m}\,\mathrm{Tate}_\mathfrak{m}^* J)$ is dual to $J[\mathfrak{m}]$. If $J[\mathfrak{m}] = 0$ then this quotient is 0, so Nakayama's lemma would imply that $\mathrm{Tate}_\mathfrak{m}^* J = 0$. This would contradicts Weil's assertion. Therefore $J[\mathfrak{m}] \neq 0$. □

18.3 Proof of the Gorenstein property

We are considering the situation with respect to $J_0(N)$ although we could consider $J_1(N)$. Let $\mathbf{T} \subset \mathrm{End}\, J_0(N)$ be the Hecke algebra and let $\mathfrak{m} \subset \mathbf{T}$ be a maximal ideal.

Let ℓ be the characteristic of the residue class field $\mathbf{T}/\mathfrak{m}$. Let $\mathbf{T}_{\mathfrak{m}} = \varprojlim \mathbf{T}/\mathfrak{m}^i\mathbf{T}$. Then $\mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Z}_\ell = \prod_{\mathfrak{m}|\ell} \mathbf{T}_{\mathfrak{m}}$. [[I want to put a good reference for this Atiyah-Macdonald like fact here.]] Each $\mathbf{T}_{\mathfrak{m}}$ acts on $\mathrm{Tate}_\ell J_0(N)$ so we obtain a product decomposition

$$\mathrm{Tate}_\ell J_0(N) = \prod_{\mathfrak{m}|\ell} \mathrm{Tate}_{\mathfrak{m}} J_0(N).$$

We have the following two facts:

1. $\mathrm{Tate}_{\mathfrak{m}} J_0(N) \neq 0$

2. $\mathrm{Tate}_\ell J_0(N)$ is $\mathbf{T} \otimes \mathbf{Z}_\ell$-autodual and each $\mathrm{Tate}_{\mathfrak{m}} J_0(N)$ is $\mathbf{Z}_\ell$-autodual.

[[autoduality for which dual? I think it is the linear dual since this is used.]]

Let $W = J[\mathfrak{m}]$, then the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on W gives a representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over the field $\mathbf{T}/\mathfrak{m}$. We compared W with a certain two dimensional representation $\rho_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to V$ over $\mathbf{T}/\mathfrak{m}$. Assume unless otherwise stated that V is irreducible as a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. Let $\mathrm{Tate}_\ell = \mathrm{Tate}_\ell J_0(N)$ and $\mathrm{Tate}_{\mathfrak{m}} = \mathrm{Tate}_{\mathfrak{m}} J_0(N)$. A formal argument due to Mazur showed that

$$W^{\text{ß}} \cong V \times \cdots \times V = V^{\oplus t}.$$

We have not yet determined t but we would like to show that $t = 1$.

Definition 18.3.1. The **Pontrjagin dual** of a module M is the module $M^\wedge := \mathrm{Hom}(M, \mathbf{Q}/\mathbf{Z})$ where M is viewed as an abelian group (if M is topological, only take those homomorphisms whose kernel is compact). The **linear dual** of a module M over a ring R is the module $M^* = \mathrm{Hom}_R(M, R)$.

Exercise 18.3.2. Note that $(\mathbf{Q}_\ell/\mathbf{Z}_\ell)^\wedge = \mathbf{Z}_\ell$ and $\mathbf{Z}_\ell^\wedge = \mathbf{Q}_\ell/\mathbf{Z}_\ell$.

Solution.. We can think of $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ as

$$\{ \sum_{n=-k}^{-1} a_n \ell^n : k > 0 \text{ and } 0 \leq a_i < \ell \}.$$

Let $(b_i) \in \mathbf{Z}_\ell$ so $b_i \in \mathbf{Z}/\ell^i\mathbf{Z}$ and $b_{i+1} \equiv b_i \pmod{\ell^i}$. Define a map $\mathbf{Q}_\ell/\mathbf{Z}_\ell \to \mathbf{Q}/\mathbf{Z}$ by $1/\ell^i \mapsto b_i/\ell^i$. To check that this is well-defined it suffices to check that $1/\ell^i$ maps to the same place as $\ell \cdot 1/\ell^{i+1}$. Now $1/\ell^i \mapsto b_i/\ell^i$ and

$$\ell \cdot 1/\ell^{i+1} \mapsto \ell \cdot b_{i+1}/\ell^{i+1} = b_{i+1}/\ell^i.$$

So we just need to check that

$$b_{i+1}/\ell^i \equiv b_i/\ell^i \pmod{\mathbf{Z}}.$$

This is just the assertion that $(b_{i+1} - b_i)/\ell^i \in \mathbf{Z}$ which is true since $b_{i+1} \equiv b_i \pmod{\ell^i}$. □

Proposition 18.3.3. *Let the notation be as above, then $t > 0$.*

Proof. The idea is to use Nakayama's lemma to show that if $t = 0$ and hence $W = 0$ then $\mathrm{Tate}_{\mathfrak{m}} = 0$ which is clearly false. But the relation between W and $\mathrm{Tate}_{\mathfrak{m}}$ is rather convoluted. In fact $J[\ell^\infty]$ is the Pontrjagin dual of Tate_ℓ^*, that is,

$$J[\ell^\infty]^\wedge = \mathrm{Tate}_\ell^* = \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell, \mathbf{Z}_\ell)$$

and

$$(\mathrm{Tate}_\ell^*)^\wedge = \mathrm{Hom}(\mathrm{Tate}_\ell^*, \mathbf{Q}_\ell/\mathbf{Z}_\ell) = J[\ell^\infty].$$

[[First: Why are they dual? Second: Why are we homing into $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ instead of $\mathbf{Q}/\mathbf{Z}$?]] Looking at the $\mathfrak{m}$-adic part shows that

$$J[\mathfrak{m}^\infty] = \mathrm{Hom}(\mathrm{Tate}_{\mathfrak{m}}^*, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$$

and hence

$$J[\mathfrak{m}] = \mathrm{Hom}(\mathrm{Tate}_{\mathfrak{m}}^* /\mathfrak{m}\,\mathrm{Tate}_{\mathfrak{m}}^*, \mathbf{Z}/\ell\mathbf{Z}).$$

Thus if $J[\mathfrak{m}] = 0$ then Nakayama's lemma implies $\mathrm{Tate}_{\mathfrak{m}}^* = 0$. By autoduality this implies $\mathrm{Tate}_{\mathfrak{m}} = 0$. □

We have two goals. The first is to show that $t = 1$, i.e., that $J[\mathfrak{m}]$ is 2-dimensional over $\mathbf{T}/\mathfrak{m}$. The second is to prove that $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein, i.e., that $\mathbf{T}_{\mathfrak{m}} \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_{\mathfrak{m}}, \mathbf{Z}_\ell)$. This is one of the main theorems in the subject. We are assuming throughout that $\rho_{\mathfrak{m}}$ is irreducible and $\ell \nmid 2N$. Loosely speaking the condition that $\ell \not| 2N$ means that $J[\mathfrak{m}]$ has good reduction at ℓ and that $J[\mathfrak{m}]$ can be understood just by understanding $J[\mathfrak{m}]$ in characteristic ℓ. We want to prove that $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein because this property plays an essential role in proving that $\mathbf{T}_{\mathfrak{m}}$ is a local complete intersection.

Example 18.3.4. Let

$$T = \{(a, b, c, d) \in \mathbf{Z}_p^4 : a \equiv b \equiv c \equiv d \pmod{p}\}.$$

Then T is a local ring that is not Gorenstein.

For now we temporarily postpone the proof of the first goal and instead show that the first goal implies the second.

Theorem 18.3.5. *Suppose $J[\mathfrak{m}]$ is two dimensional over $\mathbf{T}/\mathfrak{m}$ (thus $t = 1$). Then $\mathbf{T}_{\mathfrak{m}}$ is Gorenstein.*

Proof. We have seen before that

$$\begin{aligned} J[\mathfrak{m}] &= \mathrm{Hom}_{\mathbf{Z}/\ell\mathbf{Z}}(\mathrm{Tate}_{\mathfrak{m}}^* /\mathfrak{m}\,\mathrm{Tate}_{\mathfrak{m}}^*, \mathbf{Z}/\ell\mathbf{Z}) \\ &= \mathrm{Hom}_{\mathbf{T}/\mathfrak{m}}(\mathrm{Tate}_{\mathfrak{m}}^* /\mathfrak{m}\,\mathrm{Tate}_{\mathfrak{m}}^*, \mathbf{T}/\mathfrak{m}). \end{aligned}$$

Thus the dual of $\mathrm{Tate}_{\mathfrak{m}}^* /\mathfrak{m}\,\mathrm{Tate}_{\mathfrak{m}}^*$ is two dimensional over $\mathbf{T}/\mathfrak{m}$ and hence $\mathrm{Tate}_{\mathfrak{m}}^* /\mathfrak{m}\,\mathrm{Tate}_{\mathfrak{m}}^*$ itself is two dimensional over $\mathbf{T}/\mathfrak{m}$. By Nakayama's lemma and autoduality of $\mathrm{Tate}_{\mathfrak{m}}$ this implies $\mathrm{Tate}_{\mathfrak{m}}$ is generated by 2 elements over $\mathbf{T}_{\mathfrak{m}}$. There is a surjection

$$\mathbf{T}_{\mathfrak{m}} \times \mathbf{T}_{\mathfrak{m}} \twoheadrightarrow \mathrm{Tate}_{\mathfrak{m}}\,.$$

In fact it is true that $\mathrm{rank}_{\mathbf{Z}_\ell} \mathrm{Tate}_{\mathfrak{m}} = 2\,\mathrm{rank}_{\mathbf{Z}_\ell} \mathbf{T}_{\mathfrak{m}}$. We temporarily postpone the proof of this claim. Assuming this claim and using that a surjection between $\mathbf{Z}_\ell$-modules of the same rank is an isomorphism implies that $\mathrm{Tate}_{\mathfrak{m}} \cong \mathbf{T}_{\mathfrak{m}} \times \mathbf{T}_{\mathfrak{m}}$. Now $\mathbf{T}_{\mathfrak{m}}$ is a direct summand of the free $\mathbf{Z}_\ell$ module $\mathrm{Tate}_{\mathfrak{m}}$ so $\mathbf{T}_{\mathfrak{m}}$ is projective. A

projective module over a local ring is free. Thus $\mathbf{T}_\mathfrak{m}$ is free of rank 1 and hence autodual (Gorenstein). [[This argument is an alternative to Mazur's – it seems too easy... maybe I am missing something.]]

We return to the claim that

$$\operatorname{rank}_{\mathbf{Z}_\ell} \operatorname{Tate}_\mathfrak{m} = 2 \operatorname{rank}_{\mathbf{Z}_\ell} \mathbf{T}_\mathfrak{m}.$$

This is equivalent to the assertion that

$$\dim_{\mathbf{Q}_\ell} \operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell = 2 \dim_{\mathbf{Q}_\ell} \mathbf{T}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell.$$

The module $\operatorname{Tate}_\ell J_0(N)$ is the projective limit of the ℓ-power torsion on the Jacobian

$$J(\mathbf{C}) = \frac{\operatorname{Hom}_\mathbf{C}(S_2(\Gamma_0(N), \mathbf{C}), \mathbf{C})}{H_1(X_0(N), \mathbf{Z})}.$$

Let $L = H_1(X_0(N), \mathbf{Z})$ be the lattice. Then L is a $\mathbf{T}$-module and $\operatorname{Tate}_\ell = L \otimes_\mathbf{Z} \mathbf{Z}_\ell$ (since $L/\ell^i L \cong (\frac{1}{\ell^i} L)/L$). Tensoring with $\mathbf{R}$ gives

$$\begin{aligned} L \otimes_\mathbf{Z} \mathbf{R} &= \operatorname{Hom}_\mathbf{C}(S_2(\Gamma_0(N), \mathbf{C}), \mathbf{C}) \\ &= \operatorname{Hom}_\mathbf{R}(S_2(\Gamma_0(N), \mathbf{R}), \mathbf{C}) \\ &= \operatorname{Hom}_\mathbf{R}(S_2(\Gamma_0(N), \mathbf{R}), \mathbf{R}) \otimes_\mathbf{R} \mathbf{C} = (\mathbf{T} \otimes_\mathbf{Z} \mathbf{R}) \otimes_\mathbf{R} \mathbf{C} \end{aligned}$$

Thus $L \otimes_\mathbf{Z} \mathbf{R}$ is free of rank 2 over $\mathbf{T} \otimes_\mathbf{Z} \mathbf{R}$ and $L \otimes_\mathbf{Z} \mathbf{C}$ is free of rank 2 over $\mathbf{T} \otimes_\mathbf{Z} \mathbf{C}$. Next choose an embedding $\mathbf{Q}_\ell \hookrightarrow \mathbf{C}$. Now Tate_ℓ is a module over $\mathbf{T} \otimes \mathbf{Z}_\ell = \prod_{\mathfrak{m}|\ell} \mathbf{T}_\mathfrak{m}$ so we have a decomposition $\operatorname{Tate}_\ell = \prod_{\mathfrak{m}|\ell} \operatorname{Tate}_\mathfrak{m}$. Since

$$\operatorname{Tate}_\ell \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell = \prod_\mathfrak{m} \operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$$

we can tensor with $\mathbf{C}$ to see that

$$\operatorname{Tate}_\ell \otimes_{\mathbf{Z}_\ell} \mathbf{C} = \prod_\mathfrak{m} \operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C}.$$

But $\operatorname{Tate}_\ell \otimes_{\mathbf{Z}_\ell} \mathbf{C} = L \otimes_\mathbf{Z} \mathbf{C}$ is free of rank 2 over $\mathbf{T} \otimes \mathbf{C}$. Therefore the product $\prod_\mathfrak{m} \operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C}$ is free of rank 2 over $\mathbf{T} \otimes \mathbf{C}$. Since

$$\mathbf{T} \otimes \mathbf{C} = (\mathbf{T} \otimes_\mathbf{Z} \mathbf{Z}_\ell) \otimes_{\mathbf{Z}_\ell} \mathbf{C} = \prod_\mathfrak{m} (\mathbf{T}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C})$$

we conclude that for each $\mathfrak{m}$, $\operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C}$ is free of rank 2 over $\mathbf{T}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C}$. This implies

$$\dim_\mathbf{C} \operatorname{Tate}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C} = 2 \dim_\mathbf{C} \mathbf{T}_\mathfrak{m} \otimes_{\mathbf{Z}_\ell} \mathbf{C}$$

which completes the proof. □

18.3.1 Vague comments

Ogus commented that this same proof shows that $\mathbf{T} \otimes_\mathbf{Z} \mathbf{C}$ is Gorenstein. Then he said that something called "faithfully flat descent" could then show that $\mathbf{T} \otimes_\mathbf{Z} \mathbf{Q}$ is Gorenstein.

We have given the classical argument of Mazur that $\mathbf{T}_\mathfrak{m}$ is Gorenstein, but we still haven't shown that $J[\mathfrak{m}]$ has dimension 2. This will be accomplished next time using Dieudonné modules.

18.4 Finite flat group schemes

Definition 18.4.1. Let S be a scheme. Then a **group scheme over** S is a group object in the category of S-schemes.

Thus a group scheme over S is a scheme G/S equipped with S-morphisms $m : G \times G \to G$, $\mathrm{inv} : G \to G$ and a section $1_G : S \to G$ satisfying the usual group axioms.

Suppose G is a group scheme over the finite field $\mathbf{F}_q$. If R is an $\mathbf{F}_q$-algebra then $G(R) = \mathrm{Mor}(\mathrm{Spec}\, R, G)$ is a group. It is the group of R-valued points of G.

We consider several standard examples of group schemes.

Example 18.4.2. The multiplicative group scheme G_m is $G_m = \mathrm{Spec}\, \mathbf{Z}[x, \frac{1}{x}]$ with morphisms. [[give maps, etc.]] The additive group scheme is $\mathrm{Spec}\, \mathbf{Z}[x]$...

Example 18.4.3. The group scheme μ_p is the kernel of the morphism $G_m \to G_m$ induced by $x \mapsto x^p$. Thus $\mu_p = \mathrm{Spec}\, \mathbf{Z}[x]/(x^p - 1)$ and so for any $\mathbf{F}_q$-algebra R we have that $\mu_p(R) = \{r \in R : r^p = 1\}$. The group scheme α_p is the kernel of the morphism $G_a \to G_a$ induced by [[what!! what is alphap?? it should be the additive group scheme of order p, no?]].

Let A be a finite algebra over $\mathbf{F}_p$ and suppose $G = \mathrm{Spec}\, A$ affine group scheme (over $\mathbf{F}_p$). Then the **order** of G is defined to be the dimension of A as an $\mathbf{F}_p$ vector space.

Example 18.4.4. Let $E/\mathbf{F}_p$ be an elliptic curve. Then $G = E[p]$ is a group scheme of order p^2 [[why is this true?]]. This is wonderful because this is the order that $E[p]$ should have in analogy with the characteristic 0 situation. When we just look at points we have

$$\#G(\overline{\mathbf{F}_p}) = \begin{cases} 1 & \text{supersingular} \\ p & \text{ordinary} \end{cases}.$$

18.5 Reformulation of $V = W$ problem

Let $J = J_0(N)$ be the Jacobian of $X_0(N)$. Then J is defined over $\mathbf{Q}$ and has good reduction at all primes not dividing N. Assume ℓ is a prime not dividing N. $J[\ell]$ extends to a finite flat group scheme over $\mathbf{Z}[\frac{1}{N}]$. This is a nontrivial result of Grothendieck (SGA 7I, LNM 288). Since $\ell \nmid N$, $J[\ell]$ gives rise to a group scheme over $\mathbf{F}_\ell$.

We have "forcefully" constructed a Galois representation $\rho_{\mathfrak{m}} : G \to V$ of dimension 2 over $\mathbf{T}/\mathfrak{m}$. Our goal is to show that this is isomorphic to the naturally defined Galois representation $W = J[\mathfrak{m}]$. So far we know that

$$0 \subset V \subset W \subset J[\ell].$$

Our assumptions are that $\ell \nmid N$, V is irreducible, and $\ell \neq 2$.

Let $\underline{J}$ be J thought of as a scheme over $\mathbf{Z}_\ell$. Grothendieck showed that $\underline{J}$ is the spectrum of a free finite $\mathbf{Z}_\ell$-module. Raynaud (1974) showed that if $\ell \neq 2$ then essentially everything about $\underline{J}[\ell]$ can be seen in terms of $J[\ell](\overline{Q}_\ell)$. He goes on to construct group scheme $\underline{V}$ and $\underline{W}$ over $\mathbf{Z}_\ell$ such that

$$\underline{V} \subset \underline{W} \subset \underline{J}[\ell].$$

Our goal is to prove that the inclusion $V \hookrightarrow W$ of Galois modules is an isomorphism. Raynaud showed noted that the category of finite flat group schemes over $\mathbf{Z}_\ell$ is an abelian category so the cokernel $\underline{Q} = \underline{W}/\underline{V}$ is defined. Furthermore, $V = W$ if and only if $\underline{Q} = 0$. Since $\underline{Q}$ is flat $\underline{Q}_{\mathbf{F}_\ell}$ has the same dimension over $\mathbf{F}_\ell$ as $\underline{Q}_{\mathbf{Q}_\ell}$ has over $\mathbf{Q}_\ell$. Passing to characteristic ℓ yields an exact sequence

$$0 \to \underline{V}_{\mathbf{F}_\ell} \to \underline{W}_{\mathbf{F}_\ell} \to \underline{Q}_{\mathbf{F}_\ell} \to 0.$$

Thus $V \hookrightarrow W$ is an isomorphism if and only if $\underline{V}_{\mathbf{F}_\ell} \hookrightarrow \underline{W}_{\mathbf{F}_\ell}$ is an isomorphism. Since $\underline{V}$, $\underline{W}$, and $\underline{Q}$ have an action of $k = \mathbf{T}/\mathfrak{m}$ that are k-vector space schemes. This leads us to Dieudonné theory.

18.6 Dieudonné theory

Let G/k be a finite k-vector space scheme where k is a finite field of order q. Suppose G has order q^n so G is locally the spectrum of a rank n algebra over k. The Dieudonné functor contravariantly associates to G an n dimensional k-vector space $D(G)$. Let Frob : $G \to G$ be the morphism induced by the pth power map on the underlying rings and let Ver be the dual of Frob. Let $\varphi = D(\text{Frob})$ and $\nu = D(\text{Ver})$, then it is a property of the functor D that $\varphi \circ \nu = \nu \circ \varphi = 0$. The functor D is a fully faithful functor.

Example 18.6.1. Let $k = \mathbf{F}_p$. If G is either μ_p, α_p, or $\mathbf{Z}/p\mathbf{Z}$ then $D(G)$ is a one-dimensional vector space over k. In the case of α_p, $\varphi = \nu = 0$. For μ_p, $\varphi = 0$ and $\nu \neq 1$ and for $\mathbf{Z}/p\mathbf{Z}$, $\varphi \neq 1$ and $\nu = 0$. [[The latter two could be reversed!]]

Let $G^\vee = \text{Hom}(G, \mu_p)$ denote the Cartier dual of the scheme G. Then

$$D(G^\vee) = \text{Hom}_k(D(G), k)$$

(φ and ν are switched.)

Example 18.6.2. Let $A/\mathbf{F}_\ell$ be an abelian variety and let $G = A[\ell]$. Then G is an $\mathbf{F}_\ell$-vector space scheme of order ℓ^{2g}. Thus $D(G)$ is a $2g$-dimensional $\mathbf{F}_\ell$-vector space and furthermore $D(G) = H^1_{DR}(A/\mathbf{F}_\ell)$. The Hodge filtration on H^1_{DR} of the abelian variety A gives rise to a diagram

$$\begin{array}{ccccccc}
 & & & & \text{Hom}(H^0(A^\vee, \Omega^1), \mathbf{F}_\ell) & = & \text{Tan}(A^\vee) \\
 & & & & & & \| \\
H^0(A, \Omega^1) & \hookrightarrow & D(G) & \longrightarrow & H^1_{DR}(A/\mathbf{F}_\ell) & \twoheadrightarrow & H^1(A, \mathcal{O}) \\
 & & & & \| & & \| \\
 & & & & D(A[\ell]) & \twoheadrightarrow & D(A[\text{Ver}])
\end{array}$$

There is an exact sequence

$$0 \to W_{\mathbf{F}_\ell} \to J_{\mathbf{F}_\ell}[\ell] \xrightarrow{\mathfrak{m}} J_{\mathbf{F}_\ell}[\ell]$$

so because D is an exact functor the sequence

$$D(J_{\mathbf{F}_\ell}[\ell]) \xrightarrow{\mathfrak{m}} D(J_{\mathbf{F}_\ell}[\ell]) \to D(W_{\mathbf{F}_\ell}) \to 0$$

is exact. Following Fontaine we consider

$$D(W_{\mathbf{F}_\ell}[\mathrm{Ver}]) = H^1(J, \mathcal{O})/\mathfrak{m}H^1(J, \mathcal{O}).$$

18.7 The proof: part II

[[We all just returned from the Washington D.C. conference and will now resume the proof.]]

Let $J_0(N)$ be the Jacobian of $X_0(N)$. Let $\mathfrak{m} \subset \mathbf{T}$ be a maximal ideal and suppose $\mathfrak{m}|\ell$. Assume that $\ell \neq 2$ and $\ell \nmid N$. The assumption that $\ell \neq 2$ is necessary for Raynaud's theory and we assume that $\ell \nmid N$ so that our group schemes will have good reduction. We attach to $\mathfrak{m}$ a 2 dimensional semisimple representation $\rho_\mathfrak{m} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to V$ and only consider the case that $\rho_\mathfrak{m}$ is irreducible.

The $\mathfrak{m}$-torsion of the Jacobian, $W = J_0(\overline{\mathbf{Q}})[\mathfrak{m}]$, is naturally a Galois module. We have shown that $W \neq 0$. By [BpR91] $W \cong V \times \cdots \times V$ (the number of fractions is not determined). We proved that $W^{s.s.} \cong V \times \cdots \times V$. Choose an inclusion $V \hookrightarrow W$ and let $Q = W/V$ be the cokernel.

Theorem 18.7.1. *$Q = 0$ so $\dim_{\mathbf{T}/\mathfrak{m}} W = 2$*

To prove the theorem we introduce the "machine" of finite flat group schemes over $\mathbf{Z}_\ell$. For example, W extends to a finite flat group scheme $W_{\mathbf{Z}_\ell}$ which is defined to be the Zariski closure of W in $J_{\mathbf{Z}_\ell}[\ell]$. Passing to group schemes yields an exact sequence

$$0 \to V_{\mathbf{Z}_\ell} \to W_{\mathbf{Z}_\ell} \to Q_{\mathbf{Z}_\ell} \to 0.$$

Reducing mod ℓ then yields an exact sequence of $\mathbf{F}_\ell$-group schemes

$$0 \to V_{\mathbf{F}_\ell} \to W_{\mathbf{F}_\ell} \to Q_{\mathbf{F}_\ell} \to 0.$$

The point is that $Q = 0$ if and only if $Q_{\mathbf{Z}_\ell} = 0$ if and only if $Q_{\mathbf{F}_\ell} = 0$.

Next we introduced the exact contravariant Dieudonné functor

$$D : (\text{ Groups Schemes } /\mathbf{F}_\ell\ ) \longrightarrow (\text{ Linear Algebra }).$$

D sends a group scheme G to a $\mathbf{T}/\mathfrak{m}$ vector space equipped with two endomorphisms $\varphi = \mathrm{Frob}$ and $\nu = \mathrm{Ver}$. Applying D gives an exact sequence of $\mathbf{T}/\mathfrak{m}$-vector spaces

$$0 \to D(Q) \to D(W) \to D(V) \to 0$$

where everything is now viewed over $\mathbf{F}_\ell$.

Lemma 18.7.2. *$D(W[\mathrm{Ver}]) = (H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathfrak{m}])^*$*

Proof. We have the diagram

$$\begin{array}{ccc} D(J_{\mathbf{F}_\ell}[\ell]) & = & H^1_{DR}(J_{\mathbf{F}_\ell}) \\ \downarrow & & \downarrow \\ D(J_{\mathbf{F}_\ell}[\mathrm{Ver}]) & = & H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_{J_{\mathbf{F}_\ell}}) \\ \downarrow & & \downarrow \\ D(W[\mathrm{Ver}]) & = & H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J)/\mathfrak{m}H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J) \end{array}$$

Furthermore we have the identifications

$$\begin{aligned} H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J) &= \mathrm{Tan}(J^\vee_{\mathbf{F}_\ell}) = \mathrm{Cot}(J^\vee_{\mathbf{F}_\ell})^* \\ &= H^0(J^\vee, \Omega^1)^* = H^0(X_0(N), \Omega^1)^* \end{aligned}$$

For the last identification we must have $J^\vee = \mathrm{Alb}(X_0(N))$. Finally

$$\begin{aligned} D(W[\mathrm{Ver}]) &= H^1(J, \mathcal{O}_J)/\mathfrak{m}H^1(J, \mathcal{O}_J) \\ &= (H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathfrak{m}])^*. \end{aligned}$$

□

Lemma 18.7.3. *$H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathfrak{m}]$ has $\mathbf{T}/\mathfrak{m}$ dimension ≤ 1.*

Proof. Let $S = H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathfrak{m}]$. Then $S \hookrightarrow \mathbf{F}_\ell[[q]]$. We defined the Hecke operators T_n on S via the identification $S \cong H^1(J, \mathcal{O}_J)$ so that they act on $S \subset \mathbf{F}_\ell[[q]]$ in the standard way. Let $\mathbf{T}(S)$ be the subalgebra of $\mathrm{End}(S)$ generated by the images of the T_n in $\mathrm{End}(S)$. ($\mathbf{T}(S)$ is not a subring of $\mathbf{T}$.) There is a perfect pairing

$$\begin{aligned} \mathbf{T}(S) \times S &\longrightarrow \mathbf{F}_\ell \\ (T, f) &\mapsto a_1(f|T) \end{aligned}$$

Thus $\dim_{\mathbf{F}_\ell} \mathbf{T}(S) = \dim_{\mathbf{F}_\ell} S$ and so $\dim_{\mathbf{T}(S)} S \leq 1$. Since $\mathfrak{m}$ acts trivially on S there is a surjection $\mathbf{T}/\mathfrak{m} \twoheadrightarrow \mathbf{T}(S)$. Thus

$$\dim_{\mathbf{T}/\mathfrak{m}} S \leq \dim_{\mathbf{T}(S)} S \leq 1$$

as desired. □

An application of the above lemma shows that $D(W[\mathrm{Ver}])$ has $\mathbf{T}/\mathfrak{m}$ dimension ≤ 1.

Lemma 18.7.4. *$D(W[\mathrm{Ver}]) \cong D(V[\mathrm{Ver}])$*

Proof. Consider the following diagram.

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & D(Q) & \longrightarrow & D(W) & \longrightarrow & D(V) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \mathrm{Ver}} & & \downarrow{\scriptstyle \mathrm{Ver}} & & \downarrow{\scriptstyle \mathrm{Ver}} & & \\ 0 & \longrightarrow & D(Q) & \longrightarrow & D(W) & \longrightarrow & D(V) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & D(Q)/\operatorname{Ver} D(Q) & \longrightarrow & D(W)/\operatorname{Ver} D(W) & \xrightarrow{?\cong?} & D(V)/\operatorname{Ver} D(V) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array}$$

$D(V)$ has dimension 2 so since $\mathrm{Ver} \circ \mathrm{Frob} = \mathrm{Frob} \circ \mathrm{Ver} = 0$ and Ver, Frob are both nonzero they must each have rank 1 (in the sense of undergraduate linear algebra). Since D is exact, $D(V[\mathrm{Ver}]) = D(V)/\operatorname{Ver} D(V)$ and $D(W[\mathrm{Ver}]) =$

$D(W)/\operatorname{Ver} D(W)$. By the previous lemma $\dim D(W)/\operatorname{Ver} D(W) = 1$. Thus $D(W)/\operatorname{Ver} D(W) \to D(V)/\operatorname{Ver} D(V)$ is a map of 1 dimensional vector spaces so to show that it is an isomorphism we just need to show that it is surjective. This follows from the commutativity of the square

$$\begin{array}{ccccc} D(W) & \longrightarrow & D(V) & \longrightarrow & 0 \\ \downarrow & & \downarrow & & \\ D(W)/\operatorname{Ver} D(W) & \longrightarrow & D(V)/\operatorname{Ver} D(V) & & \\ & & \downarrow & & \\ & & 0 & & \end{array}$$

□

Suppose for the moment that we admit [BpR91]. Then

$$W = V \times \cdots \times V = V^{\oplus t}$$

so

$$D(W[\mathrm{Ver}]) \cong D(V[\mathrm{Ver}])^{\oplus t}$$

and hence $t = 1$.

Alternatively we can avoid the use of [BpR91]. Suppose $Q \neq 0$. Then there is an injection $V \hookrightarrow Q$. [[I can't see this without using B-L-R. It isn't obvious to me from $0 \to V \to W \to Q \to 0$.]] Thus over $\mathbf{F}_\ell$, $V[\mathrm{Ver}] \hookrightarrow Q[\mathrm{Ver}]$. Since $V[\mathrm{Ver}] \neq 0$ this implies $Q[\mathrm{Ver}] \neq 0$. Thus $D(Q)/\operatorname{Ver} D(Q) = D(Q[\mathrm{Ver}]) \neq 0$. But the bottom row of the above diagram implies $D(Q)/\operatorname{Ver} D(Q) = 0$ so $Q = 0$.

18.8 Key result of Boston-Lenstra-Ribet

Let G be a group (i.e., $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$), let k be a field (i.e., $k = \mathbf{T}/\mathfrak{m}$), and let V be a two dimensional k-representation of G given by

$$\rho : k[G] \to \mathrm{End}_k(V) = M_2(k).$$

The key hypothesis is that V is absolutely irreducible, i.e., that ρ is surjective. For each $g \in G$ consider

$$p_g = g^2 - g \operatorname{tr} \rho(g) + \det \rho(g) \in k[G].$$

By the Cayley-Hamilton theorem $\rho(p_g) = 0$. Let J be the two-sided ideal of $k[G]$ generated by all p_g such that $g \in G$. Since $J \subset \ker \rho$, ρ induces a map

$$\sigma : k[G]/J \to \mathrm{End}_k(V).$$

Theorem 18.8.1 (Boston-Lenstra-Ribet)**.** *If σ is surjective then σ is an isomorphism.*

In particular if V is absolutely irreducible then σ is surjective. The theorem can be false when $\dim V > 2$.

Suppose W is a second representation of G given by $\mu : k[G] \to \mathrm{End}(W)$ and that $\mu(J) = \{0\} \subset \mathrm{End}(W)$. Then W is a module over $k[G]/J = \mathrm{End}(V)$. But $\mathrm{End}(V)$ is a semisimple ring so any $\mathrm{End}(V)$ module is a direct sum of simple $\mathrm{End}(V)$ modules. The only simple $\mathrm{End}(V)$ module is V. Thus $W \cong V^{\oplus^n}$ for some n.

19
Local Properties of ρ_λ

Let f be a newform of weight 2 on $\Gamma_0(N)$. To f we have associated an abelian variety $A = A_f$ furnished with an action of $E = \mathbf{Q}(\ldots, a_n(f), \ldots)$. Let $\mathcal{O}_E$ be the ring of integers of E and $\lambda \subset \mathcal{O}_E$ a prime. Then we obtain a λ-adic representation

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(E_\lambda)$$

on the Tate module $\mathrm{Tate}_\ell A = \varprojlim A[\lambda^i]$. We will study the local properties of ρ_λ at various primes p.

19.1 Definitions

To view ρ_λ locally at p we restrict to the decomposition group $D_p = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ at p. Recall the definition of D_p. Let K be a finite extension of $\mathbf{Q}$ and let w be a prime of K lying over p. Then the decomposition group at w is defined to be

$$D_w = \{\sigma \in \mathrm{Gal}(K/\mathbf{Q}) : \sigma w = w\}.$$

Proposition 19.1.1. $D_w \cong \mathrm{Gal}(K_w/\mathbf{Q}_p)$

Proof. Define a map $\mathrm{Gal}(K_w/\mathbf{Q}_p) \to D_w$ by $\sigma \mapsto \sigma_{|K}$. Since $\sigma_{|K}$ fixes $\mathbf{Q}$ this restriction is an element of $\mathrm{Gal}(K/Q)$. Since $w\mathcal{O}_{K_w}$ is the unique maximal ideal of $\mathcal{O}_{K_w}$ and σ induces an automorphism of $\mathcal{O}_{K_w}$, it follows that $\sigma(w\mathcal{O}_{K_w}) = w\mathcal{O}_{K_w}$. Thus $\sigma_{|K}(w) = w$ so $\sigma_{|K} \in D_w$. The map $\sigma \mapsto \sigma_{|K}$ is bijective because K is dense in K_w. □

Let

$$D_p = \varprojlim_{w|p} D_w = \varprojlim_{w|p} \mathrm{Gal}(K_w/\mathbf{Q}_p) = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p).$$

For each $w|p$ let the inertia group I_w be the kernel of the map from D_w into $\mathrm{Gal}(\mathcal{O}_K/w, \mathbf{Z}/p\mathbf{Z})$. Let $I_p = \varprojlim_{w|p} I_w$.

19.2 Local properties at primes $p \nmid N$

Next we study local properties of ρ_λ at primes $p \nmid N$. Thus suppose $p \nmid N$ and $p \neq \ell = \mathrm{char}(\mathcal{O}_E/\lambda)$. Let $D_p = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. Then

1) $\rho_\lambda|D_p$ is unramified (i.e., $\rho_\lambda(I_p) = \{1\}$) thus $\rho_\lambda|D_p$ factors through D_p/I_p so $\rho_\lambda(\mathrm{Frob}_p)$ is defined.

2) $\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_p)) = a_p(f)$

2+) We can describe $\rho_\lambda|D_p$ up to isomorphism. It is the unique semisimple representation satisfying 1) and 2).

19.3 Weil-Deligne Groups

Notice that everything is sort of independent of λ. Using Weil-Deligne groups we can summarize all of these λ-adic representations in terms of data which makes λ disappear.

We have an exact sequence

$$1 \to I_p \to \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \to 0.$$

Since $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) = \hat{\mathbf{Z}}$ there is an injection $\mathbf{Z} \hookrightarrow \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. Define the Weil group $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \subset \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ to be the set of elements of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ mapping to $\mathbf{Z} \subset \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. W fits into the exact sequence

$$1 \to I_p \to W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathbf{Z} \to 1.$$

There is a standard way in which the newform f gives rise to a representation of W. Factor the polynomial $x^2 - a_p(f)x + p$ as a product $(x - r)(x - r')$ with $r, r' \in \mathbf{C}$. Define maps α, β by

$$\alpha : \mathbf{Z} \to \mathbf{C}^* : 1 \mapsto r$$

$$\beta : \mathbf{Z} \to \mathbf{C}^* : 1 \mapsto r'$$

Combining α and β and the map $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathbf{Z}$ yields a map

$$\alpha \oplus \beta : W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{GL}_2(\mathbf{C})$$

$$\sigma \mapsto \begin{pmatrix} \alpha(\sigma) & 0 \\ 0 & \beta(\sigma) \end{pmatrix}.$$

Moreover $\alpha \oplus \beta$ gives rise via some construction to all the λ-adic representations ρ_λ.

19.4 Local properties at primes $p \mid N$

Suppose $p|N$ but $p \neq \ell$. Carayol was able to generalize 1) in his thesis which builds upon the work of Langlands and Deligne in the direction of Deligne-Rapaport and Katz-Mazur. The idea is that the abelian variety A has a conductor which is a positive integer divisible by those primes of bad reduction. The conductor of A satisfies $\mathrm{cond}(A) = M^g$ where $g = \dim A$ and M is the reduced conductor of A.

Theorem 19.4.1 (Carayol)**.** $M = N$.

How can we generalize 2) or 2+)? For each p dividing N there is a representation σ_p of $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ over $\mathbf{C}$ such that σ_p gives rise to $\rho_\lambda|D_p$ for all $\lambda \nmid p$. Here $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is the Weil-Deligne group which is Deligne's generalization of the Weil group. $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is an extension of $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. What is σ_p supposed to be? The point is that σ_p is determined by f. Thus every f gives rise to a family $(\sigma_p)_{p \text{ prime}}$. To really think about σ_p we must think about modular forms in an adelic context instead of viewing them as holomorphic functions on the complex upper halfplane.

If $p^2|N$ and $f = \sum a_n q^n$ is a newform of level p^2 then it is a classical fact that $a_p = 0$. But the study of $\rho_\lambda|D_p$ is rich and "corresponds to a rather innocuous looking crystal".

19.5 Definition of the reduced conductor

We now define the reduced conductor. Let λ be a prime of E and p a prime of $\mathbf{Q}$ such that $\lambda \nmid p$. We want to define some integer $e(p)$ so that $p^{e(p)}$ is the p-part of the reduced conductor. We will not define $e(p)$ but what we will do is define an integer $e(p, \lambda)$ which is the p part of the conductor. $e(p, \lambda)$ is independent of λ but this will not be proved here.

Consider

$$\rho_\lambda|D_p : \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Aut}_{E_\lambda} V.$$

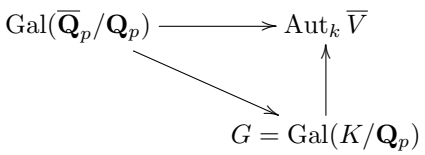
Let $V^I \subset V$ be the inertia invariants of V, i.e.,

$$V^I = \{v \in V : \rho_\lambda(\sigma)(v) = v \text{ for all } \sigma \in I\}.$$

Since ρ_λ is unramified at p if and only if $\rho_\lambda(I_p) = \{1\}$ we comment that ρ_λ is unramified at p if and only if $V^I = V$. Let

$$e(p, \lambda) = \dim V/V^I + \delta(p, \lambda)$$

where $\delta(p, \lambda)$ is the Swan conductor. By working with finite representations we define $\delta(p, \lambda)$ as follows. Choose a D_p-stable lattice L in V by first choosing an arbitrary one then taking the sum of its finitely many conjugates. Let $\overline{V} = L/\lambda L$ which is a 2 dimensional vector space over $k = \mathcal{O}_E/\lambda$. Let G be the quotient of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ by the kernel of the map $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Aut}_K \overline{V}$. Thus $G = \mathrm{Gal}(K/\mathbf{Q}_p)$ for some finite extension $K/\mathbf{Q}_p$. The extension K is finite over $\mathbf{Q}_p$ since $G \subset \mathrm{Aut}_k \overline{V}$ and $\mathrm{Aut}_k \overline{V}$ is a 2×2 matrix ring over a finite field. The corresponding diagram is



Consider in G the sequence of "higher ramification groups"

$$G = G_{-1} \supset G_0 \supset G_1 \supset \cdots .$$

Here G_0 is the inertia group of $K/\mathbf{Q}_p$ and G_1 is the p-sylow subgroup of G_0 [[the usages of "the" in this sentence makes me nervous.]] Let $G_i = \{g \in G_0 : \mathrm{ord}(g\pi - \pi) \geq i + 1\}$ where π is some kind of uniformizing parameter [[I missed this – what is π?]] Let

$$\delta(p, \lambda) = \sum_{i=1}^{\infty} \frac{1}{(G_0 : G_i)} \dim(\overline{V}/\overline{V}^{G_i}).$$

It is a theorem that $\delta(p, \lambda)$ is an integer and does not depend on λ.

If to start with we only had $\overline{V}$ and not V we could define

$$\mathrm{cond}(\overline{V}) = \sum_{i=0}^{\infty} \frac{1}{(G_0 : G_i)} \dim(\overline{V}/\overline{V}^{G_i}).$$

Then

$$e(p, \lambda) = \mathrm{cond}(\overline{V}) + (\dim \overline{V}^I - \dim V^I).$$

A reference for much of this material is Serre's *Local Factors of L-functions of λ-adic Representations.*

20
Adelic Representations

Our goal is to study local properties of the λ-adic representations ρ_λ arising from a weight 2 newform of level N on $\Gamma_0(N)$. There is a theorem of Carayol which states roughly that if $p \neq \ell$ then $\rho_\lambda | D_p$ is predictable from "the component at p of f". To understand this theorem we must understand what is meant by "the component at p of f". If $p^2 \nmid N$ this component is easy to determine but if $p^2 | N$ it is harder. One reason is that when $p^2 | N$ then $a_p(f) = 0$. [[this should be easy to see so there should be an argument here.]] If $a_p(f) = 0$ and $\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(V_\lambda)$ then $V_\lambda^{I_p} = 0$. This means that there is no ramification going on at p. See Casselman, "On representations of $GL(2)$ and the arithmetic of modular curves", Antwerp II.

20.1 Adelic representations associated to modular forms

Let R be a subring of $\mathcal{A}^2 = \mathbf{R}^2 \times (\hat{\mathbf{Z}} \otimes_{\mathbf{Z}} \mathbf{Q})^2$, suppose that $R \cong \mathbf{Q}^2$ and that $\mathbf{R} \otimes_{\mathbf{Q}} \mathcal{A} \cong \mathcal{A}^2$. Let $L = R \cap (\mathbf{R}^2 \times \hat{\mathbf{Z}}^2)$. Then the natural map $L \otimes \hat{\mathbf{Z}} \to \mathbf{Z}^2$ is an isomorphism. [[Is the isomorphism implied by the definition of L or is it part of the requirement for L to actually form an adelic lattice?]] L is called an *adelic lattice*.

The space of modular forms $S_2(\Gamma_0(N))$ is isomorphic to a certain space of functions on $G(\mathcal{A}) = \mathrm{GL}(2, \mathcal{A})$. See Borel-Jacquet [[Corvallis?]] or Diamond-Taylor, Inventiones Mathematica, 115 (1994) [[what is title?]] We will describe this isomorphism.

Write $\mathcal{A} = \mathbf{R} \times \mathcal{A}_f = \mathcal{A}_\infty \times \mathcal{A}^\infty$ where $\mathcal{A}_f = \prod \mathbf{Q}_p$ (restricted product) is the ring of finite adeles. $\mathcal{A}_\infty$ denotes the adeles with respect to the place ∞ so $\mathcal{A}_\infty = \mathbf{R}$, and $\mathcal{A}^\infty$ denotes the adeles away from the place ∞ so $\mathcal{A}^\infty = \mathcal{A}_f$.

$S_2(\Gamma_0(N))$ is isomorphic to the set of functions $\varphi : G(\mathcal{A}) \to \mathbf{C}$ which satisfy

0) φ is left invariant by $G(\mathbf{Q})$, i.e., $\varphi(x) = \varphi(gx)$ for all $g \in G(\mathbf{Q})$ and all $x \in G(\mathcal{A})$,

1) $\varphi(xu^\infty) = \varphi(x)$ for all $x \in G(\mathcal{A})$ and all $u^\infty \in U^\infty$,

2) $\varphi(xu_\infty) = \varphi(x)j(u_\infty, i)^{-k} \det(u_\infty)$ for all $x \in G(\mathcal{A})$ and $u_\infty \in U_\infty$,

3) holomorphy, cuspidal, and growth conditions.

U^∞ is the adelic version of $\Gamma_0(N)$. Thus U^∞ is the compact open subgroup

$$U^\infty = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\hat{\mathbf{Z}}) : c \equiv 0 \pmod{N}\} \subset G(\mathcal{A}_f).$$

(For $\Gamma_1(N)$ the condition is that $c \equiv 0 \pmod{N}$ and $d \equiv 1 \pmod{N}$ but a is not restricted.)

Next we describe $U_\infty \subset \mathrm{GL}_2(\mathbf{R})$. $\mathrm{GL}_2(\mathbf{R})$ operates on $\mathfrak{h}^\pm = \mathbf{C} - \mathbf{R}$ by $z \mapsto \frac{az+b}{cz+d}$. Let U_∞ be the stabilizer of i.

The third condition involves the *automorphy factor* j defined by

$$j(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, z) = cz + d.$$

To explain the holomorphy condition 3) we define, for any $g \in G(\mathcal{A}_f)$ a map

$$\alpha_{g,\varphi} : \mathfrak{h}^\pm \to \mathbf{C}$$

$$hi \mapsto \varphi(gh)j(h, i)^k(\det(h))^{-1}.$$

Here $h \in G(\mathcal{A}_f)$ so $hi \in \mathfrak{h}^\pm$. There may be several different $h \in G(\mathcal{A}_f)$ which give the same $hi \in \mathfrak{h}^\pm$ so it must be checked that $\alpha_{g,\varphi}$ is well-defined. Suppose $hi = h'i$, then $h^{-1}h'i = i$. Thus $h^{-1}h' \in U_\infty$, so by 2),

$$\varphi(gh^{-1}h') = \varphi(g)j(h^{-1}h', i)^{-k} \det(h^{-1}h').$$

Thus

$$\varphi(gh^{-1}h') \det(h')^{-1} = \varphi(g)j(h^{-1}h', i)^{-k} \det(h)^{-1}.$$

Substituting gh for g yields

$$\varphi(gh')(\det(h'))^{-1} = \varphi(gh)j(h^{-1}h', i)^{-k}(\det(h))^{-1}.$$

[[The the automorphy factor work out right. Why?]] The holomorphy condition is that the family of maps $\alpha_{g,\varphi}$ are all holomorphic.

The cuspidal condition is that for all $g \in G(\mathcal{A})$, the integral

$$\int_{u \in \mathcal{A}/\mathbf{Q}} \varphi(\left(\begin{smallmatrix} 1 & u \\ 0 & 1 \end{smallmatrix}\right), g)du = 0$$

vanishes. Since $\mathcal{A}/\mathbf{Q}$ is compact it has a Haar measure defined modulo k^* which induces du. Although the integral is not well-defined the vanishing or non-vanishing of the integral is.

We can now describe the isomorphism between $S_k(\Gamma_0(N))$ and the space of such functions φ on $G(\mathcal{A})$.

$$\{ \text{ space of } \varphi \text{ satisfying 0-3}\} \to S_k(\Gamma_0(N))$$

$$\varphi \mapsto f$$

where f is the restriction to $\mathfrak{h} = \mathfrak{h}^+$ of the function

$$hi \mapsto \varphi(h)j(h,i)^k(\det h)^{-1}.$$

Now we can associate to a newform f a representation of $G(\mathcal{A})$. We can weaken condition 1) to get

1-) $\varphi(xu^\infty) = \varphi(x)$ for all $x \in G(\mathcal{A})$ and all $u^\infty \in U^\infty$ where U^∞ is *some* compact open subgroup of $G(\mathcal{A}_f)$.

[[can the U^∞ vary for each $x \in G(\mathcal{A})$ or are they fixed throughout?]] Let $\boldsymbol{\mathcal{S}}$ be the space of all functions satisfying all conditions except 1-) replaces 1). This space has a left action of $G(\mathcal{A})$:

$$(g * \varphi)(x) = \varphi(xg)$$

If f is a newform corresponding to some φ via the above isomorphism then via this action f gives rise to an infinite dimensional representation π of $G(\mathcal{A})$. In fact we obtain, for each prime p, a representation π_p of $\mathrm{GL}(2, \mathbf{Q}_p)$. The representation space is $\sum_{g \in \mathrm{GL}_2(\mathbf{Q}_p)} \mathbf{C} \cdot g * \varphi$. Our immediate goal is to understand π_p for as many p as possible. [[spherical representations have something to do with this. are the π_p spherical reps?]]

We are studying local properties of the λ-adic representations ρ_λ associated to a newform f of weight 2, level N and character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ (with $\mathrm{cond}(\varepsilon)|N$). Let $\ell \in \mathbf{Z}$ be the prime over which λ lies. We look at ρ_λ locally at p, $p \neq \ell$.

As we saw last time f gives rise to an irreducible representation of $\mathrm{GL}(2, \mathcal{A})$. An irreducible representation of $GL(2, \mathcal{A})$ gives rise to a family of representations (π_v) where v is a prime or ∞ and π_v is an irreducible representation of $\mathrm{GL}(2, \mathbf{Q}_v)$. This is because $\mathbf{Q}_v \subset \mathcal{A}$ so $\mathrm{GL}(2, \mathbf{Q}_v) \subset \mathrm{GL}(2, \mathcal{A})$.

Carayol proved that if $p \neq \ell$ then $\rho_\lambda|_{D_p}$ depends, up to isomorphism, only on π_p. The most difficult case in the proof of this theorem is when $p^2 \nmid N$. The tools needed to obtain a proof were already available in the work of Langlands [1972].

To get an idea of what is going on we will first consider the case when $p \nmid N$. The characteristic polynomial of Frobenious (at least psychologically) under the representation $\rho_\lambda|_{D_p}$ is

$$x^2 - a_p x + p\varepsilon(p) = (x - r)(x - s).$$

Because of Weil's proof of the Riemann hypothesis for abelian varieties [over finite fields?] one knows that $|r| = |s| = \sqrt{p}$. Since ρ_λ arises from the action of Galois on an abelian variety which has good reduction at p (since $p \nmid N$) it follows that $\rho_\lambda|_{D_p}$ is unramified. [[Is this in Serre-Tate, 1968?]] We also know that $\rho_\lambda(\mathrm{Frob}_p)$ has characteristic polynomial

$$x^2 - a_p x + p\varepsilon(p) \in E[x].$$

In this situation one also knows that $\rho_\lambda(\mathrm{Frob}_p)$ is semisimple [[proof: Ribet nodded at Coleman who smiled at nodded back.]] Thus

$$\rho_\lambda \sim \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}.$$

In this situation what is the representation π_p of $\mathrm{GL}(2, \mathbf{Q}_p)$? There are two characters

$$\alpha, \beta : \mathbf{Q}_p^* \to \mathbf{C}^*$$

(called "Grössencharacters of type (a,0)") such that

1. α and β are unramified in the sense that

$$\alpha|_{\mathbf{Z}_p^*} = \beta|_{\mathbf{Z}_p^*} = 1.$$

   This is a reasonable condition since under some sort of local class field theory $\mathbf{Q}_p^*$ embeds as a dense subgroup of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ and under this embedding the inertia subgroup $I \subset \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ corresponds to $\mathbf{Z}_p^*$. [[This could be wrong. Also, is $I \cap \mathbf{Q}_p^* = \mathbf{Z}_p^*$?]]

2. $\alpha(p^{-1}) = r$, $\beta(p^{-1}) = s$.

In the 1950's Weil found a way under which α and β correspond to continuous characters α_λ, β_λ on $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ with values in $\overline{E}_\lambda^*$ such that

1. α_λ and β_λ are unramified.

2. $\alpha_\lambda(\mathrm{Frob}_p) = r$ and $\beta_\lambda(\mathrm{Frob}_p) = s$.

One has that $\rho_\lambda = \alpha_\lambda \oplus \beta_\lambda$. See [ST68]. [[Why see this?]]

Define a character Θ on the Borel subgroup

$$B = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \subset \mathrm{GL}(2, \mathbf{Q}_p)$$

by

$$\Theta\begin{pmatrix} x & y \\ 0 & z \end{pmatrix} = \alpha(x)\beta(z) \in \mathbf{C}^*.$$

Then

$$\pi_p = \mathrm{Ind}_B^{\mathrm{GL}(2,\mathbf{Q}_p)} := \mathbf{C}[\mathrm{GL}(2, \mathbf{Q}_p)] \otimes_{\mathbf{C}[B]} \mathbf{C}.$$

We call this induced representation π_p the unramified principal series representation associated to α and β and write $\pi_p = \mathrm{PS}(\alpha, \beta)$. People say π_p is spherical in the sense that there is a vector in the representation space invariant under the maximal compact subgroup of $\mathrm{GL}(2, \mathbf{Q}_p)$. [[Ribet was slightly unsure about the correct definition of spherical.]] [[For some mysterious reason]] since $\pi_p = \mathrm{PS}(\alpha, \beta)$ it follows that α, β and hence $\rho_\lambda|D_p$ is completely determined by π_p. [[This is the point and i don't see this.]]

Next we consider the more difficult case when $p||N$ (p divides N exactly). There are two cases to consider

a ε is ramified at p ($p|\operatorname{cond}(\varepsilon)$)

b ε is unramified at p ($p \nmid \operatorname{cond}(\varepsilon)$).

20.2 More local properties of the ρ_λ.

Let f be a newform of level N. Then f corresponds to a representation $\pi = \otimes\pi_v$. If $\lambda|\ell$ and $\ell \neq p$ then $\rho_\lambda|D_p$ corresponds to π_p under the Langlands correspondence. The details of this correspondence were figured out by Philip Kutzko, but Carayol completed it in the exceptional case (to be defined later). There are three cases to consider.

a) $p \nmid N$

b) $p||N$

c) $p^2 \mid N$

We considered the first two cases last time. The third case is different because the same sort of analysis as we applied to the first two cases no longer works in the sense that we no longer know what π_p looks like.

20.2.1 Possibilities for π_p

Case 1 (principal series) In this case, $\pi_p = \mathrm{PS}(\alpha, \beta)$. Here $\alpha, \beta : \mathbf{Q}_p^* \to \mathbf{C}^*$ are unramified characters. $\mathbf{Q}_p^*$ corresponds to a dense subgroup of the abelian Galois group of $\mathbf{Q}_p$ under the correspondence elucidated in Serre's "Local Classfield Theory" (in Cassels and Frohlich).

$$\begin{array}{ccc} \mathbf{Q}_p^* & \longrightarrow & \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)^{A\cap B} \\ \downarrow & & \downarrow \\ \mathbf{Z} & \longrightarrow & \hat{\mathbf{Z}} = \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \end{array}$$

Under this correspondence α and β correspond to Galois representations α_λ and β_λ and $\rho_\lambda|D_p \sim \alpha_\lambda \oplus \beta_\lambda$.

Case 2 (special) In this case, π_p is the *special* automorphic representation corresponding to the Galois representation $\kappa \otimes \mathrm{st}$ where st is the Steinberg representation which arises somehow from a split-multiplicative reduction elliptic curve and κ is a Dirichlet character. In this case

$$\rho_\lambda|D_p = \kappa \otimes \begin{pmatrix} \chi_\ell & * \\ 0 & 1 \end{pmatrix}.$$

Case 3 (cuspidal) Case 3 occurs when π_p does not fall into either of the previous cases. Such a π_p is called cuspidal or super-cuspidal. Some of these π_p come from the following recipe. Fix an algebraic closure $\overline{\mathbf{Q}}_p$ of $\mathbf{Q}_p$. Let K be a quadratic extension of $\mathbf{Q}_p$. Let $\psi : K^* \to \mathbf{C}^*$ be a Grössencharacter. Then ψ gives rise to a character

$$\psi_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}_p/K) \to \overline{E}_\lambda^*$$

which induces $\rho_\lambda|D_p$. That is,

$$\rho_\lambda|D_p = \mathrm{Ind}_{\mathrm{Gal}(\overline{\mathbf{Q}}_p/K)}^{\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)} \psi_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{GL}(2, \overline{E}_\lambda).$$

This representation is irreducible if and only if ψ is not invariant under the canonical conjugation of $K/\mathbf{Q}_p$. The pair ψ, K gives rise (via the construction of Jacquet-Langlands) to a representation of $\pi_{\psi,K}$ of $\mathrm{GL}(2, \mathbf{Q}_p)$. Those representations which do not come from this recipe and which do not fall into case 1 or case 2 above are called *extraordinary*. They can only occur when $p \leq 2$.

When can the various cases occur?

Case 1) occurs, e.g, if $p \nmid N$, and also if $p||N$ and ε (the character of f) is ramified at p.

Case 2) occurs if $p||N$ and ε is unramified at p.

20.2.2 The case $\ell = p$

We consider $\rho_\lambda | D_p$ where $\lambda | p$ and $p = \ell$. Write $f = \sum a_n q^n$. The case when $\lambda \nmid a_p$ is called the ordinary case. This case is very similar to the case for an ordinary elliptic curve. In other words,

$$\rho_\lambda | D_p = \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

[[Ribet mentioned that there is a paper on this by Mazur and Wiles in the American Journal of Mathematics. Look the reference up.]] An important point is that β is unramified so it makes sense to consider $\beta(\mathrm{Frob}_p) = a_p \in E_\lambda^*$. Since $\alpha\beta$ is the determinant, $\alpha\beta = \chi_p^{k-1}\varepsilon = \chi_p \varepsilon$ (after setting $k = 2$ to fix ideas), this gives some description of what is going on. The obvious question to ask is whether or not * is nontrivial. That is, is $\rho_\lambda | D_p$ semisimple or not?

When f has weight 2, then f gives rise to an abelian variety $A = A_f$. Then ρ_λ is defined by looking at the action of Galois on the λ-adic division points on A. If none of the λ lying over p divide a_p then A is ordinary at p. A stronger statement is that the p-divisible group $A[p^\infty]$ has good ordinary reduction.

One simple case is when f has CM. By this we mean that there is a character $\kappa \neq 1$ of order 2 such that $a_n = \kappa(n) a_n$ for all n prime to $\mathrm{cond}(\kappa)$. [[This is not a typo, I do not mean $\overline{a_n} = \kappa(n) a_n$. Coleman said that $a_n = \kappa(n) a_n$ is just a funny way to say that half of the a_n are 0.]] It is easy to prove that f has CM if and only if the ρ_λ become abelian on some open subgroup of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of finite index. Ribet explains this in his article in [Rib77]. If f has CM then since the representation ρ_λ is almost abelian one can show that * is trivial. Ribet said he does not know whether the converse is true. Note that f has CM if and only if $A_f/\overline{\mathbf{Q}}$ has CM. If all $\lambda | p$ are ordinary (i.e., they do not divide a_p) and if * is trivial for every $\rho_\lambda | D_p$ it is easy to show using [Ser98] that A has CM.

Next we will say something more about representations which appear to be ordinary. Consider the situation in which f has weight 2 and p exactly divides the level N of f. Suppose furthermore that the character ε of f is unramified at p. Then π_p is a special representation. The λ-adic representations for $\ell \neq p$ are (up to characters of finite order) like representations attached to some Tate curve. The situation is similar when $\ell = p$ since

$$\rho_\lambda | D_p = \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

As in the case of a Tate curve $\alpha/\beta = \chi_p$. Up to a quadratic character we know the situation since $\alpha\beta = \chi_p \varepsilon$. Also β is still unramified and $\beta(\mathrm{Frob}_p) = a_p$ is a unit. We know that $a_p^2 = \varepsilon(p)$ is a root of unity. When $k = 2$ the case of a spherical representation mimics what happens for ordinary reduction. The upper right hand entry * is never trivial in this case [[because of something to do with extensions and Kummer theory]].

Ribet said he knows nothing about the situation when $k > 2$. If $p || N$ and ε is unramified at p then π_p is special so ρ_λ, $\ell \neq p$ are again of the form $\left(\begin{smallmatrix} \alpha & * \\ 0 & \beta \end{smallmatrix}\right)$. So we still know everything up to a quadratic character. But if $\ell = p$ some characters are Hodge-Tate [[what does that mean?]] so by a theorem of Faltings they can not be bizarre powers of the cyclotomic character. The multiplicative case like the ordinary case is very special to the case $k = 2$. Wiles uses this heavily in his proof.

If k is arbitrary then $a_p^2 = \varepsilon(p)p^{\frac{k}{2}-1}$ so a_p is not a unit for $k > 2$ so there is no invariant line. So the representation is probably irreducible in the case $k > 2$. [[Echos of "yeah", "strange" and "very strange" are heard throughout the room.]]

20.2.3 Tate curves

Suppose $E/\mathbf{Q}_p$ is an elliptic curve with multiplicative reduction at p and that $j \in \mathbf{Q}_p$ is the j-invariant of E. Using a formula which can be found in [Sil94, V] one finds $q = q(j)$ with $|q| < 1$. The Tate curve is $E(\overline{\mathbf{Q}}_p) = \overline{\mathbf{Q}}_p/q^{\mathbf{Z}}$. The p torsion on the Tate curve is $\{\zeta_p^a(q^{1/p})^b : 0 \leq a, b \leq p-1\}$. Galois acts by $\zeta_p \mapsto \zeta_p^a$ and $q^{1/p} \mapsto \zeta_p^a q^{1/p}$. Thus the associated Galois representation is $\left(\begin{smallmatrix} \chi_p & * \\ 0 & 1 \end{smallmatrix}\right)$.

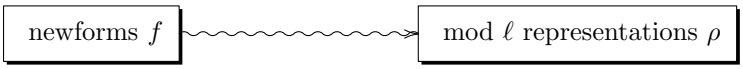
21
Serre's Conjecture

This is version 0.2 of this section. I am still unsatisfied with the organization, level of detail, and coherence of the presentation. A lot of work remains to be done. Some of Ken's lectures in Utah will be integrated into this chapter as well.

Let ℓ be a prime number. In this chapter we study certain mod ℓ Galois representations, by which we mean continuous homomorphisms ρ

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \xrightarrow{\quad\rho\quad} \mathrm{GL}(2, \mathbf{F})$$

where $\mathbf{F}$ is a finite field of characteristic ℓ. Modular forms give rise to a large class of such representations.



The motiving question is:

> Given a mod ℓ Galois representation ρ, which newforms f if any, of various weight and level, give rise to ρ?

We will assume that ρ is irreducible. It is nevertheless sometimes fruitful to consider the reducible case (see [SW97] and forthcoming work of C. Skinner and A. Wiles).

Serre [Ser87] has given a very precise conjectural answer to our motivating question. The result, after much work by many mathematicians, is that certain of Serre's conjectures are valid in the sense that if ρ arises from a modular form at all, then it arises from one having a level and weight as predicted by Serre. The main trends in the subject are "raising" and "lowering."

Our motivating question can also be viewed through the opposite optic. What is the most *bizarre* kind of modular form that gives rise to ρ? A close study of the ramification behavior of ρ allows one to at least obtain some sort of control over the possible weights and levels. This viewpoint appears in [Wil95].

In this chapter whenever we speak of a *Galois representation* the homomorphism is assumed to be continuous. The reader is assumed to be familiar with local fields, some representation theory, and some facts about newforms and their characters.

21.1 The Family of λ-adic representations attached to a newform

First we briefly review the representations attached to a given newform. Let

$$f = \sum_{n \geq 1} a_n q^n \in S_k(N, \varepsilon)$$

be a newform of level N, weight k, and character ε. Set $K = \mathbf{Q}(\ldots, a_n, \ldots)$ and let $\mathcal{O}$ be the ring of integers of K. If λ is a nonzero prime ideal of $\mathcal{O}$ we always let ℓ be the prime of $\mathbf{Z}$ over which λ lies, so $\lambda \cap \mathbf{Z} = (\ell)$. Let K_λ denote the completion of K at λ, thus K_λ is a finite extension of $\mathbf{Q}_\ell$. The newform f gives rise to a system $(\rho_{f,\lambda})$ of λ-adic representations

$$\rho_{f,\lambda} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, K_\lambda)$$

one for each λ.

Theorem 21.1.1 (Carayol, Deligne, Serre)**.** *Let f be as above and ℓ a prime. There exists a Galois representation*

$$\rho_{f,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, K \otimes \mathbf{Q}_\ell)$$

with the following property: If $p \nmid \ell N$ is a prime, then $\rho_{f,\ell}$ is unramified at p, and the image under $\rho_{f,\ell}$ of any Frobenius element for p is a matrix with trace a_p and determinant $\varepsilon(p)p^{k-1}$.

The actual construction of $\rho_{f,\ell}$ won't be used in what follows. Since $K \otimes \mathbf{Q}_\ell$ is a product $\prod_\lambda K_\lambda$ of the various completions of K at the primes λ of K lying over ℓ, we have a decomposition

$$\mathrm{GL}(2, K \otimes \mathbf{Q}_\ell) = \prod_{\lambda | \ell} \mathrm{GL}(2, K_\lambda)$$

and projection onto $\mathrm{GL}(2, K_\lambda)$ gives $\rho_{f,\lambda}$.

By Lemma 18.1.1 $\rho_{f,\lambda}$ is equivalent to a representation taking values in $\mathrm{GL}(2, \mathcal{O})$ where $\mathcal{O}$ is the ring of integers of K. Since λ is a prime of $\mathcal{O}$ reduction modulo λ defines a map $\mathrm{GL}(2, \mathcal{O}) \to \mathrm{GL}(2, \mathbf{F})$ where $\mathbf{F} = \mathcal{O}/\lambda$ is the residue class field of λ, and we obtain a mod ℓ Galois representation

$$\overline{\rho}_{f,\lambda} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, \mathbf{F}).$$

21.2 Serre's Conjecture A

Serre [Ser87] conjectured that certain mod ℓ representations arise from modular forms, and then gave a precise recipe for which type of modular form would give

rise to the representation. We refer to the first part of his conjecture as "Conjecture A" and to the second as "Conjecture B".

Let $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F})$ be a Galois representation with $\mathbf{F}$ a finite field. We say that ρ *arises from a modular form* or that ρ is *modular* if there is some newform $f = \sum a_n q^n$ and some prime ideal λ of $K = \mathbf{Q}(\ldots, a_n, \ldots)$ such that ρ is isomorphic to $\overline{\rho}_{f,\lambda}$ over $\overline{\mathbf{F}}$.

$$\begin{array}{ccc} \mathcal{O}/\lambda & \hookrightarrow & \overline{\mathbf{F}} \\ & & \big| \\ & & \mathbf{F} \end{array}$$

Recall that a Galois representation ρ is *odd* if $\det(\rho(c)) = -1$ where $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a complex conjugation. Galois representation arising from modular forms are always odd. [More detail?]

Conjecture 21.2.1 (Serre's Conjecture A)**.** *Suppose*

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F})$$

is an odd irreducible (continuous) Galois representation with $\mathbf{F}$ *a finite field. Then* ρ *arises from a modular form.*

21.2.1 The Field of definition of ρ

One difficulty is that ρ sometimes takes values in a slightly smaller field than $\mathcal{O}/\lambda$. We illustrate this by way of an example. Let f be one of the two conjugate newforms of level 23, weight 2, and trivial character. 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 $\mathcal{O}/\lambda \cong \mathbf{F}_4$ and so $\overline{\rho}_{f,\lambda}$ is a homomorphism to $\mathrm{GL}(2, \mathbf{F}_4)$.

Proposition 21.2.2. *If* $p \neq 2$ *then* $a_p \in \mathbf{Z}[\sqrt{5}]$.

Proof. We have $f = f_1 + \alpha f_2$ with

$$\begin{aligned} f_1 &= q - q^3 - q^4 + \cdots \\ f_2 &= q^2 - 2q^3 - q^4 + 2q^5 + \cdots \end{aligned}$$

Because $S_2(\Gamma_0(23))$ has dimension two, it is spanned by f_1 and f_2. Let $\eta(q) = q^{\frac{1}{24}} \prod_{n \geq 1}(1 - q^n)$. By Proposition **??**, $g = (\eta(q)\eta(q^{23}))^2 \in S_2(\Gamma_0(23))$. An explicit calculation shows that $g = q^2 - 2q^3 + \cdots$ so we must have $g = f_2$. 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. □

Thus those a_p with $p \neq 2$ map modulo λ to $\mathbf{F}_2 \subset \mathbf{F}_4$, and hence the traces and determinants of Frobenius, at primes where Frobenius is defined, take values in $\mathbf{F}_2$. This is enough to imply that $\overline{\rho}_{f,\lambda}$ is isomorphic over $\overline{\mathbf{F}}$ to a representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_2)$. Geometrically, $\mathrm{GL}(2, \mathbf{F}_2) \cong S_3$ and the representation ρ is one in which Galois acts via S_3 on three points of $X_0(23)$.

Thus in general, if we start with ρ and wish to see that ρ satisfies Conjecture A, we may need to pass to an algebraic closure of $\mathbf{F}$. In our example, starting with $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F}_2)$ we say that "ρ is modular" because $\rho \cong \overline{\rho}_{\lambda,f}$, but keep in mind that this particular isomorphism only takes place over $\mathbf{F}_4$.

At this point K. Buzzard comments, "Maybe there is some *better* modular form so that all of the a_p actually lie in $\mathbf{F}_2$ and the associated representation is isomorphic to ρ. That would be a stronger conjecture." Ribet responds that he has never thought about that question.

21.3 Serre's Conjecture B

Serre's second conjecture asserts that ρ arises from a modular form in a particular space. Suppose

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F})$$

is an odd irreducible Galois representation with $\mathbf{F}$ a finite field.

Conjecture 21.3.1 (Serre's Conjecture B)**.** *Suppose ρ arises from some modular form. Then ρ arises from a modular form of level $N(\rho)$, weight $k(\rho)$ and character $\varepsilon(\rho)$. The exact recipe for $N(\rho)$ and $k(\rho)$ will be given later.*

Conjecture B has largely been proven.

Theorem 21.3.2. *Suppose ℓ is odd. If the mod ℓ representation ρ is irreducible and modular, then ρ arises from a newform f of level $N(\rho)$ and weight $k(\rho)$.*

21.4 The Level

The level $N(\rho)$ is a conductor of ρ, essentially the *Artin conductor* except that we omit the factor corresponding to ℓ. The level is a product

$$N(\rho) = \prod_{p \neq \ell} p^{e(p)}.$$

We now define the $e(p)$. Let V be the representation space of ρ, so V is a two dimensional $\mathbf{F}$-vector space and we view ρ as a homomorphism

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(V),$$

or equivalently view V as an $\mathbf{F}[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module. Let $K = \overline{\mathbf{Q}}^{\ker(\rho)}$ be the field cut out by ρ, it is a finite Galois extension of $\mathbf{Q}$ with Galois group which we call G. Choose a prime $\wp$ of K lying over p and let

$$D = \{\sigma \in G : \sigma(\wp) \subset \wp\}$$

denote the decomposition group at $\wp$. Let $\mathcal{O}$ be the ring of integers of K and recall that the higher ramification groups $G_{-1} \supset G_0 \supset G_1 \supset G_2 \supset \cdots$ are

$$G_i = \{\sigma \in D : \sigma \text{ acts trivially on } \mathcal{O}/\wp^{i+1}\}.$$

Thus $G_{-1} = D$ and G_0 is the inertia group. Each G_i is a normal subgroup of D because G_i is the kernel of $D \to \mathrm{Aut}(\mathcal{O}/\wp^{i+1})$. Let

$$V^{G_i} = \{v \in V : \sigma(v) = v \text{ for all } \sigma \in G_i\}.$$

Lemma 21.4.1. *The subspace V^{G_i} is invariant under D.*

Proof. Let $g \in D$, $h \in G_i$ and $v \in V^{G_i}$. Since G_i is normal in D, there exists $h' \in G_i$ so that $g^{-1}hg = h'$. Then $h(gv) = gh'v = gv$ so $gv \in V^{G_i}$. □

By [Frö67, §9] that there is an integer i so that $G_i = 0$. We can now define

$$e(p) = \sum_{i=0}^{\infty} \frac{1}{[G_0 : G_i]} \dim(V/V^{G_i}).$$

Thus $e(p)$ depends only on $\rho|I_p$ where $I_p = G_0$ is an inertia group at p. In particular, if ρ is unramified at p then all G_i for $i \geq 0$ vanish and $e(p) = 0$. Separating out the term $\dim(V/V^{I_p})$ corresponding to $i = 0$ allows us to write

$$e(p) = \dim(V/V^{I_p}) + \delta(p) \leq 2 + \delta(p)$$

since $\dim V = 2$. The term $\delta(p)$ is called the *Swan conductor*. By [Ser77, 19.3] $\delta(p)$ is an integer, hence $e(p)$ is an integer. We call ρ *tamely ramified at* p if all G_i for $i > 0$ vanish, in which case the Swan conductor is 0.

Suppose $\rho \cong \overline{\rho}_{f,\lambda}$ (over $\overline{\mathbf{F}}$) for some newform f of level N. There is a relationship between $e(p)$ and something involving only f. Let V_λ be the representation space of $\rho_{f,\lambda}$, so V_λ is a vector space over an extension of $\mathbf{Q}_\ell$. It turns out that

$$\mathrm{ord}_p(N) = \dim(V_\lambda/V_\lambda^{I_p}) + \delta(p).$$

Thus

$$\mathrm{ord}_p(N) = e(p) + \text{ error term}$$

where the error term is $\dim(V^{I_p}) - \dim(V_\lambda^{I_p}) \leq 2$. The point is that more can become invariant upon reducing modulo λ. Thus if f gives rise to ρ then the power of p in the level N of f is constrained as it is given by a certain formula only depending on $e(p)$ and an error term which has magnitude at most 2.

As we will see, the weight $k(\rho)$ depends only on $\rho|I_\ell$. Thus $k(\rho)$ can be viewed as an analogue of $e(\ell)$.

21.4.1 Remark on the case $N(\rho) = 1$

One consequence of Conjecture B is that every modular mod ℓ representation ρ must come from a newform f of level prime to ℓ. Suppose that f is a modular form of level ℓ. Consider the corresponding mod ℓ representation. It is unramified outside ℓ so $N(\ell) = 1$. The conjecture then implies that this mod ℓ representation

comes from a level 1 modular form, i.e., a modular form on $\mathrm{SL}_2(\mathbf{Z})$, of possibly higher weight. This is a classical result.

For the rest of this subsection we assume that $\ell \geq 11$. There is a relationship between mod ℓ forms on $\mathrm{SL}_2(\mathbf{Z})$ of weight $\ell+1$ and mod ℓ forms on $\Gamma_0(\ell)$ of weight 2. The dimensions of each of these spaces are the same over $\mathbf{F}_\ell$ or over $\mathbf{C}$, i.e.,

$$\dim_{\mathbf{F}_\ell} S_2(\Gamma_0(\ell); \mathbf{F}_\ell) = \dim_{\mathbf{C}} S_2(\Gamma_0(\ell); \mathbf{C}).$$

$$\dim_{\mathbf{F}_\ell} S_{\ell+1}(\mathrm{SL}_2(\mathbf{Z}); \mathbf{F}_\ell) = \dim_{\mathbf{C}} S_{\ell+1}(\mathrm{SL}_2(\mathbf{Z}); \mathbf{C}).$$

We now describe a map

$$F : S_2(\Gamma_0(\ell); \mathbf{F}_\ell) \longrightarrow S_{\ell+1}(\mathrm{SL}_2(\mathbf{Z}); \mathbf{F}_\ell).$$

The weight k Eisenstein series

$$E_k = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n$$

is a modular form for $\mathrm{SL}_2(\mathbf{Z})$. Here the Bernoulli numbers B_k are defined by

$$\frac{t}{e^t - 1} = \sum_{k=0}^{\infty} B_k \frac{t^k}{k!},$$

so $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{12}$, $B_3 = -\frac{1}{720}, \ldots$.

Proposition 21.4.2. *If $\ell \geq 5$, then*

$$E_{\ell-1} \equiv 1 \pmod{\ell}.$$

Proof. We must check that $\ell \mid \frac{2(\ell-1)}{B_{\ell-1}}$, or equivalently that $\ell \mid \frac{1}{B_{\ell-1}}$. Set $n = \ell - 1$, $p = \ell$ and apply [Lan95, X.2.2]. □

Suppose $\overline{f} \in S_2(\Gamma_0(\ell); \mathbf{F}_\ell)$ is the reduction of $f \in S_2(\Gamma_0(\ell); \mathbf{F}_\ell)$. Multiplication by $E_{\ell-1}$ gives a mod ℓ form $\overline{f} \cdot E_{\ell-1}$ of weight $\ell + 1$ on $\Gamma_0(\ell)$. We have

$$F(\overline{f}) = \mathrm{tr}(\overline{f} \cdot E_{\ell-1}) \in S_{\ell+1}(\mathrm{SL}_2(\mathbf{Z}); \mathbf{F}_\ell)$$

where tr is induced by $X_0(\ell) \to X_0(1)$. The map F was discovered by Serre [DK73]. That F has the properties necessary to deduce the prime level case of Serre's Conjecture B was proved by explicit computation in the Berkeley Ph.D. thesis of C. Queen [Que77]. Katz believes that it is easy to prove the formula from the right magical moduli point of view, but neither author has seen this. The first author gave a concrete proof for $\ell > 2$ in [Rib94]. It would be nice if someone would construct a clear proof for the case $\ell = 2$.

21.4.2 Remark on the proof of Conjecture B

We now give a very brief outline of the proof of Conjecture B when $\ell > 2$. Start with a representation ρ which comes from some (possibly terrible) newform f of level $N(f)$ and weight $k(f)$.

Step 1. Using a concrete argument replace f by another newform giving rise to ρ so that $\ell \nmid N(f)$.

Step 2. Compare $N(\rho)$ and $N(f)$. These are two prime to ℓ numbers. What if $p \mid \frac{N(f)}{N(\rho)}$? Carayol separated this into several cases. In each case replace f by a better form so that the ratio has a smaller power of p in it. This is level lowering. Eventually we get to the case $N(\rho) = N(f)$.

Step 3. Using a paper of Edixhoven [Edi92b] one shows that f can be replaced with a another form of the same level so that the weight k is equal to $k(\rho)$.

21.5 The Weight

21.5.1 The Weight modulo $\ell - 1$

We first give some background which motivates Serre's numerical recipe for the weight. We start with a newform f of low weight k and consider the behavior of the representation $\rho = \overline{\rho}_{f,\lambda}|I_\ell$, where I_ℓ is an inertia group at ℓ. Assume that we are in the following situation:

- $N = N(f)$ is prime to ℓ,
- $2 \leq k \leq \ell + 1$.

There are other cases to consider but we consider this one first. Let ε be the character of f and recall that for $p \nmid \ell N$

$$\det(\rho_{f,\ell}(\mathrm{Frob}_p)) = \varepsilon(p)p^{k-1}.$$

The mod ℓ cyclotomic character $\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$ is the homomorphism sending Frob_p to $p \in \mathbf{F}_\ell^*$. Thus

$$\det(\rho) = \varepsilon\chi_\ell^{k-1}.$$

Since ε is a Dirichlet character mod N and $\ell \nmid N$, $\varepsilon|I_\ell = 1$ and so

$$\det(\rho|I_\ell) = \chi_\ell^{k-1}.$$

We see immediately that $\det(\rho|I_\ell)$ determines k modulo $\ell - 1$. Since $2 \equiv \ell + 1 \pmod{\ell - 1}$, this still doesn't distinguish between 2 and $\ell + 1$.

21.5.2 Tameness at ℓ

Let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a modular mod ℓ Galois representation. Let $D = D_\ell \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be a decomposition group at ℓ. Let σ be the semisimplification of $\rho|D$. Thus σ is either a direct sum of two characters or $\sigma = \rho|D$ depending on whether or not $\rho|D$ is irreducible. By [Frö67, 8.1] the ramification group $P = G_1$ is the unique Sylow ℓ-subgroup of $I_\ell = G_0$.

Lemma 21.5.1. *The semisimplification σ of ρ is tame, i.e., $\sigma|P = 0$ where P is the Sylow ℓ-subgroup of the inertia group I_ℓ.*

Proof. Since I_ℓ is normal in D and any automorphism of I_ℓ sends P to some Sylow ℓ-subgroup and P is the only such, it follows that P is normal in D. Let $W = \mathbf{F}_{\ell^\nu} \times \mathbf{F}_{\ell^\nu}$ be the representation space of σ. Then

$$W^P = \{w \in W : \sigma(\tau)w = w \quad \text{for all} \quad \tau \in P\}$$

is a subspace of W invariant under the action of D. For this let $\alpha \in D$ and suppose $w \in W^P$. Since P is normal in D, $\alpha^{-1}\tau\alpha = \tau'$ for some $\tau' \in P$. Therefore

$$\sigma(\alpha)^{-1}\sigma(\tau)\sigma(\alpha)w = \sigma(\tau')w = w$$

so $\sigma(\tau)\sigma(\alpha)w = \sigma(\alpha)w$ hence $\sigma(\alpha)w \in W^P$.

But $W^P \neq 0$. To see this write W as a disjoint union of its orbits under the action of P. Since P is an ℓ-Sylow group and W is finite we see that the size of each orbit is either 1 or a positive power of ℓ. Now $\{0\}$ is a singleton orbit, W has ℓ-power order, and all non-singleton orbits have order a positive power of ℓ so there must be at least $\ell - 1$ other singleton orbits. Each of these other singleton orbits gives a nonzero element of W^P.

If $W^P = W$ then P acts trivially so we are done. If $W^P \neq W$, then since W^P is nonzero it is a one dimensional subspace invariant under D, so by semisimplicity σ is diagonal. Let $\tau \in P$, then τ has order ℓ^n for some n. Write

$$\sigma(\tau) = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix},$$

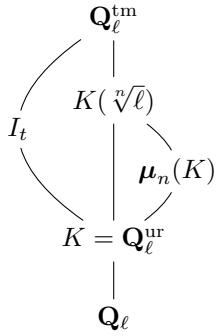
then $\alpha^{\ell^n} = \beta^{\ell^n} = 1$. Since $\alpha, \beta \in \mathbf{F}_{\ell^\nu}$ they have order dividing $|\mathbf{F}^*_{\ell^\nu}| = \ell^\nu - 1$. But $\gcd(\ell^\nu - 1, \ell^n) = 1$ from which it follows that $\alpha = \beta = 1$. Thus $P = \{1\}$ and again P acts trivially, as claimed. □

21.5.3 Fundamental characters of the tame extension

The lemma implies $\sigma|I_\ell$ factors through the tame quotient $I_t = I_\ell/P$. We now describe certain characters of I_t more explicitly. Denote by $\mathbf{Q}_\ell^{\mathrm{tm}}$ the maximal tamely ramified extension of $\mathbf{Q}_\ell$, it is the fixed field of P. Let $K = \mathbf{Q}_\ell^{\mathrm{ur}}$ be the maximal unramified extension, i.e., the fixed field of the inertia group I_ℓ. By Galois theory

$$I_t = \mathrm{Gal}(\mathbf{Q}_\ell^{\mathrm{tm}}/\mathbf{Q}_\ell^{\mathrm{ur}}).$$

For each positive integer n coprime to ℓ there is a tower of Galois extensions



Since $\mathbf{Q}(\zeta_n)/\mathbf{Q}$ is unramified at ℓ the nth roots of 1 are contained in K so by Kummer theory

$$\mathrm{Gal}(K(\sqrt[n]{\ell})/K) \cong \boldsymbol{\mu}_n(K)$$

where $\boldsymbol{\mu}_n(K)$ denotes the group of nth roots of unity in K. Thus for each n prime to ℓ we obtain a map $I_t \to \boldsymbol{\mu}_n(K)$. They are compatible so upon passing to the limit we obtain a map

$$I_t \longrightarrow \varprojlim \boldsymbol{\mu}_n(K) = \prod_{r\neq\ell} \varprojlim \boldsymbol{\mu}_{r^a}(K) = \prod_{r\neq\ell} \mathbf{Z}_r(1).$$

In fact [Frö67, 8, Corollary 3] the above maps are isomorphisms. Viewed more cleverly mod ℓ we obtain a map

$$I_t \longrightarrow \varprojlim \boldsymbol{\mu}_n(\overline{\mathbf{F}}_\ell) = \varprojlim \mathbf{F}^*_{\ell^i}.$$

Thus for each i we have a map

$$I_t \to \mathbf{F}^*_{\ell^i}.$$

which is called the *fundamental character of level i*. The construction of this character on I_t is somewhat unnatural because we had to choose an embedding $\mathbf{F}_{\ell^i} \hookrightarrow \overline{\mathbf{F}}_\ell$. Instead Serre begins with a "disembodied" field F of order ℓ^i. There are i different maps $\mathbf{F}_{\ell^i} \to F$ corresponding to the i automorphisms of $\mathbf{F}_{\ell^i}$. Restricting these maps to $\mathbf{F}^*_{\ell^i}$ and composing with $I_t \to \mathbf{F}^*_{\ell^i}$ gives the i fundamental characters of level i. The unique fundamental character of level 1 is the mod ℓ cyclotomic character χ_ℓ.

21.5.4 The Pair of characters associated to ρ

Recall that we have a representation ρ whose semisimplification gives a representation which we denote by σ:

$$\sigma : I_t \longrightarrow \mathrm{GL}(2, \mathbf{F}_{\ell^\nu}).$$

Since $\sigma(I_t)$ is a finite abelian group and the elements of I_t have order prime to ℓ, this representation is semisimple and can be diagonalized upon passing to $\overline{\mathbf{F}}_\ell$. In fact, since the characteristic polynomials all have degree two, σ can be diagonalized over $\mathbf{F}_{\ell^{2\nu}}$. Thus σ corresponds to a pair of characters

$$\alpha,\, \beta : I_t \longrightarrow \mathbf{F}^*_{\ell^{2\nu}}.$$

These characters have some stability properties since σ is the restriction of a homomorphism from the full decomposition group. Consider the tower of fields

$$\begin{array}{c} K(\sqrt[n]{\ell}) \\ | \\ K = \mathbf{Q}_\ell^{\mathrm{ur}} \\ | \\ \mathbf{Q}_\ell \end{array}$$

Let $G = \mathrm{Gal}(K(\sqrt[n]{\ell})/\mathbf{Q}_\ell)$. Recall that $\mathrm{Gal}(K(\sqrt[n]{\ell})/K) \cong \boldsymbol{\mu}_n(K)$ and $\mathrm{Gal}(K/\mathbf{Q}_\ell)$ is topologically generated by Frob_ℓ. If $h \in \boldsymbol{\mu}_n(K)$ and $g \in G$ restricts to Frob_ℓ then we have the conjugation formula:

$$ghg^{-1} = h^\ell.$$

Applying this to our representation σ with $h \in I_t$ we find that

$$\sigma(ghg^{-1}) = \sigma(h^\ell) = \sigma(h)^\ell$$

so

$$\sigma(g)\sigma(h)\sigma(g^{-1}) = \sigma(ghg^{-1}) = \sigma(h)^\ell.$$

The point is that the representation $h \mapsto \sigma(h)^\ell$ is equivalent to $h \mapsto \sigma(h)$ via conjugation by $\sigma(g)$. We conclude that the pair of characters $\{\alpha, \beta\}$ is stable under ℓ-th powering, i.e., as a set

$$\{\alpha, \beta\} = \{\alpha^\ell, \beta^\ell\}.$$

What does this mean? There are two possibilities:

- *Level 1:* $\alpha^\ell = \alpha$ and $\beta^\ell = \beta$.
- *Level 2:* $\alpha^\ell = \beta$ and $\beta^\ell = \alpha$, but $\alpha \neq \beta$.

Note that in the level 1 case, α and β take values in $\mathbf{F}_{\ell^\nu}^*$.

21.5.5 Recipe for the weight

We will play a carnival game, "guess your weight." First we consider the level 2 case. Our strategy is to express α and β in terms of the two fundamental characters of level 2. We then observe how the characters associated to a newform are expressed in terms of the two fundamental characters.

The remainder of this chapter is devoted to motivating the following definition.

Definition 21.5.2. Let ρ and σ be as above. Let ψ, ψ' be the two fundamental characters of level 2 and χ the fundamental character of level 1 (the cyclotomic character). Serre's recipe for $k(\rho)$ is as follows.

1. Suppose that α and β are of level 2. We have

$$\rho|I_\ell = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}.$$

   After interchanging α and β if necessary, we have (uniquely) $\alpha = \psi^{r+\ell q} = \psi^r(\psi')^q$ and $\beta = (\psi')^r\psi^q$ with $0 \leq r < q \leq \ell - 1$. We set $k(\rho) = 1 + \ell r + q$.

2. Suppose that α and β are of level 1. We have

$$\rho|I_\ell = \begin{pmatrix} \chi^r & * \\ 0 & \chi^q \end{pmatrix}.$$

   (a) If $* = 0$, normalize and reorder r, q so that $0 \leq r \leq q \leq \ell - 2$. We set $k(\rho) = 1 + \ell r + q$.

   (b) If $* \neq 0$, normalize so that $0 \leq q \leq \ell - 2$ and $1 \leq r \leq \ell - 1$. We set $a = \min(r, q)$, $b = \max(r, q)$. If $\chi^{r-q} = \chi$ and $\rho \otimes \chi^{-q}$ is not finite at ℓ then we set $k(\rho) = 1 + \ell a + b + \ell - 1$; otherwise we set $k(\rho) = 1 + \ell a + b$.

21.5.6 The World's first view of fundamental characters

First some background. Suppose $E/\mathbf{Q}$ is an elliptic curve and ℓ is a prime (2 is allowed). Assume E has good *supersingular* reduction at ℓ, so $\tilde{E}(\overline{\mathbf{F}}_\ell)[\ell] = \{0\}$. Then there is a Galois representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{Aut}(E[\ell])$$

which, *a priori*, may or may not be irreducible. As above, ρ gives rise to two characters

$$\alpha,\, \beta :\ \ I_t \to \mathbf{F}_{\ell^2}^*.$$

Serre (see [V$\acute{7}$2]) proved that α, β are the two fundamental characters of level 2 and that I_t acts irreducibly over $\mathbf{F}_\ell$. [[Is this right?]] He also observed that there is a map

$$I_t \to \mathbf{F}_{\ell^2}^* \subset \mathrm{GL}(2, \mathbf{F}_\ell)$$

where $\mathbf{F}_{\ell^2}^*$ sits inside $\mathrm{GL}(2, \mathbf{F}_\ell)$ via the action of the multiplicative group of a field on itself after choice of a basis, and the map to $\mathbf{F}_{\ell^2}^*$ is through one of the two fundamental characters of level 2.

21.5.7 Fontaine's theorem

Serre next asked Fontaine to identify the characters α and β attached to more general modular representations. Fontaine's answer was published in [Edi92b]. Suppose $f = \sum a_n q^n$ is a newform of weight k such that $2 \le k \le \ell$ and the level $N(f)$ of f is prime to ℓ. Let $\rho = \overline{\rho}_{f,\lambda}$ where λ is a prime of $\mathbf{Q}(\ldots, a_n, \ldots)$ lying over ℓ. Assume we are in the *supersingular* case, i.e.,

$$a_\ell \equiv 0 \mod \lambda.$$

Semisimplifying as before gives a pair of characters

$$\alpha,\, \beta : I_t \longrightarrow \mathbf{F}_{\ell^{2\nu}}^*.$$

Let $\psi, \psi' : I_t \to \mathbf{F}_{\ell^2}^*$ be the two fundamental characters of level 2 so $\psi = (\psi')^\ell$ and $\psi' = \psi^\ell$.

Theorem 21.5.3. *With the above notation and hypothesis, the characters α, β arising by semisimplifying and restricting $\overline{\rho}_{f,\lambda}$ satisfy*

$$\{\alpha,\, \beta\} = \{\psi^{k-1},\, (\psi')^{k-1}\}.$$

21.5.8 Guessing the weight (level 2 case)

We now try to guess the weight in the level 2 case. We begin with a representation ρ whose semisimplification is a representation σ, which in turn gives rise to a pair of level 2 characters

$$\{\alpha,\ \beta\} = \{\psi^a,\ (\psi')^a = \psi^{\ell a}\}.$$

Since ψ takes values in $\mathbf{F}_{\ell^2}^*$, we may think of a as a number modulo $\ell^2 - 1$. Note also that the pair is unchanged upon replacing a by ℓa. The condition that we are not in level 1 is that a is not divisible by $\ell + 1$ since

$$\psi\psi' = \psi^{\ell+1} : I_t \to \mathbf{F}_\ell^*$$

is the unique fundamental character of level 1, i.e., the mod ℓ cyclotomic character.

Normalize a so that $0 \le a < \ell^2 - 1$ and write $a = q\ell + r$. What are the possible values for q and r? By the Euclidean algorithm $0 \le r \le \ell - 1$ and $0 \le q \le \ell - 1$. If $r = q$ then a is a multiple of $\ell + 1$, so $r \neq q$. If $r > q$, multiply the above relation by ℓ to obtain $a\ell = q\ell^2 + r\ell$. But we are working mod $\ell^2 - 1$ so this becomes $a\ell = q + r\ell \pmod{\ell^2 - 1}$. Thus if we replace a by $a\ell$ then the roles of q and r are swapped in the Euclidean division. Thus we can *assume* that $0 \le r < q \le \ell - 1$. Now

$$\alpha = \psi^a = \psi^{q\ell + r} = (\psi')^q \psi^r = (\psi\psi')^r (\psi')^{q-r},$$

so

$$\{\alpha, \beta\} = \{(\psi\psi')^r (\psi')^{q-r}, (\psi\psi')^r \psi^{q-r}\}.$$

Since $\psi\psi' = \chi$ is the mod ℓ cyclotomic character, we can view $\{\alpha, \beta\}$ as a pair of characters $(\psi')^{q-r}$ and ψ^{q-r} which has been multiplied, as a pair, by χ^r.

What weight do we guess if $r = 0$? In this case

$$\{\alpha, \beta\} = \{(\psi')^{k-1}, \psi^{k-1}\}$$

where $k = 1 + q$. So in analogy with Theorem 21.5.3 we guess that

$$k(\rho) = 1 + q \qquad (r = 0, \text{ supersingular case}).$$

What do we guess in general? Suppose $f = \sum a_n q^n$ is a modular form thought of mod ℓ which gives rise to ρ, and that ℓ does not divide the level of f. We might as well ask what modular form gives rise to $\rho \otimes \chi$. In [**?**] we learn that

$$\theta f = \sum n a_n q^n \pmod{\ell}$$

is a mod ℓ eigenform, and it evidently gives rise to $\rho \otimes \chi$. Furthermore, if k is the weight of f, then θf has weight $k + \ell + 1$. Since

$$\{\alpha, \beta\} = \{\psi^{q-r}, (\psi')^{q-r}\} \cdot \chi^r$$

we guess that

$$k(\rho) = q - r + 1 + (\ell + 1)r = 1 + \ell r + q \qquad \text{(supersingular case)}$$

But be careful! the *minimal* weight k does not have to go up by $\ell - 1$, though it usually does. This is described by the theory of θ-cycles which we will review shortly.

21.5.9 θ-cycles

The theory of the θ operator was first developed by Serre and Swinnerton–Dyer and then later jazzed up by Katz in [**?**]. There is a notion of modular forms mod ℓ and of q-expansion which gives a map

$$\alpha : \bigoplus_{k \ge 0} M_k(\Gamma_1(N); \mathbf{F}_\ell) \longrightarrow \mathbf{F}_\ell[[q]].$$

This map is not injective. The kernel is the ideal generated by $A - 1$ where A is the Hasse invariant.

Suppose $f \in \mathbf{F}_\ell[[q]]$ is in the image of α. If $f \neq 0$ let $w(f)$ denote the smallest k so that f comes from some M_k. If f does not coming from any single M_k do not define $w(f)$. Define an operator θ on $\mathbf{F}_\ell[[q]]$ by

$$\theta(\sum a_n q^n) = q\frac{d}{dq}(\sum a_n q^n) = \sum n a_n q^n.$$

Serre and Swinnerton-Dyer showed that θ preserves the image of α.

Theorem 21.5.4. *Suppose $f \neq 0$ is a mod ℓ modular form as above. If $\ell \nmid w(f)$ then $w(\theta f) = w(f) + \ell + 1$.*

Associated to f we have a sequence of nonnegative integers

$$w(f),\ w(\theta f),\ w(\theta^2 f),\ \ldots$$

Fermat's little theorem implies that this sequence is periodic because $\theta^\ell f = \theta f$ and so $w(\theta^\ell f) = w(\theta f)$. We thus call the cyclic sequence $w(f)$, $w(\theta f)$, $w(\theta^2 f), \ldots$ the *θ-cycle* of f. Tate asked:

> What are the possible θ-cycles?

This question was answered in [Joc82] and [Edi92b]. We now discuss the answer in a special case.

Let f be an eigenform such that $2 \leq k = w(f) \leq \ell$ and f is supersingular, i.e., $a_\ell(f) = 0$. Since f is an eigenform the a_n are multiplicative ($a_{nm} = a_n a_m$ for $(n, m) = 1$) and if ε denotes the character of f then

$$\begin{aligned} a_{\ell^i} &= a_{\ell^{i-1}} \cdot a_\ell - \varepsilon(\ell)\ell^{k-1} a_{\ell-2} \\ &= a_{\ell^{i-1}} \cdot a_\ell = 0 \end{aligned}$$

since we are working in characteristic ℓ and $k \geq 2$. Thus $a_n(f) = 0$ whenever $\ell \mid n$ and so $\theta^{\ell-1} f = f$ hence $w(\theta^{\ell-1} f) = w(f)$.

If we apply θ successively to f what happens? Before proceeding we remark that it can be proved that there is at most one drop in the sequence

$$k = w(f),\ w(\theta f),\ w(\theta^2 f), \ldots, w(\theta^{\ell-2} f).$$

First suppose $k = 2$. The θ-cycle must be

$$2,\ 2 + (\ell + 1),\ 2 + 2(\ell + 1),\ \ldots,\ 2 + (\ell - 2)(\ell + 1).$$

This is because, by Theorem 21.5.4, applying θ raises the weight by $\ell + 1$ so long as the weight is not a multiple of ℓ. Only the last term in the above sequence is divisible by ℓ. There are $\ell - 1$ terms so this is the full θ-cycle.

Next suppose $k = \ell$. The θ-cycle is

$$\ell,\ 3,\ 3 + (\ell + 1),\ \ldots,\ 3 + (\ell + 1)(\ell - 3).$$

The last term is divisible by ℓ, no earlier term after the first is, and there are $\ell - 1$ terms so this is the full θ-cycle. We know that the second term must be 3 since it is the only number so that the θ-cycle works out right, i.e., so that the $(\ell - 1)$st

term is divisible by ℓ but no earlier term except the first is. For example, if we would have tried 2 instead of 3 we would have obtained the sequence

$$\ell,\ 2,\ 2+(\ell+1),\ \ldots,\ 2+(\ell+1)(\ell-3),\ 2+(\ell+1)(\ell-2).$$

This sequence has one too many terms.

Now we consider the remaining values of k: $2 < k < \ell$. The θ-cycle is

$$k,\ k+(\ell+1),\ \ldots,\ k+(\ell+1)(\ell-k),\ k',\ k'+\ell+1,\ \ldots,\ k'+(\ell+1)(k-3).$$

The first $\ell-k+1$ terms of the sequence are obtained by adding $\ell+1$ successively until obtaining a term $k+(\ell+1)(\ell-k)$ which is divisible by ℓ. Applying θ to a form of weight $k+(\ell+1)(\ell-k)$ causes the weight to drop to some k'. How can we guess k'? It must be such that $k'+(\ell+1)(k-3)$ is divisible by ℓ. Thus the correct answer is

$$k' = \ell+3-k.$$

21.5.10 Edixhoven's paper

Suppose that ρ is an irreducible mod ℓ representation so that the level of the characters α, β associated to the semisimplification of ρ are level 2. To avoid problems in a certain exceptional case assume $\ell \geq 3$. Edixhoven [Edi92b] proved that if ρ arises from an eigenform in $S_k(\Gamma_1(N))$ with $(N,\ell)=1$, then ρ arises from an eigenform in $S_{k(\rho)}(\Gamma_1(N))$. What are the elements of the proof?

1. The behavior of $\rho|I_\ell$ when ρ arises from an f with $2 \leq w(f) \leq \ell+1$.
2. The fact that every modular representation ρ has the form $\overline{\rho}_{f,\lambda} \otimes \chi^i$ where $i \in \mathbf{Z}/(\ell-1)\mathbf{Z}$, and $2 \leq w(f) \leq \ell+1$.

Serre knew the second element but never published a proof. How did Serre talk about his result before Edixhoven's paper? The eigenform f correponds to a system of eigenvalues in $\overline{\mathbf{F}}_\ell$ of the Hecke operators T_r, $r \nmid \ell N$. Eigenforms of weight at most $\ell+1$ give, up to twist, all systems of Hecke eigenvalues. A possible proof of this uses the construction of ρ_f in terms of certain étale cohomology groups.

21.6 The Character

Let

$$\rho : G_{\mathbf{Q}} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a Galois representation. Assume that ρ is irreducible, modular $\rho \cong \overline{\rho}_{\lambda,f}$, and $\ell > 2$. The *degree* of a character $\varphi : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ is the cardinality $|\varphi((\mathbf{Z}/N\mathbf{Z})^*)|$ of its image.

Theorem 21.6.1. *Under the above hypothesis, ρ comes from a modular form f of weight $k(\rho)$, level $N(\rho)$, and under a certain extra assumption, character ε of degree prime to ℓ.*

Extra Assumption: Not all of the following are true.

1. $\ell = 3$,

2. $\rho|\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\sqrt{-3}))$ is abelian, and

3. $\det(\rho)$ is *not* a power of the mod 3 cyclotomic character χ.

Example 21.6.2. If ρ comes from the Galois representation on the 3-torsion of an elliptic curve, then $\det(\rho) = \chi$ is the mod 3 cyclotomic character, so the extra assumption does hold and the theorem applies.

Let ρ be a representation as above. Then it is likely that

$$\det\rho : G_{\mathbf{Q}} \to \mathbf{F}_{\ell^\nu}$$

is ramified at ℓ. Let χ be the mod ℓ cyclotomic character.

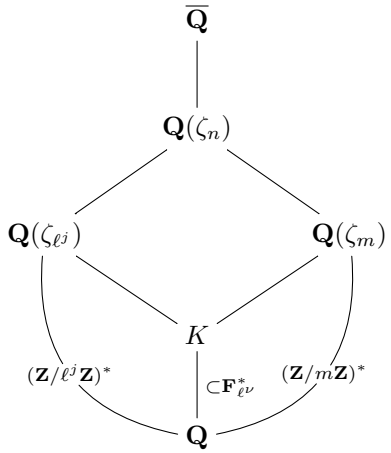
Proposition 21.6.3. *Let $\varphi : G_{\mathbf{Q}} \to \mathbf{F}_{\ell^\nu}^*$ be a continuous homomorphism. Then $\varphi = \theta\chi^i$ for some i and some $\theta : G_{\mathbf{Q}} \to \mathbf{F}_{\ell^\nu}^*$ which is unramified at ℓ.*

Proof. Since φ is continuous and $\mathbf{F}_{\ell^\nu}^*$ is finite, the subfield K of $\overline{\mathbf{Q}}$ fixed by $\ker(\varphi)$ is a finite Galois extension of $\mathbf{Q}$. Since the image of φ is abelian, the Galois group of K over $\mathbf{Q}$ is abelian. By the Kronecker-Weber theorem [Lan94, X.3] there is a cyclotomic extension $\mathbf{Q}(\zeta_n) = \mathbf{Q}(\exp^{2\pi i/n})$ which contains K. Since $\operatorname{Gal}(\mathbf{Q}(\zeta_n)/\mathbf{Q}) \cong (\mathbf{Z}/n\mathbf{Z})^*$, the character φ gives a homomorphism $\varphi' : (\mathbf{Z}/n\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*$. Write $n = m\ell^j$ with m coprime to ℓ. Then

$$(\mathbf{Z}/n\mathbf{Z})^* \cong (\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/\ell^j\mathbf{Z})^* \cong (\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/\ell^j\mathbf{Z})^*$$

which decomposes φ' as a product $\theta' \times \psi'$ where

$$\begin{aligned} \theta' &: (\mathbf{Z}/m\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^* \\ \psi' &: (\mathbf{Z}/\ell^j\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*. \end{aligned}$$



The character θ is obtained by lifting θ'. It is unramified at ℓ because it factors through $\operatorname{Gal}(\mathbf{Q}(\zeta_m)/\mathbf{Q})$ and $\ell \nmid m$. The cardinality of $(\mathbf{Z}/\ell^j\mathbf{Z})^*$ is $(\ell-1)\ell^{j-1}$ whereas the cardinality of $\mathbf{F}_{\ell^\nu}^*$ is $(\ell-1)(\ell^{\nu-1}+\cdots+1)$ so the image of ψ' lies in $\mathbf{F}_\ell^*$. Thus ψ' lifts to a power χ^i of the cyclotomic character. □

Using the proposition write $\det(\rho) = \theta\chi^i$ with θ unramified at ℓ. As in the proof of the proposition we can write θ as a Dirichlet character $(\mathbf{Z}/m\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*$. It can

be shown using properties of conductors that m can be chosen so that $m|N(\rho)$. Thus we view θ as a Dirichlet character

$$\theta : (\mathbf{Z}/N(\rho)\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*.$$

Let $H = \ker\theta \subset (\mathbf{Z}/N(\rho)\mathbf{Z})^*$. Define a congruence subgroup $\Gamma_H(N)$ by

$$\Gamma_H(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a, d \in H \right\}.$$

We have the following theorem.

Theorem 21.6.4. *Suppose $\ell > 2$ and ρ satisfies the above assumptions including the extra assumption. Then ρ arises from a form in*

$$S_{k(\rho)}(\,\Gamma_H(N(\rho))\,).$$

In particular, the theorem applies if ρ comes from the ℓ-torsion representation on an elliptic curve, since then $\det = \chi$ so the extra assumption is satisfied.

21.6.1 A Counterexample

One might ask if the extra assumption in Theorem 21.6.4 is really necessary. At first Serre suspected it was not. But he was surprised to discover the following example which shows that the extra assumption can not be completely eliminated. The space $S_2(\Gamma_1(13))$ is 2 dimensional, spanned by the eigenform

$$f = q + \alpha q^2 + (-2\alpha - 4)q^3 + (-3\alpha - 7)q^4 + (2\alpha + 3)q^5 + \cdots$$

and the $\mathrm{Gal}(\mathbf{Q}(\sqrt{-3})/\mathbf{Q})$-conjugate of f, where $\alpha^2 + 3\alpha + 3 = 0$. The character $\varepsilon : (\mathbf{Z}/13\mathbf{Z})^* \to \mathbf{C}^*$ of f has degree 6. Let $\lambda = (\sqrt{3})$ and let

$$\overline{\rho}_{f,\lambda} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_3)$$

be the associated Galois representation. Then $\det(\overline{\rho}_{f,\lambda}) = \chi\theta$ where χ is the mod 3 cyclotomic character and $\theta \equiv \varepsilon \pmod 3$. In particular θ has order 2. Thus $H = \ker(\theta) \subset (\mathbf{Z}/13\mathbf{Z})^*$ is exactly the index two subgroup of squares in $(\mathbf{Z}/13\mathbf{Z})^*$. The conclusion of the theorem can not hold since $S_2(\Gamma_H(13)) = 0$. This is because any form would have to have a character whose order is at most two since it must be trivial on H, but $S_2(\Gamma_H(13)) \subset S_2(\Gamma_1(13))$ and $S_2(\Gamma_1(13))$ is spanned by f and its Galois conjugate, both of which have character of order 6. In this example

- $\ell = 3$,
- $\overline{\rho}_{\lambda,f}|\,\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\sqrt{-3}))$ is abelian, and
- $\det\rho$ is not a power of χ.

A good way to see the second assertion is to consider the following formula:

$$f \otimes \varepsilon^{-1} = \overline{f}$$

(up to an Euler factor at 13) in the sense that

$$\rho_{f,\lambda} \otimes \varepsilon^{-1} = \rho_{\overline{f}}.$$

Now reduce mod 3 to obtain

$$\overline{\rho}_{f,\lambda} \otimes \varepsilon^{-1} \cong \overline{\rho}_{f,\lambda}$$

since $f \equiv \overline{f} \pmod{\sqrt{-3}}$ (since 3 is ramified). Thus $\overline{\rho}_{f,\lambda}$ is isomorphic to a twist of itself by a complex character so $\overline{\rho}_{f,\lambda}$ is reducible and abelian over the field corresponding to its kernel. In fact, by the same argument, $\overline{\rho}_{f,\lambda}$ is also abelian when restricted to the Galois groups of $\mathbf{Q}(\sqrt{13})$ and $\mathbf{Q}(\sqrt{39})$.

[[Give more details and describe David Jones's thesis.]]

21.7 The Weight revisited: level 1 case

We are interested in the recipe for $k(\rho)$ in the level 1 case. If we semisimplify and restrict to inertia we obtain a direct sum of two representations. In the level 1 case both representations are powers of the cyclotomic character. There are thus two possibilities for $\rho|I$:

$$\rho|I = \begin{pmatrix} \chi^\alpha & * \\ 0 & \chi^\beta \end{pmatrix} \quad \text{or} \quad \rho|I = \begin{pmatrix} \chi^\alpha & 0 \\ 0 & \chi^\beta \end{pmatrix}.$$

In the second case we guess $k(\rho)$ by looking at the exponents and normalizing as best we can. Since α and β are only defined mod $\ell - 1$ we may, after relabeling if necessary, assume that $0 \le \alpha \le \beta \le \ell - 2$. Factoring out χ^α we obtain

$$\chi^\alpha \otimes \begin{pmatrix} 1 & 0 \\ 0 & \chi^{\beta-\alpha} \end{pmatrix}.$$

Next (secretly) recall that if f is ordinary of weight k then f gives rise to the representation

$$\begin{pmatrix} \chi^{k-1} & * \\ 0 & 1 \end{pmatrix}$$

(here $*$ can be trivial). If f is of weight $\beta - \alpha + 1$, applying the θ operator α times gives the desired representation. Thus the recipe for the weight is

$$w(\rho) = (\ell + 1)\alpha + \beta - \alpha + 1 = \beta + \ell\alpha + 1.$$

There is one caveat: Serre was uncomfortable with weight 1 forms, so if $\alpha = \beta = 0$ he defines $k(\rho) = \ell$ instead of $k(\rho) = 1$.

Ogus asks what is wrong with weight 1, and Ribet replies that Serre didn't know a satisfactory way in which to define modular forms in weight 1. Merel then adds that Serre was frustrated because he could not do explicit computations in weight 1.

21.7.1 Companion forms

Suppose f is ordinary ($a_\ell \notin \lambda$) of weight k, $2 \le k \le \ell + 1$, and let $\overline{\rho}_{f,\lambda}$ be the associated representation. Then

$$\overline{\rho}_{f,\lambda}|I_\ell = \begin{pmatrix} \chi^{k-1} & * \\ 0 & 1 \end{pmatrix}.$$

To introduce the idea of a companion form suppose that somehow by chance $* = 0$. Twisting ρ by χ^{1-k} gives

$$\rho \otimes \chi^{1-k}|I_\ell = \rho \otimes \chi^{\ell-k}|I_\ell = \begin{pmatrix} 1 & 0 \\ 0 & \chi^{\ell-k} \end{pmatrix}.$$

What is the minimum weight of a newform giving rise to such a representation? Because the two characters 1 and χ take values in $\mathbf{F}_\ell^*$ we are in the level 1 situation. The representation is semisimple, reducible, and $\alpha = 0$, $\beta = \ell - k$, so the natural weight is $\ell+1-k$. Thus Conjecture B predicts that there should exist another form g of weight $\ell + 1 - k$ such that $\rho_g \cong \rho_f \otimes \chi^{\ell-k}$ (over $\overline{\mathbf{F}}_\ell$). Such a form g is called a *companion* of f. In characteristic ℓ, we have $g = \theta^{\ell-k} f$. This conjecture was for the most part proved by Gross [Gro90] when $k \neq 2, \ell$, and by Coleman-Voloch [CV92] when $k = \ell$. [[It would be nice to say something here about the *subtleties* involved in going from this mod ℓ form g to the companion form g produced by Gross-Coleman-Voloch. Roughly, how are the two objects linked?]]

21.7.2 The Weight: the remaining level 1 case

Let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a Galois representation with $\ell > 2$. Assume that ρ is irreducible and modular. Then ρ comes from a modular form in $S_{k(\rho)}(\Gamma_1(N\ell))$, with $2 \leq k(\rho) \leq \ell^2 - 1$. We still must define $k(\rho)$ in the remaining level 1 case in which

$$\rho|I_\ell = \begin{pmatrix} \chi^\alpha & * \\ 0 & \chi^\beta \end{pmatrix}.$$

If $\alpha \neq \beta + 1$ then

$$k(\rho) = 1 + \ell a + b$$

where $a = \min(\alpha, \beta)$ and $b = \max(\alpha, \beta)$. Now assume $\alpha = \beta + 1$. Then

$$\rho|I_\ell = \chi^\beta \otimes \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}.$$

Define a representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ by

$$\sigma = \rho \otimes \chi^{-\beta} \cong \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}.$$

We now give a motivated recipe for $k(\sigma)$. Granted that "finite at ℓ" is defined in the next section, the recipe is

$$k(\sigma) = \begin{cases} 2 & \text{if } \sigma \text{ is finite at } \ell, \\ \ell + 1 & \text{otherwise.} \end{cases}$$

This is enough to determine $k(\rho)$ giving

$$k(\rho) = k(\chi^\beta \otimes \sigma) = (\ell + 1)\beta + k(\sigma).$$

21.7.3 Finiteness

We continue with the notation of the previous section. Let

$$D = D_\ell = \mathrm{Gal}(\overline{\mathbf{Q}}_\ell/\mathbf{Q}_\ell) \hookrightarrow \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$$

denote a decomposition group at ℓ.

Definition 21.7.1. We say that $\sigma|D$ is **finite** if it is equivalent to a representation of the form $\mathcal{G}(\overline{\mathbf{Q}}_\ell)$, where $\mathcal{G}$ is a finite flat $\mathbf{F}_{\ell^\nu}$-vector space scheme over $\mathbf{Z}_\ell$.

This definition may not be terribly enlightening, so we consider the following special case. Suppose $E/\mathbf{Q}$ is an elliptic curve with semistable (=good or multiplicative) reduction at ℓ. Then $E[\ell]$ defines a representation

$$\sigma : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}\, E[\ell] \cong \mathrm{GL}(2, \mathbf{F}_\ell).$$

Let Δ_E be the minimal discriminant of E.

Proposition 21.7.2. *With notation as above, σ is finite at ℓ if and only if*

$$\mathrm{ord}_\ell\, \Delta_E \equiv 0 \pmod{\ell}.$$

If $p \neq \ell$ (and E has semistable reduction at p) then σ is unramified at p if and only if

$$\mathrm{ord}_p\, \Delta_E \equiv 0 \pmod{\ell}.$$

We give some hint as to how the proof goes when $p \equiv 1 \pmod{\ell}$. Let E be an elliptic curve with multiplicative reduction at p. Set $V = E[\ell] = \mathbf{F}_\ell \oplus \mathbf{F}_\ell$. We have a representation

$$\sigma : D = \mathrm{Gal}(\overline{\mathbf{Q}}_\ell/\mathbf{Q}_\ell) \to \mathrm{Aut}\, V.$$

The theory of Tate curves gives an exact sequence of D-modules

$$0 \to X \to V \xrightarrow{\alpha} Y \to 0.$$

Each of these terms is a D module and X and Y are 1-dimensional as $\mathbf{F}_\ell$-vector spaces. The action of D on Y is given by an unramified character ε of degree dividing 2. The action of D on X is given by $\chi\varepsilon$.

Next we define an element of $H^1(D, \mathrm{Hom}_{\mathbf{F}_\ell}(Y, X))$. A *splitting* $s : Y \to V$ is an $\mathbf{F}_\ell$-linear map (not necessarily a map of D-modules) such that $\alpha s = 1$. Choose such a splitting. For each $d \in D$ consider the twisting ${}^d s : Y \to V$ defined by ${}^d s(y) = ds(d^{-1}y)$. Since

$$\alpha(ds(d^{-1}y)) = d(\alpha(s(d^{-1}y))) = d(d^{-1}y) = y$$

if follows that ${}^d s$ is again a splitting. Thus ${}^d s - s : Y \to V$ followed by $\alpha : V \to Y$ is the zero map. Since ${}^d s - s$ is a linear map, ${}^d s - s \in \mathrm{Hom}_{\mathbf{F}_\ell}(Y, X)$. The map $d \mapsto {}^d s$ defines a 1-cocycle which gives an element of $H^1(D, \mathrm{Hom}_{\mathbf{F}_\ell}(Y, X))$. There is an isomorphism of D-modules $\mathrm{Hom}_{\mathbf{F}_\ell}(Y, X) \cong \mu_\ell$ so we have isomorphisms

$$H^1(D, \mathrm{Hom}_{\mathbf{F}_\ell}(Y, X)) \cong H^1(D, \mu_\ell) \cong \mathbf{Q}_p^*/(\mathbf{Q}_p^*)^\ell.$$

The last isomorphism follows from Kummer theory since $\mathbf{Q}_p$ is assumed to contain the ℓ-th roots of unity (our assumption that $p \equiv 1 \pmod{\ell}$). Thus σ defines an

element of $\mathbf{Q}_p^*/(\mathbf{Q}_p^*)^\ell$. Serre proved that τ is finite if and only if the corresponding element of $\mathbf{Q}_p^*/(\mathbf{Q}_p^*)^\ell$ is in the image of $\mathbf{Z}_p^*$.

If $E/\mathbf{Q}_p$ is an elliptic curve with multiplicative reduction then there exists a Tate parameter $q \in \mathbf{Q}_p^*$ with $\mathrm{Val}_p(q) > 0$ such that

$$E \cong E_q := \mathbf{G}_m/q^{\mathbf{Z}}$$

over the unique quadratic unramified extension of $\mathbf{Q}_p$. The kernel of multiplication by ℓ gives rise to an exact sequence as above, which is obtained by applying the snake lemma connecting $E[\ell]$ to Y in the following diagram:

$$\begin{array}{ccccccc}
0 & \longrightarrow & q^{\mathbf{Z}} & \xrightarrow{\ell} & q^{\mathbf{Z}} & \longrightarrow & Y \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
X & \longrightarrow & \mathbf{G}_m & \xrightarrow{\ell} & \mathbf{G}_m & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \\
E[\ell] & \longrightarrow & E & \xrightarrow{\ell} & E & &
\end{array}$$

Furthermore, the element of $\mathbf{Q}_p^*/(\mathbf{Q}_p^*)^\ell$ defined by the representation comming from $E[\ell]$ is just the image of q. One has

$$\Delta_E = q \prod_{n=1}^{\infty} (1-q^n)^{24}.$$

Note that the product factor is a unit in $\mathbf{Q}_p$, so $\mathrm{Val}_p \Delta_E = \mathrm{Val}_p q$.

22
Fermat's Last Theorem

22.1 The application to Fermat

"As part of this parcel, I can sketch the application to Fermat."

Suppose that semistable elliptic curves over $\mathbf{Q}$ are modular. Then FLT is true. Why? "As I have explained *so many times...*" Suppose $\ell > 5$ and

$$a^\ell + b^\ell + c^\ell = 0$$

with $abc \neq 0$, all relatively prime, and such that $A = a^\ell \equiv -1 \pmod 4$, $B = b^\ell$ is even and $C = c^\ell \equiv 1 \pmod 4$. Then we consider the elliptic curve

$$E : y^2 = x(x - A)(x - B).$$

The *minimal discriminant* is

$$\Delta_E = \frac{(ABC)^2}{2^8}$$

as discussed in Serre's [Ser87] and [DK95]. The conductor N_E is equal to the product of the primes dividing ABC (so in particular N_E is square-free). Furthermore, E is semistable – the only hard place to check is at 2. Diamond-Kramer checks this explicitly.

Here is how to get Fermat's theorem. View $E[\ell]$ as a $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. The idea is to show that this representation must come from a modular form of weight 2 and level 2. This will be a contradiction since there are no modular forms of weight 2 and level 2. But to apply the level and weight theorem we need to know that $E[\ell]$ is irreducible. The proof of this is due to Mazur.

Let $\rho : G \to \mathrm{Aut}\, E[\ell]$ be the Galois representation on the ℓ torsion of E. Since E is semistable for $p \neq \ell$,

$$\rho | I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

Assume ρ is reducible. Then ρ has an invariant subspace so

$$\rho \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

Then from the form of $\rho|I_p$ we see that the characters α and β could be ramified only at ℓ. Thus $\alpha = \chi^i$ and $\beta = \chi^{1-i}$ where χ is the mod ℓ cyclotomic character. The exponents are i and $1-i$ since $\alpha\beta$ is the determinant which is χ. [[Why is χ supposed to be the only possible unramified character? probably since whatever the character is, it is a product of χ times something else, and the other factor is ramified.]]

What happens to ρ at ℓ, i.e., what is $\rho|I_\ell$? There are only two possibilities. Either $\rho|I_\ell$ is the direct sum of the two fundamental characters or it is the sum of the trivial character with χ. The second possibility must be the one which occurs. [[I do not understand why... something about "characters globally are determined by local information."]] So either $i = 0$ or $i = 1$. If $i = 0$,

$$\rho = \begin{pmatrix} 1 & * \\ 0 & \chi \end{pmatrix}.$$

This means that there is an element of $E[\ell]$ whose subspace is left invariant under the action of Galois. Thus E has a point of order ℓ rational over $\mathbf{Q}$. If $i = 1$ then

$$\rho = \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}.$$

Therefore $\mu_\ell \hookrightarrow E[\ell]$. Divide to obtain $E' = E/\mu_\ell$. The representation on $E[\ell]/\mu_\ell$ is constant (this is basic linear algebra) so the resulting representation on $E'[\ell]$ has the form $\left(\begin{smallmatrix} 1 & * \\ 0 & \chi \end{smallmatrix}\right)$ so E' has a rational point of order ℓ.

Now $E[2]$ is a trivial Galois module since it contains 3 obvious rational points, namely $(0,0), (A,0)$, and $(B,0)$. Thus the group structure on the curve E (or E') (which we constructed from a counterexample to Fermat) contains

$$\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/\ell\mathbf{Z}.$$

In Mazur's paper [["rational isogenies of prime degree"]] he proves that $\ell \leq 3$. Since we assumed that $\ell > 5$, this is a contradiction. Notice that we have *not* just proved FLT. We have demonstrated the *irreducibility* of the Galois representation on the ℓ torsion of the elliptic curve E arising from a hypothetical counterexample to FLT.

We now have a representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell) = \mathrm{Aut}\, E[\ell]$$

which is irreducible and modular of weight 2 and level $N = N_E$ (the conductor of E). Because ρ is irreducible we conclude that ρ is modular of weight $k(\rho)$ and level $N(\rho)$. Furthermore

$$\mathrm{ord}_\ell\, \Delta_E = \mathrm{ord}_\ell (ABC)^2 = 2\ell\, \mathrm{ord}_\ell\, abc \equiv 0 \pmod{\ell}$$

so $k(\rho) = 2$.

We can also prove that $N(\rho) = 2$. Clearly $N(\rho)|N_E$. This is because $N(\rho)$ computed locally at $p \neq \ell$ divides the power of p in the conductor of the ℓ-adic representation for E at p. [[I do not understand this.]] Since ρ is only ramified at 2 or maybe ℓ, $N(\rho)$ must be a power of 2. For $p \neq \ell$, ρ is ramified at p if and only if $\mathrm{ord}_p \Delta_E \not\equiv 0 \pmod{\ell}$ which does happen when $p = 2$. Since N_E is square free this implies that $N(\rho) = 2$. But $S_2(\Gamma_1(2)) = 0$, which is the ultimate contradiction!

But how do we know semistable elliptic curves over $\mathbf{Q}$ are modular?

22.2 Modular elliptic curves

Theorem 22.2.1 (Theorem A)**.** *Every semistable elliptic curve over* $\mathbf{Q}$ *is modular.*

There are several ways to define a modular elliptic curve.

Definition 22.2.2. Let E be an elliptic curve. Then E is *modular* if there is a prime $\ell > 2$ such that the associated ℓ-adic Galois representation

$$\rho_{E,\ell^\infty} : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

defined by the ℓ-power division points on E is modular (i.e., it is a $\rho_{f,\lambda}$).

Definition 22.2.3. Let E be an elliptic curve of conductor N_E. For each prime p not dividing N_E one associates a number a_p related [[in a simple way!]] to the number of points on the reduction of E modulo p. Then E is *modular* if there exists a cusp form $f = \sum b_n q^n$ which is an eigenform for the Hecke operators such that $a_p = b_p$ for almost all p. In the end one deduces that f can be chosen to have weight 2, trivial character, and level N_E.

Definition 22.2.4 (Shimura)**.** An elliptic curve E is modular if there is nonconstant map $X_0(N) \to E$ for some N.

Theorem 22.2.5 (Theorem B. Wiles, Taylor-Wiles)**.** *Suppose E is a semistable elliptic curve over* $\mathbf{Q}$ *and suppose ℓ is an odd prime such that $E[\ell]$ is irreducible and modular. Then ρ_{E,ℓ^∞} is modular and hence E is modular.*

Now we sketch a proof that theorem B implies theorem A. First take $\ell = 3$. If $E[3]$ is irreducible then by work of Langlands-Tunnel we win. The idea is to take

$$\rho : G \to \mathrm{GL}(2, \mathbf{F}_3) \hookrightarrow \mathrm{GL}(2, \mathbf{Z}[\sqrt{-2}]) \subset \mathrm{GL}(2, \mathbf{C}).$$

The point is that there are two maps $\mathbf{Z}[\sqrt{-2}] \to \mathbf{F}_3$ given by reduction modulo each of the primes lying over 3. Choosing one gives a map $\mathrm{GL}(2, \mathbf{Z}[\sqrt{-2}]) \to \mathrm{GL}(2, \mathbf{F}_3)$ which, for some amazing reason [[which is?]], has a section. Then $\rho : G \to \mathrm{GL}(2, \mathbf{C})$ is a continuous representation with odd determinant that must still be irreducible. By Langlands-Tunnel we know that ρ is modular and in fact ρ comes from a weight 1 cusp form f of level a power of 3 times powers of most primes dividing $\mathrm{cond}(E)$. Reducing f modulo some prime of $\mathbf{Z}[\sqrt{-2}]$ lying over 3 we obtain a mod 3 modular form which corresponds to $\rho : G \to E[3]$. The proof of all this uses the immense base-change business in Langlands' book. [[Ribet next says: "have to get 3's out of the level! This jacks up the weight, and the level is still not square free. Then have to adjust the weight again." I do not know what the point of this is.]]

Kevin Buzzard asked a question relating to how one knows the hypothesis needed for the theorem on weights and levels applies in our situation. To answer this, suppose $\ell = 3$ or 5. Form the associated representation $\rho : G \to \mathrm{GL}(2, \mathbf{F}_\ell)$ coming from $E[\ell]$ and assume it is irreducible, modular and semistable.

Definition 22.2.6. A mod ℓ Galois representation is *semistable* if for all $p \neq \ell$, the inertia group I_p acts unipotently and the conjectured weight is 2 or $\ell + 1$.

Note that $\det \rho = \chi$.

Lemma 22.2.7. *Under the above assumptions, ℓ divides the order of the image of ρ.*

Proof. If not, then ρ is finite at all primes p, since for primes $p \neq \ell$ inertia acts trivially [[some other argument for ℓ]]. Inertia acts trivially because if the order of the image of ρ is prime to ℓ then ρ acts diagonally. For if not then since $\rho|I_p$ is unipotent (hence all eigenvalues are 1), in a suitable basis something like $\left(\begin{smallmatrix} 1 & \psi \\ 0 & 1 \end{smallmatrix}\right)$ is in the image of ρ and has order ℓ, a contradiction. Because of finiteness $k(\rho) = 2$ and $N(\rho) = 1$ which is a contradiction since there are no forms of weight 2 and level 1. □

Next we consider what happens if $E[3]$ is reducible. There are two cases to consider. First suppose $E[5]$ is also reducible. Then E contains a rational subgroup of order 15. We can check by hand that all such curves are modular. The key result is that $X_0(15)(\mathbf{Q})$ is finite.

The second possibility is that $E[5]$ is irreducible. In this case we first find a curve E' which is semistable over $\mathbf{Q}$ such that

- $E'[5] \cong E[5]$ (this is easy to do because of the lucky coincidence that $X_0(5)$ has genus 0)
- $E'[3]$ is irreducible

Next we discover that E' is modular since $E'[3]$ is irreducible. This implies $E'[5]$ is modular hence $E[5]$ is modular. Theorem B then implies E is modular.

23
Deformations

23.1 Introduction

For the rest of the semester let ℓ be an odd prime. Let $\rho : G \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$ be such that

- ρ is modular
- ρ is irreducible
- ρ is semistable

Definition 23.1.1. The representation ρ is *semistable* if

- for all $p \neq \ell$,
$$\rho|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix},$$
($*$ is typically trivial since most primes are unramified.)
- $k(\rho) = 2$ or $\ell + 1$.

This means that there are 2 possibilities.

1. ρ is finite at D_ℓ.
2. $$\rho|D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$
where β is unramified. (Since $\det(\rho) = \chi$ we can add that $\alpha|I_\ell = \chi$.)

If $k(\rho) = 2$ then possibility 1 occurs. If $k(\rho) = \ell + 1$ then we are in case 2, but being in case 2 does not imply that $k(\rho) = \ell + 1$. If ρ comes from an elliptic curve $E/\mathbf{Q}$, then case 1 occurs if E has good reduction at ℓ whereas case 2 occurs if

E has ordinary multiplicative reduction [[I am not sure about this last assertion because I missed it in class.]].

What is a deformation of ρ and when can we prove that it is modular?

Let A be a complete local Noetherian ring with maximal ideal $\mathfrak{m}$ and residue field $\mathbf{F}_{\ell^\nu}$ (so A is furnished with a map $A/\mathfrak{m} \xrightarrow{\sim} \mathbf{F}_{\ell^\nu}$). Let

$$\tilde{\rho}_A : G \to \mathrm{GL}(2, A)$$

be a representation which is ramified at only finitely many primes. Assume $\tilde{\rho} = \tilde{\rho}_A$ lifts ρ, i.e., the reduction of $\tilde{\rho}$ mod $\mathfrak{m}$ gives ρ.

Theorem 23.1.2. *Let the notation be as above. Then $\tilde{\rho}$ is modular if and only if it satisfies $(*)$.*

Neither of the terms in this theorem have been defined yet.

23.2 Condition $(*)$

Let

$$\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a Galois representation.

The statement we wish to understand is

> "All $\tilde{\rho}$ which satisfy $(*)$ are modular."

Let A be a complete local Noetherian ring with residue field $\mathbf{F}_{\ell^\nu}$. This means that we are given a map $A/\mathfrak{m} \cong \mathbf{F}_{\ell^\nu}$. Suppose $\tilde{\rho} : G \to \mathrm{GL}(2, A)$ satisfies the following conditions:

- $\tilde{\rho}$ lifts ρ,
- $\det \tilde{\rho} = \tilde{\chi}$,
- $\tilde{\rho}$ is ramified at only finitely many primes, and
- condition $(*)$.

What is condition $(*)$? It the requirement that $\tilde{\rho}$ have the same qualitative properties as ρ locally at ℓ. There are two cases to consider.

Case 1. (arising from supersingular reduction at ℓ) Suppose ρ is finite and flat at ℓ. Then $\rho | I_\ell$ is given by the 2 fundamental characters

$$I_\ell \to \mathbf{F}_{\ell^2}^*$$

of level 2 (instead of from powers of these characters because of the semistability assumption). Condition $(*)$ is that the lift of ρ is also constrained to be finite and flat. This means that for every $n \geq 1$,

$$\tilde{\rho} | D_\ell \mod \mathfrak{m}^n : D_\ell \to \mathrm{GL}(2, A/\mathfrak{m}^n)$$

is finite and flat, i.e., it comes from a finite flat group scheme over $\mathbf{Z}_\ell$ which is provided with an action of $A/\mathfrak{m}^n$ as endomorphisms.

Case 2. (bad multiplicative reduction at ℓ or good ordinary reduction at ℓ) In case 2

$$\rho|D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

where β is unramified (equivalently $\alpha|I_\ell = \chi$). [[In the case of good ordinary reduction β is given by the action of Galois on $E[\ell]$ in characteristic ℓ.]] Condition $(*)$ is the requirement that

$$\tilde{\rho}|D_\ell \cong \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix},$$

where $\tilde{\beta}$ is unramified, $\tilde{\alpha}|I_\ell = \tilde{\chi}$. It follows automatically that $\tilde{\beta}$ lifts β.

23.2.1 Finite flat representations

In general what does it mean for ρ to be finite and flat? It means that there exists a finite flat group scheme $\mathcal{G}$ over $\mathbf{Z}_\ell$ such that $\mathcal{G}(\overline{\mathbf{Q}}_\ell)$ is the representation space of ρ. This definition is *subtle*.

Coleman asked if there is a way to reformulate the definition without mentioning group schemes. Ribet mentioned Hopf algebras but then stopped. Buzzard suggested some of the subtlety of the definition by claiming that in some situation χ is finite flat whereas χ^2 is not. Ogus vaguely conjectured that Fontaine's language is the way to understand this.

It is possible in case 2 above for $\rho|D_\ell$ to be finite flat without $\tilde{\rho}$ finite flat. The quintessential example is an elliptic curve E with supersingular reduction at ℓ such that $\ell | \operatorname{ord}_\ell \Delta_E$.

23.3 Classes of liftings

Let Σ be a finite set of prime numbers. We characterize a class of liftings $\tilde{\rho}$ which depends on Σ. What does it mean for $\tilde{\rho}$ to be in the class of deformations corresponding to Σ?

23.3.1 The case $p \neq \ell$

First we talk about the case when $p \neq \ell$. If $p \in \Sigma$ then there is no special condition on $\tilde{\rho}|I_p$. If $p \notin \Sigma$ one requires that $\tilde{\rho}$ is qualitatively the same as ρ. This means

1. If ρ is unramified at p (which is the usual case), then we just require that $\tilde{\rho}$ is unramified at p.
2. If ρ is ramified but unipotent at p so

   $$\rho|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$$

   we require that

   $$\tilde{\rho}|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

   This situation can occur with an elliptic curve which has multiplicative reduction at p and for which $\ell \nmid \operatorname{ord}_p \Delta_E$.

At this point I mention the prime example.

Example 23.3.1. Suppose $\tilde{\rho} = \rho_{f,\lambda}$ is the λ-adic representation attached to f. What can we say about $\operatorname{ord}_p N(f)$? We know that

$$\operatorname{ord}_p N(f) = \operatorname{ord}_p N(\rho) + \dim(\rho)^{I_p} - \dim(\tilde{\rho})^{I_p}$$

where $(\rho)^{I_p}$ means the inertia invariants in the representation space of ρ. If ρ is semistable then

$$\operatorname{ord}_p N(\rho) + \dim(\rho)^{I_p} = 2.$$

Since we are assuming ρ is semistable, $\operatorname{ord}_p N(f) \leq 2$. Furthermore, the condition $p \notin \Sigma$ is a way of saying

$$\operatorname{ord}_p N(f) = \operatorname{ord}_p N(\rho).$$

Thus the requirement that $p \notin \Sigma$ is that the error term $\dim(\rho)^{I_p} - \dim(\tilde{\rho})^{I_p}$ vanish.

Note that $\operatorname{ord}_p N(\tilde{\rho}) = \operatorname{ord}_p N(f)$ by Carayol's theorem. Thus $\operatorname{ord}_p N(f)$ is just a different way to write $\operatorname{ord}_p N(\tilde{\rho})$.

Example 23.3.2. Imagine ρ is ramified at p and $\operatorname{ord}_p N(\rho) = 1$. Then $\operatorname{ord}_p N(f)$ is either 1 or 2. The requirement that $p \notin \Sigma$ is that $\operatorname{ord}_p N(f) = 1$.

23.3.2 The case $p = \ell$

Next we talk about the case $p = \ell$. There are two possibilities: either $\ell \in \Sigma$ or $\ell \notin \Sigma$. If $\ell \in \Sigma$ then we impose no *further* condition on $\tilde{\rho}$ (besides the already imposed condition $(*)$, semistability at ℓ, etc.). If $\ell \notin \Sigma$ and ρ if finite and flat (which is not always the case) then we require $\tilde{\rho}$ to be finite and flat. If $\ell \notin \Sigma$ and ρ is not finite flat then no further restriction (this is the Tate curve situation).

Suppose $\tilde{\rho} = \rho_{f,\lambda}$, and $\tilde{\rho}$ belongs to the class defined by Σ. We want to *guess* (since there is no Carayol theorem) an integer N_Σ such that $N(f) | N_\Sigma$. What is N_Σ? It will be

$$N_\Sigma = \prod_{\substack{p \neq \ell \\ p \in \Sigma}} p^2 \cdot \prod_{\substack{p \neq \ell \\ p \notin \Sigma}} p^{\operatorname{ord}_p N(\rho)} \cdot \ell^\delta.$$

Here $\operatorname{ord}_p N(\rho)$ is 1 if and only if ρ is ramified at p. The exponent δ is either 0 or 1. It is 1 if and only if $k(\rho) = \ell + 1$ or $\ell \in \Sigma$ and ρ is ordinary at ℓ, i.e.,

$$\rho | D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

[[This definition could be completely wrong. It was definitely not presented clearly in class.]]

There is an exercise associated with this. It is to justify *a priori* the definition of δ. Suppose, for example, that $\rho_{f,\lambda}$ satisfies $(*)$, then we want to show that $\ell^2 \nmid N(f)$.

Theorem 23.3.3. *Every $\tilde{\rho}$ of class Σ comes from $S_2(\Gamma_0(N_\Sigma))$.*

Define an approximation $\mathbf{T}$ to the Hecke algebra by letting

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \operatorname{End}(S_2(\Gamma_0(N_\Sigma)))$$

where we adjoin only those T_n for which $(n, \ell N_\Sigma) = 1$. For some reason there exists a map

$$\mathbf{T} \to \mathbf{F}_{\ell^\nu} : \quad T_r \mapsto \operatorname{tr} \rho(\mathrm{Frob}_r).$$

Why should such a map exist? The point is that we know by the theorem that ρ comes by reduction from a $\rho_{f,\lambda}$ with $f \in S_2(\Gamma_0(N_\Sigma))$.

But there is a wrinkle. [[I do not understand: He says: "Clearly $N(\rho)|N_\Sigma$. $k(\rho) = 2$ or $\ell + 1$. If $k(\rho) = \ell + 1$ then $\delta = 1$ so $\ell|N_\Sigma$. Have to slip over to weight 2 in order to get f." This does not make any sense to me.]]

23.4 Wiles's Hecke algebra

Composing the map $\mathbf{T} \to \mathcal{O}_{E_f}$ with reduction mod λ from $\mathcal{O}_{E_f}$ to $\mathbf{F}_{\ell^\nu}$ we obtain a map $\mathbf{T} \to \mathbf{F}_{\ell^\nu}$. Let $\mathfrak{m}$ be the kernel. Then $\mathfrak{m}$ is a maximal ideal of $\mathbf{T}$. Wiles's Hecke algebra is the completion $\mathbf{T}_\mathfrak{m}$ of $\mathbf{T}$ at $\mathfrak{m}$. [[For some reason]] there exists

$$\tilde{\rho} : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{T}_\mathfrak{m})$$

such that $\operatorname{tr}(\tilde{\rho}(\mathrm{Frob}_r)) = T_r$. What makes this useful is that $\tilde{\rho}$ is *universal* for lifts of type Σ. This means that given any lift τ of type Σ there exists a map

$$\varphi : \mathrm{GL}(2, \mathbf{T}_\mathfrak{m}) \to \mathrm{GL}(2, A)$$

such that $\tau = \varphi\tilde{\rho}$.

Another key idea is that the approximation $\mathbf{T}$ obtained by just adjoining those T_n with $(n, \ell N_\Sigma) = 1$ is, after completing at certain primes, actually equal to the whole Hecke algebra.

24

The Hecke Algebra T_Σ

24.1 The Hecke algebra

Throughout this lecture $\ell \neq p$ and $\ell \geq 3$. We are studying the representation $\rho : G \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$. This is an irreducible representation, ℓ is odd, ρ is semistable, and $\det \rho = \chi$. To single out certain classes of liftings we let Σ be a finite set of primes. Let A be a complete local Noetherian ring with residue filed $\mathbf{F}_{\ell^\nu}$. We take liftings $\tilde{\rho} : G \to \mathrm{GL}(2, A)$ such that $\tilde{\rho}$ reduces down to ρ, $\det \tilde{\rho} = \tilde{\chi}$, and $\tilde{\rho}$ is "like" ρ away from Σ. For example, if ρ is unramified at p we also want $\tilde{\rho}$ unramified at p, etc.

Assume $\tilde{\rho}$ is modular. We guess the serious divisibility condition that $N(f)|N_\Sigma$. Recall that

$$N_\Sigma = \prod_{\substack{p \neq \ell \\ p \in \Sigma}} p^2 \cdot \prod_{\substack{p \neq \ell \\ p \notin \Sigma}} p^{\mathrm{ord}_p N(\rho)} \cdot \ell^\delta.$$

To define δ consider two cases.

- *level 1 case.* Take $\delta = 1$ if $\ell \in \Sigma$ or if ρ is not finite at ℓ. Take $\delta = 0$ otherwise.
- *level 2 case.* This is the case when $\rho | I_\ell$ has order $\ell^2 - 1$. Take $\delta = 0$.

A priori nobody seems to know how to prove that $N(f)|N_\Sigma$ given only that $\tilde{\rho}$ is modular. In the end we will show that all modular $\tilde{\rho}$ in fact come from

$$S_2(\Gamma_0(N_\Sigma)).$$

This can be regarded as a proof that $N(\rho)|N_\Sigma$.

Last time we tried to get things going by defining the *anemic Hecke algebra*

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N_\Sigma)))$$

where we only adjoin those T_n for which $(n, \ell N_\Sigma) = 1$. By some level lowering theorem there exists an $f \in S_w(\Gamma_0(N_\Sigma))$ giving $\tilde{\rho}$ so we obtain a map $\mathbf{T} \to \mathbf{F}$. The

map sends T_n to the reduction modulo λ of its eigenvalue on f. (λ is a prime of the ring of integers of E_f lying over ℓ.) [[... something about needing an f of the right level $= N(\rho)$.]]

Let $\mathfrak{m} \subset \mathbf{T}$ be the kernel of the above defined map $\mathbf{T} \to \mathbf{F}$. Then $\mathfrak{m}$ is a maximal ideal. Let $\mathbf{T}_\mathfrak{m}$ be the completion of $\mathbf{T}$ at $\mathfrak{m}$ and note that

$$\mathbf{T}_\mathfrak{m} \hookrightarrow \mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Z}_\ell.$$

We need to know that $\mathbf{T}_\mathfrak{m}$ is Gorenstein. This is done by comparing $\mathbf{T}_\mathfrak{m}$ to some full Hecke ring. Thus

$$\mathbf{T} \subset R = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N_\Sigma)))$$

where the full Hecke ring R is obtained by adjoining the T_n for *all* integers n. We should think of R as

$$R = \mathbf{T}[T_\ell, \{U_P : p | N_\Sigma\}].$$

Note that T_ℓ may or may not be a U_ℓ depending on if $\ell | N_\Sigma$.

Lemma 24.1.1. *If $\ell \nmid N_\Sigma$ then $R = \mathbf{T}[U_p, \ldots]$. Thus if T_ℓ is not a U_ℓ then we do not need T_ℓ.*

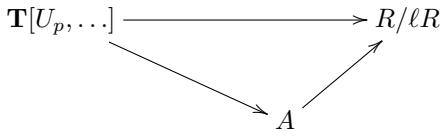
Proof. "This lemma is an interesting thing and the proof goes as follows. Oooh. Sorry, this is not true. Ummm.... hmm."

The ring $\mathbf{T}[U_p, \ldots]$ is clearly of finite index in R since: if q is a random prime number consider $R \otimes_{\mathbf{Z}} \mathbf{Q}_q$ compared to $\mathbf{T}[U_p, \ldots] \otimes_{\mathbf{Z}} \mathbf{Q}_q$. [[I do not know how to do this argument. The lemma as stated above probably isn't really true. The point is that the following lemma is the one we need and it is true.]] □

Let $\mathbf{T}[U_p, \ldots]$ be the ring obtained by adjoining to $\mathbf{T}$ just those U_p with $p | N_\Sigma$.

Lemma 24.1.2. *If $\ell \nmid N_\Sigma$ then the index $(R : \mathbf{T}[U_p, \ldots])$ is prime to ℓ. Note that we assume $\ell \geq 3$.*

Proof. We must show that the map $\mathbf{T}[U_p, \ldots] \to R/\ell R$ is surjective. [[I thought about it for a minute and did not see why this suffices. Am I being stupid?]] Let $A = \mathbf{F}_\ell[T_n : (n, \ell) = 1]$ be the image of $\mathbf{T}[U_p, \ldots]$ in $R/\ell R$ so we have a diagram



We must show that $A = R/\ell R$. There is a beautiful duality

$$R/\ell R \times S_2(\Gamma_0(N_\Sigma); \mathbf{F}_\ell) \longrightarrow \mathbf{F}_\ell \quad \text{(perfect pairing)}.$$

Thus $A^\perp = 0$ if and only if $A = R/\ell R$.

Suppose $0 \neq f \in A^\perp$, then $a_n(f) = 0$ for all n such that $(n, \ell) = 1$. Thus $f = \sum a_{n\ell} q^{n\ell}$. Let $\theta = q\frac{d}{dq}$ be the theta operator. Since the characteristic is ℓ, $\theta(f) = 0$. On the other hand $w(f) = 2$ and since $\ell \geq 3$, $\ell \nmid 2$. Thus $w(\theta f) = w(f) + \ell + 1 = 2 + \ell + 1$ which is a contradiction since $\theta f = 0$ and $w(0) = 0 \neq 3 + \ell$. [[The weight is an integer *not* a number mod ℓ, right?]] □

Example 24.1.3. The lemma only applies if $\ell \geq 3$. Suppose $\ell = 2$ and consider $S_2(\Gamma_0(23))$. Then

$$\mathbf{T}[U_p, \ldots] = \mathbf{Z}[\sqrt{5}] \subset R = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$$

so $(R : \mathbf{T}[U_p, \ldots])$ is not prime to 2.

Remark 24.1.4. If $N = N_\Sigma$ then

$$\operatorname{rank}_{\mathbf{Z}} \mathbf{T} = \sum_{M|N} S_2(\Gamma_0(M))^{\text{new}}$$

and

$$\operatorname{rank}_{\mathbf{Z}} R = \dim S_2(\Gamma_0(N)).$$

There is an injection

$$\bigoplus_{M|N} S_2(\Gamma_0(M))^{\text{new}} \hookrightarrow S_2(\Gamma_0(N))$$

but $\oplus S_2(\Gamma_0(M))^{\text{new}}$ is typically much smaller than $S_2(\Gamma_0(N))$.

24.2 The Maximal ideal in R

The plan is to find a special maximal ideal $\mathfrak{m}_R$ of R lying over $\mathfrak{m}$.

$$\begin{array}{ccc} \mathfrak{m}_R & \text{——} & R \\ | & & | \\ \mathfrak{m} & \text{——} & \mathbf{T} \end{array}$$

Once we finally find the correct $\mathfrak{m}$ and $\mathfrak{m}_R$ we will be able to show that the map $\mathbf{T}_{\mathfrak{m}} \to R_{\mathfrak{m}_R}$ is an isomorphism. In finding $\mathfrak{m}_R$ we will not invoke some abstract going up theorem but we will "produce" $\mathfrak{m}_R$ by some other process. The ideal $\mathfrak{m}$ was defined by a newform f level $M|N_\Sigma$. The coefficients of f lie in

$$\mathcal{O}_\lambda = (\mathcal{O}_{E_f})_\lambda$$

where E_f is the coefficient ring of f and λ is a prime lying over ℓ. Thus $\mathcal{O}_\lambda$ is an ℓ-adic integer ring. Composing the residue class map $\mathcal{O}_\lambda \to \overline{\mathbf{F}}$ with the eigenvalue map $\mathbf{T} \to \mathcal{O}_\lambda$ we obtain the map $\mathbf{T} \to \overline{\mathbf{F}}$.

In order to obtain the correct $\mathfrak{m}_R$ we will make a sequence of changes to f to make some good newform. [[Is the motivation for all this that the lemma will not apply if $\ell|N_\Sigma$?]]

24.2.1 Strip away certain Euler factors

Write $f = \sum a_n q^n$. Replace f by

$$h = \sum_{\text{certain } n} a_n q^n$$

where the sum is over those n which are prime to each $p \in \Sigma$. What does this mean? If we think about the L-function $L(f,s) = \prod_p L_p(f,s)$, then h has L-function

$$L(h,s) = \prod_{p \notin \Sigma} L_p(f,s).$$

Furthermore making this change does not take us out of $S_2(\Gamma_0(N_\Sigma))$, i.e., $h \in S_2(\Gamma_0(N_\Sigma))$. [[He explained why but my notes are very incomplete. They say: Why. Because $h = (f \otimes \varepsilon) \otimes \varepsilon$, ($\varepsilon$ Dirichlet character ramified at primes in N_Σ). Get a form of level $\mathrm{lcm}(\prod_{p\in\Sigma} p^2, N(f))|N_\Sigma$. Can strip and stay in space since N_Σ has correct squares built into it.]]

Remark 24.2.1. Suppose that $p||N_\Sigma$. Then $p \notin \Sigma$ and $p||N(\rho)$. The level of f is

$$N(f) = \begin{cases} N(\rho), & \text{if } k(\rho) = 2 \\ N(\rho)\ell, & \text{if } k(\rho) = \ell + 1 \end{cases}.$$

If $p||N(\rho)$ then $f|T_p = a_p(f)f$. Thus h is already an eigenform for T_p unless $p \in \Sigma$ in which the eigenvalue is 0.

[[I do not understand this remark. Why would the eigenvalue being 0 mean that T_p is not an eigenform? Furthermore, what is the point of this remark in the wider context of transforming f into a good newform.]]

24.2.2 Make into an eigenform for U_ℓ

We perform this operation to f to obtain a form g then apply the above operation to get the ultimate h having the desired properties. Do this if $\ell|N_\Sigma$ but $\ell \nmid N(f)$. This happens precisely if $\ell \in \Sigma$ and ρ is good and ordinary at ℓ. Then $f|U_\ell$ is just some random junk. Consider $g = f + *f(q^\ell)$ where $*$ is some coefficient. We see that if $*$ is chosen correctly then $g|U_\ell = Cg$ for some constant C. There are 2 possible choices for $*$ which lead to 2 choices for C. Let $a_\ell = a_\ell(f)$, then C can be either root of

$$X^2 + a_\ell X + \ell = 0.$$

This equation has exactly one unit root in $\mathcal{O}_\lambda$. The reason is because we are in the ordinary situation so a_ℓ is a unit. But ℓ is not a unit. The sum of the roots is a unit but the product is not. [[Even this is not clear to me right now. Definitely check this later.]] Make the choice of $*$ so that real root is C. Then we obtain a g such that

$$g|U_\ell = (\text{unit}) \cdot g.$$

Next apply the above procedure to strip g and end up with an h such that

- $h|U_\ell = (\text{unit}) \cdot h$,
- $h|U_p = 0$ for $p \in \Sigma$ ($p \neq \ell$), and
- $h|T_p = a_p(f) \cdot h$ for $p \notin \Sigma$.

Now take the form h. It gives a map $R \to \mathcal{O}_\lambda$ which extends $\mathbf{T} \to \mathcal{O}_\lambda$. Let $\mathfrak{m}_R$ be the kernel of the map $R \to \overline{\mathbf{F}}$ obtained by composing $R \to \mathcal{O}_\lambda$ with $\mathcal{O}_\lambda \to \overline{\mathbf{F}}$. A lot of further analysis shows that $\mathbf{T}_\mathfrak{m} \to R_{\mathfrak{m}_r}$ is an isomorphism. We end up having to show separately that the map is injective and surjective.

[[In this whole lecture $p \neq \ell$.]]

[[Wiles's notation: His $\mathbf{T}_{\mathfrak{m}_R}$ is my $\mathbf{R}$ and his $\mathbf{T}'$ is my $\mathbf{T}_\mathfrak{m}$.

24.3 The Galois representation

We started with a representation ρ, chose a finite set of primes Σ and then made the completed Hecke algebra $\mathbf{T}_{\mathfrak{m}}$. Our goal is to construct the universal deformation of ρ of type Σ. The universal deformation is a representation

$$\tilde{\rho} : G \to \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}})$$

such that for all primes p with $p \nmid \ell N_\Sigma$, $\tilde{\rho}(\mathrm{Frob}_p)$ has trace $T_p \in \mathbf{T}_{\mathfrak{m}}$ and determinant p.

We now proceed with the construction of $\tilde{\rho}$. Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots], \quad (n, \ell N_\Sigma) = 1$$

be the anemic Hecke algebra. Then $\mathbf{T} \otimes \mathbf{Q}$ decomposes as a product of fields

$$\mathbf{T} \otimes \mathbf{Q} = \prod_f E_f$$

where the product is over a set of representatives for the Galois conjugacy classes of newforms of weight 2, trivial character, and level dividing N_Σ. Since $\mathbf{T}$ is integral (it is for example a finite rank $\mathbf{Z}$-module), $\mathbf{T} \hookrightarrow \prod_f \mathcal{O}_f$. Since $\mathbf{Z}_\ell$ is a flat $\mathbf{Z}$-module,

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod_f \mathcal{O}_f \otimes \mathbf{Z}_\ell = \prod_{f,\lambda} \mathcal{O}_{f,\lambda}$$

where the product is over a set of representatives f and all $\lambda | \ell$.

$\mathbf{T}_{\mathfrak{m}}$ is a direct factor of $\mathbf{T} \otimes \mathbf{Z}_\ell$. [[This is definitely not the assertion that $\mathbf{T}_{\mathfrak{m}}$ is an $\mathcal{O}_{f,\lambda}$. What exactly is it the assertion of really?]]

We can restrict the product to a certain finite set S and still obtain an injection

$$\mathbf{T}_{\mathfrak{m}} \hookrightarrow \prod_{(f,\lambda) \in S} \mathcal{O}_{f,\lambda}.$$

The finite set S consists of those (f, λ) such that the prime λ of $\mathcal{O}_f$ pulls back to $\mathfrak{m}$ under the map $\mathbf{T} \to \mathcal{O}_f$ obtained by composing $\mathbf{T} \to \prod_f \mathcal{O}_f$ with the projection onto $\mathcal{O}_f$. [[Why is this enough so that $\mathbf{T}_{\mathfrak{m}}$ still injects in?]] Restricting to a finite product is needed so that

$$[\prod_{(f,\lambda) \in S} \mathcal{O}_{f,\lambda} : \mathbf{T}_{\mathfrak{m}}] < \infty.$$

Given f and λ there exists a representation

$$\rho_{f,\lambda} : G \to \mathrm{GL}(2, \mathcal{O}_{f,\lambda}).$$

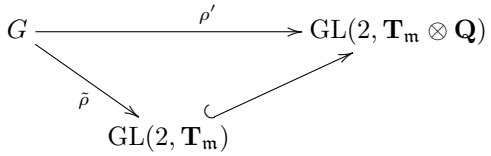
It is such that $\mathrm{tr}\, \rho_{f,\lambda}(\mathrm{Frob}_p) = a_p$ is the image of T_p under the inclusion

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod_{(f,\lambda) \in S} \mathcal{O}_{f,\lambda}.$$

Put some of these $\rho_{f,\lambda}$ together to create a new representation

$$\prod_{(f,\lambda) \in S} \rho_{f,\lambda} : G \xrightarrow{\quad \rho' \quad} \mathrm{GL}(2, \prod \mathcal{O}_{f,\lambda}) \subset \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}} \otimes \mathbf{Q}).$$

The sought after universal deformation $\tilde{\rho}$ is a map making the following diagram commute



Theorem 24.3.1. *ρ' is equivalent to a representation taking values in* $\mathrm{GL}(2, \mathbf{T}_\mathfrak{m})$.

One way to [[try to]] prove this theorem is by invoking a general theorem of Carayol. [[and then what? does this way work? why is it not a good way?]] But the right way to prove the theorem is Wiles's way.

24.3.1 The Structure of $\mathbf{T}_\mathfrak{m}$

Just as an aside let us review the structure of $\mathbf{T}_\mathfrak{m}$.

- $\mathbf{T}_\mathfrak{m}$ is local.
- $\mathbf{T}_\mathfrak{m}$ is not necessarily a discrete valuation ring.
- $\mathbf{T}_\mathfrak{m} \otimes \mathbf{Q}$ is a product of finite extensions of $\mathbf{Q}_\ell$.
- $\mathbf{T}_\mathfrak{m}$ is not necessarily a product of rings $\mathcal{O}_{f,\lambda}$.
- $\mathbf{T}_\mathfrak{m}$ need not be integral.

24.3.2 The Philosophy in this picture

Choose c to be a complex conjugation in $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Since ℓ is odd $\det(c) = -1$ is a very strong condition which rigidifies the situation.

24.3.3 Massage ρ

Choose two 1-dimensional subspaces so that

$$\rho(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

with respect to any basis consisting of one vector from each subspace. For any $\sigma \in G$ write

$$\rho(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}.$$

Then $a_\sigma d_\sigma$ and $b_\sigma c_\sigma$ are somehow intrinsically defined. This is because

$$\rho(\sigma c) = \begin{pmatrix} -a_\sigma & ? \\ ? & d_\sigma \end{pmatrix}$$

so

$$a_\sigma = \frac{\mathrm{tr}(\rho(\sigma)) - \mathrm{tr}(\rho(\sigma c))}{2}$$

and

$$d_\sigma = \frac{\mathrm{tr}(\rho(\sigma)) + \mathrm{tr}(\rho(\sigma c))}{2}.$$

Since we know the determinant it follows that $b_\sigma c_\sigma$ is also intrinsically known. [[The point is that we know certain things about these matrices in terms of their traces and determinants.]]

Proposition 24.3.2. *There exists $g \in G$ such that $b_g c_g \neq 0$.*

Proof. Since ρ is irreducible there exists σ_1 such that $b_{\sigma_1} \neq 0$ and there exists σ_2 such that $c_{\sigma_2} \neq 0$. If $b_{\sigma_2} \neq 0$ or $c_{\sigma_1} \neq 0$ then we are done. So the only problem case is when $b_{\sigma_1} = 0$ and $c_{\sigma_2} = 0$. Easy linear algebra shows that in this situation $g = \sigma_1\sigma_2$ has the required property. □

Now rigidify by choosing a basis so that $b_g = 1$. Doing this does not fix a basis because they are many ways to choose such a basis.

24.3.4 Massage ρ'

Choose a basis of $(\mathbf{T}_\mathfrak{m} \otimes \mathbf{Q})^2$ so that

$$\rho'(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For any $\sigma \in G$ write

$$\rho'(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}$$

Using an argument as above shows that $a_\sigma, d_\sigma \in \mathbf{T}_\mathfrak{m}$ since the traces live in $\mathbf{T}_\mathfrak{m}$. Furthermore $b_\sigma c_\sigma \in \mathbf{T}_\mathfrak{m}$ since the determinant is in $\mathbf{T}_\mathfrak{m}$.

The *key observation* is that $b_\sigma c_\sigma$ reduces mod $\mathfrak{m}$ to give the previous $b_\sigma c_\sigma \in \mathbf{F}$ corresponding to $\rho(\sigma)$. This is because the determinants and traces of ρ' are lifts of the ones from ρ. Since $\mathfrak{m}$ is the maximal ideal of a local ring and $b_g c_g$ reduces mod $\mathfrak{m}$ to something nonzero it follows that $b_g c_g$ is a unit in $\mathbf{T}_\mathfrak{m}$.

Choose a basis so that

$$\rho'(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and also so that

$$\rho'(g) = \begin{pmatrix} a_g & 1 \\ u & d_g \end{pmatrix} \in \mathrm{GL}(2, \mathbf{T}_\mathfrak{m}).$$

Here u is a unit in $\mathbf{T}_\mathfrak{m}$.

Proposition 24.3.3. *Write*

$$\rho'(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}$$

with respect to the basis chosen above. Then $a_\sigma, b_\sigma, c_\sigma, d_\sigma \in \mathbf{T}_\mathfrak{m}$.

Proof. We already know that $a_\sigma, d_\sigma \in \mathbf{T}_\mathfrak{m}$. The question is how to show that $b_\sigma, c_\sigma \in \mathbf{T}_\mathfrak{m}$. Since

$$\rho'(\sigma g) = \begin{pmatrix} a_\sigma a_g + b_\sigma u & ? \\ ? & c_\sigma + d_\sigma d_g \end{pmatrix}$$

we see that $a_\sigma a_g + b_\sigma u \in \mathbf{T}_\mathfrak{m}$. Since $a_\sigma a_g \in \mathbf{T}_\mathfrak{m}$ it follows that $b_\sigma u \in \mathbf{T}_\mathfrak{m}$. Since u is a unit in $\mathbf{T}_\mathfrak{m}$ this implies $b_\sigma \in \mathbf{T}_\mathfrak{m}$. Similarly $c_\sigma + d_\sigma d_g \in \mathbf{T}_\mathfrak{m}$ so $c_\sigma \in \mathbf{T}_\mathfrak{m}$. □

As you now see, in this situation we can prove Carayol's theorem with just some matrix computations. [[This is basically a field lowering representation theorem. The thing that makes it easy is that there exists something (namely c) with distinct eigenvalues which is rational over the residue field. Schur's paper, models over smaller fields. "Schur's method".]]

24.3.5 Representations from modular forms mod ℓ

If you remember back in the 70's people would take an $f \in S_2(\Gamma_0(N); \mathbf{F})$ which is an eigenform for almost all the Hecke operators

$$T_p f = c_p f \quad \text{for almost all } p, \text{ and } c_p \in \mathbf{F}.$$

The question is then: Can you find

$$\rho : G \to \mathrm{GL}(2, \mathbf{F})$$

such that

$$\mathrm{tr}(\rho(\mathrm{Frob}_p)) = c_p \text{ and } \det(\rho(\mathrm{Frob}_p)) = p$$

for all but finitely many p? The answer is yes. The idea is to find ρ by taking $\rho_{f,\lambda}$ [[which was constructed by Shimura?]] for some f and reducing mod λ. The only special thing that we need is a lemma saying that the eigenvalues in characteristic ℓ lift to eigenvalues in characteristic 0.

24.3.6 Representations from modular forms mod ℓ^n

Serre and Deligne asked: "What happens mod ℓ^n?"

More precisely, let R be a local finite Artin ring such that $\ell^n R = 0$ for some n. Take $f \in S_2(\Gamma_0(N); R)$ satisfying the hypothesis

$$\{r \in R : rf = 0\} = \{0\}.$$

This is done to insure that certain eigenvalues are unique. Assume that for almost all p one has $T_p f = c_p f$ with $c_p \in R$. The problem is to find $\rho : G \to \mathrm{GL}(2, R)$ such that

$$\mathrm{tr}(\rho(\mathrm{Frob}_p)) = c_p \text{ and } \det(\rho(\mathrm{Frob}_p)) = p$$

for almost all p.

The big stumbling block is that ρ need not be the reduction of some $\rho_{g,\lambda}$ for any g, λ. [[I couldn't understand why – I wrote "can mix up f's from characteristic 0 so can not get one which reduces correctly."]]

Let $\mathbf{T} = \mathbf{Z}[\ldots, T_p, \ldots]$ where we only adjoin those T_p for which f is an eigenvector [[I made this last part up, but it seems very reasonable]]. Then f obviously gives a rise to a map

$$\mathbf{T} \to R : T \mapsto \text{ eigenvalue of } T \text{ on } f \ .$$

The strange hypothesis on f insures that the eigenvalue is unique. Indeed, suppose $Tf = af$ and $Tf = bf$, then $0 = Tf - Tf = af - bf = (a-b)f$ so by the hypothesis $a - b = 0$ so $a = b$.

Since the pullback of the maximal ideal of R is a maximal ideal of $\mathbf{T}$ we get a map $\mathbf{T}_{\mathfrak{m}} \to R$ for some $\mathfrak{m}$. [[I do not understand why we suddenly get this map and I do not know why the pullback of the maximal ideal is maximal.]]

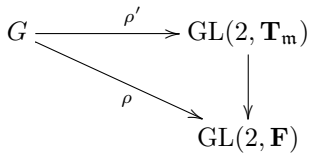
Now the problem is solved. Take $\rho' : G \to \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}})$ with the sought after trace and determinant properties. Then let ρ be the map obtained by composing with the map $\mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}}) \to \mathrm{GL}(2, R)$.

24.4 ρ' is of type Σ

Let ρ be modular irreducible and semistable mod ℓ representation with $\ell > 2$. Let Σ be a finite set of primes. Then $N(\rho)|N_\Sigma$. We constructed the anemic Hecke algebra $\mathbf{T}$ which contains a certain maximal ideal $\mathfrak{m}$. We then consider the completion $\mathbf{T}_{\mathfrak{m}}$ of $\mathbf{T}$ at $\mathfrak{m}$. Next we constructed

$$\rho' : G \to \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}})$$

lifting ρ. Thus the diagram



commutes.

Some defining properties of ρ' are

- $\det \rho' = \tilde{\chi}$.
- $\mathrm{tr}\, \rho'(\mathrm{Frob}_r) = T_r$. Since topologically $\mathbf{T}_{\mathfrak{m}}$ is generated by the Frob_r this is a tight condition.
- ρ' is a lift of type Σ.

To say ρ' is a lift of type Σ entails that ρ' is unramified outside primes $p|N_\Sigma$. This is true because ρ' is constructed from various $\rho_{f,\lambda}$ with $N(f)|N_\Sigma$. If $p \neq \ell$ and $p|N(\rho)$ then $p||N_\Sigma$. Recall that

$$\rho|D_p \sim \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

where $\alpha = \beta\chi$ and α and β are unramified. For ρ' to be a lift of type Σ we require that

$$\rho'|D_p \sim \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are unramified lifts and $\tilde{\alpha} = \tilde{\beta}\chi$. [[I find it mighty odd that α is χ times an unramified character and yet α is not ramified! How can that be?

Restricted to inertia β and α would be trivial but χ would not be.]] Is this true of ρ'? Yes since by a theorem of Langlands the factors

$$\rho_{f,\lambda}|D_p \sim \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix}.$$

[[Ribet said more about this but it does not form a cohesive whole in my mind. Here is what I have got. Since $\rho_{f,\lambda}$ obviously ramified at p, $p||N(f)|N_\Sigma$. $\rho_{f,\lambda}|D_p$ is like an elliptic curve with bad multiplicative reduction at p. That $\rho_{f,\lambda}|D_p \sim \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \tilde{\alpha} * 0 \tilde{\beta}$ really comes down to Deligne-Rapaport. If write $f = \sum a_n q^n$, then $a_p \neq 0$ and $\tilde{\beta}(\mathrm{Frob}_p) = a_p$, $\tilde{\alpha}(\mathrm{Frob}_p) = p a_p$. Thus $a_p^2 = 1$ since $\tilde{\alpha}\tilde{\beta} = \chi$. Thus $a_p = \pm 1$ and $a_p \mod \lambda = \beta(\mathrm{Frob}_p) = \pm 1$ independent of (f, λ). So we have these numbers $a_p = a_p(f) = \pm 1$, independent of λ.]]

24.5 Isomorphism between $\mathbf{T}_\mathfrak{m}$ and $R_{\mathfrak{m}_R}$

Let $\mathbf{T} \subset R = \mathbf{Z}[\ldots, T_n, \ldots]$ be the anemic Hecke algebra with maximal ideal $\mathfrak{m}$. The difference between $\mathbf{T}$ and R is that R contains all the Hecke operators whereas $\mathbf{T}$ only contains the T_p with $p \nmid \ell N_\Sigma$. Wiles proved that the map $\mathbf{T}_\mathfrak{m} \to R_{\mathfrak{m}_R}$ is an isomorphism. Which Hecke operators are going to hit the missing T_p? If we do the analysis in $R_{\mathfrak{m}_R}$ we see that [[I think for $p \neq \ell$!]]

$$T_p = \begin{cases} \pm 1, & \text{for } p|N_\Sigma,\ p \notin \Sigma \\ 0, & \text{for } p \in \Sigma \end{cases}.$$

This takes care of everything except T_ℓ. In proving the surjectivity of $\mathbf{T}_\mathfrak{m} \to R_{\mathfrak{m}_R}$ we are quite happy to know that $T_p = \pm 1$ or 0. The nontrivial proof is given in [DDT94].

Consider the commuting diagram

$$\begin{array}{ccc} \mathbf{T}_\mathfrak{m} & \hookrightarrow & \prod \mathcal{O}_{f,\lambda} \\ & \searrow & \uparrow \\ & & R_{\mathfrak{m}_R} \end{array}$$

The map $R_{\mathfrak{m}_R} \to \prod \mathcal{O}_{f,\lambda}$ is constructed by massaging f by stripping away certain Euler factors so as to obtain an eigenvector for all the Hecke operators. This diagram forces $\mathbf{T}_\mathfrak{m} \to R_{\mathfrak{m}_R}$ to be injective.

[[Ogus: Is it clearly surjective on the residue field? Ribet: Yes. Ogus: OK, then we just need to prove it is ètale.]]

From the theory of the θ operator we already know two-thirds of the times that $\mathbf{T}$ contains T_ℓ.

Suppose $\ell|N_\Sigma$. This entails that we are in the ordinary case, ρ is not finite at ℓ, or $\ell \in \Sigma$. We did not prove in this situation that $T_\ell \in \mathbf{Z}[\ldots, T_n, \ldots : (n, \ell) = 1]$.

Using generators and relations and brute force one shows that $R_{\mathfrak{m}_R} \to \prod \mathcal{O}_{f,\lambda}$ is an injection. Then we can compare everything in $\prod \mathcal{O}_{f,\lambda}$. Now

$$\rho_{f,\lambda}|D_\ell \sim \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix}$$

and

$$\tilde{\beta}(\mathrm{Frob}_\ell) = T_\ell \in \prod \mathcal{O}_{f,\lambda}.$$

Using arguments like last time one shows that $\tilde{\beta}(\mathrm{Frob}_\ell)$ can be expressed in terms of the traces of various operators. This proves surjectivity in this case.

Ultimately we have $\mathbf{T}_\mathfrak{m} \cong R_{\mathfrak{m}_R}$. The virtue of $\mathbf{T}_\mathfrak{m}$ is that it is generated by traces. The virtue of $R_{\mathfrak{m}_R}$ is that it is Gorenstein. We have seen this if $\ell \nmid N_\Sigma$. In fact it is Gorenstein even if $\ell || N_\Sigma$. [[Ribet: As I stand here today I do not know how to prove this last assertion in exactly one case. Ogus: You mean there is another gap in Wiles's proof. Ribet: No, it is just something I need to work out.]] When $\ell || N_\Sigma$ there are 2 cases. Either ρ is not finite at ℓ or it is. The case when ρ is not finite at ℓ was taken care of in [MR91]. A proof that $R_{\mathfrak{m}_R}$ is Gorenstein when ρ is finite at ℓ ($\ell \in \Sigma$) is not in the literature.

Now forget $R_{\mathfrak{m}_R}$ and just think of $\mathbf{T}_\mathfrak{m}$ in both ways: trace generated and Gorenstein.

24.6 Deformations

Fix an absolutely irreducible modular mod ℓ representation ρ and a finite set of primes Σ. Consider the category $\mathcal{C}$ of complete local Noetherian $W(\mathbf{F})$-algebras A (with $A/\mathfrak{m} = \mathbf{F}$). Here $\mathbf{F} = \mathbf{F}_{\ell^\nu}$ and $W(\mathbf{F})$ is the ring of Witt vectors, i.e., the ring of integers of an unramified extension of $\mathbf{Q}_\ell$ of degree ν.

Define a functor $\mathcal{F} : \mathcal{C} \to \mathrm{Set}$ by sending A in $\mathcal{C}$ to the set of equivalence classes of lifts

$$\tilde{\rho} : G \to \mathrm{GL}(2, A)$$

of ρ of type Σ. The equivalence relation is that $\tilde{\rho_1} \sim \tilde{\rho_2}$ if and only if there exists $M \in \mathrm{GL}(2, A)$ with $M \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{\mathfrak{m}}$ such that $\tilde{\rho_1} = M^{-1}\tilde{\rho_2}M$.

Mazur proved [Maz89] that $\mathcal{F}$ is representable.

Theorem 24.6.1. *There exists a lift*

$$\rho_{univ} : G \to \mathrm{GL}(2, R_\Sigma)$$

of type Σ such that given any lift $\tilde{\rho} : G \to \mathrm{GL}(2, A)$ there exists a unique homomorphism $\varphi : R_\Sigma \to A$ such that $\varphi \circ \rho_{univ} \sim \tilde{\rho}$ in the sense of the above equivalence relation.

Lenstra figured out how to concretely construct R_Σ.

Back in the student days of Ribet and Ogus, Schlessinger wrote a widely quoted thesis which gives conditions under which a certain class of functors can be representable. Mazur checks these conditions in his paper.

[[Buzzard: What happened to Schlessinger anyways? Ribet: He ended up at University of North Carolina, Chapel Hill.]]

We have constructed $\tilde{\rho} = \rho' : G \to \mathrm{GL}(2, \mathbf{T}_\mathfrak{m})$. By the theorem there exists a unique morphism $\varphi : R_\Sigma \to \mathbf{T}_\mathfrak{m}$ such that $\rho' = \varphi \circ \rho_{\mathrm{univ}}$.

Theorem 24.6.2. *φ is an isomorphism thus ρ' is the universal deformation and $\mathbf{T}_\mathfrak{m}$ is the universal deformation ring.*

This will imply that any lift of type Σ is modular.

The morphism φ is surjective since

$$T_p = \operatorname{tr} \tilde{\rho}(\mathrm{Frob}_p) = \operatorname{tr} \varphi \circ \rho_{\mathrm{univ}}(\mathrm{Frob}_p) = \varphi(\operatorname{tr}(\rho_{\mathrm{univ}}(\mathrm{Frob}_p))).$$

We have two very abstractly defined local Noetherian rings. How would you prove they are isomorphic? Most people would be terrified by this question. Wiles dealt with it.

24.7 Wiles's main conjecture

> "We are like a train which is trying to reach Fermat's Last Theorem. Of course it has not made all of its scheduled stops. But it is on its way."

We have a representation $\rho : G \to \mathrm{GL}(2, \mathbf{F})$. Take $\mathbf{F} = \mathbf{F}_\ell$ for our applications today. Then the ring of Witt vectors is $W(\mathbf{F}) = \mathbf{Z}_\ell$. The Hecke algebra can be embedded as

$$\mathbf{T}_\mathfrak{m} \subset \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu}.$$

The Hecke algebra $\mathbf{T}_\mathfrak{m}$ has the following properties.

- The index of $\mathbf{T}_\mathfrak{m}$ in $\prod \mathcal{O}_{g,\mu}$ is finite.
- Gorenstein as a $\mathbf{Z}_\ell$-module, i.e., there exists an isomorphism $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_\mathfrak{m}, \mathbf{Z}_\ell) \cong \mathbf{T}_\mathfrak{m}$.
- $\mathbf{T}_\mathfrak{m}$ is generated by the T_r with r prime and $(r, \ell N_\Sigma) = 1$.

We have constructed a representation

$$\rho' : G \to \mathrm{GL}(2, \mathbf{T}_\mathfrak{m}).$$

Composing appropriately with the map $\mathbf{T}_\mathfrak{m} \hookrightarrow \prod \mathcal{O}_{g,\mu}$ gives a map

$$G \to \prod_{(g,\mu)} \mathrm{GL}(2, \mathcal{O}_{g,\mu}).$$

This is the product of representations $\prod \rho_{g,\mu}$. The triangle is

$$\begin{array}{ccc} G & \xrightarrow{\rho'} & \mathrm{GL}(2, \mathbf{T}_\mathfrak{m}) \\ & \searrow^{\prod \rho_{g,\mu}} & \downarrow \\ & & \prod \mathrm{GL}(2, \mathcal{O}_{g,\mu}) \end{array}$$

Moreover, ρ' is a deformation of ρ of type Σ so it lifts ρ and satisfies certain "nice as ρ" properties at primes $p \notin \Sigma$.

Let

$$\rho_{\mathrm{univ}} : G \to \mathrm{GL}(2, R_\Sigma)$$

be the universal deformation of ρ of type Σ. Lenstra gave a very concrete paper [dSL97] constructing this ρ_{univ}. Before his paper there was only Schlesinger's thesis. By the definition of ρ_{univ} there exists a unique map $\varphi : R_\Sigma \to \mathbf{T}_\mathfrak{m} = \mathbf{T}_\Sigma$ such that $\varphi \circ \rho_{\text{univ}} = \rho'$. By $\varphi \circ \rho_{\text{univ}}$ we mean the composition of ρ_{univ} with the map $\mathrm{GL}(2, R_\Sigma) \to \mathrm{GL}(2, \mathbf{T}_\mathfrak{m})$ induces by φ. As noted last time it is easy to see that φ is surjective.

Theorem 24.7.1 (Wiles's main 'conjecture')**.** *φ is an isomorphism (for each Σ).*

The theorem implies the following useful result.

Theorem 24.7.2. *Suppose E is a semistable elliptic curve over $\mathbf{Q}$ and that for some $\ell > 2$ the representation $\rho = \rho_{\ell,E}$ on $E[\ell]$ is irreducible and modular. Then E is modular.*

Proof. The representation

$$\tilde{\rho} = \rho_{\ell^\infty,E} : G \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

on the ℓ-power torsion $E[\ell^\infty] = \cup E[\ell^n]$ is a lift of

$$\rho : G \to \mathrm{Aut}(E[\ell]) = \mathrm{GL}(2, \mathbf{F}_\ell).$$

Furthermore, $\tilde{\rho}$ is a deformation of ρ of type Σ. Applying universality and using the theorem that $\mathbf{R}_\Sigma \cong \mathbf{T}_\Sigma$ we get a map

$$\mathbf{T}_\Sigma \to \mathbf{Z}_\ell : \quad T_r \mapsto a_r = a_r(E) = \mathrm{Tr}\, \tilde{\rho}(\mathrm{Frob}_r).$$

The relevant diagram is

$$\begin{array}{ccc} R_\Sigma & & \\ \downarrow & \searrow & \\ \mathbf{T}_\Sigma & \longrightarrow & \mathbf{Z}_\ell \end{array}$$

where the map $R_\Sigma \to \mathbf{Z}_\ell$ is given by

$$\mathrm{Tr}\, \rho_{\text{univ}}(\mathrm{Frob}_r) \mapsto a_r.$$

Now the full Hecke algebra $\mathbf{Z}[\ldots, T_n, \ldots]$ embeds into the completion $\mathbf{T}_\Sigma$. Composing this with the map $\mathbf{T}_\Sigma \to \mathbf{Z}_\ell$ above we obtain a map

$$\alpha : \mathbf{Z}[\ldots, T_n, \ldots] \to \mathbf{Z}_\ell.$$

Because of the duality between the Hecke algebra and modular forms there exists a modular form $h \in S_2(\Gamma_0(N_\Sigma), \mathbf{Z}_\ell)$ corresponding to α. Since α is a homomorphism h is a normalized eigenform. Furthermore $a_r(h) = a_r(E) \in \mathbf{Z}$ for all primes $r \nmid \ell N_\Sigma$. Since almost all coefficients of h are integral it follows that h is integral. Because we know a lot about eigenforms we can massage h to an eigenform in $S_2(\Gamma_0(N_E), \mathbf{Z})$. □

[[Some undigested comments follow.]]

- Once there is any connection between ρ and a modular form one can prove Taniyama-Shimura in as strong a form as desired. See the article *Number theory as Gadfly.*

- Take the abelian variety A_h attached to h. The λ-adic representation will have pieces with the same representations. Using Tate's conjecture we see that E is isogenous to A_h. Use at some points Carayol's theorem: If g is a form giving rising to the abelian variety A then the conductor of A is the same as the conductor of g.
- Tate proved that if two elliptic curves have isomorphic $\rho_{\ell\infty}$ for some ℓ then they are isogenous.

24.8 $\mathbf{T}_\Sigma$ is a complete intersection

Recall the construction of $\mathbf{T}_\Sigma = \mathbf{T}_\mathfrak{m}$. Let $\mathbf{T}$ be the anemic Hecke algebra. Then

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod \mathcal{O}_{g,\mu}$$

where the product is over a complete set of representatives (for the action of Galois on eigenforms) g and primes μ lying over ℓ. We found a specific (f, λ) for $\Sigma = \emptyset$ such that $\overline{\rho}_{\lambda,f} = \rho$. The maximal ideal $\mathfrak{m}$ was defined as follows. The form f induces a map $\mathbf{T} \to \mathcal{O}_{f,\lambda}$. Taking the quotient of $\mathcal{O}_{f,\lambda}$ by its maximal ideal we obtain a map $\mathbf{T} \to \overline{\mathbf{F}}_\ell$. Then $\mathfrak{m}$ is the kernel of this map. The diagram is

$$\begin{array}{ccc} \mathbf{T} & & \\ \downarrow & \searrow & \\ \mathcal{O}_{f,\lambda} & \longrightarrow & \overline{\mathbf{F}}_\ell \end{array}$$

The map $\mathcal{O}_{f,\lambda} \to \overline{\mathbf{F}}_\ell$ is

$$a_r(f) \mapsto \operatorname{Tr} \rho(\mathrm{Frob}_r).$$

To fix ideas we cheat and suppose $\mathcal{O}_{f,\lambda} = \mathbf{Z}_\ell$. [[In Wiles's optic this is OK since he can work this way then tensor everything at the end.]]

Now $\rho_{f,\lambda}$ is a distinguished deformation of ρ [["Distinguished" is not meant in a mathematical sense]]. The map f gives rise to a map $\mathbf{T}_\Sigma \to \mathcal{O} = \mathbf{Z}_\ell$ which we also denote by f

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi} & \mathbf{T}_\Sigma \\ & \searrow & \downarrow f \\ & & \mathcal{O} = \mathbf{Z}_\ell \end{array}$$

24.9 The Inequality $\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2 \leq \#\wp_R/\wp_R^2$

Let

$$\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell)$$

be irreducible and modular with $\ell > 2$. Let Σ be a finite set of primes. We assume there is a modular form f of weight 2 with coefficients in $\mathbf{Z}_\ell$ which gives rise ρ. Let

$$\rho_{f,\lambda} : G \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

be the representation coming from f, then $\rho_{f,\lambda}$ reduces to ρ modulo ℓ.

Let R_Σ be the universal deformation ring, so every deformation of ρ of type Σ factors through R_Σ in an appropriate sense. Let $\mathbf{T}_\Sigma$ be the Hecke ring associated to Σ. It is a $\mathbf{Z}_\ell$-algebra which is free of finite rank. Furthermore

$$\mathbf{T}_\Sigma \subset \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} = \mathcal{O}_{f,\mu} \times \prod_{(g,\mu)\in\mathcal{A}-\{(f,\mu)\}} \mathcal{O}_{g,\mu}$$

where $\mathcal{A}$ is as defined before. Define projections pr_1 and pr_2 onto the first and rest of the factors, respectively

$$\begin{aligned} \mathrm{pr}_1 &: \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} \to \mathcal{O} = \mathcal{O}_{f,\lambda} \\ \mathrm{pr}_2 &: \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} \to \prod_{(g,\mu)\neq(f,\lambda)} \mathcal{O}_{g,\mu} \end{aligned}$$

Let $\varphi : R_\Sigma \to \mathbf{T}_\Sigma$ be the map coming from the universal property of R_Σ. This map is surjective. The famous triangle which dominates all of the theory is

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi} & \mathbf{T}_\Sigma \\ & \searrow & \downarrow \mathrm{pr}_1 \\ & & \mathcal{O} = \mathbf{Z}_\ell \end{array}$$

24.9.1 The Definitions of the ideals

We now define two ideals. View $\mathbf{T}_\Sigma$ as sitting in the product $\prod \mathcal{O}_{g,\mu}$.

1. The *congruence ideal* $\eta \subset \mathcal{O}$ is
$$\eta := \mathcal{O} \cap \mathbf{T}_\Sigma = \ker\Big(\mathrm{pr}_2 : \mathbf{T}_\Sigma \to \prod_{(g,\mu)\neq(f,\lambda)} \mathcal{O}_{g,\mu}\Big)$$
2. The prime ideal $\wp_T \subset \mathbf{T}_\Sigma$ is
$$\wp_T = \ker\Big(\mathrm{pr}_1 : \mathbf{T}_\Sigma \to \mathcal{O}\Big)$$

It is true that

$$\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2.$$

The condition for equality is a theorem of Wiles.

Theorem 24.9.1. *$\mathbf{T}_\Sigma$ is a complete intersection if and only if $\#\mathcal{O}/\eta = \#\wp_T/\wp_T^2$.*

There is an analogous construction for

$$\psi = \mathrm{pr}_1 \circ \varphi : R_\Sigma \to \mathcal{O}.$$

The diagram is

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi} & \mathbf{T}_\Sigma \\ & \psi \searrow & \downarrow \mathrm{pr}_1 \\ & & \mathcal{O} \end{array}.$$

Let $\wp_R$ be the kernel of ψ. From the commutativity of the above diagram we see that ψ maps $\wp_R \twoheadrightarrow \wp_T$. Thus we have an induced map $\overline{\psi}$ on "tangent spaces"

$$\overline{\psi} : \wp_R/\wp_R^2 \twoheadrightarrow \wp_T/\wp_T^2.$$

It follows that

$$\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2 \leq \#\wp_R/\wp_R^2.$$

There is an analogous theorem.

Theorem 24.9.2. *The above inequalities are all equalities iff*

- $\varphi : R_\Sigma \to \mathbf{T}_\Sigma$ *is an isomorphism, and*
- $\mathbf{T}_\Sigma$ *is a complete intersection ring.*

24.9.2 Aside: Selmer groups

Let M be the set of matrices in $M(2, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$ which have trace 0. Then $\mathrm{GL}(2, \mathbf{Z}_\ell)$ operates on M by conjugation. Thus G acts on M via the representation $\rho' : G \to \mathrm{GL}(2, \mathbf{Z}_\ell)$. To $\wp_R/\wp_R^2$ there corresponds the *Selmer group* which is a subgroup of $H^1(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), M)$. The subgroup is

$$H^1_\Sigma(G, M) = \mathrm{Hom}_{\mathcal{O}}(\wp_R/\wp_R^2, \mathbf{Q}_\ell/\mathbf{Z}_\ell) \subset H^1(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), M).$$

[[Since Flach's thesis there has been a problem of trying to get an upper bound for the Selmer group. Wiles converted it into the above problem.]]

24.9.3 Outline of some proofs

We outline the key steps in the proof that $\#\wp_R/\wp_R^2 \leq \#\mathcal{O}/\eta$.

Step 1: $\Sigma = \emptyset$

The key step is the minimal case when $\Sigma = \emptyset$. This is done in [TW95]. They claim to be proving the apparently weaker statement that $\mathbf{T}_\Sigma$ a complete intersection implies

$$\#\wp_T/\wp_T^2 = \#\mathcal{O}/\eta.$$

But in Wiles's paper [Wil95] he obtains the inequality

$$\#\wp_R/\wp_R^2 \leq \frac{(\#\wp_T/\wp_T^2)^2}{\#\mathcal{O}/\eta}.$$

Combining these two shows that

$$\#\wp_R/\wp_R^2 \leq \#\mathcal{O}/\eta.$$

In an appendix to [TW95] Faltings proves directly that

$$\#\wp_R/\wp_R^2 \leq \#\mathcal{O}/\eta.$$

At this point there were some remarks about why Wiles might have taken a circuitous route in his Annals paper. Ribet replied,

> "As Serre says, it is sometimes better to leave out any psychological behavior related to how people did something but instead just report on what they did."

Step 2: Passage from $\Sigma = \emptyset$ to σ general

The second step is the induction step in which we must understand what happens as Σ grows. Thus Σ is replaced by $\Sigma' = \Sigma \cup \{q\}$ where q is some prime not in Σ.

We will use the following notation. The object attached to Σ' will be denoted the same way as the object attached to Σ but with a $'$. Thus $(\wp_R/\wp_R^2)'$ denotes the Selmer group for Σ'.

The change in the Selmer group when Σ is replaced by Σ' is completely governed by some local cohomology group. There is a constant c_q such that

$$\#(\wp_R/\wp_R^2)' \leq c_q \#(\wp_R/\wp_R^2) \leq c_q \#\mathcal{O}/\eta.$$

So we just need to know that

$$\#\mathcal{O}/\eta' \geq c_q \#\mathcal{O}/\eta,$$

i.e., that η' is *small* as an ideal in $\mathcal{O}$. We need a formula for the ratio of the two orders.

Let $\mathbf{T}$ be the anemic ring of Hecke operators on $S_2(\Gamma_0(N_\Sigma))$ obtained by adjoining to $\mathbf{Z}$ all the Hecke operators $\mathbf{T}_n$ with n prime to ℓN_Σ. Let $\mathbf{T}'$ be the anemic Hecke ring of Hecke operators on $S_2(\Gamma_0(N_{\Sigma'}))$.

Since $N_\Sigma | N_{\Sigma'}$ there is an inclusion

$$S_2(\Gamma_0(N_\Sigma)) \hookrightarrow S_2(\Gamma_0(N_{\Sigma'})).$$

There is one subtlety, this injection is not equivariant for all of the Hecke operators. But this is no problem because $\mathbf{T}$ and $\mathbf{T}'$ are anemic. So the inclusion induces a restriction map $r : \mathbf{T}' \to \mathbf{T}$.

We now introduce a relative version of η which is an ideal $I \subset \mathbf{T}$. One way to think of I is as $\mathbf{T} \cap \mathbf{T}'$ where $\mathbf{T}$ and $\mathbf{T}'$ are both viewed as subrings of $\prod \mathcal{O}_{g,\mu}$

$$\begin{array}{ccc}
\mathbf{T}' & \xrightarrow{\quad r \quad} & \mathbf{T} \\
\big\downarrow & & \big\downarrow \\
\mathbf{T}'_{\mathfrak{m}'} & & \mathbf{T}_{\mathfrak{m}} \\
\big\downarrow & & \big\downarrow \\
\prod_{(g,\mu)} \mathcal{O}_{g,\mu} & \hookrightarrow & \prod_{\text{more } (g,\mu)} \mathcal{O}_{g,\mu}
\end{array}$$

The definition Lenstra would give is that

$$I := r(\mathrm{Ann}_{\mathbf{T}'}(\ker(r))).$$

The amazing formula is

$$\eta' = \eta \cdot f(I)$$

where $f : \mathbf{T} \to \mathcal{O}$ is the map induced by the modular form f. [[After introducing this definition Ogus was very curious about how deep it is, in particular, about whether its proof uses the Gorensteiness of $\mathbf{T}$. Ribet said, "somehow I do not think this formula can possibly be profound."]]

We pause with an aside to consider Wiles's original definition of η. By duality the map $f : \mathbf{T}_\Sigma \to \mathbf{Z}_\ell$ induces

$$f^\vee : \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{Z}_\ell, \mathbf{Z}_\ell) \to \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_\Sigma, \mathbf{Z}_\ell) \cong \mathbf{T}_\Sigma.$$

Because $\mathbf{T}_\Sigma$ is Gorenstein there is an isomorphism $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_\Sigma, \mathbf{Z}_\ell) \cong \mathbf{T}_\Sigma$. Now $f^\vee(\mathrm{id}) \in \mathbf{T}_\Sigma$ so $f(f^\vee(\mathrm{id})) \in \mathcal{O} = \mathbf{Z}_\ell$. Wiles let $\eta = (f(f^\vee(\mathrm{id})))$ be the ideal generated by $f(f^\vee(\mathrm{id}))$.

To finish step 2 we must show that $\#\mathcal{O}/f(I) \geq c_q$, i.e., that "$I$ is small". [[I do not see how this actually finishes step 2, but it is reasonable that it should. How does this index relate to the index of $f(I)\eta$ in $\mathcal{O}$?]]

Let $J = J_0(N_\Sigma)$ and $J' = J_0(N_{\Sigma'})$. Since

$$S_2(\Gamma_0(N_\Sigma)) \hookrightarrow S_2(\Gamma_0(N_{\Sigma'}))$$

functoriality of the Jacobian induces a map $J \hookrightarrow J'$. By autoduality we also obtain an injection $J^\vee \hookrightarrow J'$ and $J \cap J^\vee$ is a finite subgroup of J'. [[I definitely do not understand why J is not just equal to $J^\perp$. Where does the other embedding $J^\perp \hookrightarrow J'$ come from?]]

It can be seen that $J \cap J^\vee = J[\delta]$ for some $\delta \in \mathbf{T}$. It turns out that

$$\mathrm{Ann}_{\mathbf{T}}(J \cap J^\vee) = \mathbf{T} \cap \delta \,\mathrm{End}(J) \supseteq \delta\mathbf{T}.$$

It is an observable fact that $f(\delta\mathbf{T})$ is an ideal of $\mathcal{O}$ of norm c_q. The *heart* of the whole matter is to see that the inclusion

$$\delta\mathbf{T} \subseteq \mathbf{T} \cap \delta \,\mathrm{End}(J)$$

is an equality after localization at $\mathfrak{m}$. To do this we have to know that $\mathrm{Tate}_{\mathfrak{m}}(J) \cong \mathbf{T}_{\mathfrak{m}}^2$. This is equivalent to the Gorenstein business. With this in hand one can just check this equality.

Unfortunately, it is the spring of 1996 and we are now 10 minutes past when the course should end. Realizing this, Ken brings the course to a close. In the grand Berkeley tradition, the room fills with applause.

25

Computing with Modular Forms and Abelian Varieties

26

The Modular Curve $X_0(389)$

Let N be a positive integer, and let $X_0(N)$ be the compactified coarse moduli space that classifies pairs (E, C) where E is an elliptic curve and C is a cyclic subgroup of order N. The space $X_0(N)$ has a canonical structure of algebraic curve over $\mathbf{Q}$, and its properties have been very well studied during the last forty years. For example, Breuil, Conrad, Diamond, Taylor, and Wiles proved that every elliptic curve over $\mathbf{Q}$ is a quotient of some $X_0(N)$.

The smallest N such that the Jacobian of $X_0(N)$ has positive Mordell-Weil rank is 37, and Zagier studied the genus-two curve $X_0(37)$ in depth in his paper [Zag85b]. From this viewpoint, the next modular curve deserving intensive investigation is $X_0(389)$, which is the first modular curve whose Jacobian has Mordell-Weil rank larger than that predicted by the signs in the functional equations of the L-series attached to simple factors of its Jacobian; in fact, 389 is the smallest conductor of an elliptic curve with Mordell-Weil rank 2. Note that 389 is prime and $X_0(389)$ has genus $g = 32$, which is much larger than the genus 2 of $X_0(37)$, which makes explicit investigation more challenging.

Work of Kolyvagin [Kol88b, Kol88a] and Gross-Zagier [GZ86] has completely resolved the rank assertion of the Birch and Swinnerton-Dyer conjecture (see, e.g., [Tat66]) for elliptic curves E with $\mathrm{ord}_{s=1} L(E, s) \leq 1$. The lowest-conductor elliptic curve E that doesn't submit to the work of Kolyvagin and Gross-Zagier is the elliptic curve E of conductor 389 mentioned in the previous paragraph. At present we don't even have a conjectural natural construction of a finite-index subgroup of $E(\mathbf{Q})$ analogous to that given by Gross and Zagier for rank 1 (but see Mazur's work on universal norms, which might be used to construct $E(\mathbf{Q}) \otimes \mathbf{Z}_p$ for some auxiliary prime p).

Inspired by the above observations, and with an eye towards providing helpful data for anyone trying to generalize the work of Gross, Zagier, and Kolyvagin, in this paper we compute everything we can about the modular curve $X_0(389)$. Some of the computations of this paper have already proved important in several other papers: the discriminant of the Hecke algebra attached to $X_0(389)$ plays

a roll in [Rib99], the verification of condition 3 in [MS01], and the remark after Theorem 1 of [**?**]; also, the arithmetic of $J_0(389)$ provides a key example in [AS02, §4.2]. Finally, this paper serves as an entry in an "encyclopaedia, atlas or hiker's guide to modular curves", in the spirit of N. Elkies (see [Elk98, pg. 22]).

We hilight several surprising "firsts" that occur at level 389. The discriminant of the Hecke algebra attached to $S_2(\Gamma_0(389))$ has the apparently unusual property that it is divisible by $p = 389$ (see Section 26.2.1). Also $N = 389$ is the smallest integer such that the order of vanishing of $L(J_0(N), s)$ at $s = 1$ is larger than predicted by the functional equations of eigenforms (see Section 26.1.3). The author conjectures that $N = 389$ is the smallest level such that an optimal newform factor of $J_0(N)$ appears to have Shafarevich-Tate group with nontrivial odd part (see Section 26.4.1). Atkin conjectures that 389 is the largest prime such that the cusp of $X_0^+(389)$ fails to be a Weierstrass point (see Section 26.4.2).

26.1 Factors of $J_0(389)$

To each newform $f \in S_2(\Gamma_0(389))$, Shimura [Shi73] associated a quotient A_f of $J_0(389)$, and $J_0(389)$ is isogeneous to the product $\prod A_f$, where the product runs over the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugacy classes of newforms. Moreover, because 389 is prime each factor A_f cannot be decomposed further up to isogeny, even over $\overline{\mathbf{Q}}$ (see [Rib75]).

26.1.1 Newforms of level 389

There are five $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugacy classes of newforms in $S_2(\Gamma_0(389))$. The first class corresponds to the unique elliptic curve of conductor 389, and its q-exansion begins

$$f_1 = q - 2q^2 - 2q^3 + 2q^4 - 3q^5 + 4q^6 - 5q^7 + q^9 + 6q^{10} + \cdots .$$

The second has coefficients in the quadratic field $\mathbf{Q}(\sqrt{2})$, and has q-expansion

$$f_2 = q + \sqrt{2}q^2 + (\sqrt{2} - 2)q^3 - q^5 + (-2\sqrt{2} + 2)q^6 + \cdots .$$

The third has coefficients in the cubic field generated by a root α of $x^3 - 4x - 2$:

$$f_3 = q + \alpha q^2 - \alpha q^3 + (\alpha^2 - 2)q^4 + (-\alpha^2 + 1)q^5 - \alpha^2 q^6 + \cdots .$$

The fourth has coefficients that generate the degree-six field defined by a root β of $x^6 + 3x^5 - 2x^4 - 8x^3 + 2x^2 + 4x - 1$ and q-expansion

$$f_4 = q + \beta q^2 + (\beta^5 + 3\beta^4 - 2\beta^3 - 8\beta^2 + \beta + 2)q^3 + \cdots .$$

The fifth and final newform (up to conjugacy) has coefficients that generate the degree 20 field defined by a root of

$$\begin{aligned} f_5 \quad = \quad & x^{20} - 3x^{19} - 29x^{18} + 91x^{17} + 338x^{16} - 1130x^{15} - 2023x^{14} + 7432x^{13} \\ & +6558x^{12} - 28021x^{11} - 10909x^{10} + 61267x^9 + 6954x^8 - 74752x^7 \\ & +1407x^6 + 46330x^5 - 1087x^4 - 12558x^3 - 942x^2 + 960x + 148. \end{aligned}$$

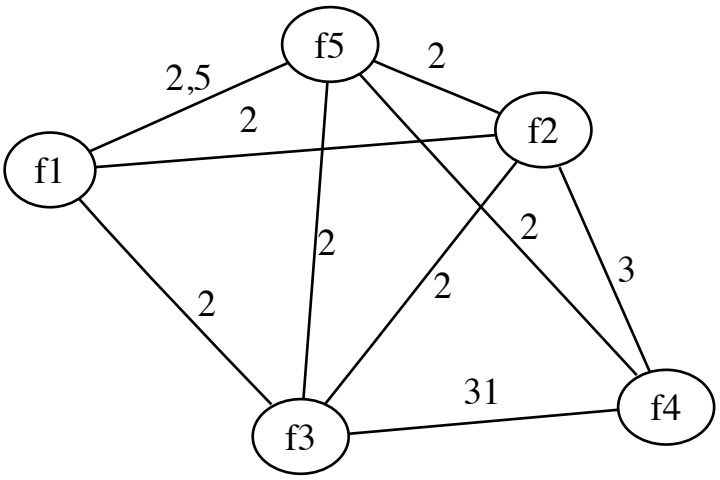


FIGURE 26.1.1. Congruences Between Newforms

Congruences

The vertices in Figure 26.1.1 correspond to the newforms f_i; there is an edge between f_i and f_j labeled p if there is a maximal ideal $\wp \mid p$ of the field generated by the Fourier coefficients of f_i and f_j such that $f_i \equiv f_j \pmod{\wp}$.

26.1.2 Isogeny structure

We deduce from the above determination of the newforms in $S_2(\Gamma_0(389))$ that $J_0(389)$ is $\mathbf{Q}$-isogenous to a product of $\overline{\mathbf{Q}}$-simple abelian varieties

$$J \sim A_1 \times A_2 \times A_3 \times A_4 \times A_5.$$

View the duals $A_i^\vee$ of the A_i as abelian subvarieties of $J_0(389)$. Using modular symbols as in [AS05, §3.4] we find that, for $i \neq j$, a prime p divides $\#(A_i^\vee \cap A_j^\vee)$ if and only if $f_i \equiv f_j \pmod{\wp}$ for some prime $\wp \mid p$ (recall that the congruence primes are given in Figure 26.1.1 above).

26.1.3 Mordell-Weil ranks

Suppose $f \in S_2(\Gamma_0(N))$ is a newform of some level N. The functional equation for $L(f,s)$ implies that $\operatorname{ord}_{s=1} L(f,s)$ is odd if and only if the sign of the eigenvalue of the Atkin-Lehner involution W_N on f is $+1$.

Proposition 26.1.1. *If $f \in S_2(\Gamma_0(N))$ is a newform of level $N < 389$, then $\operatorname{ord}_{s=1} L(f,s)$ is either 0 or 1.*

Proof. The proof amounts to a large computation, which divides into two parts:

1. Verify, for each newform f of level $N < 389$ such that $W_N(f) = -f$, that $L(f,1) = * \int_0^{i\infty} f(z)dz$ (for some nonzero $*$) is nonzero. This is a purely algebraic computation involving modular symbols.

2. Verify, for each newform f of level $N < 389$ such that $W_N(f) = f$, that $L'(f, 1) \neq 0$ (see [FpS$^+$01, §4.1], which points to [Cre97, §2.11,§2.13]). We do this by approximating an infinite series that converges to $L'(f, 1)$ and noting that the value we get is far from 0.

□

Thus $N = 389$ is the smallest level such that the L-series of some factor A_f of $J_0(N)$ has order of vanishing higher than that which is forced by the sign in the functional equation.

Proposition 26.1.2. *The following table summarizes the dimensions and Mordell-Weil ranks (over the image of the Hecke ring) of the newform factors of $J_0(N)$:*

	A_1	A_2	A_3	A_4	A_5
Dimension	1	2	3	6	20
Rank	2	1	1	1	0

Proof. The elliptic curve A_1 is $389A$ in Cremona's tables, which is the elliptic curve of smallest conductor having rank 2. For A_5 we directly compute whether or not the L-function vanishes using modular symbols, by taking an inner product with the winding element $e_w = -\{0, \infty\}$. We find that the L-function does not vanish. By Kolyvagin-Logachev, it follows that A_5 has Mordell-Weil rank 0.

For each of the other three factors, the sign of the functional equation is odd, so the analytic ranks are odd. As in the proof of Proposition 26.1.1, we verify that the analytic rank is 1 in each case. By work of Gross, Zagier, and Kolyvagin it follows that the ranks are 1. □

26.2 The Hecke algebra

26.2.1 The Discriminant is divisible by p

Let N be a positive integer. The Hecke algebra $\mathbf{T} \subset \mathrm{End}(S_2(\Gamma_0(N)))$ is the subring generated by all Hecke operators T_n for $n = 1, 2, 3, \ldots$. We are concerned with the *discriminant* of the trace pairing $(t, s) \mapsto \mathrm{Tr}(ts)$.

When N is prime, $\mathbf{T}_{\mathbf{Q}} = \mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Q}$ is a product $K_1 \times \cdots \times K_n$ of totally real number fields. Let $\tilde{\mathbf{T}}$ denote the integral closure of $\mathbf{T}$ in $\mathbf{T}_{\mathbf{Q}}$; note that $\tilde{\mathbf{T}} = \prod \mathcal{O}_i$ where $\mathcal{O}_i$ is the ring of integers of K_i. Then $\mathrm{disc}(\mathbf{T}) = [\tilde{\mathbf{T}} : \mathbf{T}] \cdot \prod_{i=1}^{n} \mathrm{disc}(K_i)$.

Proposition 26.2.1. *The discriminant of the Hecke algebra associated to $S_2(\Gamma_0(389))$ is*

$$2^{53} \cdot 3^4 \cdot 5^6 \cdot 31^2 \cdot 37 \cdot 389 \cdot 3881 \cdot 215517113148241 \cdot 477439237737571441.$$

Proof. By [**?**], the Hecke algebra $\mathbf{T}$ is generated as a $\mathbf{Z}$-module by $T_1, T_2, \ldots T_{65}$. To compute $\mathrm{disc}(\mathbf{T})$, we proceed as follows. First, compute the space $\boldsymbol{\mathcal{S}}_2(\Gamma_0(389))$ of cuspidal modular symbols, which is a faithful $\mathbf{T}$-module. Choose a random element $x \in \boldsymbol{\mathcal{S}}_2(\Gamma_0(389))_+$ of the $+1$-quotient of the cuspidal modular symbols, then compute the images $v_1 = T_1(x), v_2 = T_2(x), \ldots, v_{65} = T_{65}(x)$. If these don't span a space of dimension $32 = \mathrm{rank}_{\mathbf{Z}} \mathbf{T}$ choose a new random element x and repeat. Using the Hermite Normal Form, find a $\mathbf{Z}$-basis $b_1, \ldots, b_{32}$ for the $\mathbf{Z}$-span

of $v_1, \ldots, v_{65}$. The trace pairing on $\mathbf{T}$ induces a trace pairing on the v_i, and hence on the b_i. Then $\mathrm{disc}(\mathbf{T})$ is the discriminant of this pairing on the b_i. The reason we embed $\mathbf{T}$ in $\boldsymbol{\mathcal{S}}_2(\Gamma_0(389))_+$ as $\mathbf{T}x$ is because directly finding a $\mathbf{Z}$-basis for $\mathbf{T}$ would involve computing the Hermite Norm Form of a list of 65 vectors in a 1024-dimensional space, which is unnecessarily difficult (though possible). □

We compute this discriminant by applying the definition of discriminant to a matrix representation of the first 65 Hecke operators $T_1, \ldots, T_{65}$. Matrices representing these Hecke operators were computed using the modular symbols algorithms described in [Cre97]. [Sturm, *On the congruence of modular forms*].

In the case of $X_0(389)$, $\mathbf{T} \otimes \mathbf{Q} = K_1 \times K_2 \times K_3 \times K_6 \times K_{20}$, where K_d has degree d over $\mathbf{Q}$. We have

$$\begin{aligned}
K_1 &= \mathbf{Q}, \\
K_2 &= \mathbf{Q}(\sqrt{2}) \\
K_3 &= \mathbf{Q}(\beta), \quad \beta^3 - 4\beta - 2 = 0, \\
K_6 &= \mathbf{Q}(\gamma), \quad \gamma^6 + 3\gamma^5 - 2\gamma^4 - 8\gamma^3 + 2\gamma^2 + 4\gamma - 1 = 0, \\
K_{20} &= \mathbf{Q}(\delta), \quad \delta^{20} - 3\delta^{19} - 29\delta^{18} + 91\delta^{17} + 338\delta^{16} - 1130\delta^{15} - 2023\delta^{14} + 7432\delta^{13} \\
&\qquad +6558\delta^{12} - 28021\delta^{11} - 10909\delta^{10} + 61267\delta^9 + 6954\delta^8 - 74752\delta^7 \\
&\qquad +1407\delta^6 + 46330\delta^5 - 1087\delta^4 - 12558\delta^3 - 942\delta^2 + 960\delta + 148 = 0.
\end{aligned}$$

The discriminants of the K_i are

K_1	K_2	K_3	K_6
1	2^3	$2^2 \cdot 37$	$5^3 \cdot 3881$

and

$$\mathrm{disc}(K_{20}) = 2^{14} \cdot 5 \cdot 389 \cdot 215517113148241 \cdot 477439237737571441.$$

Observe that the discriminant of K_{20} is divisible by 389. The product of the discriminants is

$$2^{19} \cdot 5^4 \cdot 37 \cdot 389 \cdot 3881 \cdot 215517113148241 \cdot 477439237737571441.$$

This differs from the exact discriminant by a factor of $2^{34} \cdot 3^4 \cdot 5^2 \cdot 31^2$, so the index of $\mathbf{T}$ in its normalization is

$$[\tilde{\mathbf{T}} : \mathbf{T}] = 2^{17} \cdot 3^2 \cdot 5 \cdot 31.$$

Notice that 389 does not divide this index, and that 389 is not a "congruence prime", so 389 does not divide any modular degrees.

Question 26.2.2. Is there a newform optimal quotient A_f of $J_0(p)$ such that p divides the modular degree of A_f? (No, if $p < 14000$.)

26.2.2 Congruences primes in $S_{p+1}(\Gamma_0(1))$

K. Ono asked the following question, in connection with Theorem 1 of [**?**].

Question 26.2.3. Let p be a prime. Is p ever a congruence prime on $S_{p+1}(\Gamma_0(1))$? More precisely, if K is the number field generated by all the eigenforms of weight $p+1$ on $\Gamma_0(1)$, can there be a prime ideal $\wp \mid p$ for which $f \equiv g \pmod{\wp}$ for distinct eigenforms $f, g \in S_{p+1}(\Gamma_0(1))$?

The answer is "yes". There is a standard relationship between $S_{p+1}(\Gamma_0(1))$ and $S_2(\Gamma_0(p))$. As noted in Section 26.2.1, $p = 389$ is a congruence prime for $S_2(\Gamma_0(389))$, so we investigate $S_{389+1}(\Gamma_0(1))$.

Proposition 26.2.4. *There exist distinct newforms $f, g \in S_{389+1}(\Gamma_0(1))$ and a prime $\wp$ of residue characteristic* 389 *such that $f \equiv g \pmod{\wp}$.*

Proof. We compute the characteristic polynomial f of the Hecke operator T_2 on $S_{389+1}(\Gamma_0(1))$ using nothing more than [Ser73, Ch. VII]. We find that f factors modulo 389 as follows:

$$\begin{aligned}\overline{f} \;=\;& (x+2)(x+56)(x+135)(x+158)(x+175)^2(x+315)(x+342)(x^2+387)\\ & (x^2+97x+164)(x^2+231x+64)(x^2+286x+63)\\ & (x^5+88x^4+196x^3+113x^2+168x+349)\\ & (x^{11}+276x^{10}+182x^9+13x^8+298x^7+316x^6+213x^5\\ & \quad +248x^4+108x^3+283x^2+x+101)\end{aligned}$$

Moreover, f is irreducible and $389 \mid\mid \mathrm{disc}(f)$, so the square factor $(x+175)^2$ implies that 389 is ramified in the degree-32 field L generated by a single root of f. Thus there are exactly 31 distinct homomorphisms from the ring of integers of L to $\overline{\mathbf{F}}_{389}$. That is, there are exactly 31 ways to reduce the q-expansion of a newform in $S_{390}(\Gamma_0(1))$ to obtain a q-expansion in $\overline{\mathbf{F}}_{389}[[q]]$. Let K be the field generated by all eigenvalues of the 32 newforms $g_1, \ldots g_{32} \in S_{390}(\Gamma_0(1))$, and let $\wp$ be a prime of $\mathcal{O}_K$ lying over 389. Then the subset $\{g_1 \pmod{\wp}, g_2 \pmod{\wp}, \ldots, g_{32} \pmod{\wp}\}$ of $\overline{\mathbf{F}}_{389}[[q]]$ has cardinality at most 31, so there exists $i \neq j$ such that $g_i \equiv g_j \pmod{\wp}$. □

26.3 Supersingular points in characteristic 389

26.3.1 The Supersingular j-invariants in characteristic 389

Let α be a root of $\alpha^2 + 95\alpha + 20$. Then the $33 = g(X_0(389)) + 1$ supersingular j-invariants in $\mathbf{F}_{389^2}$ are

$$\begin{aligned}&0, 7, 16, 17, 36, 121, 154, 220, 318, 327, 358, 60\alpha+22, 68\alpha+166, 80\alpha+91, 86\alpha+273,\\ &93\alpha+333, 123\alpha+350, 123\alpha+375, 129\alpha+247, 131\alpha+151, 160\alpha+321, 176\alpha+188,\\ &213\alpha+195, 229\alpha+292, 258\alpha+154, 260\alpha+51, 266\alpha+335, 266\alpha+360, 296\alpha+56,\\ &303\alpha+272, 309\alpha+271, 321\alpha+319, 329\alpha+157.\end{aligned}$$

26.4 Miscellaneous

26.4.1 The Shafarevich-Tate group

Using visibility theory [AS02, §4.2], one sees that $\#Ш(A_5)$ is divisible by an odd prime, because

$$(\mathbf{Z}/5\mathbf{Z})^2 \approx A_1(\mathbf{Q})/5A_1(\mathbf{Q}) \subset Ш(A_5).$$

Additional computations suggest the following conjecture.

Conjecture 26.4.1. *$N = 389$ is the smallest level such that there is an optimal newform quotient A_f of $J_0(N)$ with $\#Ш(A_f)$ divisible by an odd prime.*

26.4.2 Weierstrass points on $X_0^+(p)$

Oliver Atkin has conjectured that 389 is the largest prime such that the cusp on $X_0^+(389)$ fails to be a Weierstrass point. He verified that the cusp of $X_0^+(389)$ is not a Weierstrass point but that the cusp of $X_0^+(p)$ is a Weierstrass point for all primes p such that $389 < p \leq 883$ (see, e.g., [Elk98, pg.39]). In addition, the author has extended the verification of Atkin's conjecture for all primes < 3000. Explicitly, this involves computing a reduced-echelon basis for the subspace of $S_2(\Gamma_0(p))$ where the Atkin-Lehner involution W_p acts as $+1$, and comparing the largest valuation of an element of this basis with the dimension of the subspace. These numbers differ exactly when the cusp is a Weierstrass point.

26.4.3 A Property of the plus part of the integral homology

For any positive integer N, let $H^+(N) = H_1(X_0(N), \mathbf{Z})^+$ be the $+1$ eigen-submodule for the action of complex conjugation on the integral homology of $X_0(N)$. Then $H^+(N)$ is a module over the Hecke algebra $\mathbf{T}$. Let

$$F^+(N) = \operatorname{coker}\left(H^+(N) \times \operatorname{Hom}(H^+(N), \mathbf{Z}) \to \operatorname{Hom}(\mathbf{T}, \mathbf{Z})\right)$$

where the map sends (x, φ) to the homomorphism $t \mapsto \varphi(tx)$. Then $\#F^+(p) \in \{1, 2, 4\}$ for all primes $p < 389$, but $\#F^+(389) = 8$.

26.4.4 The Field generated by points of small prime order on an elliptic curve

The prime 389 arises in a key way in the verification of condition 3 in [MS01].

References

[AS02] A. Agashe and W. Stein, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory **97** (2002), no. 1, 171–185.

[AS05] ———, *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

[BC04] Kevin Buzzard and Frank Calegari, *A counterexample to the Gouvêa-Mazur conjecture*, C. R. Math. Acad. Sci. Paris **338** (2004), no. 10, 751–753. MR 2059481 (2005g:11070)

[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). MR 2002d: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

[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.

[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.

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR 91i:14034

[BpR91] 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.

[Buz96] Kevin Buzzard, *On the eigenvalues of the Hecke operator T_2*, J. Number Theory **57** (1996), no. 1, 130–132. MR 96m:11033

[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

[CF99] J. B. Conrey and D. W. Farmer, *Hecke operators and the nonvanishing of L-functions*, Topics in number theory (University Park, PA, 1997), Math. Appl., vol. 467, Kluwer Acad. Publ., Dordrecht, 1999, pp. 143–150. MR 2000f:11055

[CM98] R. Coleman and B. Mazur, *The Eigencurve*, Galois representations in arithmetic algebraic geometry (Durham, 1996), Cambridge Univ. Press, Cambridge, 1998, pp. 1–113. MR 1 696 469

[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

[Con01] B. Conrad, *The shimura construction in weight 2 (appendix to Ribet-Stein, Lectures on Serre's Conjecture)*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 205–232. MR 2002h:11047

[Con03] K. Conrad, *Partial Euler products on the critical line*, Preprint (2003).

[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.

[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.

[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/`.

[CV92] R. F. Coleman and J. F. Voloch, *Companion forms and Kodaira-Spencer theory*, Invent. Math. **110** (1992), no. 2, 263–281.

[DDT94] 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.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[DK73] P. Deligne and W. Kuyk (eds.), *Modular functions of one variable. II*, Springer-Verlag, Berlin, 1973, Lecture Notes in Mathematics, Vol. 349.

[DK95] F. Diamond and K. Kramer, *Modularity of a family of elliptic curves*, Math. Res. Lett. **2** (1995), no. 3, 299–304.

[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.

[DS05] Fred Diamond and Jerry Shurman, *A first course in modular forms*, Graduate Texts in Mathematics, vol. 228, Springer-Verlag, New York, 2005.

[dSL97] B. de Smit and Jr. Lenstra, H. W., *Explicit construction of universal deformation rings*, Modular forms and Fermat's last theorem (Boston, MA, 1995), Springer, New York, 1997, pp. 313–326.

[Edi92a] B. Edixhoven, *Néron models and tame ramification*, Compositio Math. **81** (1992), no. 3, 291–306. MR 93a:14041

[Edi92b] ———, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594.

[Eis95] D. Eisenbud, *Commutative algebra with a view toward algebraic geometry*, Springer-Verlag, New York, 1995. MR 97a:13001

[Elk98] N. D. Elkies, *Elliptic and modular curves over finite fields and related computational issues*, Computational perspectives on number theory (Chicago, IL, 1995), Amer. Math. Soc., Providence, RI, 1998, pp. 21–76.

[Ell02] J. Ellenberg, **q**-*curves and Galois Representations*, `http://www.math.princeton.edu/ ellenber/papers.html#MCAV` (2002).

[ES00] J. Ellenberg and C. Skinner, *On the Modularity of* **Q**-*curves*, `http://www.math.princeton.edu/ ellenber/papers.html#QCURVE` (2000).

[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

[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.

[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). MR 1 836 926

[Frö67] A. Fröhlich, *Local fields*, Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965), Thompson, Washington, D.C., 1967, pp. 1–41.

[Ghi] Alex Ghitza, *Maeda's conjecture for weight up to 4096*, `http://wstein.org/Tables/charpoly_level1/t2/ghitza.html`.

[Gol82] D. Goldfeld, *Sur les produits partiels eulériens attachés aux courbes elliptiques*, C. R. Acad. Sci. Paris Sér. I Math. **294** (1982), no. 14, 471–474. MR 84d:14031

[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.

[Gro91] ———, *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

[Har77] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[HH] Bill Hart and David Harvey, *Flint: Fast library for number theory*, `http://www.flintlib.org/`.

[Hid00] H. Hida, *Modularity Problems of* $\mathbf{Q}$*-Motives and Base-Change*, Preprint, `http://www.math.ucla.edu/ hida/` (2000).

[Igu56] J. Igusa, *Fibre systems of Jacobian varieties*, Amer. J. Math. **78** (1956), 171–199.

[Joc82] N. Jochnowitz, *A study of the local components of the Hecke algebra mod* ℓ, Trans. Amer. Math. Soc. **270** (1982), no. 1, 253–267.

[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 MR0447119 (56 #5434)

[Kat81] N. Katz, *Serre-Tate local moduli*, Algebraic surfaces (Orsay, 1976–78), Lecture Notes in Math., vol. 868, Springer, Berlin, 1981, pp. 138–202. MR 638600 (83k:14039b)

[KM85] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

[Kna92] A. W. Knapp, *Elliptic curves*, Princeton University Press, Princeton, NJ, 1992.

[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. 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. MR 90f:11035

[KW08] C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture (i)*, Preprint (2008).

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR 93a:11048

[Lan93] ———, *Algebra*, third ed., Addison-Wesley Publishing Co., Reading, Mass., 1993.

[Lan94] ———, *Algebraic number theory*, second ed., Springer-Verlag, New York, 1994.

[Lan95] ———, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[Li75] W-C. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.

[LR11] Álvaro Lozano-Robledo, *Elliptic Curves, Modular Forms and their L-functions*, American Mathematical Society, 2011.

[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

[Maz72] B. Mazur, *Rational points of abelian varieties with values in towers of number fields*, Invent. Math. **18** (1972), 183–266.

[Maz77] ———, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[Maz78] ———, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[Maz89] ———, *Deforming Galois representations*, Galois groups over **Q** (Berkeley, CA, 1987), Springer, New York, 1989, pp. 385–437.

[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

[Mer94] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations, Springer, 1994, pp. 59–94.

[MFK94] D. Mumford, J. Fogarty, and F. Kirwan, *Geometric invariant theory*, third ed., Springer-Verlag, Berlin, 1994.

[Mil86] J. S. Milne, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[Miy89] T. Miyake, *Modular forms*, Springer-Verlag, Berlin, 1989, Translated from the Japanese by Yoshitaka Maeda.

[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). MR MR1141460 (93d:11056)

[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

[MSD74] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61. MR 50 #7152

[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.

[Ogg71] A. P. Ogg, *Rational points of finite order on elliptic curves*, Invent. Math. **12** (1971), 105–111. MR 45 #178

[Pyl] E. Pyle, *Abelian varieties over* $\mathbf{Q}$ *with large endomorphism algebras and their simple components over* $\overline{\mathbf{Q}}$, `http://math.berkeley.edu/ ribet/pyle_thesis.pdf`.

[Que77] C. Queen, *The existence of p-adic Abelian L-functions*, Number theory and algebra (New York), Academic Press, 1977, pp. 263–288.

[Rib75] K. A. Ribet, *Endomorphisms of semi-stable abelian varieties over number fields*, Ann. Math. (2) **101** (1975), 555–562. MR 51 #8120

[Rib77] ———, *Galois representations attached to eigenforms with Nebentypus*, Modular functions of one variable, V (Proc. Second Internat. Conf., Univ. Bonn, Bonn, 1976) (Berlin), Springer, 1977, pp. 17–51. Lecture Notes in Math., Vol. 601.

[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

[Rib94] ———, *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.

[Rib99] ———, *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

[Ros86] M. Rosen, *Abelian varieties over* $\mathbf{C}$, 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

[Rub89] K. Rubin, *The work of Kolyvagin on the arithmetic of elliptic curves*, Arithmetic of complex manifolds (Erlangen, 1988), Springer, Berlin, 1989, pp. 128–136.

[Sch06] 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.

[Sch65] O. F. G. Schilling (ed.), *Arithmetical algebraic geometry. (Proceedings of a Conference held at Purdue University, December 5–7, 1963)*, Harper & Row Publishers, New York, 1965.

[Sch90] A. J. Scholl, *Motives for modular forms*, Invent. Math. **100** (1990), no. 2, 419–430.

[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

[SD74] H. P. F. Swinnerton-Dyer, *Analytic theory of abelian varieties*, Cambridge University Press, London, 1974, London Mathematical Society Lecture Note Series, No. 14. MR 51 #3180

[Ser73] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[Ser77] ———, *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.

[Ser79] ———, *Local fields*, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

[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.

[Ser88] ———, *Algebraic groups and class fields*, Springer-Verlag, New York, 1988, Translated from the French.

[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.

[Shi59] G. Shimura, *Sur les intégrales attachées aux formes automorphes*, J. Math. Soc. Japan **11** (1959), 291–311.

[Shi73] ———, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544.

[Shi94] ———, *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] ———, *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.

[Ste04] W. Stein, *Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using MAGMA*, to appear in J. Cannon, ed., *Computational Experiments in Algebra and Geometry*, Springer-Verlag (2004).

[Ste07a] ———, *Explicitly computing with modular forms*, Graduate Studies in Mathematics, American Math Society, 2007.

[Ste07b] 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. MR MR2289048

[SW97] 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.

[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.

[TW95] R. Taylor and A. J. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[Vél72] J. Vélu, *Groupes de Galois attachés aux points d'ordre fini des courbes elliptiques sur un corps de nombres, d'après J. P. Serre*, Séminaire de Théorie des Nombres, 1971–1972 (Univ. Bordeaux I, Talence), Exp. No. 30, Lab. Théorie des Nombres, Centre Nat. Recherche Sci., Talence, 1972, p. 12.

[Wal85] J.-L. Waldspurger, *Quelques propriétés arithmétiques de certaines formes automorphes sur* GL(2), Compositio Math. **54** (1985), no. 2, 121–171. MR 87g:11061a

[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`.

[Zag85a] D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. **28** (1985), no. 3, 372–384. MR 86m:11041

[Zag85b] ———, *Modular points, modular curves, modular surfaces and modular forms*, Workshop Bonn 1984 (Bonn, 1984), Springer, Berlin, 1985, pp. 225–248.