Grundlehren der mathematischen Wissenschaften 222

A Series of Comprehensive Studies in Mathematics

Serge Lang

Introduction to Modular Forms

Springer-Verlag

Serge Lang

Introduction to Modular Forms

With 9 Figures

Springer-Verlag
Berlin Heidelberg New York

Serge Lang
Department of Mathematics, Yale University,
New Haven, CT 06520, U.S.A.

The AMS(MOS) classification scheme was made up before the subject of modular forms exploded. New numbers should be created for this subject.
It is impossible at present to find numbers fitting this book appropriately.

July 8, 1976

S. Lang

ISBN 3-540-07833-9 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-07833-9 Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Lang, Serge, 1927–. Introduction to modular forms. (Grundlehren der mathematischen Wissenschaften; 222). Bibliography: p. 000. Includes index.
1. Forms, Modular. I. Title. II. Series: Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen; 222. QA243.L257. 512'.74. 76-25140.

Typesetting: William Clowes & Sons Limited, London, Beccles and Colchester.
Printed and bound by Quinn-Woodbine Inc., Woodbine, New Jersey.
Printed in the United States of America.

9 8 7 6 5 4 3 2

Foreword

This book is intended as a partial survey for the elementary parts of an exceptionally active field which found a resurgence of interest over the last 8 years, after being almost forgotten for 30 years. I have attempted to put together some of the basic facts to make it easier for those who don't know the subject, to get some idea where it is going in the arithmetic direction, and how to get into it. I hope that the reader will find this book a helpful introduction to the Antwerp Conference volumes (Springer Lecture Notes).

It is unfortunate that Hecke's Institute Lecture Notes [H] never received wide distribution nor attention, and that they were omitted from his collected works. They summarize a great deal of his insights into modular forms. Ogg's book [O], for instance, follows almost the same table of contents, the main additions being the Petersson scalar product and Weil's theorem on functional equations, which Hecke did not have. Considering the progress which has been made since then, they have perhaps mostly historical interest, but I feel that even now, it is profitable to look at them. They have the merit, among many others, to be brief and accessible.

Partly because of Hitler and the war, which almost annihilated the German school of mathematics, and partly because of the great success of certain algebraic methods of Artin, Hasse, Deuring, modular forms and functions were to a large extent ignored by most mathematicians for about 30 years after the thirties. Eichler, Maass, Petersson, and Rankin were the main exceptions. It is striking that except for Petersson, the other three contributed to the International Colloquium on Zeta Functions, J. Indian Math. Soc. 1956. Maass was the first to develop a Hecke theory for non-holomorphic modular forms. In another direction Siegel in the 40's and 50's had some influence on the one variable case by his work on several variables, as well as through his Tata Institute notes. Selberg's contributions in the 50's were to have far reaching influence, but with some delays due to the lack of published proofs.

Shimura and Weil had much to do with bringing modular forms back into the forefront of mathematics. The Taniyama–Weil conjecture relating modular forms of weight 2 and elliptic curves gave impetus to the subject. Langlands gave an exceedingly broad framework for the connection between modular forms and the arithmetic of number fields, involving what can be called non-abelian class field theory as a special case. He recognized the connection between Hecke's work on Dirichlet series associated with modular forms and the Artin L-functions of finite Galois extensions of the rationals, among others. His conjectures also include the Weil conjecture as special case. In Jacquet-Langlands, it is shown how the Hecke

theory can be viewed as a vast generalization of Kronecker's theorem that every abelian extension of the rationals is cyclotomic, modulo the "Artin conjecture" (that L-functions are entire), and the theory is seen to apply as well to not necessarily holomorphic modular forms. Conversely, it was proved by Serre and Deligne that to every holomorphic form of weight 1 it is possible to associate an "odd" 2-dimensional representation of the Galois group over the rationals.

Historically, it is very interesting that Hecke noticed explicitly that by the Mellin transform, one can associate a modular form to each entire function defined by a Dirichlet series having a functional equation of standard type with one gamma factor, and conversely. He was looking for such functions. At the same time and place that he was writing this, Artin was working with his L-series. But as Tate once said, neither was digging what the other was doing, and so they did not notice that they were doing two aspects of the same thing. One had to wait till the Langlands conjectures for that.

To me, it is this direction which motivates the study of modular forms, i.e. their connections with representations of Galois groups of number fields.

The contents of this book consists mostly of lectures given at Yale in fall 1974. The first two chapters are essential to everything that follows. On the other hand, the rest of the book can be read in sections which are independent of each other. The first half is organized around Hecke operators, in various settings, mostly for $SL_2(\mathbf{Z})$, and over the complex numbers, including work of Eichler–Shimura and Manin. The second half deals with p-adic properties and the connection with Galois groups due to Serre and Swinnerton-Dyer, and distribution theory according to Iwasawa, touching on the connection with values of zeta functions, and p-adic modular forms, as developed by, among others, Klingen, Siegel, Serre, Coates, Sinnott, Katz, Manin, Mazur, etc.

I tried to select topics for which no systematic introduction is yet available. Since several introductions are available for the connection between Dirichlet series with functional equations and modular forms, this topic has been omitted.

I am much indebted to Ribet, Serre, and Zagier for their careful reading of the manuscript.

New Haven, in Summer 1976.

S. Lang

Table of Contents

Part I

Classical Theory

Chapter I. Modular Forms

For the convenience of the reader we reproduce a few facts and definitions about modular forms, although these are covered in a number of other places. However, some normalizations of terminology are not completely standardized, so it seemed preferable to spend a few pages going over these facts.

§ 1. The Modular Group

By SL_2 we mean the group of 2×2 matrices with determinant 1. We write $SL_2(R)$ for those elements of SL_2 having coefficients in a ring R. In practice, the ring R will be $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$. We call $SL_2(\mathbf{Z})$ the **modular group**.

If L is a lattice in $\mathbf{C}$, then we can always select a basis, $L=[\omega_1, \omega_2]$ such that $\omega_1/\omega_2=\tau$ is an element of the upper half plane, i.e. has imaginary part >0. Two bases of L can be carried into each other by an integral matrix with determinant ± 1, but if we normalize the bases further to satisfy the above condition, then the matrix will have determinant 1, in other words, it will be in $SL_2(\mathbf{Z})$. Conversely, transforming a basis as above by an element of $SL_2(\mathbf{Z})$ will again yield such a basis. This is based on a simple computation, as follows. If

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $GL_2(\mathbf{R})$, i.e. is a real non-singular matrix, and $\operatorname{Im}(z)>0$, then

$$\operatorname{Im} \frac{az+b}{cz+d} = \frac{(ad-bc)\operatorname{Im}(z)}{|cz+d|^2}.$$

We denote by $\mathfrak{H}$ the upper plane, i.e. the set of complex numbers z with $\operatorname{Im} z>0$. If α is a matrix as above, in $GL_2^+(\mathbf{R})$, (i.e. α has positive determinant), then we see that the element

$$\alpha(z)=\frac{az+b}{cz+d}$$

also lies in $\mathfrak{H}$, and one verifies by brute force that the association

$$(\alpha, z) \mapsto \alpha(z)=\alpha z$$

defines an operation of $GL_2^+(\mathbf{R})$ on $\mathfrak{H}$, i.e. is associative, and the unit matrix operates as the identity. In fact, all diagonal matrices $aI (a \in \mathbf{R})$ operate trivially, especially ± 1. Hence we have an operation of $SL_2(\mathbf{R})/\pm 1$ on $\mathfrak{H}$. For $\alpha \in SL_2(\mathbf{R})$, the relation mentioned above becomes

$$\operatorname{Im} \alpha(z) = \frac{\operatorname{Im} z}{|cz+d|^2} .$$

If f is a meromorphic function on $\mathfrak{H}$, then the function $f \circ \alpha$ such that

$$(f \circ \alpha)(z) = f(\alpha z)$$

is also meromorphic.

We let $\Gamma = SL_2(\mathbf{Z})$, so that Γ is a discrete subgroup of $SL_2(\mathbf{R})$. By a **(weak) fundamental domain** D for Γ in $\mathfrak{H}$ we shall mean a subset of $\mathfrak{H}$ such that every orbit of Γ has one element in D, and two elements of D are in the same orbit only if they lie on the boundary of D.

Theorem 1.1. *Let D consist of all $z \in \mathfrak{H}$ such that*

$$-\tfrac{1}{2} \leqslant \operatorname{Re} z \leqslant \tfrac{1}{2} \qquad \text{and} \qquad |z| \geqslant 1 .$$

Then D is a fundamental domain for Γ in $\mathfrak{H}$. Let

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

Then:

(i) *S and T generate Γ.*

(ii) *We have $S^2 = (ST)^3 = \pm I$, and $\{S\}$, $\{ST\}$ are the isotropy groups of i and ρ respectively in $\Gamma/\pm 1$.*

(iii) *All points in $\mathfrak{H}$ not equivalent to i or ρ have trivial isotropy group.*

The proof is standard, cf. Serre's *Course in Arithmetic* or [L 3], Chapter III, § 1. We omit it.

We do reproduce the illustration of the fundamental domain as follows.

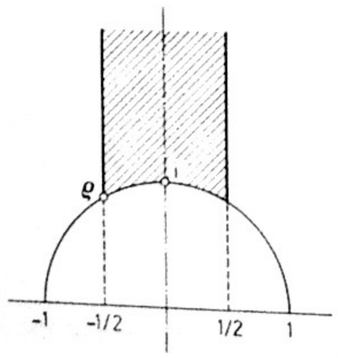

Fig. 1.

§ 2. Modular Forms

Let $\mathfrak{H}$ be the upper half plane again, let $B > 0$, and let $\mathfrak{H}_B$ be the set of complex numbers z with $\operatorname{Im} z > B$. The map

$$z \mapsto e^{2\pi i z} = q_z$$

defines a holomorphic map from $\mathfrak{H}_B$ to the punctured disc of radius $e^{-2\pi B}$, i.e. the disc from which the origin is deleted. Furthermore, if $\mathfrak{H}_B/T$ denotes the quotient space of $\mathfrak{H}_B$ modulo translations by integers (essentially a cylinder), then q induces an analytic isomorphism between $\mathfrak{H}_B/T$ and this punctured disc (trivial verification, since for $z = x + iy$, we have

$$e^{2\pi i z} = e^{2\pi i x}\, e^{-2\pi y}) .$$

Consequently a meromorphic function f on $\mathfrak{H}_B$ which has period 1, i.e. is invariant under T, induces a meromorphic function f_∞ on the punctured disc. A necessary and sufficient condition that f_∞ be also meromorphic at 0 is that there exist some positive integer N such that $f_\infty(q) q^N$ is bounded near 0. If this is the case, then f_∞ has a power series expansion

$$f_\infty(q) = \sum_{-N}^{\infty} c_n q^n .$$

We shall say that f is **meromorphic** (resp. **holomorphic**) **at infinity** if f_∞ is meromorphic (resp. holomorphic) at 0. By abuse of notation in this case, we also write

$$f = \sum_{-N}^{\infty} c_n q^n ,$$

and call this the **q-expansion** of f **at infinity**. The coefficients c_n are called the **Fourier coefficients** of f. If $c_{-N} \neq 0$, we call $-N$ the **order** of f **at infinity**, and denote it by $v_\infty f$. For any $z \in \mathfrak{H}$ we let the order of f at z be denoted by $v_z f$.

Let $\mathfrak{M}$ be the field of meromorphic functions on $\mathfrak{H}$ and let

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be in $GL_2^+(\mathbf{R})$. For $f \in \mathfrak{M}$, we define the operator $[\alpha]_k$ by

$$\boxed{f \circ [\alpha]_k = f(\alpha z)(cz+d)^{-k} (\det \alpha)^{k/2} .}$$

In particular, $[\alpha]_k$ depends only on the coset of α modulo scalar matrices, i.e. the image of α in the projective linear group.

We say that f is **modular of weight** k (or **degree** $-k$) if

$$f\circ[\alpha]_k = f$$

for all $\alpha\in\Gamma$, and if f is also meromorphic at infinity. Note that translation by 1 leaves f invariant, so our definition makes sense. Note also that a modular form of odd weight is obviously 0, so for $SL_2(\mathbf{Z})$ only modular forms of even weight will enter into consideration.

Theorem 2.1. *Let f be modular of weight k and $f\not\equiv 0$. Then*

$$v_\infty(f) + \frac{1}{3}v_p(f) + \frac{1}{2}v_i(f) + \sum_{P\neq i,\rho} v_P(f) = \frac{k}{12}.$$

The sum is taken over all points P of the upper half plane mod Γ, *not in the orbit of ρ or i.*

Proof. We integrate f'/f along the contour of Fig. 2(a), but modified by taking small arcs around the possible poles on the boundary, as on Fig. 2(b). For simplicity we phrase the proof under the assumption that f has no pole or zero on the edges other than at i or ρ, which are the most subtle possibilities. We have

$$\frac{1}{2\pi i}\int f'/f\,dz = \frac{1}{2\pi i}\int d\log f = \sum \text{Residues} = \sum_{P\neq i,\rho} v_P(f).$$

We shall now compute the integral over the top, sides, arcs around the corners, arc around i, and the main arcs on the bottom circle.

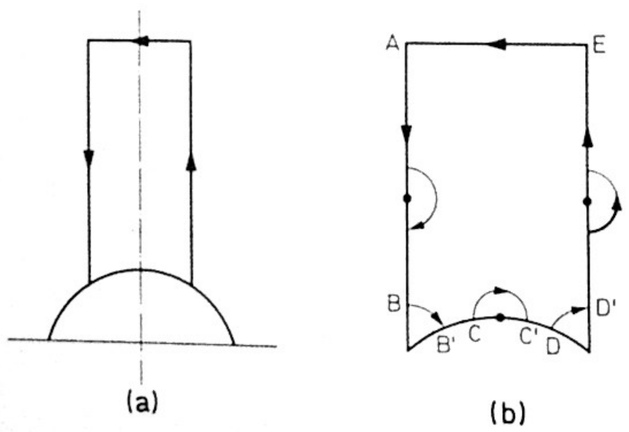

Fig. 2.

Under the q-change of variables, the top segment between E and A transforms into the circle centered around the origin, clockwise. The integral over the top therefore gives

$$-v_\infty(f).$$

The integral over the left vertical side downward, plus the integral over the right vertical side upward yields 0 by the periodicity of f.

The integral around ρ over the small arc is equal to

$$\frac{1}{2\pi i}\int_B^{B'} d\log f.$$

We make the translation of ρ to 0, and thus suppose we consider a function also denoted by f near the origin, with power series expansion

$$f(z) = cz^m(1+\cdots).$$

Then

$$\frac{f'(z)}{f(z)} = \frac{m}{z} + \text{holomorphic terms}.$$

As the radius of the small circle tends to 0, the integral of the holomorphic terms tend to 0. Integrating over an arc tending to $\pi/3$ in the clockwise direction, and taking the limit as the radius tends to 0 yields the value $-m/6$. We get a similar contribution on the small circle around $-\bar{\rho}$, whence the contributions from these two small circles yield

$$-\tfrac{1}{3}v_\rho(f).$$

The same argument for the small arc around i shows that we get a contribution of

$$-\tfrac{1}{2}v_i(f).$$

There remains to compute the integrals over the main arcs

$$\int_{B'}^{C} + \int_{C'}^{D}.$$

The map S transforms the arc $B'C$ to the arc DC'. By definition,

$$f(Sz) = z^k f(z),$$

and

$$\frac{df(Sz)}{dz} = f'(Sz)\frac{1}{z^2} = z^k f'(z) + 2kz^{k-1}f(z).$$

Since

$$\int_{C'}^{D}\frac{f'(w)}{f(w)}\,dw = \int_{C}^{B'}\frac{f'(Sz)}{f(Sz)}\,dz,$$

and

$$\frac{1}{z^2}\frac{f'(Sz)}{f(Sz)} = \frac{f'(z)}{f(z)} + \frac{k}{z},$$

we see that the integral over the second arc has one term which cancels the integral over the first arc, plus another term which is

$$\frac{1}{2\pi i}\int_{B'}^{C} \frac{k}{z}\,dz$$

and approaches $k/12$.

Putting all these contributions together proves our theorem.

Examples. They are constructed by using the following remark.

There is a bijection between functions of lattices, homogeneous of degree $-k$, *i.e. satisfying*

$$G(\lambda L) = \lambda^{-k} G(L), \qquad \lambda \in \mathbf{C},\ \lambda \neq 0,$$

and functions g *on* $\mathfrak{H}$ *satisfying the condition*

$$g(\alpha(z)) = (cz+d)^k g(z).$$

The bijection is obtained as follows. Given a function G homogeneous of degree $-k$, we let

$$g(z) = G(z, 1) = G\begin{pmatrix} z \\ 1 \end{pmatrix},$$

where by $G(z, 1)$ we mean the function G evaluated at the lattice $[z, 1]$. It then follows at once that

$$g(\alpha(z)) = (cz+d)^k g(z),$$

Conversely, given a function g satisfying this condition, define

$$G(z, 1) = G\begin{pmatrix} z \\ 1 \end{pmatrix} = g(z),$$

and for any lattice $L = [\omega_1, \omega_2]$ define

$$G(L) = \omega_2^{-k} g(\omega_1/\omega_2).$$

Then again it follows at once that $G(\lambda L) = \lambda^{-k} G(L)$.

The fact that G is a function of lattices can be written in our vertical notation as

$$G\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = G\left(\alpha\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}\right)$$

for any $\alpha \in SL_2(\mathbf{Z})$.

It is convenient to use the same symbol for the function of two variables and one variable, so that we shall also write

$$g(z) = g(z, 1) = g\begin{pmatrix} z \\ 1 \end{pmatrix}.$$

A function of weight k is called a **modular form** (of weight k) if it is *holomorphic on* $\mathfrak{H}$ *and at infinity*. The special examples we now give will be of this type. In the next section, we construct a function of weight 0, holomorphic on $\mathfrak{H}$ but not at infinity.

Consider the functions

$$S_k(L) = S_k = \sum_{\omega \neq 0} \frac{1}{\omega^k}, \qquad \text{for} \quad k > 2.$$

Then the dehomogenized function on $\mathfrak{H}$,

$$G_k(z) = \sum_{(m,n) \neq (0,0)} \frac{1}{(mz+n)^k}$$

is obviously holomorphic, and substituting $z = \infty$ formally gives

$$G_k(\infty) = \sum_{n \neq 0} \frac{1}{n^k} = 2\zeta(k).$$

We shall actually get the q-expansion for G_k later, and see that G_k is holomorphic at infinity, with the above value. Hence G_k is a modular form of weight k, and non-vanishing at infinity.

Since G_k begins with a non-zero constant term (which we shall study in Chapter V), we can define E_k as the constant multiple of G_k whose q-expansion begins with the constant 1. For now, we let

$$\Delta = \frac{1}{1728}(E_4^3 - E_6^2).$$

We note that Δ has weight 12. It is easy to determine the q-expansion (and will be done in Chapter X, § 3, § 4). We then see that Δ has order 1 at infinity.

The G_k defined above carries the power $(2\pi i)^k$ as a factor of its q-expansion. In arithmetic applications, it is convenient to consider the series where this factor has been taken off. Here we are working over $\mathbf{C}$, so this power does not matter, but we shall take it out for instance in Chapter V, § 8.

Let M_k be the set of modular forms of weight k. Then M_k is a vector space over $\mathbf{C}$. It is clear that

$$M_k M_l \subset M_{k+l} .$$

The direct sum

$$\coprod_{k=0}^{\infty} M_k$$

can therefore be viewed as a graded algebra, whose structure is given by the next theorem.

Theorem 2.2. *The functions G_4 and G_6 are algebraically independent, and*

$$\coprod_{k=0}^{\infty} M_k = \mathbf{C}[G_4, G_6] .$$

Proof. Note that G_4, G_6 generate a subalgebra of our graded algebra. To analyse M_k we shall apply the formula of Theorem 2.1 and observe that for $f \in M_k, f \neq 0$, all the orders on the left-hand side are $\geqslant 0$. We now proceed systematically.

k = 0. The right-hand side is 0, so all the terms on the left are 0. If $f \in M_0$ and f is not identically 0, then f has no zero on $\mathfrak{H}$ or at infinity. The constants are contained in M_0. Let $c = f(\infty)$. Then $g = f - c$ vanishes at infinity, hence is identically 0, so $M_0 = \mathbf{C}$.

k = 2, The right-hand side is 1/6. The left-hand side shows that this is not possible, so $M_2 = 0$.

k = 4. We prove that $M_2 = (G_4)$ is the 1-dimensional vector space generated by G_4. Let $f \in M_4, f \neq 0$. The right-hand side of the basic formula is 1/3. The only time this is compatible with the left-hand side is when all the terms on the left are 0 except for $\frac{1}{3} v_\rho(f)$, and we must have $v_\rho(f) = 1$, while f has no other zero. In particular, we have also proved:

G_4 has a zero only at ρ, and it is of order 1 .

For some constant c, $f - cG_2$ has a zero at infinity, and lies in M_4, hence is identically zero, and $f = cG_4$, thus proving what we wanted.

k = 6. We prove that $M_6 = (G_6)$. The right-hand side of the basic formula is 1/2, for f in $M_6, f \neq 0$. The only way this is possible is that $v_i(f) = 1$, and f has no other zero. In particular,

G_6 has a zero only at i, and it is of order 1 .

The same argument as before shows that $f = cG_6$ for some constant c.

k = 8. We prove that $M_8 = (G_4^2)$. The right-hand side of the formula for $f \in M_8, f \neq 0$ is 2/3, and hence $v_\rho(f) = 2$, and f has no other zero. It follows that $f = cG_4^2$ as before.

k = 10. We prove that $M_{10} = (G_4 G_6)$. In this case, the same arguments as before show that $f \in M_{10}, f \neq 0$ has a zero of order 1 at i and ρ, and no other zero, and also that $f = cG_4 G_6$.

k ⩾ 12. The right-hand side of the formula for $k = 12$ is equal to 1, and the q-expansion shows that $v_\infty(\Delta) = 1$, i.e. that Δ has a zero of order 1 at infinity. Theorem 2.1 then shows that Δ does not vanish on $\mathfrak{H}$.

Now $G_{12} \in M_{12}$ and $G_{12}(\infty) \neq 0$. If $f \in M_{12}$, then there exists a constant c such that $f - cG_{12}$ vanishes at infinity. Then

$$\frac{f - cG_{12}}{\Delta} \in M_0 = \mathbf{C} ,$$

and we see that $f = b\Delta + cG_{12}$ for some constant b. Inductively, the same technique shows that for $k \geqslant 12$, even,

$$\boxed{M_k = \Delta M_{k-12} \oplus (G_k) .}$$

We can prove by induction that any $f \in M_k$ is a polynomial in G_4 and G_6. This has already been shown for $k \leqslant 10$. If $k \geqslant 12$, we write $k = 4r$ or $k = 4r + 2$, and we can subtract cG_4^r or $cG_4^{r-1} G_6$ from f, with a suitable constant c, to get a function vanishing at infinity, so that

$$\frac{f - cG_4^r}{\Delta} \quad \text{or} \quad \frac{f - cG_4^{r-1} G_6}{\Delta}$$

lies in M_{k-12}, and our proof is complete, by induction.

There remains to prove that G_4 and G_6 are algebraically independent, to be sure we get the formal polynomial ring. First it is clear from the homogeneity property that a non-trivial linear relation among elements of distinct M_k's cannot exist, i.e. if $f_1, \ldots, f_m$ are of distinct weight, then they are linearly independent over the complex numbers. If we had an algebraic relation among G_4, G_6, then we could assume that the monomials in it have the same weight. In such a relation, if a pure power of G_4 occurs, then the relation is of the form

$$G_4^m + G_6 P(G_4, G_6) = 0$$

where P is some polynomial. Evaluating this at i shows that it is impossible because $G_6(i) = 0$ and $G_4(i) \neq 0$. Similarly, no pure power of G_6 can occur. Hence

G_4 divides each monomial, and cancelling G_4 yields a relation of lower degree, so the proof is finished by induction.

The subspace of M_k consisting of those modular forms which have a zero at infinity is called the space of **cusp forms**, and is denoted by M_k^0.

The recursive construction used to prove the theorem obviously gives the dimension of the space of cusp forms, namely:

$$\dim M_k^0 = \begin{cases} \left[\frac{k}{12}\right] & \text{if} \quad k \equiv 2 \pmod{12} \\ \left[\frac{k}{12}\right] + 1 & \text{if} \quad k \not\equiv 2 \pmod{12}. \end{cases}$$

Cusp forms and modular forms can also be defined for subgroups of $SL_2(\mathbf{Z})$, and the dimension of the corresponding spaces is easily computable. Cf. Shimura [Sh 2], Theorem 2.24.

The exposition of this section is due to Serre [Se 7]. For an interesting development of these classical computations, cf. Siegel [Si 2].

§ 3. The Modular Function j

We define the modular function

$$j = E_4^3/\Delta .$$

From the properties of G_4, G_6 proved in the preceding section, we see that j has weight 0, and since Δ is holomorphic, non-zero on $\mathfrak{H}$, we see that j has a pole at infinity, of order 1.

Theorem 3.1. *The map $j: \Gamma\backslash\mathfrak{H} \to \mathbf{C}$ is a bijection.*

Proof. We apply the basic relation of Theorem 2.1 with $k=0$, so the right-hand side is 0, to the function $j-c$ for $c \in \mathbf{C}$. Then $j-c$ has a simple pole at infinity, and

$$\tfrac{1}{3}v_\rho + \tfrac{1}{2}v_i + \sum v_P = 1 .$$

The terms on the left are all $\geqslant 0$. This is possible if and only if the order of $j-c$ at some unique z in $\Gamma\backslash\mathfrak{H}$ is $\neq 0$. The multiplicity is 1 if z is not in the orbit of ρ, i and otherwise, it is 2 at i and 3 at ρ. In any case, our theorem is proved.

§ 4. Estimates for Cusp Forms

On several occasions we want to integrate differential forms to infinity over certain regions of the upper half plane. For this purpose, we need estimates on cusp forms. Wait to read this section until you need it.

Lemma 1. *Let f be a cusp form. Then f satisfies the estimate*

$$|f(x+iy)| \ll e^{-2\pi y}$$

for all y sufficiently large, uniformly in x.

Proof. By definition, f has a q-expansion

$$f_\infty(q) = a_1 q + a_2 q^2 + \cdots ,$$

with $q = e^{2\pi i(x+iy)} = e^{2\pi i x}\, e^{-2\pi y}$. The lemma follows immediately.

Lemma 2. *Let f be a cusp form with q-expansion*

$$f = \sum a_n q^n .$$

If f has weight k, then the coefficients a_n satisfy the estimate

$$|a_n| \ll n^{k/2} .$$

Proof. The transformation law for f immediately shows that the function

$$z \mapsto y^{k/2}|f(z)| , \qquad z = x+iy ,$$

is invariant under $SL_2(\mathbf{Z})$. Hence this function is bounded on $\mathfrak{H}$, and we get the estimate

$$|f(x+iy)| \ll \frac{1}{y^{k/2}} , \qquad \text{for} \quad y \to 0 .$$

Integrating the q-expansion of f yields the Fourier coefficients,

$$e^{-2\pi n y} a_n = \int_0^1 f(x+iy)\, e^{-2\pi i n x}\, dx .$$

This is true for any value of y. Let $y = 1/n$. We obtain

$$|a_n| \ll n^{k/2} ,$$

as desired.

Hecke **associated** to a cusp form the **Dirichlet series**

$$L_f(s) = \sum a_n n^{-s} .$$

The estimate of Lemma 2 shows:

Theorem 4.1. *The associated Dirichlet series to a cusp form of weight k converges to define an analytic function in the domain* $\operatorname{Re} s > k/2+1$.

§ 5. The Mellin Transform

The formal association of a Dirichlet series at the end of the last section will be shown to be attainable by an integral transform. However, no further use will be made of this, except for one passing reference in Chapter X, so the reader may omit this section without impairing the understanding of the rest of the book.

By Lemma 1 of § 4, the integral

$$\int_{t_0}^{\infty} f(it)t^s \frac{dt}{t}, \qquad t_0 > 0,$$

converges absolutely for all complex s. Because f is rapidly decreasing on the imaginary axis, one can differentiate under the integral sign, and we see that in fact, the integral defines an entire function of s.

Let us define the **Mellin transform** by

$$\mathbf{M}f(s) = \int_0^{\infty} f(it)t^s \frac{dt}{t}.$$

Recall that

$$\Gamma(s) = \int_0^{\infty} e^{-t}t^s \frac{dt}{t}.$$

Using the fact that the integral is invariant under multiplicative translation, in particular under $t \mapsto t/2\pi n$, we find:

$$\mathbf{M}f(s) = \sum \int_0^{\infty} a_n e^{-2\pi nt}t^s \frac{dt}{t} = (2\pi)^{-s}\Gamma(s)\sum a_n n^{-s} = (2\pi)^{-s}\Gamma(s)L_f(s).$$

Thus the Mellin transform of the cusp form is the associated Dirichlet series, up to an exponential and gamma factor.

The analytic continuation and functional equation is easily obtainable. Writing

$$\int_0^{\infty} = \int_0^1 + \int_1^{\infty}$$

we find

$$\int_0^1 f(it)t^s \frac{dt}{t} = \int_1^{\infty} f(i/t)t^{-s} \frac{dt}{t}.$$

Using the functional equation of modular forms of weight k, this is

$$= \int_1^{\infty} (it)^k f(it)t^{-s} \frac{dt}{t} = i^k \int_1^{\infty} f(it)t^{k-s} \frac{dt}{t}.$$

In particular, we get the functional equation of the Mellin transform,

$$\boxed{i^k \mathbf{M}f(k-s) = \mathbf{M}f(s).}$$

Of course, this is true for all complex s.

Chapter II. Hecke Operators

Hecke operators are averaging operators similar to a trace. They operate on the space of modular forms. Let f be a modular form, $f=\sum a_n q^n$, with associated Dirichlet series

$$\varphi=\sum a_n n^{-s}.$$

It turns out that f is an eigenfunction for all Hecke operators if and only if the Dirichlet series has an Euler product. Such Euler products give relations among the coefficients, which show that they are multiplicative in n (i.e. $a_{mn}=a_m a_n$ if m, n are relatively prime), and that they satisfy certain recurrence relation for prime power indices. The reader will find applications for these in Chapter VI, § 3. One of the basic problems of the theory is to organize into a coherent role the relations satisfied by these coefficients, and their effect on the arithmetic of number fields. The Hecke ones are in a sense the oldest. Later chapters touch on congruence relations. Manin [Man 4] found some which are much more hidden. The situation is very much in flux as this book is written.

§ 1. Definitions and Basic Relations

We have seen that modular forms for $SL_2(\mathbf{Z})$ may be viewed as functions of lattices. To define operators on them, it suffices to define operators on the lattices, which we now do.

Let $\mathscr{L}$ be the free abelian group generated by the lattices in $\mathbf{C}$. We define the **Hecke operator** $T(n)$ for each positive integer n to be the map

$$T(n):\mathscr{L}\to\mathscr{L}$$

such that

$$T(n)L=\sum_{(L:L')=n} L'.$$

Thus $T(n)$ associates with L the sublattices L' of index n in L, with multiplicity 1.

We define another operator $R(n):\mathscr{L}\to\mathscr{L}$ to be such that

$$R(n)L=nL,$$

i.e. the sublattice of L consisting of all n-th multiples of elements of L. It is clear that the operators $R(n)$ and $T(m)$ commute with each other.

Theorem 1.1. (i) $T(m)T(n)=T(mn)$ *if* $(m,n)=1$.

(ii) *For a prime power,*

$$T(p^r)T(p)=T(p^{r+1})+pR(p)T(p^{r-1}).$$

(iii) *The algebra generated by the operators* $T(n)$ (*all* n) *is generated by the operators* $R(p)$ *and* $T(p)$ *for all primes* p.

Proof. We begin by proving (ii). Both the right- and left-hand side associate with L sublattices of index p^{r+1}, and we have to verify that the multiplicities are the same. Let L' be such a sublattice. If $L'\subset pL$, then the right-hand side gives L' multiplicity $1+p$. Furthermore, L' is contained in every sublattice of L of index p. The left-hand side associates all sublattices of index p^r to the lattices of index p in L. Thus L' occurs once for each such sublattice of index p, and therefore also has multiplicity $1+p$ from the left-hand side.

Suppose that L' is not contained in pL. Then it has multiplicity 1 from the right-hand side. If it had multiplicity >1 from the left-hand side, then it would be contained in at least two sublattices of index p in L, whence in their intersection, which is precisely pL. Hence L' must have multiplicity 1 also from the left-hand side, as desired.

The formula shows that $T(p^2)$ commutes with $T(p)$ because it is a polynomial in $T(p)$ and $R(p)$. Similarly by induction one sees that $T(p^{r+1})$ is a polynomial in $T(p)$ and $R(p)$, and therefore commutes with $T(p)$.

For composite n, the first assertion is clear because if L' is a sublattice of index mn, then there exists a unique lattice between L and L' of index m in L, and a unique lattice between L and L' of index n in L, say because the factor group L/L' is abelian of order mn, and decomposes into a direct sum of factors of order m and n respectively. This proves the theorem.

Although we don't need it for the sequel, the reader can verify by induction that

$$T(p^r)T(p^s)=\sum_{i\leqslant r,s} p^i R(p^i)T(p^{r+s-2i}),$$

and then the general formula by multiplicativity,

$$T(n)T(m)=\sum_{d\mid(n,m)} dR(d)T\left(\frac{mn}{d^2}\right).$$

Let now f be a modular form for $SL_2(\mathbf{Z})$, of weight k. Then we know that f corresponds to a homogeneous function F of lattices of degree $-k$. We define the

k-th **Hecke operator** $T_k(n)$ by

$$T_k(n)f(L)=n^{k-1}\sum_{(L:L')=n}F(L')\,.$$

Then $T_k(n)f$ is homogeneous of degree $-k$, and we shall see later the effect of $T_k(n)$ on the q-expansion coefficients, which will make it clear that $T_k(n)f$ has a q-expansion. It follows that $T_k(n)f$ is again a modular form of weight k.

Theorem 1.2. *On the space M_k, we have:*

(i) $T_k(m)T_k(n)=T_k(mn)$ *if m, n are relatively prime.*

(ii) *For a prime power,*

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

(iii) *The algebra generated by the operators $T_k(n)$ for all n is also generated by the operators $T_k(p)$ for all primes p, and is commutative.*

Proof. The proof is immediate from Theorem 1.1 and the definition of $T_k(n)$. Observe that

$$F\circ R(n)(L)=n^{-k}F(L)\,.$$

Thus in transferring the properties of $T(n)$ to $T_k(n)$, the operators $R(n)$ disappear into a scalar factor because of the homogeneity property of F.

Again the reader can verify the general formula

$$T_k(n)T_k(m)=\sum_{d\mid(n,m)}d^{k-1}T_k\left(\frac{mn}{d^2}\right),$$

which is not needed in the sequel.

We now want to see the effect of the Hecke operators on the Fourier coefficients of f. For this we need an explicit determination of the sublattices of index n in some lattice.

Let $L=[\omega_1,\omega_2]$. If $L'=[\omega_1',\omega_2']$ is a sublattice of index n, then there is an integral matrix

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}$$

such that

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

and we get all bases of L' by multiplying still on the left by an element of $SL_2(\mathbf{Z})$ (assuming throughout that when we write a basis $[\omega_1,\omega_2]$ the quotient ω_1/ω_2 is the upper half plane).

Let $\mathbf{M}^n$ denote the set of integral matrices with determinant n. Let

$$\mathbf{M}^n=\sum_{i=1}^{\psi(n)}\Gamma\alpha_i\,,\qquad \Gamma=SL_2(\mathbf{Z})$$

be a coset decomposition of $\mathbf{M}^n$ with respect to $SL_2(\mathbf{Z})$. Then sublattices of L of index n correspond to the representatives α_i $(i=1,\dots,\psi(n))$. Consequently, we can write the operation of the Hecke operators on the function f as follows.

Theorem 1.3. *The action of $T_k(n)$ on the modular form f is given by*

$$T_k(n)f=n^{k/2-1}\sum_{i=1}^{\psi(n)}f\circ[\alpha_i]_k$$

where α_i $(i=1,\dots,\psi(n))$ ranges over a complete set of representatives of $\mathbf{M}^n$ with respect to $SL_2(\mathbf{Z})$.

We now determine explicitly a set of representatives α_i.

Lemma. *Choosing the matrices α_i to be*

$$\begin{pmatrix}a & b\\ 0 & d\end{pmatrix}$$

with a, $d>0$, $ad=n$ and $0\leqslant b<d$ gives a complete set of coset representatives of $\mathbf{M}^n$ with respect to $SL_2(\mathbf{Z})$.

Proof. First it is clear that any coset has a representative with $c=0$. (Solve $xa+yc=0$ with relatively prime x, y, and then complete x, y to a matrix in $SL_2(\mathbf{Z})$, with (x,y) as the bottom row.) Multiplying by -1 if necessary we may assume a, $d>0$.

Next, by multiplying on the left with matrices

$$\begin{pmatrix}1 & m\\ 0 & 1\end{pmatrix}$$

we can add the bottom row of

$$\begin{pmatrix}a & b\\ 0 & d\end{pmatrix}$$

to the top row, with coefficient m, thus allowing the reduction of b to representatives mod d, i.e.

$$0\leqslant b<d\,.$$

It is now immediately verified that the matrices so obtained constitute a complete set of representatives for the cosets, so the lemma is proved.

Remark. Let $\mathbf{M}^{n*}$ be the set of integral primitive matrices of determinant n, i.e. those matrices whose entries are relatively prime. Then $\mathbf{M}^{n*}$ also has a coset decomposition, and in fact there is a unique double coset,

$$\mathbf{M}^{n*} = \Gamma(1)\alpha\Gamma(1)$$

for any $\alpha \in \mathbf{M}^{n*}$. This holds in particular when n is prime. For the proof, we note that applying α to $\mathbf{Z}^2$ may be viewed as giving a sublattice of index equal to n. Changing α on the right and left by elements of $SL_2(\mathbf{Z})$ amounts to changing bases in $\mathbf{Z}^2$ and the sublattice. The elementary divisor theorem then proves the remark.

Theorem 1.4. *Let*

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

be the q-expansion of a modular form of weight k. Then

$$T_k(m)f = \sum_{n=0}^{\infty} a_n(m) q^n$$

where

$$a_n(m) = \sum_{d \mid (n,m)} d^{k-1} a_{nm/d^2}$$

$$a_0(m) = \sigma_{k-1}(m) a_0 , \qquad a_1(m) = a_m .$$

In particular, if f is a cusp form ($a_0 = 0$), then $T_k(m)f$ is also a cusp form.

Proof. Using the representatives and the definitions, we get:

$$T_k(m)f(\tau) = m^{k-1} \sum_{a,b,d} d^{-k} \sum_{n=0}^{\infty} a_n e^{2\pi i n(a\tau+b)/d} ,$$

where the outer sum is taken over the integers a, b, d of the lemma. The sum

$$\sum_{b=0}^{d-1} e^{2\pi i b n/d}$$

is equal to d if d divides n and equal to 0 otherwise. Hence putting together all the terms with the same power of $q = e^{2\pi i \tau}$, we find

$$T_k(m)f(\tau) = m^{k-1} \sum_{\substack{ad=m \\ n \geqslant 0}} d^{1-k} a(nd) q^{an} = \sum a_n(m) q^n ,$$

where $a_n(m)$ has the value stated in the theorem. This proves what we wanted.

It also shows that $T_k(m)f$ is a modular form, i.e. has a q-expansion at infinity.

The most important case of the transformation of coefficients is when m is a prime, and explicitly we have:

$$T_k(p)f = \sum a_n(p) q^n$$

where

$$a_n(p) = \begin{cases} a_{pn} & \text{if } p \nmid n \\ a_{pn} + p^{k-1} a_{n/p} & \text{if } p \mid n. \end{cases}$$

§ 2. Euler Products

Let p be a prime number, and let λ be a function from the set of powers p^r ($r \geqslant 0$) into a commutative ring. Let X be a variable.

Lemma. *We have the relation*

$$\frac{1}{1 - \lambda(p)X + p^{k-1}X^2} = \sum_{r=0}^{\infty} \lambda(p^r) X^r$$

if and only if λ satisfies the conditions:

(i) $\lambda(1) = 1$.

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

Proof. Immediate, by multiplying

$$1 - \lambda(p)X + p^{k-1}X^2$$

with the power series $\sum \lambda(p^r)X^r$. The condition that the product is equal to 1 is trivially equivalent with (i) and (ii).

Letting $X = p^{-s}$ (viewed formally as a variable), one may form the formal Dirichlet series

$$\sum \frac{T_k(p^r)}{p^{rs}}$$

and the lemma shows that it can be expressed as a rational function of p^s as the expression on the left.

Since the Hecke operators $T_k(n)$ are multiplicative, this further means that we have the formal product

$$\sum T_k(n) n^{-s} = \prod_p (1 - T_k(p)p^{-s} + p^{k-1-2s})^{-1} ,$$

taken over all primes p.

Let $\mathscr{H} = \mathscr{H}_{\mathbf{C}}$ be the algebra generated by all operators $T_k(n)$, for all positive integers n. We call $\mathscr{H}$ the **Hecke algebra**. According to Theorem 1.2, it is also generated by the operators $T_k(p)$ for all primes p. If ψ is a homomorphism of this algebra into the complex numbers, then the values

$$\psi(T_k(p^r)) = \lambda(p^r)$$

satisfy the conditions of the lemma, and the function

$$n \mapsto \psi(T_k(n)) = \lambda(n)$$

is multiplicative.

We obtain such λ as follows. Let f be a modular form of weight k. Suppose that f is an eigenfunction of all Hecke operators $T_k(n)$, i.e. an eigenfunction of the algebra $\mathscr{H}$. Then

$$T_k(n)f = \lambda(n)f\,,$$

with an eigenvalue $\lambda(n)$, such that the function ψ is a character of $\mathscr{H}$, i.e. an algebra homomorphism into the complex numbers. In particular these values satisfy the same recurrence formula as the Hecke operators themselves.

Theorem 2.1. *Let f be a modular form of weight k, not identically zero, and eigenfunction of the Hecke algebra. Let*

$$f = \sum a_n q^n$$

be its q-expansion. Then:

(i) *We have $a_1 \neq 0$.*

(ii) *If f is normalized so that $a_1 = 1$, then*

$$a_n = \lambda(n)$$

where ψ is the eigencharacter of $\mathscr{H}$ on f.

Proof. By Theorem 1.4, the coefficient of q in $T_k(n)f$ is equal to a_n. Since f is an eigenfunction with eigenvalue $\lambda(n)$, this coefficient is also equal to

$$\lambda(n)a_1\,.$$

If $a_1 = 0$ then $a_n = 0$ for all $n > 0$, so f is constant, which is impossible. Hence $a_1 \neq 0$. We may therefore assume without loss of generality that $a_1 = 1$, in which case we get

$$a_n = \lambda(n)\,,$$

as was to be shown.

Thus we see that if f is an eigenfunction for the Hecke algebra, normalized to have $a_1 = 1$, then there is a character ψ_f of $\mathscr{H}$ such that

$$\psi_f : T_k(n) \mapsto a_n\,.$$

In particular, the Dirichlet series has an Euler product

$$\sum a_n n^{-s} = \prod (1 - a_p p^{-s} + p^{k-1-2s})^{-1}$$

which converges absolutely for $\operatorname{Re} s > k/2 + 1$ by I, § 4, Lemma 2.

Example 1. Suppose that the integer k is such that M_k^0 has dimension 1. Then any basis element of M_k^0 is necessarily an eigenfunction for the Hecke algebra, and consequently the associated Dirichlet series has an Euler product.

Example 2. *Let G_k be the function of lattices given by*

$$G_k(L) = \sum_{\substack{w \in L \\ w \neq 0}} \frac{1}{w^k}\,,$$

with some even integer $k > 2$. Then G_k is an eigenfunction of the Hecke algebra.

Proof. It suffices to prove the eigenproperty for the operators $T_k(p)$ with prime p. Then

$$T_k(p)G_k(L) = p^{k-1} \sum_{(L : L') = p} G_k(L')\,.$$

The sum $G_k(L')$ can be decomposed as the sum over those elements w which lie only in the sublattice L' and no other sublattice of index p in L; and the sum over those elements w which lie in at least two such sublattices, or equivalently, lie in pL. It then follows at once that

$$\sum_{L'} G_k(L') = G_k(L) + pG_k(pL) = (1 + p^{1-k})G_k(L)\,.$$

This proves what we wanted.

We see from this example that in the direct sum decomposition

$$M_k = M_k^0 \oplus (G_k)\,,$$

the Hecke operators leave each one of these spaces invariant.

Chapter III. The Petersson Scalar Product

We first define the Riemann surface obtained by taking the quotient of the upper half plane by a subgroup Γ of $SL_2(\mathbf{Z})$, of finite index, and we show how to complete it to a compact Riemann surface X_Γ. We then define modular forms and cusp forms for such subgroups. In a sense, these generalize the notion of differential form of the first kind on the Riemann surface defined above. Just as one can define a scalar product for differentials of the first kind on X_Γ, one can extend the definition of this product to arbitrary cusp forms. The Hecke operators act essentially as a trace mapping, from one level to another. They act as Hermitian operators with respect to this scalar product.

§ 1. The Riemann Surface $\Gamma\backslash\mathfrak{H}^*$

We are mainly concerned with subgroups of $SL_2(\mathbf{Z})$ which are of finite index. These are discrete in $SL_2(\mathbf{R})$. However, if Γ is such a subgroup and $\alpha \in GL_2^+(\mathbf{Q})$ is a rational matrix with positive determinant, then it becomes essential to consider also the conjugate subgroup

$$\alpha\Gamma\alpha^{-1}$$

which is not necessarily contained in $SL_2(\mathbf{Z})$. For instance, one wants to operate on various objects with elements α in

$$\mathbf{M}_2^+(\mathbf{Z}),$$

the set of integral matrices with positive determinant. Such operations give rise to conjugations of groups canonically associated with these objects.

Consequently, we call a subgroup of $SL_2(\mathbf{R})$ **admissible** if it is conjugate to a subgroup of finite index in $SL_2(\mathbf{Z})$ by some matrix in $GL_2^+(\mathbf{Q})$. The set of admissible subgroups is closed under conjugation by such matrices, and any admissible subgroup is discrete in $SL_2(\mathbf{R})$.

We let $\mathfrak{H}^*$ be the union of the upper half plane $\mathfrak{H}$ with a symbol ∞, and the rational numbers $\mathbf{Q}$. One sometimes writes $i\infty$ to suggest the vizualization ordinarily associated with this situation. Rational numbers or ∞ are called **cusps**.

If Γ is an admissible group, we want to make $\Gamma\backslash\mathfrak{H}^*$ into a compact Riemann surface. We first define a topology, and then define complex analytic charts as follows.

We define a fundamental system of neighborhoods of ∞ to consist of the open sets U in $\mathfrak{H}$ (together with ∞) such that U is the part of the upper half plane lying above some horizontal line. We define a fundamental system of neighborhoods of a rational number r to be the union of r and the inside of a circle in the upper half plane tangent to the real line at r. These neighborhoods look as on the figure.

We note that $SL_2(\mathbf{Z})$ operates transitively on the cusps, since any relatively prime a, c can be completed to a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

in $SL_2(\mathbf{Z})$. We shall use the standard abbreviation

$$\Gamma(1) = SL_2(\mathbf{Z}).$$

We shall first describe the situation when Γ is contained in $SL_2(\mathbf{Z})$, and then make appropriate remarks for the more general admissible groups. *Until otherwise specified, we assume* $\Gamma \subset SL_2(\mathbf{Z})$.

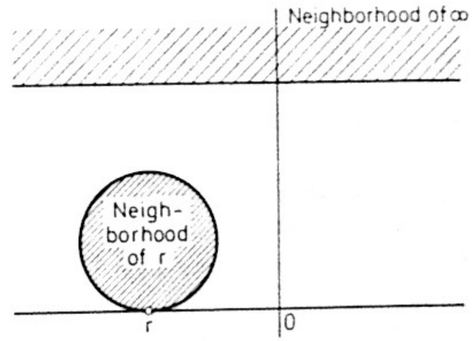

Fig. 3.

A fundamental system of neighborhoods of a point in $\mathfrak{H}$ is the usual one.

If $v \in \mathfrak{H}^*$ we let Γ_v be the isotropy group of v, that is the set of elements $\gamma \in \Gamma$ such that $\gamma v = v$. It is easily seen that $\Gamma(1)_\infty$ consists of the matrices

$$\pm\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

Thus Γ_∞ is a subgroup of finite index in $\Gamma(1)_\infty$, and there is a smallest positive integer e such that

$$\pm\begin{pmatrix} 1 & e \\ 0 & 1 \end{pmatrix}$$

lies in Γ_∞. We call e the **ramification index of Γ at ∞**.

Since $\Gamma(1)=SL_2(\mathbf{Z})$ operates transitively on the cusps, given any cusp s, there exists $\alpha\in\Gamma(1)$ such that $\alpha s=\infty$, and an element $\gamma\in\Gamma$ is such that $\gamma s=s$ if and only if

$$\alpha\gamma\alpha^{-1}(\infty)=\infty\ .$$

Thus the isotropy group of s in Γ can always be conjugated to the isotropy group of ∞ for a conjugate of Γ.

Furthermore, α transforms one of our fundamental systems of neighborhoods of s into a fundamental system of neighborhoods of ∞.

For many questions, this reduces the study of a neighborhood of s to the study of a neighborhood of ∞.

The first few pages of Shimura's book give all details for a complete description of $\Gamma\backslash\mathfrak{H}^*$. As this is exceedingly boring, we shall not reproduce the arguments here, only state the main result. Note that Shimura works with more general types of groups, and that for subgroups of $SL_2(\mathbf{Z})$, some simplifications occur. They arise from the formula

$$\operatorname{Im}\gamma z=\frac{\operatorname{Im} z}{|cz+d|^2}$$

if $\gamma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$. Since c is 0 or an integer, one sees trivially that if $c\neq 0$ then $|c|\geqslant 1$, and we thus get a bound for the imaginary part of γz in terms of that for z.

Theorem 1.1. (i) *The quotient $\Gamma\backslash\mathfrak{H}^*$ is compact Hausdorff.*
(ii) *The orbit space of the cusps under Γ is finite, of cardinality $\leqslant(\Gamma(1):\Gamma)$.*
(iii) *The set of charts defined below makes $\Gamma\backslash\mathfrak{H}^*$ into a Riemann surface, i.e. gives it a complex analytic structure.*

The charts are defined as follows. We make use of a property, which already intervenes in the proof of (i), namely:

For every element $v\in\mathfrak{H}^*$ there exists a neighborhood U such that, if $\gamma\in\Gamma$ and $\gamma U\cap U$ is not empty, then $\gamma v=v$.

This property is easily proved by distinguishing separately points of $\mathfrak{H}$ and cusps, and making use of the formula for the imaginary parts of γz and z. Hence we have an injection

$$\Gamma_v\backslash U\to\Gamma\backslash\mathfrak{H}^*\ ,$$

and $\Gamma_v\backslash U$ is an open neighborhood of the projection of v in $\Gamma\backslash\mathfrak{H}^*$.

Case 1. $z\in\mathfrak{H}$ and Γ_z contains only 1 or -1. Then the map $U\to\Gamma\backslash U$ is a homeomorphism. We take the inverse mapping as a chart.

Case 2. $z\in\mathfrak{H}$ and Γ_z contains other elements besides 1 or possibly -1. Let $\overline{\Gamma}_z=\Gamma_z$ or $\Gamma_z/\{\pm 1\}$ according to these two cases. Then $\overline{\Gamma}_z$ is cyclic of order d. In fact, we already know that $d=2$ or 3, according as z is a translate of i or $e^{2\pi i/3}$, but no matter. Let

$$\lambda:\mathfrak{H}\to D$$

be a holomorphic isomorphism of $\mathfrak{H}$ onto the unit disc such that $\lambda z=0$. For instance, if $z=i$, then

$$\lambda z=\frac{z-i}{z+i}\cdot$$

Then $\lambda\overline{\Gamma}_z\lambda^{-1}$ consists of the transformations

$$w\mapsto\zeta w\ ,$$

where ζ ranges over the d-th roots of unity. We let the chart be the map $\Gamma_z\backslash U\to\mathbf{C}$ given by

$$w\mapsto\lambda(w)^d\ .$$

This map is clearly a homeomorphism onto an open subset of $\mathbf{C}$.

Case 3. $v=s$ is a cusp. Let $\alpha\in SL_2(\mathbf{Z})$ be such that

$$\alpha s=\infty\ .$$

Let Γ_s be the isotropy group at s. Then

$$\alpha\Gamma_s\alpha^{-1}\cdot\{\pm 1\}$$

consists of

$$\pm\begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix},$$

where n ranges over all integral multiples of some fixed *positive* integer m. Then the map

$$\tau\mapsto e^{2\pi i\alpha(\tau)/m}$$

is taken as a chart in a neighborhood of s, i.e. on $\Gamma_s\backslash U$.

It is then easy to verify that the above charts give a complex analytic structure to $\Gamma\backslash\mathfrak{H}^*$.

By using a non-singular model of the modular function field as described in Shimura [Sh 2] or [L 2], Chapter 6, one can get an explicit embedding of this

Riemann surface in projective space, i.e. a realization as an algebraic curve. In fact, there exists a model, denoted by X_Γ, defined over a number field, such that we have a complex analytic isomorphism

$$X_\Gamma(\mathbf{C}) \to \Gamma\backslash\mathfrak{H}^* ,$$

where $X_\Gamma(\mathbf{C})$ is the set of complex points of X_Γ. Here we shall not deal with this algebraization, so we use X_Γ to denote $\Gamma\backslash\mathfrak{H}^*$.

We let Y_Γ be the image of $\mathfrak{H}$ itself in the natural projection

$$\pi_\Gamma : \mathfrak{H}^* \to \Gamma\backslash\mathfrak{H}^* .$$

Thus X_Γ is the union of Y_Γ and the cusps (i.e. orbits of the cusps in $\mathfrak{H}^*$ under Γ). Note that X_Γ and Y_Γ differ by a finite set of points.

We denote by $\overline{\Gamma}$ the factor group $\Gamma\{\pm 1\}/\{\pm 1\}$, so that $\overline{\Gamma}$ operates effectively on $\mathfrak{H}$.

Suppose that $\Gamma \supset \Gamma'$ are two subgroups. Then we have a natural map

$$X_{\Gamma'} \to X_\Gamma ,$$

which is a ramified covering. The only possible ramification points are the cusps, or those points which are images of elements $z \in \mathfrak{H}$ such that the isotropy group of z in $\Gamma(1)$ is larger than ± 1. Since we know these points explicitly, it is usually an easy matter to determine the ramified points and ramification index explicitly in concrete examples. We shall recall some congruence subgroups below, and the computation of the ramification is carried out explicitly in Shimura [Sh 2] and Ogg [Ogg 1] or Schoeneberg [Sch].

Similarly, one has the genus formula of Hurwitz, comparing the genus in a base and a (ramified) covering. This allows the explicit computation of the genus for the standard congruence subgroups, also computed in the above mentioned references. We note that the degree of the covering

$$X_{\Gamma'} \to X_\Gamma$$

is given by the index

$$n = (\overline{\Gamma} : \overline{\Gamma}') .$$

This degree enters in the Hurwitz formula:

$$2g' - 2 = n(2g-2) + \sum (e_Q - 1) ,$$

where:

g, g' are the genera of the curves X_Γ, $X_{\Gamma'}$ respectively;
n is the degree;
e_Q is the ramification index at a point Q of $X_{\Gamma'}$;
and the sum is taken over all points of $X_{\Gamma'}$.

For a proof, cf. [L 3], for instance.

We now make additional remarks on $\Gamma'\backslash\mathfrak{H}$ when $\Gamma' = \alpha\Gamma\alpha^{-1}$ for some $\alpha \in \mathbf{M}_2^+(\mathbf{Z})$, or what amounts to the same thing,

$$\alpha \in GL_2^+(\mathbf{Q}) ,$$

and Γ is contained in $SL_2(\mathbf{Z})$. Then

$$z \mapsto \alpha(z)$$

induces an analytic isomorphism of $\mathfrak{H}$ with itself, and the operation of α permutes the cusps. Thus we can define the complex analytic structure of $\Gamma'\backslash\mathfrak{H}^*$ as that obtained by pull back from the mapping $z \mapsto \alpha(z)$. It must then be verified that if Γ' happens to be contained in $SL_2(\mathbf{Z})$, then this pull back coincides with the analytic structure which we have already defined. This requires a specific analysis at each type of point: Ordinary in $\mathfrak{H}$, special in $\mathfrak{H}$ (i.e. having non-trivial isotropy group), and the cusps, handled by first making an isomorphism of a neighborhood of the cusp with a neighborhood of infinity, and then analysing the behavior of a local parameter at infinity. We shall omit the tedious but essentially straightforward arguments which prove this. If the foundations are covered in sufficient generality as in Shimura, then the fact is obvious.

§ 2. Congruence Subgroups

The most important subgroups of $SL_2(\mathbf{Z})$ for our purposes are the following. Let N be a positive integer.

$\Gamma(N)$ consists of those elements γ in $SL_2(\mathbf{Z})$ such that

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv I \pmod{N} ,$$

that is,

$$a \equiv d \equiv 1 \pmod{N} \qquad \text{and} \qquad c \equiv b \equiv 0 \pmod{N} .$$

$\Gamma_0(N)$ consists of the matrices γ with $c \equiv 0 \pmod{N}$.

$\Gamma_1(N)$ consists of the matrices γ with

$$\gamma \equiv \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \pmod{N} ,$$

where b is arbitrary.

In each case respectively, the quotient curve is denoted by

$$X(N), \qquad X_0(N), \qquad X_1(N)$$

and the complement of the cusps by

$$Y(N), \qquad Y_0(N), \qquad Y_1(N),$$

instead of X_Γ and Y_Γ.

We have $\Gamma(1)=SL_2(\mathbf{Z})$.

The mapping

$$j: \Gamma(1)\backslash\mathfrak{H} \to \mathbf{C}$$

of Theorem 3.1, Chapter I, can now be interpreted as giving an affine embedding for $Y(1)$ into projective space (of dimension 1!), and gives a complex analytic isomorphism. The point ∞ in $\mathfrak{H}^*$ goes to the point at infinity in the projective line.

Example. The ramification of the covering

$$X(N) \to X(1) = \text{projective } j\text{-line}$$

for $N > 1$ is easily determined to be:

of order N above infinity;
of order 2 above i;
of order 3 above $e^{2\pi i/3}$.

It is easy to compute the index

$$(\Gamma(1) : \Gamma(N)) = N^3 \prod_{p \mid N}\left(1-\frac{1}{p^2}\right).$$

Since $-1 \in \Gamma(2)$ but $-1 \notin \Gamma(N)$ for $N>2$, one finds:

$$(\overline{\Gamma}(1) : \overline{\Gamma}(N)) = \begin{cases} \frac{1}{2}N^3 \prod_{p \mid N}\left(1-\frac{1}{p^2}\right) & \text{if } N>2 \\ 6 & \text{if } N=2. \end{cases}$$

From this and the genus formula of Hurwitz, one finds the genus of $X(N)$ to be

$$g_N = 1 + \frac{N-6}{12N}(\overline{\Gamma}(1) : \overline{\Gamma}(N)), \qquad \text{for } N>1.$$

We repeat that the computations are carried out in detail in Shimura [Sh 2] and [Ogg 1].

One can also describe the cusps explicitly for $X(N)$. We note that a rational number can be expressed uniquely as a quotient

$$\frac{r}{s} = \frac{-r}{-s},$$

where r, s are relatively prime integers.

Two vectors (r, s) and (r', s') of relatively prime integers are congruent mod N *if and only if there exists an element*

$$\gamma \in \Gamma(N)$$

such that

$$\gamma\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} r' \\ s' \end{pmatrix}.$$

We leave the proof as an easy exercise to the reader, who will work it out faster than looking it up (e.g. in Shimura). Using this one concludes:

Two cusps $v = r/s$ and $v' = r'/s'$ expressed as quotients of relatively prime integers are in the same orbit of $\Gamma(N)$, if and only if

$$\pm\begin{pmatrix} r \\ s \end{pmatrix} \equiv \begin{pmatrix} r' \\ s' \end{pmatrix} \pmod{N}.$$

This essentially describes the story complex analytically for $\Gamma(N)$. For a description of the modular function field, and an algebraic description of the ramification, cf. [L 3], Chapter IX, § 3.

The analogous facts for $\Gamma_0(N)$ will be found in [Sh 2], end of Chapter 1.

A **congruence subgroup** of $SL_2(\mathbf{Z})$ is by definition a subgroup which contains $\Gamma(N)$ for some N. Non congruence subgroups are interesting both for their own sake and for their interrelation with the congruence subgroups. For a discussion of the (large) extent to which non-congruence subgroups occur, we refer to Bass-Milnor-Serre [BMS]. For a connection with congruence subgroups which may be of particular interest for diophantine analysis, we refer to Kubert-Lang [KL I].

We conclude the section by a remark which shows that the study of $\Gamma(N)$ can sometimes be reduced to the study of $\Gamma_1(N')$ for some N', and conversely. Let

$$\eta = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}.$$

Then η defines an inner automorphism of $GL_2(\mathbf{Q})$, by letting

$$\alpha^* = \eta^{-1}\alpha\eta.$$

If

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then

$$\alpha^* = \begin{pmatrix} a & b/N \\ cN & d \end{pmatrix}.$$

From this we conclude at once:

$$\Gamma(N)^* = \Gamma_0(N^2) \cap \Gamma_1(N) ,$$

or in other words, $\Gamma(N)^*$ consists of the matrices having the described congruence properties:

$$\begin{pmatrix} 1 \bmod N & * \\ 0 \bmod N^2 & 1 \bmod N \end{pmatrix}.$$

In particular, $\Gamma(N)^*$ contains $\Gamma_1(N^2)$. Thus a conjugate of $\Gamma(N)$ contains $\Gamma_1(N^2)$. Groups like $\Gamma_1(N)$ are useful because they contain elements of the form

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$$

which are easy to handle problems concerning cusps. They have the disadvantage that they do not behave nicely with respect to Galois theory since they are not normal in $\Gamma(1)$. Conversely, $\Gamma(N)$ behave very well with respect to Galois theory, but less well in the other respect.

§ 3. Differential Forms and Modular Forms

Let X be a compact Riemann surface. We let $\Omega^1(X)$ be the complex space of differentials of first kind on X. It has dimension g, where g is the genus of X. We assume known the basic facts of Riemann surfaces, cf. for instance [L 3].

Suppose that $X = X_\Gamma$ for some Γ, as before. Let $\omega \in \Omega^1(X)$. Under the map

$$\pi : \mathfrak{H} \to \Gamma \backslash \mathfrak{H} \subset X ,$$

we can take the pull back

$$\pi^* \omega, \text{ also denoted by } \omega \circ \pi .$$

Then $\omega \circ \pi$ is a holomorphic differential form on $\mathfrak{H}$, which can be written in the uniformizing parameter $z = \tau$,

$$\omega \circ \pi = f(\tau)\, d\tau ,$$

with some holomorphic function f on $\mathfrak{H}$. By abuse of notation, we often write ω instead of $\omega \circ \pi$. We note that ω (i.e. $\omega \circ \pi$) is invariant under Γ. But we have

$$\omega \circ \gamma = f(\gamma z)\, d(\gamma z) = f(\gamma z)(cz + d)^{-2}\, dz .$$

Hence

$$f \circ [\gamma]_2 = f .$$

Furthermore, let

$$T^m = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

where m is the ramification index of Γ at infinity. Since ω is invariant under T^m, we can write

$$f(z) = f_\infty(q^{1/m}) = \sum a_n q^{n/m} = \sum a_n\, e^{2\pi i n z/m} ,$$

where f_∞ is a power series in $q^{1/m} = e^{2\pi i z/m}$. Since

$$d(q^{1/m}) = q^{1/m} \frac{2\pi i}{m}\, dz ,$$

we see that

$$f(z)\, dz = f_\infty(q^{1/m}) \frac{m}{2\pi i} \frac{d(q^{1/m})}{q^{1/m}} .$$

We know that $q^{1/m}$ is a local parameter at infinity. Hence ω is holomorphic at infinity if and only if f_∞ has a zero at infinity, i.e. the power series is of the type

$$f_\infty = \sum_{n=1}^{\infty} a_n q^{n/m} .$$

Let s be a cusp for Γ, and let $\alpha \in SL_2(\mathbf{Z})$ be such that $\alpha(s) = \infty$. Then $\omega \circ \alpha^{-1}$ is holomorphic at infinity, and can be written in the form

$$\omega \circ \alpha^{-1}(z) = h(z)\, dz ,$$

for some holomorphic function h on $\mathfrak{H}$. The same analysis as above shows that there exists some positive integer m for which h has a power series expansion h_∞ in terms of $q^{1/m}$, and the same argument as before shows that the power series has a constant term equal to 0.

Let k be an integer $\geqslant 0$. We define a **modular form** of **weight** k on Γ to be a holomorphic function f on $\mathfrak{H}$, which satisfies

$$f \circ [\gamma]_k = f , \qquad \text{all} \quad \gamma \in \Gamma ,$$

and such that $f \circ \alpha$ is holomorphic at all cusps, for $\alpha \in SL_2(\mathbf{Z})$.

Warning. In making explicit the power series expansion at a cusp with respect to a chart which first maps the cusp to infinity, we don't care which fractional power of q occurs. However, one must be careful about the presence of ± 1 in Γ if one wants to get the exact denominator of this fraction. Cf. Shimura [Sh 2], Chapter 2, p. 29, where regular and irregular cusps are discussed according to the behavior

of -1 with respect to Γ. This plays no role for what we want to do in the sequel, so we omit this discussion.

We define a **cusp form** for Γ to be a modular form which has a zero at each cusp.

Theorem 3.1. *The map which to each differential of first kind $\omega = f(z)\,dz$ on X_Γ associates the function f, is an isomorphism between $\Omega^1(X_\Gamma)$ and the space of cusp forms of weight 2 with respect to Γ.*

Proof. The discussion at the beginning of this section showed that the map gives an injection of $\Omega^1(X_\Gamma)$ into the space of cusp forms. Conversely, given a cusp form f of weight 2, one puts

$$\omega(z) = f(z)\,dz\,,$$

and it is then easily verified from the definition of the analytic structure $\Gamma\backslash\mathfrak{H}$ that $f(z)\,dz$ is invariant under Γ, and gives a differential of first kind on X_Γ. We omit the details.

A similar discussion can be given for other weights, by considering

$$f(z)\,(dz)^k\,.$$

For any subgroup Γ of $SL_2(\mathbf{Z})$, of finite index, we let

$$M(\Gamma, k)$$

be the space of modular forms on Γ of weight k. Then $M(\Gamma, k)$ can be identified with a vector space as in the Riemann-Roch theorem, in a manner similar to the case when $k=2$. For the general discussion with arbitrary k, and a computation of the dimensions of the spaces $M(\Gamma, k)$ for arbitrary Γ, and $k \neq 1$, see Shimura [Sh 2], Chapter 2, Theorem 2.24 *et seq.*

It is a major problem to determine the dimension in case $k=1$. This ties up with the theory of representations and the existence of Galois extensions of the rationals, cf. Deligne-Serre [D-S].

It is sometimes useful to consider modular forms which fail to satisfy the condition of holomorphy, merely assume that they are meromorphic on $\mathfrak{H}$ and at infinity. The space of such forms is denoted by

$$\mathscr{F}(\Gamma, k)\,.$$

If $\Gamma = \Gamma(N)$, we also write this space as $\mathscr{F}(N, k)$. If $\Gamma = \Gamma_1(N)$, we write this space as $\mathscr{F}_1(N, k)$.

We let $M(k)$ and $M^0(k) = \mathbf{S}(k)$ be the unions of all the spaces $M(N, k)$ and $M^0(N, k) = \mathbf{S}(N, k)$ respectively. Then one sees at once that the operation of *rational* matrices $[\alpha]_k$, with $\alpha \in GL_2^+(\mathbf{Q})$, leaves these spaces stable, even though

such operations change the level of a given form. Indeed, instead of rational matrices, one may use only primitive integral matrices, after multiplication by an appropriate scalar, which has trivial action, and then one uses the decomposition of such a matrix in the form

$$\gamma \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

with $\gamma \in SL_2(\mathbf{Z})$. Triangular matrices as above do not change the property of being holomorphic at infinity, or of having a zero at infinity.

§ 4. The Petersson Scalar Product

We assume throughout that k is a positive integer.

If Γ is a subgroup of $SL_2(\mathbf{Z})$ we denote by $\overline{\Gamma}$ its projectivization, i.e.

$$\overline{\Gamma} = \Gamma \cdot \{\pm 1\}/\{\pm 1\}\,.$$

We first make some remarks on admissible groups, i.e. groups of the form

$$\alpha\Gamma\alpha^{-1}\,,$$

where Γ is of finite index in $SL_2(\mathbf{Z})$ and $\alpha \in M_2^+(\mathbf{Z})$.

Suppose that α has a determinant N. If $\gamma \in \Gamma(N)$, we can write

$$\gamma = 1 + N\beta\,,$$

where β is a matrix with integer coefficients. Hence

$$\alpha\gamma\alpha^{-1} = 1 + N\alpha\gamma\alpha^{-1}$$

also has integer coefficients, and therefore lies in $SL_2(\mathbf{Z})$. In other words,

$$\alpha\Gamma(N)\alpha^{-1} \subset \Gamma(1)\,,$$

or equivalently

$$\Gamma(N) \subset \alpha^{-1}\Gamma(1)\alpha\,.$$

If Γ is of finite index in $SL_2(\mathbf{Z})$ and $\alpha \in M_2^+(\mathbf{Z})$, then there exists a subgroup Γ' of finite index in $SL_2(\mathbf{Z})$ contained in $\alpha\Gamma\alpha^{-1}$.

Proof. We look at the diagram

$$\begin{array}{ccc} & & \alpha\Gamma\alpha^{-1} \to \alpha\Gamma(1)\alpha^{-1} \\ & & \uparrow \qquad\qquad \uparrow \\ \Gamma' = (\alpha\Gamma\alpha^{-1}) \cap (\alpha\Gamma(N)\alpha^{-1}) & \to & \alpha\Gamma(N)\alpha^{-1} \end{array} \qquad \text{finite index} = (\Gamma(1) : \Gamma(N))$$

and use the preceding remarks, together with the fact that

$$\Gamma \cap \Gamma(N)$$

is of finite index in $SL_2(\mathbf{Z})$ to get a proof.

Let Γ be an admissible group. A fundamental domain for Γ is a subset F of $\mathfrak{H}$ whose boundary consists of a finite number of analytic arcs, and which contains exactly one element from each Γ-orbit.

If F_Γ is a fundamental domain for Γ, if $\Gamma' \subset \Gamma$, and

$$\overline{\Gamma} = \bigcup_{i=1}^{r} \gamma_i \overline{\Gamma}'$$

is a coset decomposition, then it is clear that a fundamental domain for Γ' is given by the union

$$F_{\Gamma'} = \bigcup_{i=1}^{r} \gamma_i F_\Gamma .$$

Furthermore, αF_Γ is a fundamental domain for $\alpha \Gamma \alpha^{-1}$.

Since we know a fundamental domain for $\Gamma(1)$, it follows that we can construct a fundamental domain for an arbitrary admissible group by a succession of the above two constructions.

In the sequel, a fundamental domain for a subgroup Γ of $\Gamma(1)$ will always be assumed to consist of a finite number of translations by elements of $\Gamma(1)$ of a fundamental domain for $\Gamma(1)$.

Lemma 1. *Let f be a cusp form, for some subgroup of $\Gamma(1)$. Then there exists $c>0$ such that for $\gamma \in SL_2(\mathbf{Z})$ we have*

$$|f(\gamma(x+iy))| \ll e^{-cy}$$

for all y sufficiently large.

Proof. We have the q-expansion

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

starting with a term of order at least 1. Since $\sum a_n q^{n/m}$ converges in a neighborhood of $q=0$, it follows that

$$|f(\gamma(z))| \ll |q|^{1/m} = e^{-2\pi y/m},$$

as was to be shown.

Let f, g be cusp forms of weight k for an arbitrary admissible group. We know that this group always contains some subgroup of $SL_2(\mathbf{Z})$, say Γ. We define the **Petersson scalar product**

$$\langle f, g \rangle_\Gamma = \frac{1}{(\overline{\Gamma}(1):\overline{\Gamma})} \int_{F_\Gamma} f(z)\overline{g(z)} y^k \frac{dx\,dy}{y^2} .$$

We can write

$$dx \wedge dy = \frac{i}{2} dz \wedge d\bar{z} ,$$

and one then verifies trivially that the differential form

$$\frac{dx \wedge dy}{y^2} = \frac{i}{2} (\operatorname{Im} z)^{-2}\, dz \wedge d\bar{z}$$

is invariant under $GL_2^+(\mathbf{R})$, by substituting αz for z, and using the obvious transformation rules for $d(\alpha z)$ and $\operatorname{Im}(\alpha z)$.

The differential form under the integral sign satisfies a simple transformation law with respect to $GL_2^+(\mathbf{R})$. Let us give it a name,

$$\Omega(f, g) = f(z)\overline{g(z)} (\operatorname{Im} z)^k \frac{1}{(\operatorname{Im} z)^2} \frac{i}{2} dz \wedge d\bar{z} .$$

Then

$$\langle f, g \rangle_\Gamma = \frac{1}{(\overline{\Gamma}(1) : \Gamma)} \int_{F_\Gamma} \Omega(f, g) .$$

In particular, for $k=2$ when we can identify a cusp form of weight 2 and a differential of the first kind, if we put

$$\omega = f(z)\, dz \qquad \text{and} \qquad \eta = g(z)\, dz ,$$

then the scalar product can be written in the form

$$\langle f, g \rangle = \langle \omega, \eta \rangle = \frac{1}{(\overline{\Gamma}(1):\overline{\Gamma})} \int_{X_\Gamma} \frac{i}{2} \omega \wedge \bar{\eta} .$$

Of course, we must verify that the Petersson scalar product makes sense. In other words, we shall verify:

(i) The integral is independent of the fundamental domain.

(ii) It converges absolutely.

(iii) It is independent of the choice of Γ.

The independence of the fundamental domain will follow from the next lemma, and after we have proved the other two properties, we shall be able to write the scalar product

$$\langle f, g \rangle$$

without any index, defined for cusp forms of *any* admissible group.

Lemma 2. *If* $\alpha \in GL_2^+(\mathbf{R})$, *then*

$$\Omega(f \circ [\alpha]_k, g \circ [\alpha]_k) = \Omega(f, g) \circ \alpha .$$

Proof. This follows at once by replacing z with αz, and using the definitions:

$$f \circ [\alpha]_k(z) = f(\alpha z)(cz+d)^{-k} (\det \alpha)^{k/2}$$

$$g \circ [\alpha]_k(z) = g(\alpha z)(cz+d)^{-k} (\det \alpha)^{k/2}$$

$$d(\alpha z) = (cz+d)^2 (\det \alpha)\, dz$$

$$\operatorname{Im} \alpha z = \frac{(\operatorname{Im} z) \det \alpha}{|cz+d|^2} .$$

The relation drops out.

Taking α in Γ and using the change of variables formula

$$\int_{\alpha A} \Omega = \int_A \Omega \circ \alpha$$

for any set A, we see that the integral giving the scalar product is independent of the choice of fundamental domain.

For (ii), we note that a fundamental domain for Γ is a union of a finite number of translates of a fundamental domain for $\Gamma(1)$. It suffices to prove that the integral over the part going to infinity is convergent, and this follows from the lemma. The convergence is even very rapid, and it would suffice to assume that only one of f and g is a cusp form.

As for (iii), if we compute the scalar product with respect to two subgroups Γ_1, Γ_2 in $SL_2(\mathbf{Z})$, we can compare them with the scalar product with respect to the intersection $\Gamma_1 \cap \Gamma_2$, and the factor in front of the integral has been put there so that the final value is independent of the choice of Γ, as one verifies at once to conclude the proof.

Lemma 3. *Let* $\alpha \in GL_2^+(\mathbf{Q})$. *Assume that both* Γ *and*

$$\alpha \Gamma \alpha^{-1}$$

are contained in $SL_2(\mathbf{Z})$. *Then*

$$(\overline{\Gamma}(1):\overline{\Gamma}) = (\overline{\Gamma}(1):\alpha \overline{\Gamma} \alpha^{-1}) .$$

Proof. The measure $dx\, dy/y^2$ is invariant under $GL_2^+(\mathbf{R})$, and the total measure of $\Gamma \backslash \mathfrak{H}$ is finite. Conjugation by α preserves the measure, so that

$$\text{measure}(\Gamma \backslash \mathfrak{H}) = \text{measure}(\alpha \Gamma \alpha^{-1} \backslash \mathfrak{H}) .$$

Furthermore, the measure of $\Gamma \backslash \mathfrak{H}$ is equal to the index $(\overline{\Gamma}(1):\overline{\Gamma})$, because a fundamental domain for $\Gamma \backslash \mathfrak{H}$ consists of a finite number of translates of the fundamental domain for $\Gamma(1)$. The lemma follows trivially.

If $\alpha \in GL_2^+(\mathbf{R})$, we recall that

$$\alpha' = \alpha^{-1} \det \alpha .$$

Theorem 4.1. *Let* f, g *be cusp forms of weight* k. *Let* $\alpha \in GL_2^+(\mathbf{Q})$. *We have:*

(i) $\langle f \circ [\alpha]_k, g \circ [\alpha]_k \rangle = \langle f, g \rangle$.

(ii) $\langle f \circ [\alpha]_k, g \rangle = \langle f, g \circ [\alpha']_k \rangle$.

(iii) *The scalar product in* (ii) *depends only on the double coset* $\Gamma \alpha \Gamma$, *if* f, g *are invariant under* Γ.

Proof. We may assume $\alpha \in \mathbf{M}_2^+(\mathbf{Z})$. We have by Lemma 2,

$$\langle f \circ [\alpha]_k, g \circ [\alpha]_k \rangle = \frac{1}{(\overline{\Gamma}(1):\overline{\Gamma})} \int_F \Omega(f \circ [\alpha]_k, g \circ [\alpha]_k)$$

$$= \frac{1}{(\overline{\Gamma}(1):\overline{\Gamma})} \int_F \Omega(f, g) \circ \alpha$$

$$= \frac{1}{(\overline{\Gamma}(1):\overline{\Gamma})} \int_{\alpha F} \Omega(f, g) .$$

But αF is a fundamental domain for $\alpha \Gamma \alpha^{-1}$. The first part of the theorem now follows from Lemma 3, after picking Γ sufficiently small.

The second part follows from the first, and the fact that

$$g \circ [\alpha']_k = g \circ [\alpha^{-1}]_k .$$

The third part is obvious from (ii), since

$$f \circ [\gamma]_k = f \qquad \text{and} \qquad g \circ [\gamma^{-1}]_k = g$$

for $\gamma \in \Gamma$.

The theorem gives us explicitly the transpose of the operation on cusp forms by elements of $GL_2^+(\mathbf{Q})$. In particular, we see that this operation is *unitary* for the Petersson scalar product.

Suppose that $k=2$. Then we know that cusp forms of weight 2 can be identified with differential forms. If

$$\omega = f(z)\,dz\,, \qquad \eta = g(z)\,dz$$

are differentials of first kind on some X_Γ, expressed in terms of the complex variable in $\mathfrak{H}$, then we also write

$$\langle f, g\rangle = \langle \omega, \eta\rangle\,,$$

and the unitary property of the operators $[\alpha]_2$ can then be written in the form

$$\|\omega \circ \alpha\| = \|\omega\|\,,$$

if $\|\ \|$ is the norm associated with the Petersson hermitian product, or also in the form

$$\langle \omega \circ \alpha, \eta \circ \alpha\rangle = \langle \omega, \eta\rangle\,.$$

This is the notation which will be used in the applications of the next chapter, dealing with modular symbols.

We return to give some corollaries for arbitrary k, and $\Gamma = SL_2(\mathbf{Z})$.

Theorem 4.2. *On the space of cusp forms of weight k for $SL_2(\mathbf{Z})$, the Hecke operators are hermitian with respect to the Petersson scalar product.*

Proof. The scalar product

$$\langle f \circ [\alpha]_k, g\rangle$$

depends only on the double coset $\Gamma(1)\alpha\Gamma(1)$, by Theorem 4.1. Since the set of integral matrices $\mathbf{M}^p$ of determinant p has only one double coset, the result follows from Theorem 4.1(ii), and the fact that

$$\alpha \mapsto \alpha'$$

permutes the elements of $\mathbf{M}^p$.

Corollary 1. *The eigenvalues of $T_k(n)$ are totally real algebraic numbers.*

Proof. The space of cusp forms has a basis over $\mathbf{Q}$, so that the characteristic polynomial of $T_k(n)$ has rational coefficients. Its roots are the eigenvalues of $T_k(n)$, which are all real by elementary linear algebra. This proves the corollary. Also by linear algebra, we get:

Corollary 2. *The space of cusp forms of given dimension has a basis consisting of eigenvectors for the commutative algebra of Hecke operators.*

Proof. This is a general elementary fact about any finite dimensional hermitian vector space having a commutative algebra of hermitian operators.

For any ring R in $\mathbf{C}$ we let $M_k^0(R)$ be the module of cusp forms on $SL_2(\mathbf{Z})$ having q-expansion coefficients in R. We shall see later that $M_k^0(\mathbf{C})$ is the extension to $\mathbf{C}$ of the rational space $M_k^0(\mathbf{Q})$. We let $\mathscr{H}_\mathbf{C}$ be the algebra generated by all the operators $T_k(n)$ (or $T_k(p)$ for prime p) on $M_k^0(\mathbf{C})$. From the action of the operators $T_k(n)$ on cusp forms, it is clear that $T_k(n)$ maps $M_k^0(\mathbf{Q})$ into itself. Thus

$$\mathscr{H}_\mathbf{C} \approx \mathbf{C} \otimes \mathscr{H}_\mathbf{Q}\,,$$

where $\mathscr{H}_\mathbf{Q}$ is the algebra of Hecke operators on $M_k^0(\mathbf{Q})$.

Let $f_1, \ldots, f_r$ be normalized eigenfunctions of $\mathscr{H}_\mathbf{C}$ which form a basis for $M_k^0(\mathbf{C})$ according to Corollary 2. For each i, we have a character ψ_i of $\mathscr{H}_\mathbf{C}$ such that for all $T \in \mathscr{H}_\mathbf{C}$,

$$Tf_i = \psi_i(T) f_i\,.$$

If $\psi_i = \psi_j$ then $f_i = f_j$ because if $f = \sum a_n q^n$ is a normalized eigenfunction of $\mathscr{H}_\mathbf{C}$ with character ψ, then

$$T_k(n) f = a_n f \qquad \text{and} \qquad a_n = \psi(T_k(n))\,.$$

Theorem 4.3. *The map $T \mapsto (\psi_1(T), \ldots, \psi_r(T))$ is an isomorphism of $\mathscr{H}_\mathbf{C}$ with $\mathbf{C}^r$.*

Proof. The map is obviously injective. On the other hand, the standard Artin proof of elementary algebra shows that the distinct characters $\psi_1, \ldots, \psi_r$ are linearly independent. We reproduce the lemma below for the convenience of the reader, and this concludes the proof.

Lemma. *Let $\psi_1, \ldots, \psi_r$ be distinct characters of an algebra over a field. Then they are linearly independent.*

Proof. If there exists a shortest non-trivial linear relation, say

$$c_1\psi_1 + c_2\psi_2 + \cdots = 0\,,$$

with $c_1, c_2 \neq 0$, pick T_0 such that $\psi_1(T_0) \neq \psi_2(T_0)$. We evaluate the relation at both a variable element T, and at $T_0 T$. One of the values, say $\psi_1(T_0)$ cannot be equal to 0. We divide by $\psi_1(T_0)$ and subtract the two relations to give a shorter relation, contradiction.

Corollary. *$M_k^0(\mathbf{C})$ is a free module of rank 1 over $\mathscr{H}_\mathbf{C}$.*

Proof. If $\{f_1, \ldots, f_r\}$ is a basis of normalized eigenfunctions as above, let

$$f = f_1 + \cdots + f_r\,.$$

Then f is a basis element for $M_k^0(\mathbf{C})$ over $\mathscr{H}_\mathbf{C}$.

Remark. It is easy to see that the basis element f in the corollary in fact has Fourier coefficients in $\mathbf{Q}$, so that f is also a basis of $M_k^0(\mathbf{Q})$ over $\mathscr{H}_{\mathbf{Q}}$.

For arithmetic applications, it is useful to know that the space of cusp forms and Hecke operators over the complex numbers are extended from natural spaces over $\mathbf{Q}$, or even over $\mathbf{Z}$. We show this by using remarks from the beginning of Victor Miller's thesis, as follows.

Let

$$f = a_1 q + \cdots, \qquad a_1 = a_1(f),$$

be a cusp form. If T is in the Hecke algebra, we let

$$\langle f, T\rangle = a_1(Tf).$$

In Chapter X, Theorem 4.4, we shall prove:

Let $M_k^0(\mathbf{C})$ be the $\mathbf{C}$-vector space of cusp forms of weight k over $\mathbf{C}$, and let r be its dimension. There exists a basis $\{f_1, \ldots, f_r\}$ with q-expansion coefficients $a_i(f_j)$ such that

$$a_i(f_j) = \delta_{ij}, \qquad 1 \leqslant i, j \leqslant r,$$

and $a_n(f_j) \in \mathbf{Z}$ for all n and $j = 1, \ldots, r$.

Lemma 1. *Let $f \in M_k^0(\mathbf{C})$. Then*

$$f = a_1(f) f_1 + \cdots + a_r(f) f_r.$$

Proof. Clear, since $\{f_1, \ldots, f_r\}$ is a basis.

Corollary. *The cusp forms $f_1, \ldots, f_r$ are uniquely determined, and for any cusp form f we have*

$$\langle f, T_k(n)\rangle = a_1(T_k(n) f) = a_n(f).$$

Lemma 2. *For the basis $\{f_1, \ldots, f_r\}$ described above, we have*

$$\langle f_i, T_k(j)\rangle = \delta_{ij} \qquad \text{for} \quad i, j = 1, \ldots, r.$$

Proof. Immediate from the above.

Corollary. *The Hecke operators $1 = T_k(1), \ldots, T_k(r)$ form a $\mathbf{C}$-basis of $\mathscr{H}_{\mathbf{C}}$.*

Proof. The operators $T_k(j)$, $j = 1, \ldots, r$ are linearly independent by the above, and we can use Theorem 4.3 to conclude the proof.

Let $\mathscr{H}_k(\mathbf{Z})$ be the ring of Hecke operators generated by the operators $T_k(n)$ for all n over $\mathbf{Z}$.

Theorem 4.4. *The ring $\mathscr{H}_k(\mathbf{Z})$ is a free $\mathbf{Z}$-module with basis*

$$T_k(1), \ldots, T_k(r).$$

The association

$$(f, T) \mapsto \langle f, T\rangle$$

is a perfect pairing between $M_k^0(\mathbf{Z})$ and $\mathscr{H}_k(\mathbf{Z})$.

Proof. By the corollary of Lemma 2, there exist $b_{ij} \in \mathbf{C}$ such that

$$T_k(i) = b_{i1} T_k(1) + \cdots + b_{ir} T_k(r).$$

It suffices to prove that $b_{ij} \in \mathbf{Z}$ for the first assertion. However

$$a_i(f_j) = \langle f_i, T_k(j)\rangle = b_{i1}\langle f_j, T_k(1)\rangle + \cdots + b_{ir}\langle f_j, T_k(r)\rangle = b_{ij}$$

by Lemma 2. But $a_i(f_j) \in \mathbf{Z}$ so the first assertion is proved. The second is then immediate from the fact that $f_1, \ldots, f_r$ form a $\mathbf{Z}$-basis of $M_k^0(\mathbf{Z})$.

Appendix by D. Zagier
The Eichler–Selberg Trace Formula on $SL_2(\mathbf{Z})$

Throughout this appendix we let $\Gamma = \Gamma(1) = SL_2(\mathbf{Z})$. We let F be a fundamental domain for Γ in $\mathfrak{H}$. We fix a weight k even ≥ 4. We write $T(m)$ instead of $T_k(m)$ for the Hecke operator on the space of cusp forms $S_k = M_k^0$.

Let $h(z, z')$ be a function of two variables z, z' in $\mathfrak{H}$, and assume that h as a function of each variable is a cusp form of weight k. If $f \in S_k$ then we define $f * h$ as a function of z' by

$$f * h(z') = \int_F f(z)\overline{h(z, -\bar{z}')}((\operatorname{Im} z)^k \frac{dx\,dy}{y^2}. \tag{1}$$

Thus this operation is merely the Petersson scalar product of f and h, viewed as a function of the first variable z. The purpose of this appendix is to show that the Hecke operator $T(m)$ can be represented by a kernel h_m, and to give a formula for its trace on S_k.

We let $f_1, \ldots, f_r$ be a basis of eigenfunctions for the Hecke operators, and assume that they are normalized, i.e.

$$f_i = \sum_{n=1}^{\infty} a_n^i q^n, \qquad a_1^i = 1, \qquad T(m) f_i = a_m^i f_i\,. \tag{2}$$

Note that this basis of eigenfunctions is orthogonal for the Petersson scalar product.

For each positive integer m we define

$$h_m(z, z') = \sum_{ad-bc=m} (czz' + dz' + az + b)^{-k}\,, \tag{3}$$

where the sum is taken over all integer matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with determinant m. We may also write h_m in the form

$$h_m(z, z') = \sum_{ad-bc=m} (cz+d)^{-k}\left(z' + \frac{az+b}{cz+d}\right)^{-k}. \tag{4}$$

The imaginary part of

$$z' + \frac{az+b}{cz+d}$$

is >0, so this expression never vanishes, and each term of (4) is holomorphic in z, z'. It is easily verified that the series is absolutely convergent because $k \geq 4$. The function $h_m(z, z')$ is therefore holomorphic in z, z'. It is also immediate from (4) that it is a cusp form in each variable separately.

Theorem 1. *Let*

$$C_k = \frac{(-1)^{k/2}\pi}{2^{(k-3)}(k-1)}. \tag{5}$$

(i) *The function $C_k^{-1} m^{k-1} h_m(z, z')$ is a "kernel" for the operator*

$$T(m): S_k \to S_k\,.$$

In other words, for every $f \in S_k$ we have

$$f * h_m(z') = C_k m^{-k+1}(T(m) f)(z')\,. \tag{6}$$

(ii) *We have the identity*

$$C_k^{-1} m^{k-1} h_m(z, z') = \sum_{i=1}^{r} a_m^i \frac{1}{\langle f_i, f_i\rangle} f_i(z) f_i(z')\,. \tag{7}$$

(iii) *The trace* Tr $T(m)$ *is given by*

$$\operatorname{Tr} T(m) = C_k^{-1} m^{k-1} \int_F h_m(z, -\bar{z}) \operatorname{Im}(z)^k \frac{dx\,dy}{y^2}. \tag{8}$$

Proof. Suppose first that $m = 1$. If

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$$

then from the definition of the operation $[\gamma]_k$ we get

$$(c\bar{z} + d)^{-k} f(z) y^k = f(\gamma z) \operatorname{Im}(\gamma z)^k\,.$$

From (4) we get

$$f(z)\,\overline{h_1(z, z')} y^k = \sum_{\gamma \in \Gamma} (\bar{z}' + \gamma\bar{z})^{-k} f(\gamma z) \operatorname{Im}(\gamma z)^k$$

and therefore

$$f * h_1(z') = \int_F \sum_{\gamma \in \Gamma} (-z' + \gamma\bar{z})^{-k} f(\gamma z) \operatorname{Im}(\gamma z)^k \frac{dx\,dy}{y^2}.$$

Interchanging the integral and summation, and using the invariance of $dx\,dy/y^2$ with respect to Γ yields

$$(9)\qquad f * h_1(z') = \sum_{\gamma \in \Gamma} \int_{\gamma F} (-z' + \bar{z})^{-k} f(z)\,\mathrm{Im}\,(z)^k \frac{dx\,dy}{y^2}$$

$$= 2\int_0^\infty \int_{-\infty}^\infty (x - iy - z')^{-k} f(x + iy) y^{k-2}\,dx\,dy\,.$$

This last equality comes from the fact that the upper half plane is equal to the union of transforms of the fundamental domain under Γ, disjoint except for boundary points of measure zero, and except for the fact that $\pm\gamma$ give the same transform, whence the factor of 2. Cauchy's formula and the fact that f is holomorphic and sufficiently small at infinity imply that

$$\int_{-\infty}^\infty (x - iy - z')^{-k} f(x + iy)\,dx = \frac{2\pi i}{(k-1)!} f^{(k-1)}(2iy + z')\,.$$

Therefore the right-hand side of (9) is

$$= \frac{4\pi i}{(k-1)!} \int_0^\infty y^{k-2} f^{(k-1)}(2iy + z')\,dy$$

$$= \frac{4\pi i}{(k-1)!} \int_0^\infty \frac{1}{(2i)^{k-2}} (d/dt)^{k-2} f'(2ity + z')\Big|_{t=1} dy$$

$$= \frac{4\pi i}{(k-1)!} \frac{1}{(2i)^{k-2}} (d/dt)^{k-2} \int_0^\infty f'(2ity + z')\,dy\Big|_{t=1}$$

$$= \frac{4\pi i}{(k-1)!} \frac{1}{(2i)^{k-2}} (d/dt)^{k-2} \left(\frac{-f(z')}{2it}\right)\Big|_{t=1}$$

$$= C_k f(z')\,.$$

This proves the desired formula (6) in case $m = 1$. The general case is a consequence of the case $m = 1$, because one easily sees that

$$m^{k-1} h_m = T(m) h_1\,,$$

where $T(m)$ operates with respect to the first variable z on the right-hand side.

Part (ii) now follows essentially from elementary linear algebra. The function h_m being a cusp form with respect to each variable z, z' can be written in the form

$$h_m(z, z') = \sum_{i,j=1}^r c_{ij} f_i(z) f_j(z')\,.$$

We apply Part (i) to a function f_μ (one of the normalized eigenfunctions), and Part (ii) follows at once using the orthogonality. Part (iii) follows trivially from Part (ii). This proves the theorem.

The second theorem will give an explicit expression for the trace. We need some definitions.

We define a function $H(n)$ for integers n first by putting

$$H(n) = 0 \text{ if } n < 0 \quad \text{and} \quad H(0) = -1/12\,.$$

If $n > 0$, we let $H(n)$ be the number of equivalence classes with respect to $SL_2(\mathbf{Z})$ of positive definite binary quadratic forms

$$ax^2 + bxy + cy^2$$

with discriminant

$$b^2 - 4ac = -n\,,$$

counting forms equivalent to a multiple of $x^2 + y^2$ (resp. $x^2 + xy + y^2$) with multiplicity $\frac{1}{2}$ (resp. $\frac{1}{3}$).

If $n \equiv 1$ or $2 \pmod 4$ then $H(n) = 0$. We have the following table.

n	0	3	4	7	8	11	12	15	16	19	20	23	24
$H(n)$	$-\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{2}$	1	1	1	$\frac{4}{3}$	2	$\frac{3}{2}$	1	2	3	2

We also define a polynomial $P_k(t, N)$ ($k > 0$ even) as the coefficient of x^{k-2} in the power series development of

$$(1 - tx + Nx^2)^{-1}\,.$$

We also have

$$P_k(t, N) = \frac{\rho^{k-1} - \bar{\rho}^{k-1}}{\rho - \bar{\rho}}$$

where

$$\rho + \bar{\rho} = t \quad \text{and} \quad \rho\bar{\rho} = N\,.$$

For instance $P_2(t, N) = 1$ and $P_4(t, N) = t^2 - N$.

Theorem 2. (Trace Formula) *let $k \geqslant 4$ be an even integer and let m be an integer >0. Then the trace of the Hecke operator $T(m)$ on the space of cusp forms S_k is given by*

$$\operatorname{Tr} T(m) = -\tfrac{1}{2}\sum_{t=-\infty}^{\infty} P_k(t, m)H(4m - t^2) - \tfrac{1}{2}\sum_{dd'=m} \min(d, d')^{k-1}\,.$$

Note. The first sum is in fact finite, because $H(4m - t^2) = 0$ for $t > 2\sqrt{m}$. The second sum is taken over all factorizations of m as a product of two positive integers.

Example. For $k = 4$ the only cusp forms are 0, so the right-hand side of the formula is 0. This implies relations among the class numbers $H(m)$. For instance for $m = 5$, we find:

$$\sum (t^2 - m)H(4m - t^2) = -5H(20 - 8H(19) - 2H(16) + 8H(11) + 22H(4)$$

$$= -10 - 8 - 3 + 8 + 11 = -2\,,$$

$$\sum \min(d, d')^3 = 1^3 + 1^3 = 2\,.$$

The rest of this appendix is devoted to the proof of Theorem 2.
In Theorem 1 we have proved the identity

$$\operatorname{Tr} T(m) = C_k^{-1} m^{k-1} \int_F \sum_{ad-bc=m} \frac{y^k}{(c|z|^2 + d\bar{z} - az - b)^k} \frac{dx\,dy}{y^2}\,.$$

The sum on the right-hand side is invariant under Γ (otherwise the integral would not be independent of the choice of fundamental domain F). Looking at the terms of this sum, we observe that replacing z by γz amounts to replacing the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{by} \quad \gamma^{-1}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\gamma\,.$$

These two matrices have the same determinant and the same trace. Therefore we may decompose the sum into pieces which are Γ-invariant, characterized by the condition $a + d =$ constant:

$$\operatorname{Tr} \mathrm{T}(m) = \sum_{t=-\infty}^{\infty} I(m, t)\,,$$

where

$$I(m, t) = C_k^{-1} m^{k-1} \int_F \sum_{\substack{ad-bc=m \\ a+d=t}} \frac{y^k}{(c|z|^2 + d\bar{z} - az - b)^k} \frac{dx\,dy}{y^2}\,. \tag{10}$$

We shall prove:

$$\tfrac{1}{2}(I(m, t) + I(m, -t)) = \begin{cases} -\tfrac{1}{2}P_k(t, m)H(4m - t^2) & \text{for } t^2 - 4m < 0 \\ \dfrac{k-1}{48}m^{(k-2)/2} - \tfrac{1}{4}m^{(k-1)/2} & \text{for } t^2 - 4m = 0 \\ -\dfrac{1}{2}\left(\dfrac{|t| - u}{2}\right)^{k-1} & \text{for } t^2 - 4m = u^2,\ u > 0 \\ 0 & \text{for } t^2 - 4m > 0 \\ & \text{non-square} \end{cases} \tag{11}$$

It is clear that these formulas imply the trace formula in Theorem 2. The numbers

$$\left|\frac{t+u}{2}\right| \quad \text{and} \quad \left|\frac{t-u}{2}\right|$$

play the role of d, d' in the trace formula.

To study the integral (10), we first remark that there is a bijection between the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with determinant m and trace t, and the set of binary quadratic forms g with discriminant

$$|g| = t^2 - 4m\,.$$

The bijection is given by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto g(u, v) = cu^2 + (d - a)uv - bv^2$$

$$g(u, v) = \alpha u^2 + \beta uv + \gamma v^2 \mapsto \begin{pmatrix} \frac{1}{2}(t - \beta) & -\gamma \\ \alpha & \frac{1}{2}(t + \beta) \end{pmatrix}.$$

For every form $g(u, v) = \alpha u^2 + \beta uv + \gamma v^2$ and real t, $z = x + iy \in \mathfrak{H}$, we put

$$R_g(z, t) = \frac{y^k}{(\alpha(x^2 + y^2) + \beta x + \gamma - ity)^k}\,. \tag{12}$$

Then

$$I(m, t) = C_k^{-1} m^{k-1} \int_F \sum_{|g| = t^2 - 4m} R_g(z, t) \frac{dx\,dy}{y^2}\,. \tag{13}$$

where the sum is taken over all forms of discriminant $t^2 - 4m$. An element $\gamma \in \Gamma$ transforms a quadratic form g into a form γg having the same discriminant, and one verifies that

$$R_{\gamma g}(z, t) = R_g(\gamma z, t) . \tag{14}$$

Therefore, for each discriminant D (i.e. for each integer $D \equiv 0$ or $1 \bmod 4$) we have the equality

$$\sum_{|g|=D} R_g(z, t) = \sum_{\substack{|g|=D \\ \bmod \Gamma}} \sum_{\gamma \in \Gamma/\Gamma_g} R_{\gamma g}(z, t)$$
$$\sum_{\substack{|g|=D \\ \bmod \Gamma}} \sum_{\gamma \in \Gamma/\Gamma_g} R_g(\gamma z, t) .$$

The first sum is taken over a set of representatives for classes of quadratic forms with discriminant D, and the second sum is taken over right cosets of Γ with respect to the isotropy group Γ_g of elements leaving g fixed. For $D \neq 0$, the class number $h(D)$ is finite, and therefore the first sum is finite, giving

$$\int_F \sum_{|g|=D} R_g(z, t) \frac{dx\, dy}{y^2} = \sum_{\substack{|g|=D \\ \bmod \Gamma}} \int_{F_g} R_g(z, t) \frac{dx\, dy}{y^2} , \tag{15}$$

where

$$F_g = \bigcup_{\gamma \in \Gamma/\Gamma_g} \gamma F$$

is a fundamental domain for the operation of Γ_g on $\mathfrak{H}$. The argument is the same as that used in the proof of Theorem 1.

For $D = 0$ we can take as a system of representatives for the forms of discriminant the forms g_r ($r \in \mathbf{Z}$), where $g_r(u, v) = rv^2$. The isotropy group of g_r is equal to Γ for $r = 0$, and is equal to

$$\Gamma_\infty = \left\{ \pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\}$$

for $r \neq 0$. In this case, we find

$$\int_F \sum_{|g|=0} R_g(z, t) \frac{dx\, dy}{y^2} = \int_F R_{g_0}(z, t) \frac{dx\, dy}{y^2} + \int_{F_\infty} \sum_{r \neq 0} R_{g_r}(z \cdot t) \frac{dx\, dy}{y^2} , \tag{16}$$

where F_∞ is a fundamental domain for the operation of Γ_∞ on $\mathfrak{H}$, say the strip between 0 and 1. Here we cannot interchange the order of integration and summation, since for instance

$$\int_{F_\infty} R_{g_r}(z, t) \frac{dx\, dy}{y^2} = 0 \quad \text{for all } r ,$$

but the integral of the sum is $\neq 0$, as we shall see below.

There remains to compute the right-hand side of (15) and (16) for $D = t^2 - 4m$. We distinguish four cases.

Case 1. $D < 0$.

In this case Γ_g is finite for each form g in (15), (and one even can prove that its order is 1, 2, or 3). For a quadratic form

$$g(u, v) = \alpha u^2 + \beta uv + \gamma v^2$$

with discriminant D we therefore have

$$\int_{F_g} R_g(z, t) \frac{dx\, dy}{y^2} = \frac{1}{|\Gamma_g|} \int_{\mathfrak{H}} R_g(z, t) \frac{dx\, dy}{y^2}$$
$$= \frac{1}{|\Gamma_g|} \int_{\mathfrak{H}} \frac{y^k}{(|z|^2 - ity - \frac{1}{4}D)^k} \frac{dx\, dy}{y^2} .$$

(For this last equality, we used the substitution $z \mapsto (2z - \beta)/2\alpha$.) Let I denote the value of the integral. It depends only on D and t. The right-hand side of (15) is therefore equal to

$$\sum_{\substack{|g|=D \\ \bmod \Gamma}} \frac{1}{|\Gamma_g|} I = 2H(-D)I .$$

The factor 2 comes from the fact that in the definition of $H(n)$ we counted positive definite forms, whereas here we count all forms, positive or negative. Finally, using the formula

$$\int_{-\infty}^{\infty} (x^2 + A)^{-k}\, dx = \frac{\pi}{(k-1)!} \tfrac{1}{2} \tfrac{3}{2} \cdots (k - \tfrac{3}{2}) A^{-k+1/2} ,$$

obtained by differentiating $k - 2$ times with respect to A in the corresponding formula for $k = 2$, we obtain:

$$I = \int_0^\infty y^{k-2} \int_{-\infty}^{\infty} (x^2 + y^2 - ity - \tfrac{1}{4}D)^{-k}\, dx\, dy$$
$$= \frac{\pi}{(k-1)!} \tfrac{1}{2} \tfrac{3}{2} \cdots (k - \tfrac{3}{2}) \int_0^\infty (y^2 - ity - \tfrac{1}{4}D)^{-k+1/2} y^{k-2}\, dy$$
$$= \frac{\pi i^{k-2}}{2(k-1)!} (d/dt)^{k-2} \int_0^\infty (y^2 - ity - \tfrac{1}{4}D)^{-3/2}\, dy$$

$$= \frac{\pi i^{k-2}}{2(k-1)!}(d/dt)^{k-2}\left(\frac{4}{t^2 - D}\frac{y - \frac{1}{2}it}{\sqrt{y^2 - ity - \frac{1}{4}D}}\Bigg|_0^\infty\right)$$

$$= \frac{\pi i^{k-2}}{2(k-1)!}(d/dt)^{k-2}\left(\frac{4}{\sqrt{|D|}}\frac{1}{\sqrt{|D|} - it}\right)$$

$$= \frac{2\pi}{k-1}\frac{1}{\sqrt{|D|}}\frac{1}{(\sqrt{|D|} - it)^{k-1}}.$$

Formula (13) then gives

$$I(m, t) = C_k^{-1} m^{k-1} 2H(4m - t^2)\frac{2\pi}{k-1}\frac{1}{\sqrt{4m - t^2}}\frac{1}{(\sqrt{4m - t^2} - it)^{k-1}}$$

$$= \frac{\bar{\rho}^{k-1}}{\rho - \bar{\rho}}H(4m - t^2), \quad \text{where } \rho = \tfrac{1}{2}(t + i\sqrt{4m - t^2}).$$

This proves the first formula in (11).

Case 2. $D = 0$

We now use formula (16). The first term is equal to $(-1)^{k/2}\pi/6t^k$, because

$$R_{g_0}(z, t) = (i/t)^k \quad \text{and} \quad \int_F \frac{dx\,dy}{y^2} = \frac{\pi}{6}.$$

The second term is equal to

$$\int_0^\infty\int_0^1 y^{k-2}\sum_{\substack{r\in\mathbf{Z}\\ r\neq 0}}(r - ity)^{-k}\,dx\,dy = \frac{i^{k-2}}{(k-2)!}(d/dt)^{k-2}\sum_{\substack{r\in\mathbf{Z}\\ r\neq 0}}(r - ity)^{-2}\,dy$$

$$= \frac{i^{k-2}}{(k-1)!}(d/dt)^{k-2}\int_0^\infty\left(\frac{1}{t^2y^2} - \frac{\pi^2}{\sinh^2\pi ty}\right)dy$$

$$= \frac{i^{k-2}}{(k-1)!}(d/dt)^{k-2}\left(\frac{\pi}{|t|}\right)$$

$$= (-1)^{(k-2)/2}\frac{\pi}{k-1}|t|^{-k+1}.$$

For $t = \pm 2\sqrt{m}$ we get the value

$$I(m, t) = C_k^{-1} m^{k-1}\int_F \sum_{|g|=0} R_g(z, t)\frac{dx\,dy}{y^2} = \frac{k-1}{48}m^{(k-2)/2} - \tfrac{1}{4}m^{(k-1)/2}$$

This is precisely the second formula in (11).

Case 3. $D = u^2$ and $u > 0$.

As in the case $D < 0$ there is only a finite number of classes of forms with discriminant D, and Γ_g is a finite group. Hence

$$\int_F \sum_{|g|=D} R_g(z, t)\frac{dx\,dy}{y^2} = HI,$$

where

$$H = \sum_{\substack{|g|=D\\ \bmod \Gamma}}\frac{1}{|\Gamma_g|} \quad \text{and} \quad I = \int_{\mathfrak{H}}\frac{y^k}{(|z|^2 - ity - \frac{1}{4}D)^k}\frac{dx\,dy}{y^2}.$$

We have $H = u$, because the groups Γ_g are trivial in this case, and because there are u classes of quadratic forms with discriminant u^2. As in the case $D < 0$, we obtain

$$I = \frac{\pi i^{k-2}}{2(k-1)!}(d/dt)^{k-2}\left(\frac{4}{t^2 - D}\frac{y - \frac{1}{2}it}{\sqrt{y^2 - ity - \frac{1}{4}D}}\Bigg|_0^\infty\right)$$

but for $D > 0$ the expression in parentheses is equal to

$$\frac{-4}{\sqrt{D}}\frac{1}{\sqrt{D + |t|}} \quad \text{and not} \quad \frac{4}{\sqrt{|D|}}\frac{1}{\sqrt{|D|} + it} \quad \text{as before}.$$

(The fact that the integral here depends only on $|t|$ is due to the fact that the value of $\sqrt{y^2 - ity - \frac{1}{4}D}$ for $y = 0$ depends on the sign of t, because we must choose the branch of the square root which has positive real part for $y \to \infty$.) We therefore have

$$HI = (-1)^{(k-2)/2}\frac{2\pi}{k-1}\frac{1}{(u + |t|)^{k-1}}$$

$$I(m, t) = C_k^{-1} m^{k-1} HI = -\tfrac{1}{2}\left(\frac{|t| - u}{2}\right)^{k-1}.$$

This proves the third formula in (11).

Case 4. $D > 0$ and non-square

Here we again have only a finite number of classes of quadratic forms, but the isotropy groups are infinite cyclic. Intuitively we have

$$H = \sum\frac{1}{|\Gamma_g|} = \frac{1}{\infty} = 0,$$

and thus $HI = 0$. We now assert that for each g of discriminant D, we in fact have

$$\int_{F_g} R_g(z, t)\frac{dx\,dy}{y^2} + \int_{F_g} R_g(z, -t)\frac{dx\,dy}{y^2} = 0 . \tag{17}$$

Let $g(u, v) = \alpha u^2 + \beta uv + \gamma v^2$ be such a quadratic form, and let $w > w'$ be the roots of the equation $\alpha u^2 + \beta u + \gamma = 0$. Then the matrix

$$\gamma = (w - w')^{-1/2}\begin{pmatrix} w' & -w \\ 1 & 1 \end{pmatrix} \in SL_2(\mathbf{R})$$

transforms g into γg, with

$$\gamma g(u, v) = \sqrt{D}\, uv .$$

The conjugate of Γ_g by γ operates on the upper half plane as the infinite cyclic group generated by $z \mapsto \varepsilon^2 z$, where $\varepsilon > 1$ is the fundamental unit of the order in $\mathbf{Q}(\sqrt{D})$ associated with g. We can therefore choose the fundamental domain F_g so that $\gamma^{-1}F_g$ is an annulus defined by

$$y > 0 \quad \text{and} \quad r_0 \leqslant |z| \leqslant \varepsilon^2 r_0 .$$

Then

$$\int_{F_g} R_g(z, t)\frac{dx\,dy}{y^2} = \int_{F_g} R_{\gamma g}(\gamma^{-1}z, t)\frac{dx\,dy}{y^2} \quad \text{(by (14))}$$

$$= \iint_{\text{annulus}} (\sqrt{D}\, x - ity)^{-k} y^{k-2}\, dx\, dy .$$

We write $z = x + iy$ in polar coordinates, $z = r\, e^{i\theta}$ to obtain

$$= \int_0^{\pi}\int_{r_0}^{\varepsilon^2 r_0} (\sqrt{D}\cos\theta - it\sin\theta)^{-k}(\sin\theta)^{k-2}\frac{dr}{r}\,d\theta$$

$$= (\log \varepsilon^2)\int_0^{\pi} (\sqrt{D}\cos\theta - it\sin\theta)^{-k}(\sin\theta)^{k-2}\,d\theta .$$

To prove (17) it suffices therefore to verify that

$$\int_{-\pi}^{\pi} (\sqrt{D}\cos\theta - it\sin\theta)^{-k}(\sin\theta)^{k-2}\,d\theta = 0 ,$$

which is easily done by putting $\zeta = e^{i\theta}$ and using the residue theorem.

The last formula in (11) follows easily from (13), (15) and (17). This concludes the proof of the trace formula.

Part II
Periods of Cusp Forms

Chapter IV. Modular Symbols

The points at infinity (called cusps) on the quotient curves X_Γ of the upper half plane turn out to be especially interesting.

By using a variation of the standard Hecke operators, Manin proved a special case, generalized by Drinfeld to the general situation, of the assertion that if Γ is a congruence subgroup, then all the divisors of degree 0 whose components are points at infinity on the curve X_Γ are of finite order in the group of divisor classes. Cf. [Man 2] and [Dr 1]. We reproduce this proof. Kubert and Lang [KL I], [KL II] have shown how one can get a realization of the functions which represent appropriate multiples of these divisors at infinity by means of explicit modular forms. Special cases of the realization of the multiple of a cusp by a function were treated by Newman [N] and Ogg [O 5], who give lower bounds for the order of the cusp in the divisor class group, for curves $X_0(p)$ with special primes p.

The Kubert-Lang series also studies the points at infinity from a diophantine point of view, not touched upon in this book. They can be used to parametrize ideal classes in number fields, and one needs the explicit algebraic study of those papers, rather than the transcendental arguments of Manin and Drinfeld.

The Manin-Mumford conjecture asserts that on a curve of genus $\geqslant 2$, canonically embedded in its Jacobian, there exist only a finite number of points on the curve which are of finite order in the Jacobian. Cf. Lang [L] where this conjecture is reduced to a Galois property of torsion points. The Manin-Drinfeld theorem gives significant examples of such exceptional points.

In the last section, we point out how the action of the ordinary Hecke operators gives rise to certain recurrence formulas, which have interesting p-adic properties. This connects with the Mazur p-adic theory of distributions, discussed in Chapter VII.

Modular symbols were introduced by Birch [B] in connection with the Birch-Swinnerton-Dyer conjecture. We do not discuss this aspect of them, but refer to Manin [Man 1], [Man 2], who was the first to develop their properties systematically.

§ 1. Basic Properties

We let Γ denote a subgroup of $SL_2(\mathbf{Z})$, of finite index. As before, we let

$$\mathfrak{H}^* = \mathfrak{H} \cup \{\infty\} \cup \mathbf{Q},$$

and we use the same notation as in the previous chapter.

Let $z_1, z_2 \in \mathfrak{H}^*$. By a **path** C_{z_1,z_2} joining z_1 to z_2, we mean a piecewise C^∞ path, lying inside $\mathfrak{H}$ (except for the end points if these are cusps), and also analytic at the end points in the following sense: Suppose that $z_2 = \infty$. Then we require that the path leading to ∞ should be contained in a vertical strip, of finite width, and that under the mapping

$$z \mapsto e^{2\pi i z}$$

it should project onto an analytic arc leading to 0 in the disc. Then for any positive integer N, the projection of the path leading to ∞ under the map

$$z \mapsto e^{2\pi i z/N}$$

is also an analytic arc leading to 0. In particular, for any Γ the projection of the path leading to ∞ by the map

$$\pi_\Gamma : \mathfrak{H}^* \to X_\Gamma$$

is an analytic arc on X_Γ leading to $\pi_\Gamma(\infty)$.

For instance, a path which goes to infinity vertically on a straight line clearly satisfies the desired condition. Pictures of the portions of paths leading to ∞ are drawn on Fig. 4.

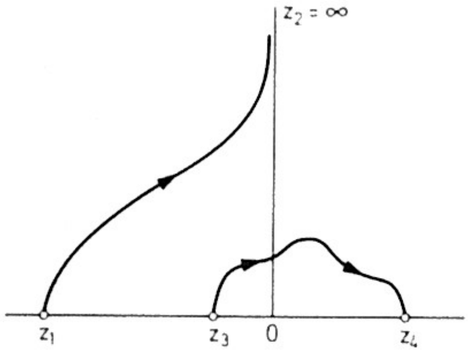

Fig. 4.

If z_1 is a cusp $\neq \infty$, so a rational number, then a neighborhood of z_1 is analytically isomorphic to a neighborhood of ∞, so the condition of analyticity can be defined again in terms of a local parameter at infinity, composed with such an analytic isomorphism. Paths which go down vertically to a rational point z_1 (or z_3, z_4 as on the figure) are analytic, and their projections on X_Γ by π_Γ for any Γ are analytic.

Lemma. *Two paths joining points z_1, z_2 in $\mathfrak{H}^*$ are homotopic, and their projections under π_Γ for any Γ are homotopic on X_Γ.*

Proof. Suppose z_1 is rational and $z_0 = \infty$. We split the paths into three portions, one lying in $\mathfrak{H}$, away from the end points, and the two tail ends leading to z_1, z_2 as on Fig. 5.

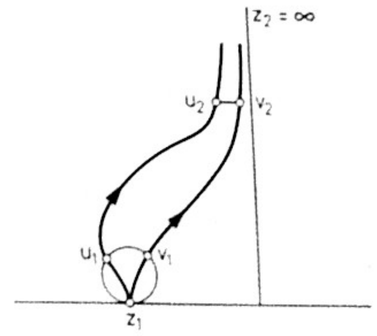

Fig. 5.

The pieces between z_1 and u_1 and between z_1 and v_1 are homotopic by a homotopy taking place inside the disc tangent to z_1. The pieces between u_1 and u_2, and between v_1 and v_2 are homotopic in the upper half plane. The pieces leading from u_2 to ∞ and v_2 to ∞ are homotopic, again by a purely local construction using the parameter $e^{2\pi i z}$ which transforms the situation to a neighborhood of 0 in the plane. So the lemma is obvious.

Let $\omega \in \Omega^1(X_\Gamma)$ be a differential of first kind on some modular curve X_Γ. Then for any z_1, z_2 in $\mathfrak{H}^*$, the integral

$$\int_{C_{z_1,z_2}} \omega \circ \pi_\Gamma = \int_{C_{z_1,z_2}} \pi_\Gamma^* \omega$$

is independent of the path C_{z_1,z_2}, and depends only on z_1, z_2. Therefore we view

$$\omega \mapsto \int_{z_1}^{z_2} \omega \circ \pi_\Gamma = \int_{C_{z_1,z_2}} \omega \circ \pi_\Gamma$$

as a functional, which we denote by

$$\{z_1, z_2\},$$

and which we call the **modular symbol**. By abuse of notation, we sometimes omit the projection π_Γ from the notation, and simply write

$$\int_{z_1}^{z_2} \omega .$$

Note that this integral is independent of the choice of Γ.

Thus if we let $\Omega^1(\mathfrak{H})$ be the union of all inverse images,

$$\Omega^1(\mathfrak{H}) = \bigcup_\Gamma \pi_\Gamma^* \Omega^1(X_\Gamma),$$

then $\{z_1, z_2\}$ is a functional on $\Omega^1(\mathfrak{H})$.

We recall some facts from the theory of compact Riemann surfaces.

The first homology group $H_1(X_\Gamma, \mathbf{R})$ with real coefficients is dual to $\Omega^1(X_\Gamma)$, as vector space over $\mathbf{R}$.

The duality is obtained as follows. We use the notation of [L 3], Chapter III, § 5. Let $a_1, \ldots, a_{2g}$ be the fundamental cycles relative to a polygonal decomposition of the Riemann surface. We may view $H_1(X_\Gamma, \mathbf{R})$ as the space of formal linear combinations

$$\sigma = \sum x_i a_i$$

with real coefficients $x_i \in \mathbf{R}$. Then the pairing between $H_1(X_\Gamma, \mathbf{R})$ and $\Omega^1(X_\Gamma)$ is given by

$$\langle \sigma, \omega \rangle = \sum x_i \int_{a_i} \omega .$$

Note that $H_1(X_\Gamma, \mathbf{R})$ is nothing but $H_1(X_\Gamma, \mathbf{Z}) \otimes \mathbf{R}$, where $H_1(X_\Gamma, \mathbf{Z})$ is the space of formal linear combinations of the elements a_i with integer coefficients. Similarly, $H_1(X_\Gamma, \mathbf{Q})$ is the space of linear combinations of the a_i with coefficients in $\mathbf{Q}$.

Lemma 4 of [L 3], Chapter III, § 5 constructs the dual basis in $\Omega^1(X_\Gamma)$ to the basis $\{a_1, \ldots, a_{2g}\}$, with respect to the real part of the above pairing, and in particular, shows that the natural map which to each σ associates the function obtained by integrating over σ is an injection of $H_1(X_\Gamma, \mathbf{R})$ into the dual space of $\Omega^1(X_\Gamma)$. Since the two spaces have the same real dimension, this injection must be an isomorphism.

It follows that we can identify the modular symbol

$$\{z_1, z_2\} = \{z_1, z_2\}_\Gamma$$

as an element of $H_1(X_\Gamma, \mathbf{R})$, and we have the formula, by definition,

$$\langle \{z_1, z_2\}, \omega \rangle = \int_{z_1}^{z_2} \omega \circ \pi_\Gamma .$$

We use the subscript Γ if we want to view the symbol as a functional on a specific $\Omega^1(X_\Gamma)$, and omit it if we view the symbol as a functional on the union $\Omega^1(\mathfrak{H})$.

The modular symbol satisfies the following formalism.

MS 1. *If $z_1, z_2, z_3 \in \mathfrak{H}^*$ then*

$$\{z_1, z_2\} + \{z_2, z_3\} = \{z_1, z_3\} .$$

Proof. Say z_1, z_3 are rational, and $z_2 = \infty$ as on Fig. 6.

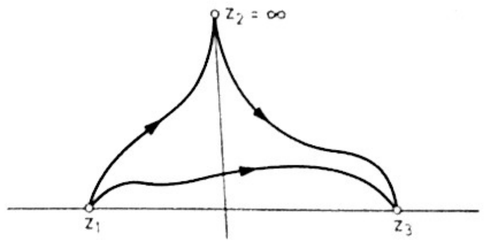

Fig. 6.

Then the integral of the holomorphic differential form over the "triangle" oriented counterclockwise is equal to 0 by Cauchy's theorem. One can work this out in the upper half plane, by cutting off for instance small neighorhoods of the three end points and analysing these separately in terms of a local parameter, or one can work on one of the Riemann surfaces X_Γ, and use Cauchy's theorem there.

MS 2. *We have $\{z_1, z_1\} = 0$.*

Proof. Obvious, because a path from z_1 to z_1 is contractible.

MS 3. *For any $\omega \in \Omega^1(\mathfrak{H})$ and $\alpha \in GL_2^+(\mathbf{Q})$ we have*

$$\langle \{\alpha z_1, \alpha z_2\}, \omega \rangle = \langle \{z_1, z_2\}, \omega \circ \alpha \rangle .$$

Proof. Let C_{z_1, z_2} be a path from z_1 to z_2. Then

$$\alpha C_{z_1, z_2} = C_{\alpha z_1, \alpha z_2}$$

is a path from αz_1 to αz_2, and the ordinary change of variables formula shows that

$$\int_{C_{z_1, z_2}} \omega \circ \alpha = \int_{\alpha C_{z_1, z_2}} \omega = \int_{\alpha z_1}^{\alpha z_2} \omega ,$$

whence the assertion follows.

In particular:

MS 4. *If $\gamma \in \Gamma$, then $\{z_1, z_2\}_\Gamma = \{\gamma z_1, \gamma z_2\}_\Gamma$.*

§ 2. The Manin-Drinfeld Theorem

We specialize to modular curves X_Γ where Γ is a *congruence subgroup*, i.e. Γ contains some $\Gamma(N)$ for some N.

We want to prove:

Theorem 2.1. *Let x, y be cusps. If Γ is a congruence subgroup, then*

$$\{x, y\}_\Gamma \in H_1(X_\Gamma, \mathbf{Q})$$

is in the homology with rational coefficients.

This can be formulated in terms of divisor classes. We recall some facts about Abel's theorem. Put $X = X_\Gamma$ and let $P_0 \in X$.

Theorem 2.2. *Let $\mathfrak{a} = \sum m_i P_i$ be a divisor of degree 0 on X. Then $\mathfrak{a}$ is the divisor of a rational function on X if and only if there exists a cycle $\sigma \in H_1(X, \mathbf{Z})$ such that*

$$\int^{\mathfrak{a}} \omega = \sum m_i \int_{P_0}^{P_i} \omega = \int_\sigma \omega .$$

for every $\omega \in \Omega^1(X)$.

Proof. We use Abel's theorem as formulated in [L 3]. Let $\Phi = (\varphi_1, \ldots, \varphi_g)$ be a basis of the differentials of first kind. Let P_0 be a fixed point on X. Then by definition,

$$\int^{\mathfrak{a}} \Phi = \sum m_i \int_{P_0}^{P_i} \Phi ,$$

and if a cycle σ exists as in the theorem, then the above integral lies in the period lattice, so the divisor is linearly equivalent to 0. The converse is equally clear.

Corollary. *The modular symbol $\{x, y\}_\Gamma$ lies in $H_1(X_\Gamma, \mathbf{Q})$ if and only if there exists a positive integer m such that*

$$m(\pi_\Gamma(x) - \pi_\Gamma(y))$$

is the divisor of a function on X_Γ. In particular, Theorem 2.1 implies that every divisor of degree 0 on X whose components are cusps is of finite order in the group of divisor classes.

Proof. Immediate from Theorem 2.2 and 2.1.

To prove Theorem 2.1, it suffices to do so when $\Gamma = \Gamma(N)$ is the principal congruence subgroup, which we assume from now on.

We shall use the existence of a set of matrices M satisfying the following properties.

(i) *M is closed under multiplication on the right and left by elements of $\Gamma(N)$, and there is a finite number of cosets*

$$M = \bigcup_{i=1}^{s} \Gamma(N)\alpha_i , \qquad \alpha_i \in M .$$

(ii) *If x is a cusp, then $\overline{\alpha x} = \bar{x}$, where the bar denotes the projection of x into $X(N)$ under $\pi_{\Gamma(N)}$.*

(iii) *Given $\omega \in \Omega^1(X(N))$, if $\omega \circ \alpha = \omega$ for all $\alpha \in M$, then $\omega = 0$.*

Example. One way to exhibit such a set M explicitly is as follows. We select a prime p sufficiently large, and such that

$$p \equiv 1 \pmod{N} .$$

We let M be the set of integral matrices which are $\equiv 1 \pmod{N}$ and whose determinant is equal to p. Then (i) is obvious. It is easy to verify (ii) from the description of cusps given in Chapter III, § 2. As for (iii), if ω is fixed by M, we must have

$$\omega(pz) = \omega(z)$$

by using the matrix

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$$

in M. Iterating yields $\omega(p^n z) = \omega(z)$ for all n. This means that the coefficients a_m of the q-expansion of ω are all 0, unless $p^n | m$ for arbitrarily large n, whence $a_m = 0$ for all m and $\omega = 0$.

Each element $\alpha \in M$ then gives an operator on differential forms, which we may denote by $[\alpha]$, namely

$$[\alpha] : \omega \mapsto \omega \circ \alpha .$$

This operator is linear on $\Omega^1(\mathfrak{H})$, but of course does not map $\Omega^1(X(N))$ into itself.

We define the **Hecke operator** T_M to be

$$T_M = \sum_{i=1}^{s} [\alpha_i] .$$

If $\omega \in \Omega^1(X(N))$, then $T_M \omega = \sum \omega \circ \alpha_i$, and consequently $T_M \omega$ is invariant under $\Gamma(N)$. Thus

$$T_M : \Omega^1(X(N)) \to \Omega^1(X(N))$$

gives a linear map of $\Omega^1(X(N))$ into itself.

Theorem 2.3. *The map $T_M - sI$ on $\Omega^1(X(N))$ is invertible.*

Proof. Let ω be a differential of first kind on $X(N)$ such that

$$\omega \circ T_M = s\omega .$$

By the properties of the Petersson product, Chapter III, Theorem 4.1, we know that

$$\|\omega \circ \alpha_i\| = \|\omega\| .$$

Thus the vectors $\omega \circ \alpha_i$ all have the same length, and the length of their sum is

$$\|s\omega\| = s\|\omega\| .$$

It follows that all vectors must point in the same direction, and have the same length, so that $\omega \circ \alpha_i = \omega$ for all i. By property (iii) we conclude that $\omega = 0$, thus proving the theorem.

By duality, T_M operates on modular symbols, i.e. from **MS 3**, the transpose is given by

$${}^tT_M\{z_1, z_2\} = \sum \{\alpha_1 z_1, \alpha_i z_2\} .$$

For simplicity, the superscipt t is sometimes omitted from the notation.

Theorem 2.4. *The operator ${}^tT_M$ maps $H_1(X(N), \mathbf{Z})$ into itself.*

Proof. Any closed curve on $X(N)$ can be lifted to $\mathfrak{H}^*$, and it will suffice to prove that the image of such a closed curve under T_M lies in $H_1(X(N), \mathbf{Z})$. The lifting can be represented by the symbol

$$\sigma = \{z_1, z_2\} ,$$

with two points $z_1, z_2 \in \mathfrak{H}^*$ such that $\overline{z_1} = \overline{z_2}$. Then

$${}^tT_M\sigma = \sum \{\alpha_i z_1, \alpha_i z_2\} .$$

There is some $\gamma \in \Gamma(N)$ such that $z_2 = \gamma z_1$, and the map

$$\Gamma\alpha_i \mapsto \Gamma\alpha_i\gamma$$

permutes the cosets of M with respect to $\Gamma(N)$. Hence

$$\sum(\overline{\alpha_i z_2}) - \sum(\overline{\alpha_i z_1}) = 0 .$$

Thus ${}^tT_M\sigma$ is represented by a cycle, as was to be shown.

From Theorem 2.3, we conclude that

$${}^tT_M - sI$$

is invertible on $H_1(X(N), \mathbf{Q})$. We may now prove the Manin–Drinfeld Theorem 2.1. Let x, y be two cusps. Then

$${}^tT_M\{x, y\} = \sum \{\alpha_i x, \alpha_i y\} ,$$

and since by Property (ii),

$$\overline{\alpha_i x} = \overline{x} \qquad \text{and} \qquad \overline{\alpha_i y} = \overline{y} ,$$

it follows that

$$({}^tT_M - sI)\{x, y\}$$

is a cycle with integer coefficients on $X(N)$. Inverting shows that $\{x, y\}$ lies in $H_1(X(N), \mathbf{Q})$, and proves the theorem.

David Rohrlich has investigated the situation for the Fermat curve, parametrized by modular functions. For the most classical parametrization, he has shown the analogous statement to be true, and determined the structure of the finite group generated by the cusps in the divisor class group. He has also given examples of curves belonging to non congruence subgroups for which the analogous statement is false. Cf. his forthcoming papers.

§ 3. Hecke Operators and Distributions

Following Manin [Man 4] and Mazur [Maz 1], we shall see how Hecke operators in certain situations lead to "distributions", discussed in Chapter XII.

We need a lemma. We let N be a positive integer, p a prime, $p \nmid N$. Let

$$\mathbf{M}_0^p(N)$$

be the set of integral matrices, with determinant p, and of the form

$$\alpha = \begin{pmatrix} a & b \\ Nc & d \end{pmatrix}$$

with c an integer. As mentioned before, $\Gamma_0(N)$ is the subgroup of $SL_2(\mathbf{Z})$ satisfying the same congruence condition as the matrices of $\mathbf{M}_0^p(N)$.

Lemma. *If*

$$\alpha_p = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}, \qquad \alpha_i = \begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix}, \quad i = 0, \ldots, p-1$$

then the α_i $(i=0,\dots,p-1)$ *form a complete set of coset representatives,*

$$\mathbf{M}_0^p(N) = \bigcup_{i=0}^{p} \pm\Gamma_0(N)\alpha_i .$$

Proof. The proof is essentially the same as the proof of the analogous statement when no congruence condition is imposed, cf. Chapter II, § 1. Given α as above, we find integers Nx, y such that

$$Nxa + Nyc = 0 ,$$

and such that x, y are relatively prime. Then Nx and y are also relatively prime, for suppose q is a prime dividing N and y. Then q divides xa, whence q divides a, contradicting the hypothesis that α has determinant p. We may then complete (Nx, y) to a matrix

$$\begin{pmatrix} z & w \\ Nx & y \end{pmatrix}$$

in $SL_2(\mathbf{Z})$, which is in fact an element of $\Gamma_0(N)$. Thus cosets are represented by matrices having 0 in the lower left corner. Using again elements of type

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

shows that representatives occur among the matrices α_i, and it is clear that they represent distinct cosets, as desired.

We may now form the Hecke operator, acting on $\Omega^1(\Gamma_0(N))$, by the formula

$$\omega \circ T(p) = \sum_{i=0}^{p} \omega \circ \alpha_i .$$

Operating on the right with $\Gamma_0(N)$ permutes the cosets of $\mathbf{M}_0^p(N)$, and consequently if $\omega \in \Omega^1(\Gamma_0(N))$, then $\omega \circ T(p)$ is invariant under $\Gamma_0(N)$.

For rational x, y the modular symbol $\{x, y\}$ depends only on the class of x, y mod $\mathbf{Z}$ because $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ lies in $\Gamma_0(N)$. Hence the modular symbol is defined on $(\mathbf{Q}/\mathbf{Z})^2$.

By duality, as before we have the action of $T(p)$ on the modular symbols, namely for any rational numbers x, y we have

$${}^tT(p)\{x, y\} = \sum \{\alpha_i x, \alpha_i y\} .$$

We look at this for the special case, where we put $\{x\} = \{0, x\}$. Then

$$\begin{aligned} {}^tT(p)\{x\} &= \sum_{i=0}^{p} \{\alpha_i 0, \alpha_i x\} \\ &= \sum_{i=1}^{p-1} \left\{\frac{i}{p}, \frac{x+i}{p}\right\} + \{px\} \\ &= \sum_{i=0}^{p-1} \left\{\frac{i}{p}, 0\right\} + \sum_{i=0}^{p-1} \left\{0, \frac{x+i}{p}\right\} + \{px\} . \end{aligned}$$

To make the notation more functional, put

$$f(x) = \{x\} ,$$

and define the **averaging operator**

$$A_1 f(x) = \sum_{i=0}^{p-1} f\left(\frac{x+i}{p}\right).$$

Then we obtain the formula (omitting the superscript t for simplicity)

$$T(p)f(x) = A_1 f(x) - A_1 f(0) + f(px) .$$

Let $\Omega^1(\Gamma_0(N), \lambda)$ for some complex number λ, be the λ-eigenspace of the space of differential forms under $T(p)$. Let

$$f_\lambda(x)$$

be the restriction of the modular symbol $\{x\}$ to this eigenspace.

Theorem 3.1. *The modular symbol* f_λ *on the* λ*-eigenspace for* $T(p)$ *satisfies the relation*

$$T(p)f_\lambda = \lambda f_\lambda = A_1 f_\lambda + f_\lambda \circ p - A_1 f_\lambda(0) ,$$

or in other words,

$$A_1 f_\lambda = \lambda f_\lambda - f_\lambda \circ p + A_1 f_\lambda(0) ,$$

This is a restatement of the formula obtained above, taking into account the eigen property of f_λ.

This fits the formalism to be studied in Chapter XII, § 2, where p-adic congruences are formally derived from such relations.

Chapter V. Coefficients and Periods of Cusp Forms on $SL_2(\mathbf{Z})$

In pioneering work, Eichler [E 2] discovered relations between periods of cusp forms, extended by Shimura [Sh 1]. Manin [Man 4] made more explicit the connection of these relations with the coefficients in the q-expansion, by using the Hecke operators and continued fractions, and in this chapter, we reproduce part of his paper, after stating the Eichler–Shimura relations.

Throughout this chapter and the next, we use the following notation.

Let $M_k^0 = M_k^0(\mathbf{R})$ be the space of cusp forms for $SL_2(\mathbf{Z}) = \Gamma(1)$, over the real numbers, of weight k. The integer $k-2$ plays a special role, so we put

$$w = k-2 .$$

Unless otherwise specified, a cusp form f is an element of M_k^0 (so $M_k^0(\mathbf{R})$ according to the convention in force).

For any integer $s = 0, \ldots, k-2 = w$ we define what we call a **period** (with moment) introduced by Eichler,

$$r_s(f) = \int_0^{i\infty} f(z) z^s \, dz .$$

The integral may be taken over the vertical axis (we discuss this in § 1, and also various convergence properties).

Since f has real coefficients, it follows that f takes on real values on the imaginary axis, from 0 to ∞. Letting

$$z = it , \qquad dz = i dt ,$$

we conclude:

If s is even, then $r_s(f)$ is pure imaginary. If s is odd, then $r_s(f)$ is real.

It is therefore natural to consider the mappings

$$r^+ : M_k^0 \to \mathbf{R}_{w+1}^+ = \mathbf{R}^{w/2+1}$$

$$r^- : M_k^0 \to \mathbf{R}_{w+1}^- = \mathbf{R}^{w/2}$$

given by

$$r^+(f) = \frac{1}{i}(r_0(f), \ldots, r_w(f))$$

$$r^-(f) = (r_1(f), \ldots, r_{w-1}(f)) .$$

In the next chapter, we determine the linear space which is the image of r^+ and r^-, in other words we determine the linear relations satisfied by r^+ and r^-, which are due to Eichler and Shimura. In this chapter, we analyse the effect of the Hecke operators on the periods, and obtain the new expressions of Manin for the coefficients of the q-expansion of f, when f is an eigenfunction of the Hecke algebra. We also show how Manin uses the Eichler–Shimura relations to obtain the rationality of period ratios over the field of coefficients of the cusp form, when the latter is an eigenfunction of the Hecke algebra. This was generalised to arbitrary levels by Razar [Raz].

Manin in his paper also discusses the p-adic theory, but it is already clear from the analogous case of Chapter IV that any formalism of Hecke operators leads to the abstract situation discussed in Chapter XII, Theorem 2.1, so that we shall omit this part of Manin's paper, for the special features pertaining to the present situation.

§ 1. The Periods and Their Integral Relations

Let $z_1, z_2 \in \mathfrak{H}^*$, so that z_1, z_2 either lie in the upper half plane, or are rational numbers, or $i\infty$. An integral

$$\int_{z_1}^{z_2}$$

will always be taken in the same manner as in Chapter IV, so that at the end points, it is taken along an analytic arc leading to the end point under a local parameter at infinity

$$e^{2\pi i z/m}$$

for some m, and otherwise, if an end point is rational, it is first transformed as an integral to infinity under an element of $SL_2(\mathbf{Z})$.

The function f being a cusp form, it follows that for any power z^s (with an integer s), the integral

$$\int_{z_1}^{z_2} f(z) z^s \, dz$$

converges. Indeed, suppose first $z_1 \in \mathfrak{H}$ and $z_2 = i\infty$. Since

$$|f(z)| \ll e^{-cy}$$

for sufficiently large y, it follows that in terms of the parameter q, the integrand has an order of magnitude bounded by

$$|q|^{c/m}(\log|q|)^s\left|\frac{dq}{q}\right|$$

for q near 0, and is consequently integrable near the end point at infinity. The same therefore holds when the end point is any cusp by a change of variables.

More generally, let

$$\sigma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in SL_2(\mathbf{Z})\,.$$

Then by the change of variables formula, we get for any integer s,

$$\int_{\sigma(z_1)}^{\sigma(z_2)} f(z)z^s\,dz=\int_{z_1}^{z_2}\sigma^*(f(z)z^s\,dz)=\int_{z_1}^{z_2}f(z)(az+b)^s(cz+d)^{w-s}\,dz$$

by using the definitions, $w=k-2$, and

$$f\circ[\sigma]_k=f(\sigma z)(cz+d)^{-k}=f(z)\,.$$

In particular, if s is an integer with

$$0\leqslant s\leqslant k-2$$

(an interval which we call the critical strip), then the powers of linear functions of z on the right-hand side are polynomials in z. In the sequel we work exclusively with such values of s.

If x_1, x_2 are cusps (so rational numbers or ∞), then we define the integral

$$\int_{x_1}^{x_2} f(z)z^s\,dz\,,\qquad s=0,\ldots,k-2$$

to be a **period** of f (with **moment** s). We are especially interested in the special periods

$$r_s(f)=\int_0^{i\infty} f(z)z^s\,dz\,.$$

To write down some relations, it is best to use matrix notation.

We use the abbreviation $z^{(w)}\,dz$ for the column vector

$$z^{(w)}\,dz=\begin{pmatrix}z^w\,dz\\ \vdots\\ z^0\,dz\end{pmatrix}.$$

We write

$$\omega(f)=f(z)z^{(w)}\,dz=\begin{pmatrix}f(z)z^w\,dz\\ \vdots\\ f(z)z^0\,dz\end{pmatrix}.$$

From the relation

$$\sigma^*(f(z)z^j\,dz)=f(z)(az+b)^j(cz+d)^{w-j}$$

we see that there exists an integral matrix $\pi(\sigma)$ such that

$$\sigma^*(f(z)z^{(w)}\,dz)=\pi(\sigma)(f(z)z^{(w)}\,dz)\,.$$

Then the change of variables formula can be written in the form

$$\pi(\sigma)\int_0^{i\infty}\omega(f)=\int_{\sigma(0)}^{\sigma(i\infty)}\omega(f)=\int_0^{i\infty}\sigma^*\omega(f)\,.$$

We let

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

be the usual generators of $SL_2(\mathbf{Z})$. We then get relations among the periods as follows. We have, suppressing the integrand $\omega(f)$:

$$(I+\pi(S))\int_0^{i\infty}=\int_0^{i\infty}+\int_{S(0)}^{S(i\infty)}=0$$

because $S(0)=i\infty$ and $S(i\infty)=0$, so the two integrals cancel. Furthermore, in a similar way,

$$(I+\pi(ST)+\pi((ST)^2)\int_0^{i\infty}=\int_0^{i\infty}+\int_{ST(0)}^{ST(i\infty)}+\int_{(ST)^2(0)}^{(ST)^2(i\infty)}=\int_0^{i\infty}+\int_{-1}^{0}+\int_{i\infty}^{-1}=0$$

again because the integrals cancel.

Using

$$\int_0^{i\infty} \sigma^*\omega(f) \text{ instead of } \pi(\sigma)\int_0^{i\infty}$$

yields a system of linear equations in the periods, with integer coefficients. Furthermore, since the periods $r_j(f)$ with even j are pure imaginary, and the periods $r_j(f)$ with odd j are real, these relations can be decomposed into two sets of relations, one involving only the even periods, and the other involving only the odd periods.

We let $\mathbf{R}^+_{w+1}$ be the euclidean space whose coordinates

$$(r_0, r_2, \ldots, r_w)$$

are indexed by the even integers from 0 to w. Similarly, we let $\mathbf{R}^-_{w+1}$ be the space whose coordinates

$$(r_1, r_3, \ldots, r_{w-1})$$

are indexed by the odd integers from 1 to $w-1$. Then the above relations with integer coefficients define two subspaces V^+ of $\mathbf{R}^+_{w+1}$ and V^- of $\mathbf{R}^-_{w+1}$ respectively, and these spaces are defined over $\mathbf{Q}$.

The basic theorem is:

Theorem 1.1. (Eichler–Shimura) *The map*

$$f \mapsto r^-(f)$$

is an isomorphism of M_k^0 with V^-. The map

$$f \mapsto r^+(f)$$

is an isomorphism of M_k^0 with a subspace of V^+ of codimension 1, *not containing the vector*

$$(1, 0, \ldots, 0, -1)\,.$$

This theorem will be proved in the next chapter. It will be used below for applications to the coefficients of cusp forms. In that application, we need not know the precise nature of the relations, merely that the spaces V^+ and V^- are defined over $\mathbf{Q}$, and that the image of r^+ does not contain $(1, 0, \ldots, 0, -1)$.

The linear relations can of course be written down explicitly. The first ones, arising from $\pi(S)$, are very simple. Indeed, using the transformation law

$$S^*(f(z)z^s\,dz) = f(z) + (-1)^s f(z) z^{w-s}\,dz\,,$$

we obtain:

ES 1. $$r_s + (-1)^s r_{w-s} = 0\,.$$

In particular, if s is even, then $r_s(f) = -r_{w-s}(f)$, and if s is odd then

$$r_s(f) = r_{w-s}(f)\,.$$

The other relations coming from ST look a little more complicated, and no use will be made of them, but we include them for the record. The reader may skip them. From the transformation formula for ST, we obtain for any s with $0 \leqslant s \leqslant w$:

$$0 = \int_0^{i\infty} f(z)[z^s + (z-1)^{w-s}(-1)^s + (z-1)^s z^{w-s}]\,dz$$

$$= r_s(f) + \sum_{j=0}^{s} \binom{w-s}{j} (-1)^{w-j+s} r_j(f) + \sum_{j=0}^{s} \binom{s}{j} (-1)^{s-j} r_{w-s+j}(f)\,.$$

Splitting these relations into real and imaginary parts, we obtain:

ES 2. $$r_s + (-1)^s \sum_{\substack{j=0 \\ j \text{ even}}}^{s} \binom{s}{j} r_{w-s+j} + (-1)^s \sum_{\substack{j=0 \\ j \equiv s \bmod 2}}^{w-s} \binom{w-s}{j} r_j = 0$$

ES 3. $$\sum_{\substack{j=1 \\ j \text{ odd}}}^{s} \binom{s}{j} r_{w-s+j} + \sum_{\substack{j=0 \\ j \not\equiv s \bmod 2}}^{w-s} \binom{w-s}{j} r_j = 0\,.$$

It looks messy, and as we said, no use will be made of such explicit formulas.

Our vector space V^- is then the subspace of $\mathbf{R}^-_{w+1}$ consisting of all vectors satisfying relations **ES 1**, **ES 2** for odd s, and **ES 3** for even s. The vector space V^+ is the subspace of $\mathbf{R}^+_{w+1}$ consisting of all vectors satisfying **ES 1**, **ES 2** for even s, and **ES 3** for odd s. In the next chapter, an equivalent formulation is given using a single space V.

§ 2. The Manin Relations

In the previous section, we have already proved:

Lemma 1. *If $0 \leqslant s \leqslant k-2$, and $\alpha \in SL_2(\mathbf{Z})$, then*

$$\int_{\alpha(0)}^{\alpha(i\infty)} f(z)z^s\,dz$$

is an integral linear combination of the periods $r_j(f)$, with

$$0 \leqslant j \leqslant k-2 .$$

We want a similar result when α is an integral matrix with positive determinant, not necessarily in $SL_2(\mathbf{Z})$. For this, we expand a cusp in a continued fraction.

Let b/d be a rational number, expressed as a fraction in lowest form, and say $d>0$. For the basic facts about continued fractions we refer to [L 4], Chapter I (all other facts will be proved). In the case of a rational number, the continued fraction

$$b/d=[a_0, a_1, \ldots, a_m]$$

may be written in two ways, differing in their last component. However, let us assume as will be the case in the applications, that $|b/d|<1$. Then of these two ways, one of them is uniquely determined by the condition that

$$a_m \geqslant 2$$

In the sequel, the continued fraction is always assumed to be so normalized. [Cf. [L 4], Chapter I, § 2, p. 7.]

Corresponding to this normalized continued fraction, we shall write the principal convergents with the notation

$$b_\nu/d_\nu \qquad \nu=0,\ldots,m ,$$

and we call $m=m(b/d)$ the **length** of the continued fraction. (The classical notation is p_n/q_n, but p, q are occupied for other purposes.) We have

$$\frac{b_0}{d_0}=\frac{a_0}{1},$$

so that $b_0=a_0$ and $d_0=1$. If $0<b/d<1$ then $a_0=0$.

Let

$$g_\nu=\begin{pmatrix} b_\nu & (-1)^{\nu-1}b_{\nu-1} \\ d_\nu & (-1)^{\nu-1}d_{\nu-1}\end{pmatrix}, \qquad \nu=0,\cdots,m .$$

Then

$$g_\nu(0)=b_{\nu-1}/d_{\nu-1} \qquad \text{and} \qquad g_\nu(i\infty)=b_\nu/d_\nu .$$

Lemma 2. *Let b/d be a fraction in lowest form, $0<b/d<1$. Then*

$$\int_0^{b/d} f(z)z^s\,dz ,$$

is a linear integral combination of the periods $r_j(f)$.

Proof. We have

$$\int_0^{b/d} f(z)z^s\,dz=\sum \int_{g_\nu(0)}^{g_\nu(i\infty)} f(z)z^s\,dz ,$$

and we apply Lemma 1 to conclude the proof.

We note that the coefficients occurring in these integral linear combinations can be computed explicitly from the continued fraction of b/d, without special difficulty. We shall compute them explicitly for the case $s=0$ in the next section.

On the other hand, we shall also be interested in the integral from 0 to $-b/d$, where again $0<b/d<1$. In that case, we use the matrices

$$h_\nu=\begin{pmatrix} -(-1)^{\nu-1}b_{\nu-1} & -b_\nu \\ (-1)^{\nu-1}d_{\nu-1} & d_\nu\end{pmatrix}$$

obtained by interchanging the columns of g_ν, and multiplying the first row by -1. The determinant is again 1. Thus we get:

$$h_\nu(0)=-b_\nu/d_\nu \qquad \text{and} \qquad h_\nu(i\infty)=-b_{\nu-1}/d_{\nu-1} .$$

Lemma 3. *With b/d as above,*

$$\int_0^{-b/d} f(z)z^s\,dz$$

is a linear integral combination of the periods $r_j(f)$.

Proof. We have

$$\int_0^{-b/d} = -\sum \int_{h_\nu(0)}^{h_\nu(i\infty)}$$

and again Lemma 1 concludes the proof.

For the applications, we determine explicitly these linear combinations when $s=0$. Actually, the integrals will occur in pairs, so we state the result in the form needed.

Lemma 4. *Let b/d be a fraction in lowest form, $0<b/d<1$. Then*

$$\int_0^{b/d} + \int_0^{-b/d} f(z)\,dz=\sum_{\nu=1}^{m(b/d)} \sum_{\substack{j=2 \\ j \text{ even}}}^{w-2} \binom{w}{j} (d_\nu^j d_{\nu-1}^{w-j}-d_{\nu-1}^j d_\nu^{w-j})\, r_j(f) + 2(1-d^w)r_0(f) .$$

Proof. Let b_ν/d_ν $(\nu=0,\dots,m)$ be the principal convergents to b/d, with $b_0=0$. Then

$$\int_0^{b/d} + \int_0^{-b/d} f(z)\,dz = \sum_{\nu=1}^{m(b/d)} \int_{g_\nu(0)}^{g_\nu(i\infty)} - \int_{h_\nu(0)}^{h_\nu(i\infty)} f(z)\,dz$$

$$= \sum_{\nu=1}^{m(b/d)} \int_0^{i\infty} f(z)[(d_\nu z+(-1)^{\nu-1}d_{\nu-1})^w - ((-1)^{\nu-1}d_{\nu-1}z+d_\nu)^w]\,dz\,.$$

We expand with the binomial theorem. Because of the symmetry relation

$$r_j(f)=r_{w-j}(f)$$

for odd j, we see that the terms with odd j will cancel, and therefore this last expression is equal to

$$\sum_{\nu=1}^{m(b/d)} \sum_{\substack{j=0\\ j\text{ even}}}^{w} \binom{w}{j} (d_\nu^j d_{\nu-1}^{w-j} - d_{\nu-1}^j d_\nu^{w-j}) r_j(f)\,.$$

Since $r_0(f)=-r_w(f)$ by the symmetry relation, it follows that there is a term with $r_0(f)$ having exactly the coefficient stated in the lemma, which is now clear.

§ 3. Action of Hecke Operators on the Periods

We let $\mathbf{M}^n$ be the set of integral matrices with determinant n. Then we have the coset decomposition

$$\mathbf{M}^n = \bigcup_{i=1}^{\psi(n)} \Gamma(1)\alpha_i\,,$$

where the matrices α_i can be chosen to be

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, \qquad a,d>0,\ ad=n,\ b \bmod d\,.$$

The corresponding Hecke operator is

$$T_k(n)=n^{k/2-1}\sum[\alpha_i]_k\,.$$

We recall that $w=k-2$. We shall state Manin's general theorem for arbitrary n at the end of the section. Here we carry out in detail the case when $n=p$ is a prime. This case already exhibits all the main features of the arguments, and if the reader does not want to work out the details for arbitrary n, he can look them up in Manin. The prime case is all we need for the applications.

As usual, we let

$$\sigma_m(n)=\sum_{d\mid n} d^m\,.$$

We let f be a cusp form for $SL_2(\mathbf{Z})$, of weight k with real coefficients.

Theorem 3.1. *Let p be a prime >2. The number*

$$\int_0^{i\infty} T_k(p)f(z)\,dz$$

is a linear combination of $r_0(f),\dots,r_{w-2}(f)$ with integer coefficients, given by the following expression, where b_ν/d_ν $(\nu=0,\dots,m(b/d))$ are the principal convergents to b/d, and $d=p$.

$$\int_0^{i\infty} T_k(p)f(z)\,dz=\sigma_{w+1}(p)r_0(f)-\sum_{b=1}^{[p/2]}\sum_{\nu=1}^{m(b/d)}\sum_{\substack{j=2\\ j\text{ even}}}^{w-2}\binom{w}{j}(d_\nu^j d_{\nu-1}^{w-j}-d_{\nu-1}^j d_\nu^{w-j})r_j(f)\,.$$

Proof. By definition, for arbitrary n,

$$\int_0^{i\infty} T_k(n)f(z)\,dz=\sum_{d\mid n}\sum_{b \bmod d}\int_0^{i\infty} f\left(\frac{n}{d^2}z+\frac{b}{d}\right)\frac{n^{k-1}}{d^k}\,dz\,.$$

We change variables, let $z\mapsto d^2z/n$, and then make a translation by $-b/d$, to see that this last expression is

$$\sum_{d\mid n}\sum_{b \bmod d}\frac{n^w}{d^w}\int_{b/d}^{i\infty} f(z)\,dz\,.$$

Writing

$$\int_{b/d}^{i\infty} = \int_0^{i\infty} - \int_0^{b/d}\,,$$

we find

$$-\int_0^{i\infty} T_k(n)f(z)\,dz=-n\sigma_{w-1}(n)r_0(f)+\sum_{d\mid n}\sum_{b \bmod d}\frac{n^w}{d^w}\int_0^{b/d} f(z)\,dz\,.$$

We now use $n=p$, so that $d=1$ or $d=p$. The term with $d=1$ is equal to 0, so all that remains is the term with $d=p$. For $p\neq 2$ we pick the residue class representatives b satisfying

$$-\frac{p}{2}<b<\frac{p}{2}\,,$$

so that the expression on the right-hand side is equal to

$$-p\sigma_{w-1}(p)r_0(f)+\sum_{1\le b<p/2}\left(\int_0^{b/p}+\int_0^{-b/p}\right)f(z)\,dz\,.$$

We may now apply Lemma 4 of § 2. The coefficient of $r_0(f)$ is equal to

$$-p(1+p^{w-1})+2(1-p^w)\frac{p-1}{1}=-(1+p^{w+1})=-\sigma_{w+1}(p)\,.$$

The other terms are precisely those stated in the theorem.

Theorem 3.2. *For $p=2$, we have*

$$\int_0^{i\infty}T_k(2)f(z)\,dz=\sigma_{w+1}(2)r_0(f)-\sum_{\substack{j=2\\ j\text{ even}}}^{w-2}\binom{w}{j}2^j r_j(f)\,.$$

Proof. We start as before, but when $n=2$ we have only the possibilities $b=0$ and $b=1$, so there is only one integral, with $b=1$, $d=2$. We thus find

$$\int_0^{i\infty}T_k(2)f(z)\,dz=2\sigma_{w-1}(2)r_0(f)-\int_0^{1/2}f(z)\,dz\,.$$

We use the matrix

$$g=\begin{pmatrix}1&0\\2&1\end{pmatrix}$$

in $SL_2(\mathbf{Z})$, which is such that $g(0)=0$ and $g(i\infty)=1/2$. By the change of variables formula, we see that the right-hand side is equal to

$$2(1+2^{w-1})r_0(f)-\sum_{j=0}^{w}\binom{w}{j}2^j r_j(f)\,.$$

Using $r_w=-r_0$ immediately yields the desired formula, where however, the sum is taken for $j=1,\dots,w-1$. But $T_k(2)f$ is a cusp form with real coefficients, and the expression on the left is $r_0(T_k(2)f)$, which is pure imaginary. Hence all the real terms with odd j on the right-hand side can be omitted from the sum, as was to be shown.

Suppose now that f is a cusp form and is an eigenfunction for all Hecke operators $T_k(p)$, normalized to have $a_1=1$. Then

$$T_k(p)f=a_p f\,,$$

and hence the left-hand side in the previous two theorems is equal to

$$a_p r_0(f)\,.$$

This immediately gives a_p as an explicit integral linear combination of the period ratios

$$r_j(f)/r_0(f)\,.$$

Note that $r_0(f)=-r_w(f)\neq 0$, because $r_w(f)$ is, up to obvious factors, the value $L_f(w+1)$, where L_f is the associated Dirichlet series, cf. Chapter I, § 5, having an Euler product (Chapter II, § 2), and therefore not vanishing at $s=w+1$.

We shall prove in a later section that the above ratios are elements of the field $\mathbf{Q}(a_1, a_2, a_3, \dots)$ generated by the coefficients of f.

The general formula for composite n is given in the next theorem. There, one has to take into account the fact that b, d may not be relatively prime, and extract a common divisor, giving rise to the sums $\sigma_w(n/d)$. Otherwise, the proof is the same. Of course, that $r_0(T_k(n)f)$ is a linear combination of periods with integral coefficients is trivial once one knows this for all $T_k(p)$, since the Hecke algebra is generated by $T_k(p)$. Only the explicit determination of the coefficients remains in question. *No use will be made of this formula for composite n, and the reader may omit it.*

Theorem 3.3. *The number*

$$\int_0^{i\infty}T_k(n)f(z)\,dz$$

is a linear combination of $r_0(f),\dots,r_{w-2}(f)$ with integer coefficients. More precisely, it is given by the following expression, where b_l/d_l ($l=0,\dots,m(b/d)$) are the principal convergents to b/d.

$$\begin{aligned}\int_0^{i\infty}T_k(n)f(z)\,dz=&\\ &\sigma_{w+1}(n)r_0(f)\\ &-\sigma_w(n/2)\sum_{i=1}^{[(w-2)/4]}\binom{w}{2i}[2^{2i}-2^{w-2i}]r_{2i}(f)\\ &-\sum_{\substack{d\mid n\\ d\ge 3}}\sigma_w(n/d)\sum_{\substack{1\le b<d/2\\ (b,d)=1}}\sum_{l=1}^{m(b/d)}\sum_{i=1}^{[(w-2)/4]}2\binom{w}{2i}(d_l^{2i}d_{l-1}^{w-2i}-d_l^{w-2i}d_{l-1}^{i})r_{2i}(f)\,.\end{aligned}$$

The term $\sigma_w(n/2)$ is 0 if n is odd, and otherwise has the usual value

$$\sigma_w(n/2)=\sum_{d\mid(n/2)}d^w\,.$$

We consider the mappings r^+ and r^- as mappings into the space of column vectors. We write T_n instead of $T_k(n)$.

Theorem 3.4. *There exists integral matrices A_n^+ and A_n^- such that for all $f \in M_k^0$ we have*

$$r^+(T_n f) = A_n^+ r^+(f) \qquad \text{and} \qquad r^-(T_n f) = A_n^- r^-(f) .$$

The matrix A_n^+ may be chosen so that its first row is equal to

$$(\sigma_{w+1}(n), *, *, *, 0) .$$

Proof. Let $0 \leqslant j \leqslant w$. We have

$$r_j(T_n(f)) = \int_0^{i\infty} T_n f(z) z^j \, dz = \sum_{d|n} \frac{n^{k-1}}{d^k} \sum_{b \bmod d} \int_0^{i\infty} f\left(\frac{n}{d^2} z + \frac{b}{d}\right) z^j \, dz$$

$$= \sum_{d|n} \left(\frac{n}{d}\right)^w \sum_{b \bmod d} \int_{b/d}^{i\infty} f(z) \left(\frac{d^2}{n}\right)^j \left(z - \frac{b}{d}\right)^j dz = \sum_{d|n} \sum_{b \bmod d} \int_0^{i\infty} - \int_0^{b/d} f(z) Q(z) \, dz$$

where $Q(z)$ is the polynomial

$$Q(z) = \frac{n^w}{d^w} \frac{d^{2j}}{n^j} \sum_{i=0}^{j} \binom{j}{i} \left(\frac{-b}{d}\right)^{j-i} z^i .$$

The coefficients of this polynomial are

$$\binom{j}{i} (-b)^{j-i} \frac{n^{w-j}}{d^{w-i-j}} ,$$

and therefore are integers. Thus our last expression is an integral linear combination of integrals

$$\int_0^{i\infty} - \int_0^{b/d} f(z) z^i \, dz ,$$

to which we can apply Lemma 1 to conclude the proof that we can choose the matrices A_n^+ and A_n^- to have integral coefficients.

Furthermore, the symmetry relation

$$r_s = -r_{w-s}$$

for even s shows that the periods

$$r_0(f), r_2(f), \ldots, r_s(f) , \qquad s \text{ even } \leqslant \frac{w-2}{2} ,$$

already generate the even periods over $\mathbf{Z}$. Consequently the matrix A_n^+ may be chosen to have zero on the upper right. The precise value for its first coefficient follows from Theorems 3.1 and 3.2 when n is prime, and Theorem 3.3 when n is not prime. However, if one uses the recursion formula for Hecke operators, the value of the first coefficient for composite n can also be deduced independently from the values for n prime.

§ 4. The Homogeneity Theorem

All we need of the Eichler–Shimura theorem is the following:

There exist linear subspaces V^+ of $\mathbf{R}_{w+1}^+$ and V^- of $\mathbf{R}_{w+1}^-$, defined over $\mathbf{Q}$, such that the map

$$r^- : M_k^0 \to V^-$$

is an isomorphism, and

$$r^+ : M_k^0 \to V^+$$

is an embedding of codimension 1, whose image does not contain

$$(1, 0, \ldots, 0, -1) .$$

Let $f \in M_k^0$, $f = \sum a_n q^n$. We say that f is **normalized** if $a_1 = 1$. We let

$$\mathbf{Q}_f = \mathbf{Q}(a_1, a_2, \ldots)$$

be the field generated over $\mathbf{Q}$ by the Fourier coefficients of f.

Theorem 4.1. *Let $f \in M_k^0$ be a normalized cusp form, which is an eigenfunction of the Hecke algebra. Then the ratios*

$$r_1(f) : \ldots : r_{w-1}(f) \qquad \text{and} \qquad r_0(f) : \ldots : r_w(f)$$

are elements of $\mathbf{Q}_f$. In other words, there exist real numbers π_f^-, π_f^+, and vectors $t^-(f), t^+(f)$ with coordinates in $\mathbf{Q}_f$ such that

$$r^-(f) = \pi_f^- t^-(f) \qquad \text{and} \qquad r^+(f) = \pi_f^+ t^+(f) .$$

Proof. We deal first with $r^-(f)$. Let E_f^- be the subspace of $\mathbf{R}_{w+1}^-$ consisting of all vectors v such that

$$A_n^- v = a_n v \qquad \text{(all } n\text{)} .$$

Then E_f^- is the intersection of all subspaces

$$\bigcap_n \mathrm{Ker}(A_n^- - a_n I) ,$$

and is defined over $\mathbf{Q}_f$. Furthermore V^- is defined over $\mathbf{Q}$. Then $E_f^- \cap V^-$ has dimension 1, because of the Eichler–Shimura isomorphism, and the fact that the Fourier coefficients determine the cusp form uniquely. Since $E_f^- \cap V^-$ is defined over $\mathbf{Q}_f$, and since $r^-(f)$ lies in E_f^-, the theorem follows for r^-.

The result also follows in the same way for r^+ once we have proved the following lemma, which was more trivial for r^- than for r^+, due to the restriction to a subspace of codimension 1 in the Eichler–Shimura isomorphism.

Lemma. *Let E_f^+ be the subspace of $\mathbf{R}_{w+1}^+$ consisting of all vectors v such that*

$$A_n^+ v = a_n v \quad \textit{for all } n .$$

Then $E_f^+ \cap V^+$ has dimension 1.

Proof. We have to show that an element $s \in E_f^+ \cap V^+$ is in the image of r^+. By the Eichler–Shimura isomorphism there exists $g \in M_k^0$ such that

$$s = r^+(g) + (x, 0, \ldots, 0, -x) , \qquad \text{some } x \in \mathbf{R} .$$

In particular, we already have

$$\frac{1}{i} r_j(g) = s_j , \qquad \text{for } \ j \text{ even}, \ 2 \leqslant j \leqslant w-2 .$$

It suffices to show that $\frac{1}{i} r_0(g) = s_0$. We let $n = p$ be prime.

By Theorem 3.4, we can take the matrix A_n^+ to have first row equal to

$$(\sigma_{w+1}(n), \ldots, 0) .$$

By definitions and hypothesis, the 0-th component of

$$(\sigma_{w+1}(n) I - A_n^+) r^+(g) \qquad \text{and} \qquad (\sigma_{w+1}(n) I - A_n^+) s$$

are the same. Hence we obtain

$$\frac{1}{i} \sigma_{w+1}(n) r_0(g) - \frac{1}{i} r_0(T_n g) = \sigma_{w+1}(n) s_0 - a_n s_0 .$$

Divide by $\sigma_{w+1}(n)$. We know from Chapter I, § 4 that

$$|a_n| \ll n^{k/2} .$$

Furthermore, g is a fixed linear combination of eigenfunctions for the Hecke algebra, and hence we also get the estimate

$$|r_0(T_n g)| \ll n^{k/2} .$$

Take the limit for $n \to \infty$ to get

$$\frac{1}{i} r_0(g) = s_0 ,$$

thereby proving the lemma, and as we have seen, also the theorem.

The above theorem is due to Manin. Special cases were obtained by Shimura [Sh 1].

Chapter VI. The Eichler–Shimura Isomorphism on $SL_2(\mathbf{Z})$

In this chapter we describe the Eichler–Shimura theory already mentioned in the preceding chapter.

The period mapping

$$f \mapsto \int_0^{i\infty} \mathrm{Re} f(z) z^j \, dz \,, \qquad j = 0, \ldots, k-2$$

where $k = w+2$ is the weight, yields an embedding of the space of cusp forms into a euclidean space, and we determine the precise image. This determines a system of linear equations satisfied by the coefficients of the cusp forms.

The first section develops the general algebraic formalism which linearizes the pull back operation

$$f(z) z^j \, dz \mapsto \sigma^*(f(z) z^j \, dz) = f(\sigma z)(\sigma z)^j \, d(\sigma z)$$

for $\sigma \in SL_2(\mathbf{Z})$. It gives rise to a representation, and we also define a scalar product which reflects the Petersson product.

We then show that the image of the period mapping lies in a certain space defined by rational linear equations, and that the map is injective. Finally we count dimensions to get the precise image.

The exposition essentially follows Shimura [Sh 1]. See especially that last part of this paper, where Shimura was already aware of certain implications of the period mapping for the rationality of the quotients of periods. However, both Shimura and Eichler formulate their map in terms of cohomology. For our purposes, it is better to deal directly with the period mapping, from the cusp forms into euclidean space, and I am indebted to David Rohrlich for the arrangement of the proof which eliminates the cohomology except for one brief appearance of the cocycle relation, to prove that the period map is injective. This makes it clearer just how the cocycle relation is used: If two cocycles have the same values on a set of generators of the group, then they are equal. The relation between the period mapping and the Eichler cohomology mapping is given as a commutative diagram in § 5.

In no way should the above arrangement be interpreted as a desire to hide the cohomology. In fact, there are obvious reasons to give presentations emphasizing it, especially as connects with de Rham cohomology, and the image of a cusp

form f under the Eichler–Shimura map can also be written

$$f(\tau)(X - \tau Y)^k \, d\tau \,,$$

as a differential form on the modular curve.

§ 1. The Polynomial Representation

Let w be a positive integer which will eventually be assumed to be even to fit with the notation of the preceding chapter. Let $\mathbf{P}_w$ be the vector space of real homogeneous polynomials in two variables X, Y of degree w over the real numbers. Thus

$$\mathbf{P}_w = \sum_{j=0}^{w} \mathbf{R} X^j Y^{w-j}.$$

Then $\dim \mathbf{P}_w = w+1$. If

$$\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $GL_2(\mathbf{R})$, then we can define a linear endomorphism $\mathbf{P}(\sigma)$ of $\mathbf{P}_w$ by prescribing for any polynomial $\varphi(X, Y)$ that

$$\mathbf{P}(\sigma)\varphi(X, Y)) = \varphi((X, Y)\sigma)$$

where

$$(X, Y)\sigma = (X, Y)\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (aX + cY, bX + dY) \,.$$

In fact, the association $(X, Y) \mapsto (X, Y)\sigma$ extends to an algebra automorphism of the polynomial algebra $\mathbf{R}[X, Y]$, and $\mathbf{P}(\sigma) = \mathbf{P}_w(\sigma)$ is the restriction to the homogeneous subspace of degree w. Note that

$$\mathbf{P}(\sigma\tau) = \mathbf{P}(\sigma)\mathbf{P}(\tau) \,.$$

Thus we get a representation of $GL_2(\mathbf{R})$ on $\mathbf{P}_w$.

We can also define

$$\mathbf{P}^*(\sigma)(\varphi(X, Y)) = \varphi((X, Y)^t\sigma) \,.$$

Then $\mathbf{P}^*$ is an antirepresentation. We define the matrix $\pi(\sigma)=\pi_w(\sigma)$ by the condition

$$\begin{pmatrix} \mathbf{P}^*(\sigma)X^wY^0 \\ \vdots \\ \mathbf{P}^*(\sigma)X^jY^{w-j} \\ \vdots \\ \mathbf{P}^*(\sigma)X^0Y^w \end{pmatrix} = \pi(\sigma) \begin{pmatrix} X^wY^0 \\ \vdots \\ X^jY^{w-j} \\ \vdots \\ X^0Y^w \end{pmatrix}.$$

Then π is covariant, i.e.

$$\pi(\sigma\tau)=\pi(\sigma)\pi(\tau).$$

Thus $\sigma \mapsto \pi(\sigma)$ is a representation.

If

$$u=\sum u_{w-j}X^{w-j}Y^j$$

is an element of $\mathbf{P}_w$, we let

$$C(u)={}^t(u_w, u_{w-1}, \ldots, u_0)$$

be the column vector of coordinates of u.

Let s_1, s_2 be numbers. We write

$$u(s)=u(s_1, s_2)=u(s_1, s_2; X, Y)=\sum s_1^{w-j}s_2^jX^{w-j}Y^j,$$

so that the coordinate vector $C(u(s_1, s_2))$, also written $C(s_1, s_2)$, is given by

$$C(s_1, s_2)=\begin{pmatrix} s_1^ws_2^0 \\ \vdots \\ s_1^0s_2^w \end{pmatrix}.$$

We define the linear map

$$L:\mathbf{P}_w \to \mathbf{P}_w \quad \text{such that} \quad L(X^{w-j}Y^j)=\binom{w}{j}^{-1}X^{w-j}Y^j.$$

Then

$$u(s_1, s_2)=L(s_1X+s_2Y)^w=L((s_1, s_2)^t(X, Y))^w.$$

We have

$$\mathbf{P}(\sigma)u(s_1, s_2; X, Y)=u(s_1, s_2; (X, Y)\sigma)=u((s_1, s_2)^t\sigma; X, Y).$$

From this we immediately get the general formula

$$\boxed{C(\mathbf{P}(\sigma)u)=\pi(\sigma)C(u)}$$

for all $u \in \mathbf{P}_w$.

Theorem 1.1. *Let w be even. There exists a unique symmetric bilinear form $[u, v]$ on $\mathbf{P}_w$ such that $\mathbf{P}(\sigma)$ is an automorphism of the form for all $\sigma \in SL_2(\mathbf{R})$, and such that for elements $u(s)$, $u(t)$, we have*

$$[u(s), u(t)]=(s_1t_2-s_2t_1)^w.$$

If Q is the matrix of the form with respect to the basis already used, this means:

$${}^t\pi(\sigma)Q\pi(\sigma)=Q, \quad \text{i.e.} \quad [\pi(\sigma)C, \pi(\sigma)C']=[C, C']$$

$${}^tC(s_1, s_2)QC(t_1, t_2)=(s_1t_2-s_2t_1)^w.$$

Proof. We let Q be the unique matrix with integer coefficients such that

$$(s^w, s^{w-1}, \ldots, 1)Q \begin{pmatrix} t^w \\ t^{w-1} \\ \vdots \\ 1 \end{pmatrix} = (s-t)^w.$$

This is an inhomogeneous form of the relation

$$(s_1^ws_2^0, \ldots, s_1^0s_2^w)Q \begin{pmatrix} t_1^wt_2^0 \\ \vdots \\ t_1^0t_2^w \end{pmatrix} = (s_1t_2-s_2t_1)^w.$$

We then define the scalar product on $\mathbf{P}_w$ in terms of the coordinates by

$$[u, v]={}^tC(u)QC(v).$$

Note that ${}^tQ=(-1)^wQ$ and so is equal to Q since w is assumed to be even. Hence Q is symmetric, and defines a symmetric bilinear form, which clearly has the desired value on the special elements $u(s)$, $u(t)$. Furthermore, if

$$(s_1', s_2')=(s_1, s_2)^t\sigma \qquad \text{and} \qquad (t_1', t_2')=(t_1, t_2)^t\sigma,$$

then
$$(s_1't_2' - s_2't_1')^w = (\det \sigma)^w (s_1t_2 - s_2t_1) .$$

From this and the transformation formula for $\mathbf{P}(\sigma)u(s_1, s_2)$ and $\mathbf{P}(\sigma)u(t_1, t_2)$ we immediately see that
$$[\mathbf{P}(\sigma)u, \mathbf{P}(\sigma)v] = [u, v]$$

for the special case $u = u(s)$, $v = u(t)$, whence also for all u, v. This proves the theorem.

Remark. We have oiled the machinery which allows us to go back and forth between the operation on the polynomial algebra, and the operation on column vectors. Depending on the occasion, one is more useful than the other. For instance, in the next section, the operation of the matrix $\pi(\sigma)$ on coordinate vectors is the appropriate one. In § 4, the operation on polynomials is more efficient.

§ 2. The Shimura Product on Differential Forms

Let f be a cusp form for $SL_2(\mathbf{Z})$ of weight $k = w+2$. Then from the definitions, if $\sigma \in SL_2(\mathbf{Z})$, we get
$$\sigma^*(f(z)z^j\, dz) = f(z)(az+b)^j(cz+d)^{w-j}\, dz ,$$

replacing z by σz. Consequently we see that the representation $\sigma \mapsto \pi(\sigma)$ was chosen so that

$$\boxed{\sigma^*(f(z)z^{(w)}\, dz) = \pi(\sigma)(f(z)z^{(w)}\, dz) .}$$

Furthermore, since $\pi(\sigma)$ is a real matrix, we also get the formula

$$\boxed{\sigma^*(\operatorname{Re} f(z)z^{(w)}\, dz) = \pi(\sigma)(\operatorname{Re} f(z)z^{(w)}\, dz) .}$$

The properties of the scalar product of § 1 can now be used for a similar product on differential forms.

If Q is the symmetric matrix of § 1, and
$$\omega = {}^t(\omega_w, \ldots, \omega_0) , \qquad \eta = {}^t(\eta_w, \ldots, \eta_0)$$

are column vectors whose components are C^∞ differential forms on the upper half plane, we define their product
$$[\omega, \eta] = {}^t\omega \wedge Q\bar{\eta} = {}^t\omega Q \wedge \bar{\eta} .$$

Then the following properties hold trivially from the definition of the matrix Q and Theorem 1.1, for the differential product just defined, and $\sigma \in SL_2(\mathbf{R})$.

DP 1. $\sigma^*[\omega, \eta] = [\sigma^*\omega, \sigma^*\eta]$

DP 2. $[\pi(\sigma)\omega, \pi(\sigma)\eta] = [\omega, \eta]$

DP 3. $d[\omega, \eta] = [d\omega, \eta] + (-1)^{\deg \omega}[\omega, d\eta]$

DP 4. $[\omega, \eta] = (-1)^{\deg \omega \deg \eta}[\eta, \omega]$

For the next property which relates the differential product with the Petersson product, we use the notation
$$\langle f, g \rangle = f\bar{g}y^{w+2}\frac{dx \wedge dy}{y^2}$$

for the integrand of the Petersson product. For the next two properties, we do not need to assume that f, g are cusp forms, they may be arbitrary C^∞ functions. The properties have only to do with the linear algebra of the matrix Q.

DP 5. $[f(z)z^{(w)}\, dz, g(z)z^{(w)}\, dz] = -(2i)^{w+1}\langle f, g \rangle$

DP 6.
$$\begin{aligned} 4[\operatorname{Re} f(z)z^{(w)}\, dz, \operatorname{Re} if(z)z^{(w)}\, dz] &= -(2i)^{w+1}(\langle f, if \rangle - \langle if, f \rangle) \\ &= -(2i)^{w+2}\langle f, f \rangle . \end{aligned}$$

Observe that in this last property, up to a constant factor we get precisely the Petersson product of f with itself. In particular, if f is a cusp form for $SL_2(\mathbf{Z})$ of weight $k = w+2$, and the integral of the left-hand side in **DP 6** is equal to 0, then $f = 0$. Again **DP 5** follows at once from the definition of Q, and **DP 6** is then obvious.

§ 3. The Image of the Period Mapping

We let $w = k-2$ as before. We let
$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

be the usual elements of $SL_2(\mathbf{Z})$. We let π be the matrix representation π_w of the first two sections. We define two subspaces of $\mathbf{R}^{w+1}$:
$$V = \operatorname{Ker}(I + \pi(S)) \cap \operatorname{Ker}(I + \pi(ST) + \pi((ST)^2))$$
$$U = (I - \pi(S))\operatorname{Ker}(I - \pi(T)) .$$

We let
$$r : M_k^0(\mathbf{C}) \to \mathbf{R}^{w+1}$$

be the map given by

$$r(f) = \int_0^{i\infty} \operatorname{Re} f(z) z^{(w)}\, dz\ .$$

Theorem 3.1. *The period map r has an image contained in V, and induces an* **R***-isomorphism of $M_k^0(\mathbf{C})$ onto V/U.*

In this section we prove the first statement, and the fact that r induces an injection of $M_k^0(\mathbf{C})$ into V/U. In the next section, we count dimensions to show that r is an isomorphism.

We let

$$\rho = \rho(f) = \operatorname{Re} f(z) z^{(w)}\, dz\ ,$$

viewed as a column vector. For any $\sigma \in SL_2(\mathbf{Z})$, we have for any z_1, z_2 in the upper half plane or cusps:

$$\pi(\sigma) \int_{z_1}^{z_2} \rho(f) = \int_{z_1}^{z_2} \pi(\sigma)\rho(f) = \int_{z_1}^{z_2} \sigma^* \rho(f) = \int_{\sigma(z_1)}^{\sigma(z_2)} \rho(f)\ .$$

To show that the image of r is in V, we have

$$(I+\pi(S)) \int_0^{i\infty} = \int_0^{i\infty} + \int_{S(0)}^{S(i\infty)} = \int_0^{i\infty} + \int_{i\infty}^{0} = 0\ .$$

Secondly,

$$(I+\pi(ST)+\pi((ST)^2)) \int_0^{i\infty} = \int_0^{i\infty} + \int_{ST(0)}^{ST(i\infty)} + \int_{(ST)^2(0)}^{(ST)^2(i\infty)} = \int_0^{i\infty} + \int_{-1}^{0} + \int_{i\infty}^{-1}$$

and again we find 0. This proves that the image of r lies in V.

Next we show that if $r(f)$ lies in U, then $f=0$. We suppose that there is a vector $v \in \mathbf{R}^{w+1}$ such that

$$r(f) = (I-\pi(S))v \qquad \text{and} \qquad (I-\pi(T))v = 0\ ,$$

Lemma. *If $\sigma \in SL_2(\mathbf{Z})$ then*

$$\int_0^{\sigma(0)} \rho(f) = (I-\pi(\sigma))(-v+r(f))\ .$$

Proof. By a cocycle $\varphi: SL_2(\mathbf{Z}) \to \mathbf{R}^{w+1}$ we mean a map satisfying the relation

$$\varphi(\sigma) + \pi(\sigma)\varphi(\tau) = \varphi(\sigma\tau)\ .$$

It is clear that two cocycles which have the same value on S and T (which generate $SL_2(\mathbf{Z})$) are equal. It is also clear that the left-hand side and the right-hand side of the equation to be proved are cocycles, as functions of σ. Therefore it suffices to prove that they coincide for $\sigma = S$ and $\sigma = T$.

If $\sigma = S$ then the left-hand side is $r(f)$, and the right-hand side is (after the usual change of variables)

$$-r(f) + r(f) + r(f) = r(f)\ ,$$

which is what we want.

If $\sigma = T$, then the left-hand side is equal to

$$\int_0^1 \rho(f)\ .$$

The right-hand side is equal to

$$\int_0^{i\infty} \rho - \int_1^{i\infty} \rho = \int_0^1 \rho\ ,$$

as desired. This proves the lemma.

Now let us define

$$u = -v + r(f) = -\pi(S)v\ ,$$

and put for z in the upper half plane:

$$F(z) = -u + \int_0^z \rho(f)$$

$$G(z) = \int_0^z \operatorname{Re} if(z) z^{(w)}\, dz = \int_0^z \rho(if)\ .$$

Then F, G are vector valued holomorphic on the upper half plane, and

$$dF = \rho(f)\ , \qquad dG = \rho(if)\ .$$

Furthermore, we have

$$\pi(\sigma)F = F \circ \sigma \qquad \text{and} \qquad \pi(\sigma)\, dG = \sigma^*\, dG\ .$$

Proof. The second relation is standard by the definition of $\pi(\sigma)$. The first has a short proof as follows:

$$\pi(\sigma)F(z) = \pi(\sigma)\left(\int_0^z \rho - u\right) = \int_{\sigma(0)}^{\sigma(z)} \rho - \pi(\sigma)u = -\int_0^{\sigma(0)} \rho + \int_0^{\sigma(z)} \rho - \pi(\sigma)u$$

$$= -(I-\pi(\sigma))u + F(\sigma z) + u - \pi(\sigma)u \qquad \text{(by the lemma)}$$

$$= F(\sigma z)\,.$$

as was to be shown.

The hypothesis concerning $r(f)$ in U was used for the lemma and the formalism $\pi(\sigma)F = F\circ\sigma$. We now conclude the proof of injectivity, that $f=0$. In view of **DP 6**, it suffices to prove that if D is a fundamental domain for $\Gamma(1)\backslash\mathfrak{H}$, then

$$\int_D [\rho(f), \rho(if)] = 0\,,$$

because the integral on the left is the integral of the Petersson scalar product of f with itself, which is positive definitive. But we have:

$$\int_D [\rho(f), \rho(if)] = \int_D [dF, dG] = \int_D d[F, dG] = \int_{\partial D} [F, dG]\,.$$

The boundary of D consists of four pieces as shown on Fig. 7.

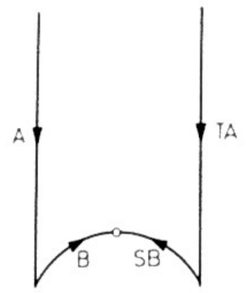

Fig. 7.

The integral over the vertical pieces of the boundary is equal to

$$\int_A - \int_{TA} [F, dG] = \int_A [F, dG] - \int_A [T^*F, T^*G]$$

$$= \int_A [F, dG] - \int_A [\pi(T)F, \pi(T)\,dG] = 0\,.$$

A similar argument shows that the integral over the other two pieces is equal to zero. This concludes the proof of the injectivity of the period mapping.

§ 4. Computation of Dimensions

The proof of Theorem 3.1 will be complete if we show that the image of the period mapping has the same dimension as V/U. In other words, it suffices to prove:

$$\boxed{\dim_{\mathbf{R}} M^0_{w+2}(\mathbf{C}) = \dim V - \dim U}$$

We do this in the present section. It is now convenient to view the spaces V and U as subspaces of the homogeneous polynomial space $\mathbf{P}_w$. We recall the operation

$$\mathbf{P}(\sigma)(\varphi(X, Y)) = \varphi((X, Y)\sigma)\,, \qquad \sigma\in SL_2(\mathbf{Z})\,.$$

In particular:

$$\mathbf{P}(S)X = -Y\,, \qquad \mathbf{P}(S)Y = X$$
$$\mathbf{P}(T)X = X\,, \qquad \mathbf{P}(T)Y = Y+X\,.$$

The spaces V, U are then defined by:

$$V = \mathrm{Ker}\,(I+\mathbf{P}(S)) \cap \mathrm{Ker}\,(I+\mathbf{P}(ST)+\mathbf{P}((ST)^2))$$
$$U = (I-\mathbf{P}(S))\mathrm{Ker}(I-\mathbf{P}(T))\,.$$

Lemma 1. *The kernel of $I-\mathbf{P}(T)$ is the one-dimensional space generated by X^w. Hence U is the one-dimensional space*

$$U = (X^w - Y^w)\,,$$

generated by the vector with coordinates $(1, 0, \ldots, 0, -1)$.

Proof. If a polynomial in X, Y is invariant under $Y\mapsto Y+X$, and so invariant under $Y\mapsto Y+nX$ for all n, then it cannot contain Y. This proves the first assertion. The second is obvious from the definitions.

We define the subspaces

$$E = \mathrm{Ker}\,(I+\mathbf{P}(S))$$
$$F = \mathrm{Ker}\,(I+\mathbf{P}(ST)+\mathbf{P}((ST)^2))\,.$$

Since $\mathbf{P}(S^2)=\mathbf{P}(-I)=I$, the characteristic polynomial of $\mathbf{P}(S)$ is X^2-1. These subspaces are chosen so that $V=E\cap F$. We get

$$\dim V = \dim E\cap F = \dim E+\dim F-\dim (E+F)\,.$$

The dimensions on the right are computed in the next lemma, and it is then immediate that the values found prove what we want, taking into account the dimensions found in Chapter I:

$$\dim_{\mathbf{C}} M^0_{w+2}(\mathbf{C})=\begin{cases}\left[\frac{w+2}{12}\right] & \text{if } w\not\equiv 0 \bmod 12\\ \left[\frac{w+2}{12}\right]-1 & \text{if } w\equiv 0 \bmod 12\,.\end{cases}$$

Since $X+1$ and $X-1$ are relatively prime, it follows that

$$E=\mathrm{Im}\,(I-\mathbf{P}(S))\,.$$

Since the characteristic polynomial of $\mathbf{P}(ST)$ is X^3-1, it follows that

$$F=\mathrm{Im}\,(I-\mathbf{P}(ST))\,.$$

Lemma 2. (i) *We have* $E+F=\mathbf{P}_w$, *so that* $\dim (E+F)=w+1$.

(ii) $$\dim E=\begin{cases}\frac{w}{2}+1 & \text{if } \frac{w}{2} \text{ is odd}\\ \frac{w}{2} & \text{if } \frac{w}{2} \text{ is even}\,.\end{cases}$$

(iii) $$\dim F=w+1-\begin{cases}[w/3]+1 & \text{if } w\equiv 0,2 \bmod 3\\ [w/3] & \text{if } w\equiv 1 \bmod 3\,.\end{cases}$$

Proof. As to (i), we have

$$\begin{aligned}\mathbf{P}(S)(E+F) &= \mathrm{Im}\,(\mathbf{P}(S)-I)\ +\ \mathrm{Im}(\mathbf{P}(S)-\mathbf{P}(T))\,.\\ &\supset \mathrm{Im}(I-\mathbf{P}(T))\,.\end{aligned}$$

We know by Lemma 1 that $\mathrm{Ker}(I-\mathbf{P}(T))$ is one-dimensional, and is generated by X^w-Y^w. On the other hand, we note that

$$\mathrm{Im}(\mathbf{P}(S)-I) \text{ contains } X^w-Y^w\,.$$

Hence $\mathbf{P}(S)(E+F)$ contains both the image of $I-\mathbf{P}(T)$ and X^w-Y^w, and is therefore equal to the whole space, thus proving (i).

As for (ii), an element

$$\sum r_j X^j Y^{w-j}$$

is in the kernel of $I+\mathbf{P}(S)$ if and only if

$$\sum r_j X^j Y^{w-j} + \sum (-1)^{w-j} r_j Y^j X^{w-j}=0\,,$$

which is equivalent with

$$r_j+(-1)^j r_{w-j}=0\,.$$

This proves the second assertion.

For (iii), we note that

$$\dim F=w+1-\dim \mathrm{Ker}\,(I-\mathbf{P}(ST))\,.$$

The dimensions are the same whether we view the operation of $SL_2(\mathbf{Z})$ on $\mathbf{P}_w(\mathbf{R})$, i.e. on the space of polynomials with real coefficients, or on $\mathbf{P}_w(\mathbf{C})$, the space of polynomials with complex coefficients. It is now easier to deal with the complex space and complex dimension. We have

$$ST=\begin{pmatrix}0 & 1\\ -1 & 1\end{pmatrix}.$$

The operator $\mathbf{P}(ST)$ has two eigenspaces on the homogeneous polynomials of degree 1, namely

$$\mathbf{C}(X-\rho Y),\ \text{the } \rho\text{-eigenspace}$$
$$\mathbf{C}(X-\bar\rho Y),\ \text{the } \bar\rho\text{-eigenspace}\,.$$

We have a direct sum decomposition

$$\mathbf{P}_w(\mathbf{C}) = \bigoplus \mathbf{C}(X-\rho Y)^j(X-\bar\rho Y)^{w-j}$$

into eigenspaces for $\mathbf{P}(ST)$, with eigenvalues $\rho^j\bar\rho^{w-j}$ respectively. The kernel of $I-\mathbf{P}(ST)$ consists of the sum of those eigenspaces for which

$$\rho^j\bar\rho^{w-j}=1\,,$$

and its dimension is equal to the number of such j with $0\leqslant j\leqslant w$. We have $\bar\rho=\rho^5$. So the desired indices j are exactly those such that

$$j+5(w-j)\equiv 0 \pmod 6$$

or equivalently,

$$j\equiv 2w \pmod 3\,.$$

This is trivially seen to give the values stated in the theorem, and concludes the proof of Theorem 3.1.

§ 5. The Map into Cohomology

For our purposes, it was more convenient to deal directly with the period mapping. However, it is worth while to show how we get the Eichler map into cohomology, and how it relates to the period mapping explicitly.

Let $\Gamma = SL_2(\mathbf{Z})$. Let

$$\pi : \Gamma \to GL_n(\mathbf{R})$$

be a representation. A cocycle

$$\varphi : \Gamma \to \mathbf{R}^n$$

is called **cuspidal** if for each $\tau \in \Gamma_\infty$ there exists a vector $v_\tau \in \mathbf{R}^n$ such that

$$\varphi(\tau) = (I - \pi(\tau))v_\tau .$$

The group of cuspidal cocycles is denoted by $Z^1(\pi)$. It contains the subgroup of coboundaries $B^1(\pi)$, i.e. maps φ such that there exists $v \in \mathbf{R}^n$ for which

$$\varphi(\sigma) = (I - \pi(\sigma))v , \qquad \text{all} \quad \sigma \in \Gamma .$$

The factor group

$$H^1(\pi) = Z^1(\pi)/B^1(\pi)$$

is called the **Eichler cohomology group**.

If $\tau_0 \in \Gamma_\infty$, we define:

$$Z(\pi, \tau_0) = \{\varphi \in Z^1(\pi),\ \varphi(\tau_0) = 0\}$$
$$B(\pi, \tau_0) = \{\varphi \in B^1(\pi),\ \varphi(\tau_0) = 0\}$$
$$H^1(\pi, \tau_0) = Z(\pi, \tau_0)/B(\pi, \tau_0) .$$

Lemma 1. *The natural map $H^1(\pi, \tau_0) \to H^1(\pi)$ is an isomorphism.*

Proof. Obviously injective, the map is seen to be surjective by making a translation of cocycles by

$$(I - \pi(\sigma))v_{\tau_0} .$$

In practice, we would of course select $\tau_0 = T$.

We now let $\pi = \pi_w$ be the representation defined for an *even* integer w, as in § 1. We let $V = V_w$, $U = U_w$ be the spaces defined in § 3. It is immediate from the definition of V_w that the map

$$Z(\pi, T) \to \mathbf{R}^{w+1}$$

given by

$$\varphi \mapsto \varphi(S)$$

has its image contained in V. If $\varphi(S) = 0$, then $\varphi = 0$ because φ is a cocycle. Hence the map is injective. It is also clear from the definitions that the map sends

$$B(\pi, T) \to U .$$

Lemma 2. *The map $\varphi \mapsto \varphi(S)$ induces an injection of $Z(\pi, T)$ into V, and a bijection of $B(\pi, T)$ with U, whence an injection*

$$H^1(\pi) \approx H^1(\pi, T) \to V/U .$$

Proof. Suppose that $\varphi(S) = u$ lies in U. By definition there exists a vector v such that

$$u = (I - \pi(S))v \qquad \text{and} \qquad (I - \pi(T))v = 0 .$$

Then the maps

$$\sigma \mapsto \varphi(\sigma) \qquad \text{and} \qquad \sigma \mapsto (I - \pi(\sigma))v$$

define cocycles which agree on T, S, whence are equal. Therefore φ is a coboundary. The rest of the lemma has already been proved.

On the other hand, we have a homomorphism

$$M_k^0(\mathbf{C}) \to H^1(\pi)$$

induced by

$$f \mapsto \varphi_f ,$$

where φ_f is the function

$$\varphi_f(\sigma) = \int_0^{\sigma(0)} \rho(f) ,$$

which is obviously a cocycle.

Lemma 3. *The cocycle is cuspidal, and the diagram*

$$\begin{array}{ccc} M_k^0(\mathbf{C}) & \longrightarrow & H^1(\pi) \\ & \searrow \quad \swarrow & \\ & V/U & \end{array}$$

is commutative, with sign -1.

Proof. We have

$$\varphi_f(S) = \int_0^{i\infty} \rho(f) = r(f) \qquad \text{and} \qquad \varphi_f(T) = \int_0^1 \rho(f)\,.$$

But

$$\int_0^1 \rho(f) = \int_0^{T(0)} \rho(f) = (I - \pi(T)) \int_0^{i\infty} \rho(f) = (I - \pi(T))r(f)\,.$$

Considering the equivalent cocycle

$$\varphi_f(\sigma) - (I - \pi(\sigma))r(f)$$

instead of φ_f proves the (-1)-commutativity, and the cuspidality.

Theorem 5.1. *All the arrows in the diagram*

$$\begin{array}{ccc} M_k^0(\mathbf{C}) & \longrightarrow & H^1(\pi) \\ & \searrow \quad \swarrow & \\ & V/U & \end{array}$$

are isomorphisms.

Proof. This is immediate, from the injectivity of

$$H^1(\pi) \to V/U\,,$$

and the isomorphism $M_k^0(\mathbf{C}) \to V/U$ proved in the preceding section.

Part III
Modular Forms for Congruence Subgroups

Chapter VII. Higher Levels

For the most part we have considered modular forms on $SL_2(\mathbf{Z})$. Incidentally in dealing with Hecke operators on such forms, we needed to pass to congruence subgroups. We want to return more systematically to modular forms on such subgroups. As already mentioned, there are three important such subgroups, which we called $\Gamma_0(N)$, $\Gamma_1(N)$ and $\Gamma(N)$. We treat $\Gamma_1(N)$ in some detail as an example. By conjugation, one can reduce the theory on $\Gamma(N)$ to that of $\Gamma_1(N)$.

The theory of Hecke operators is given an exposition in Shimura [Sh 2], in the style of double cosets, and total generality, extending to higher dimensional situations.

In addition to using Atkin–Lehner [A–L] and [Li], the exposition of this chapter and the next also benefited from a course of Katz.

§ 1. The Modular Set and Modular Forms

For level 1 we defined Hecke operators via lattices and functions of lattices, homogeneous of degree $-k$. For higher level, we have to consider something a little more general than a lattice. We consider pairs (t, L) where L is a lattice, and t is a (complex) point of (exact) period N with respect to L.

In the terminology of elliptic curves, the lattice determines a complex torus $\mathbf{C}/L$, and $t \pmod{L}$ is then a point of period N on $\mathbf{C}/L$. Parametrizing by the Weierstrass functions gives the elliptic curve, and a point of period N on the curve. We shall make further remarks about this later. The set of all such pairs is called the **modular set** for $\Gamma_1(N)$. A pair (t, L) is called a **modular point**.

Let k be an integer. We denote by $\mathscr{F}_1(N, k)$ the vector space of functions F of pairs (t, L), where L is a lattice and t a point of period N with respect to L, satisfying the conditions:

F is homogeneous of degree $-k$, i.e.

$$F(\lambda t, \lambda L) = \lambda^{-k} F(t, L), \qquad \lambda \in \mathbf{C}^*,$$

and F depends only on the class of $t \bmod L$.

We shall see in a moment how to identify elements of $\mathscr{F}_1(N, k)$ with functions on the upper half plane, and will then make an additional requirement of meromorphy, in which case we call elements of $\mathscr{F}_1(N, k)$ **meromorphic modular forms**.

We recall that $\Gamma_1(N)$ is the subgroup of elements in $SL_2(\mathbf{Z})$ satisfying

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} (\text{mod } N) .$$

Thus

$$a \equiv d \equiv 1 , \qquad c \equiv 0, \text{ and } b \text{ is arbitrary (mod } N) .$$

We consider the vector space of functions

$$F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

of pairs (ω_1, ω_2) of complex numbers such that ω_1/ω_2 is in $\mathfrak{H}$, and satisfying:

F_1 is homogeneous of degree $-k$, that is

$$F_1\begin{pmatrix} \lambda\omega_1 \\ \lambda\omega_2 \end{pmatrix} = \lambda^{-k} F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

and F_1 is invariant under $\Gamma_1(N)$, i.e.

$$F_1\left(\gamma\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}\right) = F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}, \qquad \text{all } \gamma \in \Gamma_1(N) .$$

This vector space is isomorphic to $\mathscr{F}_1(N, k)$. The correspondence is obtained as follows. Given F we define F_1 by

$$F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = F\left(\frac{\omega_2}{N}, [\omega_1, \omega_2]\right),$$

where $[\omega_1, \omega_2]$ is the lattice generated by ω_1 and ω_2. Inversely, given F_1, and a pair (t, L), we select any element $\omega_2 \in L$ such that

$$t = \omega_2/N ,$$

and then any element ω_1 such that $\omega_1/\omega_2 \in \mathfrak{H}$, and such that ω_1, ω_2 form a basis of L. We define

$$F(t, L) = F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}.$$

Using the invariance of F_1 under $\Gamma_1(N)$ shows immediately that the value of F is well defined, i.e. independent of the choices made, and one also sees at once that the two maps are inverse to each other.

Instead of saying that F (or F_1) have degree $-k$ above, we also say that they are of **weight** k.

As the group $\Gamma_1(N)$ is fixed throughout this discussion, we omit references to it, and also to k, thus speaking only of modular forms.

The above vector spaces are isomorphic to a third space:

Functions f of a variable $\tau \in \mathfrak{H}$ such that

$$f \circ [\gamma]_k = f$$

for all $\gamma \in \Gamma_1(N)$.

The correspondence between f and F_1 is given by

$$\omega_2^k F_1\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = f(\omega_1/\omega_2) .$$

The first two descriptions of modular forms will be called **homogeneous**, and the third will be called **inhomogeneous**.

Under the preceding correspondence, a modular form will be viewed as a function of $\tau \in \mathfrak{H}$. *We now make the additional requirement that for all $\gamma \in SL_2(\mathbf{Z})$, the function*

$$f(\gamma\tau)$$

is meromorphic at infinity, in the local parameter

$$q_N = e^{-2\pi i\tau/N} .$$

Since $f(\tau + 1) = f(\tau)$, it follows for f itself that it has a power series expansion

$$f_\infty(q) = \sum a_n q^n ,$$

with only a finite number of negative terms, and without any fractional power of q.

We shall now define the first of various operations on modular forms by defining operations on the modular set, and then on modular forms by composition.

Let a be an integer prime to N. We let $[a]$ be the operation on the modular set given by

$$[a] : (t, L) \mapsto (at, L) .$$

This defines an operation of $(\mathbf{Z}/N\mathbf{Z})^*$ on the modular set, whence on $\mathscr{F}_1(N, k)$ by letting

$$([a]_k F)(t, L) = F(at, L) .$$

We sometimes omit the subscript k for the operation on the modular forms if the context is clear.

Let ε be a Dirichlet character mod N, i.e. a homomorphism

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

extended to $\mathbf{Z}/N\mathbf{Z}$ by putting $\varepsilon(m)=0$ if $(m, N)\neq 1$. We let

$$\mathscr{F}_1(N, k, \varepsilon)$$

be the subspace of $\mathscr{F}_1(N, k)$ consisting of those modular forms F such that for all a prime to N we have

$$F(at, L)=\varepsilon(a)F(t, L)\ .$$

Then $\mathscr{F}_1(N, k, \varepsilon)$ is the ε-eigenspace of this operation. In particular, we have a direct sum decomposition

$$\mathscr{F}_1(N, k)=\bigoplus_{\varepsilon}\mathscr{F}_1(N, k, \varepsilon)\ .$$

Theorem 1.1. *Let a be an integer prime to N. Let σ_a be a matrix in $SL_2(\mathbf{Z})$ such that*

$$\sigma_a\equiv\begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} \pmod N\ .$$

Then for a modular form $f\in\mathscr{F}_1(N, k)$, we have

$$[a]_k f\ =\ f\circ[\sigma_a]_k\ .$$

In other words, $[a]_k=[\sigma_a]_k$.

Proof. Let

$$\sigma_a=\begin{pmatrix} a_1 & b \\ c & d \end{pmatrix}$$

with $a_1\equiv a^{-1}$ and $d\equiv a \pmod N$. We have:

$$[a]_k f(\tau) = [a]_k F\left(\frac{1}{N}, [\tau, 1]\right) = F\left(\frac{a}{N}, [\tau, 1]\right) = F\left(\frac{c\tau+d}{N}, [a_1\tau+b, c\tau+d]\right)$$
$$= F\left((c\tau+d)\left(\frac{1}{N}, [\sigma_a\tau, 1]\right)\right) = f(\sigma_a\tau)(c\tau+d)^{-k}\ .$$

This proves the theorem.

Since the map $SL_2(\mathbf{Z})\to SL_2(\mathbf{Z}/N\mathbf{Z})$ is surjective, there always exists a matrix σ_a in $SL_2(\mathbf{Z})$ as prescribed in the theorem.

If $f\in\mathscr{F}_1(N, k, \varepsilon)$ then ε is called the **character** of f. Since later we deal with characters of the Hecke algebra, we sometimes need to distinguish characters of $(\mathbf{Z}/N\mathbf{Z})^*$ from characters of this algebra, and hence we shall specify that ε is the **Dirichlet character** of f. Note that when $\varepsilon\neq 1$, Hecke called ε the **nebentypus** of f.

From Theorem 1.1 we see that a modular form $f\in\mathscr{F}_1(N, k)$ has Dirichlet character ε if and only if it satisfies the formula

$$\boxed{f\circ[\gamma]_k=\varepsilon(d)f}$$

for all $\gamma=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\Gamma_0(N)$.

§ 2. Hecke Operators

We let $\mathscr{L}_1(N)$ be the free vector space over $\mathbf{Q}$ generated by the pairs (t, L). We define an endomorphism

$$T(n):\mathscr{L}_1(N)\to\mathscr{L}_1(N)$$

for each positive integer n, by the formula:

$$nT(n):(t, L)\mapsto\sum_{\substack{(L':L)=n\\(t:L')=N}}(t, L')\ .$$

The sum is taken over the lattices L' containing L such that L has index n in L', and such that t still has period N with respect to L'. This second condition was abbreviated by writing $(t:L')=N$.

Let us put $A=\mathbf{C}/L$. The subgroup of torsion points of A is the group $\mathbf{Q}L/L$. We may think of t as generating a cyclic subgroup of order N in $A_{\text{tor}}=\mathbf{Q}L/L$. Let $C=(t)$ be this cyclic subgroup. Then a lattice L' determines a subgroup

$$H=L'/L$$

of order n such that H and C are independent, i.e. $H\cap C=\{0\}$.

It will also be convenient notationally to write the pair (t, L) as

$$(t, A)\ ,$$

viewing t as a point of order N on A.

For an integer n *prime to* N, we also define an operator $T(n, n)$ by the formula

$$n^2T(n, n):(t, L)\mapsto(t, n^{-1}L)\ .$$

The lattice $n^{-1}L$ is simply the set of all elements $n^{-1}\omega$, with $\omega\in L$.

It is immediate that the operators $T(n)$ and $T(m, m)$ commute.

Theorem 2.1. (i) *If* $(n_1, n_2) = 1$ *then* $T(n_1 n_2) = T(n_1)T(n_2)$.

(ii) *If* p *is prime,* $p|N$, *then* $p^n T(p^n) = (pT(p))^n$.

(iii) *For* $n \geq 2$ *and* $(p, N) = 1$,

$$T(p^{n-1})T(p) = T(p^n) + pT(p^{n-2})T(p, p)\,.$$

Proof. The first part is trivial. For the second part, suppose that $p|N$. An abelian p-group is cyclic if and only if it does not contain a subgroup of type (p, p). If H is a subgroup of A of order p^n, independent of $C = (t)$ (notation as above), then H is necessarily cyclic. Let H_i be the subgroups of A of order p, independent of C. Let H_{ij} be the subgroups of order p^n containing H_i and independent of C. If $i \neq i'$ then

$$H_{ij} \neq H_{i'j'}\,.$$

Furthermore, every cyclic subgroup H of order p^n independent of C is equal to some H_{ij}. This proves the formula

$$p^n T(p^n) = p^{n-1}T(p^{n-1})pT(p)\,,$$

whence (ii) follows by induction.

As for (iii), the standard proof of Chapter II, § 1 for the analogous fact works also in this case, because in effect, in the situation of (iii), the point t is irrelevant.

From Theorem 2.1 we see that all the operators $T(n)$ and $T(m, m)$ commute with each other, and generate a commutative subalgebra of End $\mathscr{L}_1(N)$. The relations of Theorem 2.1 can then be expressed in terms of power series relations with coefficients in that algebra, namely:

(ii)
$$\sum_{n\geq 0} T(p^n)X^n = \frac{1}{1 - T(p)X} \qquad \text{if } p|N$$

and

(iii)
$$\sum_{n\geq 0} T(p^n)X^n = \frac{1}{1 - T(p)X + pT(p, p)X^2} \qquad \text{if } p\nmid N.$$

These are merely reformulations of (ii) and (iii). If we use the multiplicativity property of (i), and put $X = p^{-s}$, we may then form the Euler products, and get the relation

(iv)
$$\sum_{n\geq 1} T(n)n^{-s} = \prod_{p|N} \frac{1}{1 - T(p)p^{-s}} \prod_{p\nmid N} \frac{1}{1 - T(p)p^{-s} + T(p, p)p^{1-2s}}\,.$$

The Hecke operators $T(n)$ and $T(n, n)$ can then be made to operate on $\mathscr{F}_1(N, k)$ by putting

$$nT_k(n)F(t, L) = \sum_{\substack{(L' : L) = n \\ (t : L') = N}} F(t, L')\,,$$

and similarly for $n^2 T_k(n, n)$. As the operators $T_k(n)$ and $T_k(n, n)$ obviously commute with the operation of $(\mathbf{Z}/N\mathbf{Z})^*$, it follows that they map $\mathscr{F}_1(N, k, \varepsilon)$ into itself, thus giving rise to operators

$$T_{k,\varepsilon}(n) \quad \text{and} \quad T_{k,\varepsilon}(n, n)$$

Let $\mathscr{H}$ be the subalgebra of End $\mathscr{L}_1(N)$ generated by the operators $T(n)$, $T(n, n)$, and $[a]$. Then there is a homomorphism

$$\mathscr{H} \to \text{End } \mathscr{F}_1(N, k, \varepsilon)$$

which sends

$$T(n) \mapsto T_{k,\varepsilon}(n), \quad T(n, n) \mapsto T_{k,\varepsilon}(n, n), \quad \text{and} \quad [a] \mapsto [a]_k\,.$$

The image of $\mathscr{H}$ under this homomorphism is denoted by

$$\mathscr{H}_1(N, k, \varepsilon)\,.$$

Similarly the image of $\mathscr{H}$ in End $\mathscr{F}_1(N, k)$ is denoted by $\mathscr{H}_1(N, k)$. These algebras $\mathscr{H}_1(N, k)$ and $\mathscr{H}_1(N, k, \varepsilon)$ are called the **Hecke algebras**.

From the definitions, we find at once that for $F \in \mathscr{F}_1(N, k, \varepsilon)$,

$$pT_{k,\varepsilon}(p, p)F = p^{k-1}\varepsilon(p)F\,.$$

Again in this formula we assume $p\nmid N$. Indeed,

$$p^2 T_{k,\varepsilon}(p, p)F(t, L) = F(t, p^{-1}L) = F(p^{-1}(pt, L)) = p^k \varepsilon(p)F(t, L)\,.$$

From this and the homomorphism from $\mathscr{H}$ into $\mathscr{H}_1(N, k, \varepsilon)$ we get the corresponding Euler product relations for the Hecke operators on forms of weight k, namely:

($\text{ii}_{k,\varepsilon}$)
$$\sum_{n\geq 0} T_{k,\varepsilon}(p^n)X^n = \frac{1}{1 - T_{k,\varepsilon}(p)X} \qquad \text{if } p|N$$

($\text{iii}_{k,\varepsilon}$)
$$\sum_{n\geq 0} T_{k,\varepsilon}(p^n)X^n = \frac{1}{1 - T_{k,\varepsilon}(p)X + p^{k-1}\varepsilon(p)X^2} \qquad \text{if } p\nmid N.$$

Remark. Recalling the convention that $\varepsilon(p) = 0$ if $p|N$, we note that this last formula also applies to the case $p|N$.

($\text{iv}_{k,\varepsilon}$)
$$\sum_{n\geq 1} T_{k,\varepsilon}(n)n^{-s} = \prod_{p|N} \frac{1}{1 - T_{k,\varepsilon}(p)p^{-s}} \prod_{p\nmid N} \frac{1}{1 - T_{k,\varepsilon}(p)p^{-s} + \varepsilon(p)p^{k-1-2s}}\,.$$

§ 3. Hecke Operators on q-expansions

We define two operators on the power series field in q. For each positive integer d, we let

$$V_d(\sum a_n q^n) = \sum a_n q^{dn}$$

and

$$U_d(\sum a_n q^n) = \sum_{d|n} a_n q^{n/d} .$$

Then

$U_d \circ V_d = \mathrm{id}$
$V_d \circ U_d =$ projection on the part of the power series with indices divisible by d.

These operators can also be expressed in terms of the variable $\tau \in \mathfrak{H}$. If now $f = \sum a_n q^n$ and $q = e^{2\pi i \tau}$, then

$$V_d f(\tau) = f(d\tau)$$

$$U_d f(\tau) = \frac{1}{d} \sum_{b \bmod d} f\left(\frac{\tau+b}{d}\right).$$

The first expression is obvious. The second follows from the orthogonality relation for roots of unity, namely

$$\sum_{b \bmod d} f\left(\frac{\tau+b}{d}\right) = \sum_n \sum_{b \bmod d} e^{2\pi i n\tau/d}\, e^{2\pi i n b/d} = d \sum_{d|n} a_n q^{n/d} ,$$

as desired.

Theorem 3.1. *On modular forms viewed as power series in q, the effect of $T_{k,\varepsilon}(p)$ is given as*

$$T_{k,\varepsilon}(p) = U_p + \varepsilon(p) p^{k-1} V_p .$$

Equivalently, if

$$T_{k,\varepsilon}(p) f_\infty(q) = \sum b_n q^n ,$$

then

$$b_n = a_{pn} + \varepsilon(p) p^{k-1} a_{n/p} \qquad \text{if } p \nmid N,\ p \mid n$$
$$= a_{pn} \qquad \text{if } p \mid N, \text{ or } p \nmid n.$$

Proof. The lattices L' such that $(L' : L) = p$ are of the form

$$\left[\frac{\tau+b}{p}, 1\right], b = 0, \ldots, p-1 \quad \text{and} \quad [\tau, 1/p]$$

if $L = [\tau, 1]$. If $p \mid N$, the only lattices L' such that $1/N$ has still period N with respect to L' are the first ones. Hence

$$T(p) f(\tau) = \frac{1}{p} \sum_{b=0}^{p-1} F\left(1/N, \left[\frac{\tau+b}{p}, 1\right]\right) = \frac{1}{p} \sum_{b=0}^{p-1} \sum_n e^{2\pi i n\tau/p}\, e^{2\pi i n b/p} = \sum_{p|n} a_n q^{n/p} .$$

If $p \mid N$ then $\varepsilon(p) = 0$. Hence the formula of the theorem is proved in this case.

Suppose next that $p \nmid N$. Then $T(p) f(\tau)$ has the same terms as before, plus one additional term, namely $1/p$ times:

$$F\left(\frac{1}{N}, \left[\tau, \frac{1}{p}\right]\right) = F\left(p^{-1}\left(\frac{p}{N}, [p\tau, 1]\right)\right) = p^k \varepsilon(p) F\left(\frac{1}{N}, [p\tau, 1]\right) = p^k \varepsilon(p) \sum a_n q^{np} .$$

This gives precisely the second term appearing in the expression of the theorem, as was to be shown.

If we substitute the expressions found in the preceding theorem for $T_{k,\varepsilon}(p)$ into the power series and formal Dirichlet series found in the preceding section, we obtain first the factorization

$$1 - T_{k,\varepsilon}(p) X + \varepsilon(p) p^{k-1} X^2 = (1 - U_p X)(1 - \varepsilon(p) p^{k-1} V_p X) .$$

The coefficients of powers of X lie in the endomorphism algebra of the power series in q. Hence we find:

Theorem 3.2. *We have the identities*

$$\sum_{n \geq 1} T_{k,\varepsilon}(n) n^{-s} = \left(\sum_n \varepsilon(n) n^{k-1} V_n n^{-s}\right)\left(\sum_n U_n n^{-s}\right)$$

and

$$T_{k,\varepsilon}(n) = \sum_{d|n} \varepsilon(d) d^{k-1}\, V_d \circ U_{n/d} .$$

Proof. This comes from

$$T_{k,\varepsilon}(p^n) X^n = (1 - T_{k,\varepsilon}(p) X + \varepsilon(p) p^{k-1} X^2)^{-1}$$
$$= (1 - \varepsilon(p) p^{k-1} V_p X)^{-1} (1 - U_p X)^{-1} .$$

We then put $X = p^{-s}$, and take the product over all p. The operators U_p, V_p commute with U_l, V_l for $p \neq l$, and hence

$$\sum T_{k,\varepsilon}(n) n^{-s} = \prod_p (1 - \varepsilon(p) p^{k-1} V_p p^{-s})^{-1} \prod_p (1 - U_p p^{-s})$$
$$= \left(\sum \varepsilon(n) n^{k-1} V_n n^{-s}\right)\left(\sum U_n n^{-s}\right)$$

which proves the theorem.

In terms of the coefficients of the q-expansion, if

$$f = \sum a_n q^n$$

and

$$T_{k,\varepsilon}(m) f = \sum b_n q^n$$

then the expression of the theorem shows that

$$\boxed{b_n = \sum_{d \mid (m,n)} \varepsilon(d) d^{k-1} a_{mn/d^2}.}$$

Theorem 3.3. *Let $f \in \mathscr{F}_1(N, k, \varepsilon)$. Assume that f is a non-zero eigenfunction of the Hecke algebra generated by all operators $T_{k,\varepsilon}(m)$ for all positive integers m. Let*

$$T_{k,\varepsilon}(n) f = \lambda_n f.$$

Then:

(i) *f is holomorphic at infinity.*

(ii) *We have $a_n = \lambda_n a_1$.*

(iii) *If $k \neq 0$ then $a_1 \neq 0$.*

(iv) *If $a_0 \neq 0$ then*

$$\lambda_n = \sum_{d \mid n} \varepsilon(d) d^{k-1}.$$

Proof. For a prime $p \nmid N$, and

$$f = \frac{a_{-r}}{q^r} + \cdots$$

we find that

$$T_{k,\varepsilon}(p) f = \varepsilon(p) p^{k-1} f(q^p) + U_p f(q),$$

and

$$U_p f(q) = \begin{cases} \dfrac{a_{-r/p}}{q^{r/p}} + \cdots & \text{if } p \mid r \\ \text{terms of higher order} & \text{if } p \nmid r. \end{cases}$$

Hence $T_{k,\varepsilon}(p) f$ has a pole of higher order than f if f has a pole, and $T_{k,\varepsilon}(p) f = \lambda_p f$. These are contradictory, so f is holomorphic at infinity.

The formula $a_n = \lambda_n a_1$ follows at once from Theorem 3.2. If $k \neq 0$, and $a_1 = 0$ then f is constant, which is impossible. Finally, if $a_0 \neq 0$, then again from Theorem 3.2 we get

$$\lambda_n a_0 = \sum_{d \mid n} \varepsilon(d) d^{k-1} a_0,$$

thereby proving the theorem.

§ 4. The Matrix Operations

Let $\mathbf{M}_1^n(N)$ be the set of matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$$

having determinant n, and such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} (\mathrm{mod}\ N).$$

Then $\mathbf{M}_1^n(N)$ is stable under multiplication on the left and right by $\Gamma_1(N)$.

Let $\alpha(a, b)$ be the matrix

$$\alpha(a, b) = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

with $a, d > 0$, $ad = n$, $(a, N) = 1$, $b = 0, \ldots, d-1$. For each $a \mid n$ and $(a, N) = 1$ choose

$$\sigma_a \in SL_2(\mathbf{Z}), \qquad \sigma_a \equiv \begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} (\mathrm{mod}\ N).$$

Lemma 1. *The matrices $\sigma_a \alpha(a, b)$ with a, b, d satisfying the above conditions form distinct coset representatives,*

$$\mathbf{M}_1^n(N) = \bigcup_{(a,b)} \Gamma_1(N) \sigma_a \alpha(a, b).$$

Proof. Any matrix σ of determinant n can be written in the form

$$\sigma = \gamma \alpha(a, b)$$

with some $\gamma \in SL_2(\mathbf{Z})$. If $\sigma \in \mathbf{M}_1^n(N)$ then we write

$$\sigma = \gamma \sigma_a^{-1} \sigma_a \alpha(a, b).$$

Then $\sigma_a \alpha(a, b)$ lies in $\mathbf{M}_1^n(N)$, and it is immediately verified that $\gamma \sigma_a^{-1}$ is in $\Gamma_1(N)$. Finally, it is also immediately verified that the elements $\sigma_a \alpha(a, b)$ represent distinct cosets, as was to be shown.

Theorem 4.1. *Let $f \in \mathscr{F}_1(N, k)$. Then*

$$T_k(n) f = n^{k/2-1} \sum_{\substack{ad = n,\ a > 0 \\ (a,N) = 1 \\ b \bmod d}} f \circ [\sigma_a \alpha(a, b)]_k.$$

Proof. We had already seen that

$$V_d f(\tau) = f(d\tau)$$

$$U_d f(\tau) = \frac{1}{d} \sum_{b \bmod d} f\left(\frac{\tau+b}{d}\right).$$

By Theorem 3.2 we obtain, for $f \in \mathscr{F}_1(k, N, \varepsilon)$:

$$\begin{aligned} T_{k,\varepsilon}(n) f(\tau) &= \sum_{d|n} \varepsilon(d) d^{k-1} (V_d \circ U_{n/d} f)(\tau) \\ &= \sum_{d|n} \varepsilon(d) d^{k-1}\, U_{n/d} f(d\tau) \\ &= \sum_{d|n} \varepsilon(d) \frac{d}{n} \sum_{i=0}^{n/d-1} f\left(\frac{d\tau+i}{n/d}\right) \\ &= \sum_{ad=n} \sum_{i \bmod d} \varepsilon(a) a^{k-1} \frac{1}{d} f\left(\frac{a\tau+i}{d}\right) \\ &= n^{k-1} \sum_{\substack{ad=n \\ (a,N)=1}} \varepsilon(a) \sum_{i \bmod d} d^{-k} f\left(\frac{a\tau+i}{d}\right) \\ &= n^{(k/2)-1} \sum_{\substack{ad=n \\ b \bmod d \\ (a,N)=1}} \varepsilon(a) f \circ [\alpha(a,b)]_k . \end{aligned}$$

Using $f \circ [\sigma_a]_k = \varepsilon(a) f$ concludes the proof.

§ 5. Petersson Product

Theorem 5.1. *Let $f, g \in \mathscr{F}_1(N, k, \varepsilon)$ be cusp forms, and let $(n, N) = 1$. Then*

$$\langle T_{k,\varepsilon}(n) f, g \rangle = \varepsilon(n) \langle f, T_{k,\varepsilon}(n) g \rangle .$$

Proof. Using Theorem 4.1, we find, with the same notation as in the preceding section for $\alpha(a, b)$:

$$\begin{aligned} \langle f \circ [\sigma_a \alpha(a,b)]_k, g \rangle &= \varepsilon(a) \left\langle f \circ \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}_k, g \right\rangle \\ &= \varepsilon(a) \left\langle f, g \circ \begin{bmatrix} d & b' \\ 0 & a \end{bmatrix} \right\rangle \end{aligned}$$

with an appropriate b'. We insert $\sigma_d^{-1} \sigma_d$ between g and the matrix, and use the fact that

$$g \circ [\sigma_d^{-1}]_k = \varepsilon(d)^{-1} g .$$

We can then extract $\varepsilon(d)^{-1}$ in front of the scalar product, with a complex conjugate, which combined with $\varepsilon(a)$ yields $\varepsilon(n)$. Summing over the distinct cosets proves the theorem.

Corollary. *The Hecke algebra $\mathscr{H}_1(N, k, \varepsilon)$ is stable under taking adjoints with respect to the Petersson scalar product, in other words, it is star closed.*

Proof. Clear from the theorem.

For abbreviation of notation the space of cusp forms is also denoted $\mathbf{S}_1(N, k)$, and the eigenspace having character ε is denoted by

$$\mathbf{S}_1(N, k, \varepsilon) .$$

Theorem 5.2. (i) *Let $(a, N) = 1$. The adjoint of $[a]_k$ is $[a^{-1}]_k$.*

(ii) *If f, g are cusp forms with distinct characters, then they are orthogonal with respect to the Petersson scalar product.*

Proof. The first statement is immediate from Theorem 1.1, that $[a]_k = [\sigma_a]_k$, and the fact that the operation of rational matrices is unitary on the Petersson product, Chapter III, Theorem 4.1. The second statement is standard from the first. We select a prime to N such that if ε and ε' are the characters of f and g respectively, then $\varepsilon'(a) \neq \varepsilon(a)$. We then compute

$$\bar{\varepsilon}'(a) \varepsilon(a) \langle f, g \rangle$$

putting $\varepsilon(a)$ near f, $\varepsilon'(a)$ near g, to see that this expression is equal to

$$\langle f \circ [\sigma_a]_k, g \circ [\sigma_a]_k \rangle = \langle f, g \rangle ,$$

a contradiction unless $\langle f, g \rangle = 0$, as was to be shown.

For each positive integer D we let $\mathscr{H}_1^{(D)}(N, k)$ be the algebra generated by the Hecke operators

$$T_k(n),\ T_k(n, n) \text{ with } n \text{ prime to } D ,$$

and by the operators $[a]_k$, with $a \in (\mathbf{Z}/N\mathbf{Z})^*$. We may view $T_{k,\varepsilon}(n)$ with $(n, N) = 1$ as being equal to 0 on the eigenspaces $\mathbf{S}_1(N, k, \varepsilon')$ for $\varepsilon' \neq \varepsilon$. Indeed, the projection operator on $\mathbf{S}_1(N, k, \varepsilon)$ is the group ring element

$$\mathrm{pr}_\varepsilon = \frac{1}{\phi(N)} \sum_a \bar{\varepsilon}(a) [a]_k .$$

Thus

$$T_{k,\varepsilon}(n) = T_k(n) \circ \mathrm{pr}_\varepsilon .$$

We have seen in Theorem 5.1 that we get the adjoint relation

$$T_{k,\varepsilon}(n)^* = \bar{\varepsilon}(n) T_{k,\varepsilon}(n) .$$

It follows that the adjoint $T_k(n)^*$ is a linear combination of operators $T_k(n) \circ [a]_k$. Thus we have shown:

Theorem 5.3. *The algebra $\mathscr{H}_1^{(DN)}(N, k)$ is star-closed, and so is $\mathscr{H}_1^{(DN)}(N, k, \varepsilon)$.*

§ 6. The Involution

On each quotient $\mathbf{C}/L$ one can construct (essentially by Kummer theory) a skew-symmetric non-degenerate pairing into the roots of unity. For an analytic description, cf. [L 2], Chapter 8. We give here the axioms needed for the sequel, without giving the construction again. In fact, it is convenient to deal with a slightly more general context.

Let $L \subset L'$ be two lattices. Then they give rise to a homomorphism, called an isogeny,

$$\lambda : \mathbf{C}/L \to \mathbf{C}/L' .$$

Writing $A = \mathbf{C}/L$ and $B = \mathbf{C}/L'$, we can write the isogeny

$$\lambda : A \to B .$$

One can define the transpose $\lambda' : B \to A$ to be

$$\lambda'(b) = \sum a_i$$

where the points a_i are the points of $\lambda^{-1}(b)$, and the sum is taken on A. Then there exists a pairing

$$e_\lambda : (\operatorname{Ker} \lambda) \times (\operatorname{Ker} \lambda') \to \mu_m$$

into the roots of unity of order m, where m is the exponent of $\operatorname{Ker} \lambda$, satisfying the following properties.

E 1. *The pairing is skew-symmetric, in the sense that*

$$e_\lambda(a, b) = e_{\lambda'}(b, a)^{-1} .$$

E 2. *If $\varphi : B \to C$ is another isogeny, and $a \in \operatorname{Ker} \varphi\lambda$, $b \in \operatorname{Ker} \lambda'$, then*

$$e_{\varphi\lambda}(a, b) = e_\varphi(\lambda a, b) .$$

E 3. *Let $A = \mathbf{C}/[\tau, 1]$, $B = A/(1/N)$, $\lambda : A \to A/(1/N) = B$ the canonical map. Then*

$$\lambda' : B \to A/[\tau/N, 1/N] .$$

In this case, we view τ/N as an element of $\operatorname{Ker} \lambda'$, *and the condition is that*

$$e_\lambda\left(\frac{1}{N}, \frac{\tau}{N}\right) = e^{2\pi i/N} .$$

Suppose for instance that $\lambda : A \to A$ is multiplication by N. Then $\operatorname{Ker} \lambda = A_N$ is the group of points of order N, and $\operatorname{Ker} \lambda' = A_N$ also. Then by definition,

$$e_N(t, u) = e_\lambda(t, u) .$$

Another example which occurs frequently is when we are given a point t of order N in A, and

$$\lambda : A \to A/(t)$$

is the canonical map. The pairing e_λ will sometimes be denoted by e_t in this case.

Relative to such a pairing, we are going to define a map which is essentially an involution,

$$W_N : \mathscr{L}_1(N) \to \mathscr{L}_1(N) .$$

To each pair $(t, A) = (t, \mathbf{C}/L)$, we consider the canonical map

$$\lambda : A \to A/(t) = \mathbf{C}/L_t ,$$

where $L_t = L + \mathbf{Z}t$ is the lattice generated by L and t, while (t) is the cyclic group generated by t in A. We let t' be the point in $\operatorname{Ker} \lambda'$ such that

$$e_\lambda(t, t') = \zeta_N , \qquad \text{where } \zeta_N = e^{2\pi i/N} .$$

We then define

$$W_N(t, A) = (t', A/(t)) = (t', L_t) .$$

We can then define an action on modular forms. If $F \in \mathscr{F}_1(N, k)$ then we define $W_{N,k}$ by

$$W_{N,k} F(t, L) = F(t', L_t) .$$

Lemma 1. *Let f be the function on the upper half plane corresponding to F. Let*

$$w_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} .$$

Then

$$W_{N,k} = N^{k/2}[w_N]_k .$$

In other words,

$$W_{N,k} f = N^{k/2} f \circ [w_N]_k .$$

Proof. By definition, we have

$$f(\tau) = F\left(\frac{1}{N}, [\tau, 1]\right)$$

$$W_{N,k} f(\tau) = F\left(\frac{\tau}{N}, \left[\tau, \frac{1}{N}\right]\right).$$

Note that $[\tau, 1/N] = [\tau, -1/N]$. Therefore

$$W_{N,k} f(\tau) = F\left(\tau\left(\frac{1}{N}, [1, -1/N\tau]\right)\right) = \tau^{-k} F\left(\frac{1}{N}, \left[\frac{-1}{N\tau}, 1\right]\right) = \tau^{-k} f(-1/N\tau)$$

On the other hand, since w_N has determinant N, we have:

$$f \circ [w_N]_k(\tau) = f(-1/N\tau)(N\tau)^{-k} N^{k/2} = f(-1/N\tau)\, \tau^{-k} N^{-k/2} .$$

This proves the lemma.

Lemma 2. *We have* $W_{N,k}^2 = (-N)^{k/2}$

Proof. The square of the matrix is

$$w_N^2 = \begin{pmatrix} -N & 0 \\ 0 & -N \end{pmatrix}.$$

The lemma is then immediate from Lemma 1 and the definition of the operation of a rational matrix $[\alpha]_k$.

Let a be an integer prime to N. We had defined in § 2 the action $[a]$ and $[a]_k$ on the modular set and modular forms respectively. We now see how these commute with the involutions.

Lemma 3. *Let* $(a, N) = 1$. *Then*

$$W_N \circ [a] = [a^{-1}] \circ W_N .$$

Proof. The operation of $W_N \circ [a]$ yields the following sequence of maps:

$$(t, L) \xmapsto{[a]} (at, L) \xmapsto{W_N} ((at)', L_t) ,$$

where $(at)'$ is such that

$$e_t(at, (at)') = \zeta_N .$$

Going around the other way yields the maps

$$(t, L) \xmapsto{W_N} (t', L_t) \xmapsto{[a^{-1}]} (a^{-1}t', L_t) .$$

But the linearity of the pairing yields

$$e_t(at, (at)') = e_t(t, (at)')^a .$$

We let a^{-1} be a positive integer which is the inverse of a mod N. Then we find that both expressions

$$e_t(t, a^{-1}t') \quad \text{and} \quad e_t(t, (at)')$$

are equal to $\zeta_N^{a^{-1}}$, and hence equal to each other, thereby proving the lemma.

Warning. When applying Lemma 3 to modular forms, one has to reverse the order of the composition of mappings, viewing the operators as operating on the left. Thus the formula on $\mathscr{F}_1(N, k)$ reads:

$$\boxed{[a]_k \circ W_{N,k} = W_{N,k} \circ [a^{-1}]_k .}$$

In this case, since we can replace a by a^{-1} (mod N), the order does not matter, but it will in later commutation rules.

Chapter VIII. Atkin–Lehner Theory

Atkin-Lehner [A-L] showed how to construct in a natural way a basis for the space of modular forms of given level which are eigenfunctions for the Hecke operators prime to that level, satisfying the same formalism as for level 1. They worked on $\Gamma_0(N)$. Miyake [Mi] extended this to the general case, including the modular forms in the sense of Langlands in the context of representation theory. See also Casselman [Ca]. More recently, Li [Li] reconsidered the matter in the style of Atkin-Lehner, following [A–L] very closely.

Since the study of $\Gamma(N)$ can be reduced to that of $\Gamma_1(N)$ by conjugation, we give an exposition of the Atkin-Lehner theory on $\Gamma_1(N)$.

§ 1. Changing Levels

Let d be a positive integer, $d>1$, and $d \mid N$. We want to define two maps from $\mathscr{F}_1(N/d, k)$ to $\mathscr{F}_1(N, k)$. The sum of the images of these maps will be called the **non-primitive subspace** ("old" subspace) of $\mathscr{F}_1(N, k)$. As usual, the maps are defined first on the modular set.

We define $\pi_1(d): \mathscr{L}_1(N) \to \mathscr{L}_1(N/d)$ by

$$\pi_1(d): (t, L) \longmapsto (dt, L) .$$

Then on modular forms we have

$$\pi_1(d)_k F(t, L) = F(dt, L) .$$

Since $d/N = 1/(N/d)$, it follows immediately from the definitions that on the corresponding function $f(\tau)$, the operator $\pi_1(d)_k$ acts like the identity mapping. In other words, any function on the upper half plane invariant under $\Gamma_1(N/d)$ is also invariant under $\Gamma_1(N)$, and thus our operator $\pi_1(d)_k$ corresponds to this natural injection. In particular, on the q-expansion, the map is also the identity. We may write this in the form

$$\pi_1(d)_\infty = \mathrm{id} .$$

We have a second map

$$\pi_2(d): \mathscr{L}_1(N) \to \mathscr{L}_1(N/d)$$

given by

$$(t, L) \longmapsto (t, L_{(N/d)t}) ,$$

where $L_{(N/d)t}$ is the lattice generated by L and $(N/d)t$. Thus t has period N/d with respect to $L_{(N/d)t}$. Then the corresponding map on modular forms is determined by the computation:

$$\begin{aligned} \pi_2(d)_k F\left(\frac{1}{N}, [\tau, 1]\right) &= F\left(\frac{1}{N}, \left[\tau, \frac{1}{d}\right]\right) \\ &= F\left(\frac{1}{N}, \frac{1}{d}[d\tau, 1]\right) \\ &= d^k F\left(\frac{1}{Nd}, [d\tau, 1]\right) . \end{aligned}$$

Thus we find:

Lemma 1. *Let f be the function of τ associated with F. Then*

$$\pi_2(d)_k f(\tau) = d^k f(d\tau) ,$$

and thus on the q-expansions,

$$\pi_2(d)_\infty = d^k V_d ,$$

or in other words,

$$(\pi_2(d)_k F)_\infty(q) = d^k F_\infty(q^d) .$$

The operator V_d had been defined in Chapter VII, § 3.

Lemma 2. (i) *Let $d_1, d_2 > 1$ be such that $d_1 d_1 \mid N$. Then:*

$$\pi_i(d_1 d_2) = \pi_i(d_1)\pi_i(d_2) \quad \textit{for } i = 1, 2 .$$

(ii) *If $d, d' > 1$ and dd' divides N, then $\pi_1(d)$ commutes with $\pi_2(d')$.*

Proof. Clear.

The lemma also applies to the modular forms by composition.

From Lemma 1, we see that the operators $\pi_1(d)_k$ and $\pi_2(d)_k$ map cusp forms into cusp forms. We let

$$\mathbf{S}_1^-(N, k)$$

be the sum of the images of $\pi_1(d)_k$ and $\pi_2(d)_k$ applied to the space of cusp forms

$$M_1^0(N/d, k) = \mathbf{S}_1(N/d, k) ,$$

for all $d > 1$, $d \mid N$, and call $\mathbf{S}_1^-(N, k)$ the space of **old forms**, or **non-primitive** cusp forms on $\Gamma_1(N)$. We let

$$\mathbf{S}_1^+(N, k)$$

be the orthogonal complement of $\mathbf{S}_1^-(N, k)$ in the space of cusp forms on $\Gamma_1(N)$, i.e. in $M_1^0(N, k)$, and call $\mathbf{S}_1^+(N, k)$ the space of **new forms**, or **primitive forms**, on $\Gamma_1(N)$. This is an *analytic* characterization. *Algebraic* characterizations will be given in Theorems 2.2 and 3.3.

Lemma 3. *Let $d \mid N$, $d > 1$. Let $(a, N) = 1$. Let $(n, N) = 1$. Let $i = 1, 2$. The operators $\pi_i(d)$ commute with the action of $(\mathbf{Z}/N\mathbf{Z})^*$ and with the Hecke operator $T(n)$, in other words:*

$$\pi_i(d) \circ [a] = [a] \circ \pi_i(d)$$
$$\pi_i(d) T(n) = T(n) \pi_i(d) .$$

Proof. The commutative diagram for $\pi_2(d)$ and $[a]$ is as follows, and is trivially verified.

$$\begin{array}{ccc} (t, L) & \xrightarrow{[a]} & (at, L) \\ {\scriptstyle \pi_2(d)}\downarrow & & \downarrow{\scriptstyle \pi_2(d)} \\ (t, L_{(N/d)t}) & \xrightarrow[[a]]{} & (at, L_{(N/d)t}) \end{array}$$

We leave the diagram for $\pi_1(d)$ to the reader.

The diagram for the commutation with Hecke operators is as follows, say with $\pi_2(d)$.

$$\begin{array}{ccc} (t, L) & \xrightarrow{nT(n)} & \sum (t, L') \\ {\scriptstyle \pi_2(d)}\downarrow & & \downarrow{\scriptstyle \pi_2(d)} \\ (t, L_{(N/d)t}) & \xrightarrow[nT(n)]{} & \sum (t, L'_{(N/d)_t}) \end{array}$$

It is immediately verified to be commutative. We leave the diagram for $\pi_1(d)$ to the reader.

Theorem 1.1. *The Hecke algebra $\mathscr{H}_1^{(N)}(N, k)$ maps the space of primitive forms into itself.*

Proof. Its elements commute with the maps $\pi_i(d)$, and thus map $\mathbf{S}_1^-(N, k)$ into itself. Since the Hecke algebra is star-closed, it must map the orthogonal complement into itself, as was to be shown.

More significant is the commutation rule between the maps $\pi_i(d)$ and the involutions W_N.

Theorem 1.2. *We have*

$$\pi_1(d) \circ W_N = W_{N/d} \circ \pi_2(d) ,$$

and hence

$$W_{N,k} \circ \pi_1(d)_k = \pi_2(d)_k \circ W_{N/d,k} .$$

Proof. We look at the effect of the composite maps on pairs $(t, L) = (t, A)$, and have to verify that the diagram is commutative:

$$\begin{array}{ccc} (t, A) & \xrightarrow{W_N} & (t', A/t) \\ {\scriptstyle \pi_2(d)}\downarrow & & \downarrow{\scriptstyle \pi_1(d)} \\ (\lambda t, A/dt) & \xrightarrow[W_{N/d}]{} & (dt', A/t) \end{array}$$

where λ, φ are the homomorphisms

$$A \xrightarrow{\lambda} A/dt \xrightarrow{\varphi} A/t$$

and $\varphi\lambda : A \to A/t$ is the composite. By definition,

$$e_{\varphi\lambda}(t, t') = \zeta_N ,$$

whence

$$e_{\varphi\lambda}(t, dt') = \zeta_N^d = \zeta_{N/d} .$$

On the other hand, also by definition,

$$e_\varphi(\lambda t, (\lambda t)') = \zeta_{N/d} = \zeta_N^d .$$

Using Property **E2** of the pairing shows that $(\lambda t)' = dt'$, and proves the theorem.

Theorem 1.3. *The operator $W_{N,k}$ maps the primitive space $\mathbf{S}_1^+(N, k)$ into itself. It gives an isomorphism*

$$\mathbf{S}_1^+(N, k, \varepsilon) \to \mathbf{S}_1^+(N, k, \bar{\varepsilon}) .$$

Proof. Let f be primitive, and let g have level N/d. Then

$$\langle W_{N,k} f, \pi_1(d)_k g\rangle = \langle W_{N,k} f, g\rangle = 0 .$$

Since $W_{N/d,k}$ is an automorphism of the space of cusp forms of level N/d, it suffices next to prove that $W_{N,k} f$ is orthogonal to $\pi_2(d)_k W_{N/d,k} g$. We have by the preceding theorem:

$$\begin{aligned}\langle W_{N,k} f, \pi_2(d)_k W_{N/d,k} g\rangle &= \langle W_{N,k} f, W_{N,k}\pi_1(d)_k g\rangle \\ &= N^k \langle f\circ[w_N]_k, (\pi_1(d)_k g)\circ[w_N]_k\rangle \\ &= N^k\langle f, \pi_1(d)_k g\rangle \\ &= 0 ,\end{aligned}$$

thereby proving the first part of the theorem. The second part concerning the ε-eigenspace is left as an exercise to the reader. (Cf. Lemma 3 of Chapter VII, § 6.)

§ 2. Characterization of Primitive Forms

Let f be a modular form, as a function of $\tau\in\mathfrak{H}$, on $\Gamma_1(N)$. Let $\{\alpha_i\}$ be right coset representatives of $\Gamma_1(N)$ in $\Gamma_1(N/d)$, for some divisor d of N, $d>1$. We let $\mathrm{Tr}=\mathrm{Tr}_{N,N/d}$ be the trace operator,

$$\mathrm{Tr} = \frac{1}{m}\sum_{i=1}^{m} [\alpha_i]_k ,$$

where m is the index of $\Gamma_1(N)$ in $\Gamma_1(N/d)$. Then

$$\mathrm{Tr} f = \frac{1}{m}\sum f\circ[\alpha_i]_k$$

is invariant under $\Gamma_1(N/d)$, and so is a modular form on $\Gamma_1(N/d)$.

Theorem 2.1 *Let $g\in\mathbf{S}_1(N/d, k)$ and $f\in\mathbf{S}_1(N, k)$ be cusp forms of weight k on $\Gamma_1(N/d)$ and $\Gamma_1(N)$ respectively. Then for the Petersson scalar product, the adjoint of $\pi_1(d)_k$ is the trace, that is:*

$$\langle \pi_1(d)_k g, f\rangle = \langle g, \mathrm{Tr} f\rangle .$$

Proof. We know that $[\alpha_i]_k$ is unitary for the scalar product. Hence

$$\begin{aligned}\langle \pi_1(d)_k g, f\rangle &= \langle g, f\rangle \\ &= \frac{1}{m}\sum\langle g\circ[\alpha_i]_k, f\circ[\alpha_i]_k\rangle \\ &= \frac{1}{m}\sum\langle g, f\circ[\alpha_i]_k\rangle \\ &= \langle g, \mathrm{Tr} f\rangle ,\end{aligned}$$

as was to be shown.

Theorem 2.2. (i) *The operator $W_{N,k}$ maps $\mathbf{S}_1(N, k)$ into itself.*

(ii) *A cusp form f on $\Gamma_1(N)$ is orthogonal to the image of $\pi_1(d)_k$ if and only if* $\mathrm{Tr} f = 0$.

(iii) *A cusp form f on $\Gamma_1(N)$ is orthogonal to the image of $\pi_2(d)_k$ if and only if* $\mathrm{Tr}(f\circ[w_N]_k)=0$.

Proof. Since $W_{N,k}=N^{k/2}[w_N]_k$ by Lemma 1 of Chapter VII, § 6, it maps cusp forms into cusp forms, and its definition shows that the image is again on $\Gamma_1(N)$, so $W_{N,k}$ maps $\mathbf{S}_1(N, k)$ into itself. We have

$$\begin{aligned}&\langle f, \pi_1(d) g\rangle = 0 \quad \text{for all } g\in\mathbf{S}_1(N/d, k) \\ \Leftrightarrow\ &\langle \mathrm{Tr} f, g\rangle = 0 \quad \text{for all } g\in\mathbf{S}_1(N/d, k) ,\end{aligned}$$

which is equivalent to $\mathrm{Tr} f = 0$, thereby proving (ii).

As for (iii), we have

$$\begin{aligned}&\langle f, \pi_2(d) g\rangle = 0 \quad \text{for all } g \\ \Leftrightarrow\ &\langle f, \pi_2(d) W_{N/d,k} g\rangle = 0 \quad \text{for all } g\end{aligned}$$

(because $W_{N/d,k}$ is invertible, according to Lemma 2, Chapter VII, § 6)

$$\Leftrightarrow \langle f, W_{N,k}\pi_1(d) g\rangle = 0 \quad \text{for all } g$$

(by Theorem 1.1)

$$\begin{aligned}&\Leftrightarrow \langle f, (\pi_1(d) g)\circ[w_N]_k)\rangle = 0 \quad \text{for all } g \\ &\Leftrightarrow \langle f\circ[w_N]_k, \pi_1(d)_k g\rangle = 0 \quad \text{for all } g \\ &\Leftrightarrow \mathrm{Tr}(f\circ[w_N]_k) = 0,\end{aligned}$$

as was to be shown.

Serre [S 5] originally remarked the characterization of the orthogonal complement by means of the trace for $\Gamma_0(p)$, p prime. This was extended in general by Ogg and Li, cf. [Li], Theorem 4.

§ 3. The Structure Theorem

For each positive integer D we let as before $\mathscr{H}_1^{(ND)}(N, k)$ be the algebra operating on the cusp forms of weight k, level N, on $\Gamma_1(N)$, generated by $(\mathbf{Z}/N\mathbf{Z})^*$ and the Hecke operators $T_k(n)$ with n prime to ND. Since the level N remains fixed, and also the weight, we write simply $\mathscr{H}^{(ND)}$ for this algebra.

If f is an eigenfunction, then for each operator T we have

$$Tf = \psi_f(T) f ,$$

where $\psi_f(T)$ is a scalar, and $T \mapsto \psi_f(T)$ is a character. Conversely, it is easy to see that each character is associated with an eigenfunction. Since $(\mathbf{Z}/N\mathbf{Z})^*$ also lies in the Hecke algebra, there is also a Dirichlet character ε associated with f.

By all the commutativity theorems, and the fact that the algebra is star closed, we know that $\mathscr{H}^{(ND)}$ maps the primitive space $\mathbf{S}^+(N) = \mathbf{S}_1^+(N, k)$ into itself. The image of $\mathscr{H}^{(ND)}$ in the algebra of endomorphisms of $\mathbf{S}^+(N)$ will be denoted by

$$\mathscr{H}^{(ND),+} \quad \text{or} \quad \mathscr{H}^{(ND),\text{prim}} .$$

Since we deal only with modular forms from now on, we omit usually the subscript k, and write for instance $\pi_1(d)g$ instead of $\pi_1(d)_k g$.

For convenience of notation in the next lemma, we agree to the convention that if $d = 1$ then $\pi_1(d) = \pi_2(d) = \mathrm{id}$.

Lemma 1. *There exists a basis $\{f_1, \ldots, f_r\}$ of $\mathbf{S}_1(N, k)$ such that*

$$f_i = \pi_1(d_i)\pi_2(d_i')g_i ,$$

for some positive integers d_i, d_i', $d_i d_i' \mid N$, and g_i equal to a primitive form of level $N/d_i d_i'$ which is an eigenfunction of $\mathscr{H}^{(N)}$.

Proof. By definition, $\mathbf{S}_1(N, k)$ is the orthogonal sum of the primitive space and the old space. By induction on the level, and Lemma 2 of § 1, we see that forms of type

$$\pi_1(d)\pi_2(d')g ,$$

with g primitive of level N/dd', generate the old space. Since the primitive space of any level admits a basis consisting of eigenfunctions for the Hecke algebra (because the Hecke algebra is star closed and commutative), we can pick a basis from these generators, as was to be shown.

Remark. Let $f \in \mathbf{S}_1(N, k)$. We write f as a linear combination of the basis of Lemma 1,

$$f = \sum c_i f_i .$$

If $c_i \neq 0$, then we say that f_i **occurs** in f. From the linear independence and the eigenproperty, it is clear that if f is an eigenfunction of the Hecke algebra $\mathscr{H}^{(ND)}$, with character ψ, then each basis element f_i occurring in f also has this same character ψ. In particular, we can find an element g_i of level $N/d_i d_i'$ having the same character, and which is primitive.

In the next section we shall prove the following theorem. This is the basic result of Atkin–Lehner, who state the theorem of $\Gamma_0(N)$, but the proof for $\Gamma_1(N)$ is along the same lines in [Li].

Theorem 3.1. *Let $f \in \mathscr{F}_1(N, k)$, and $f = \sum a_n q^n$. Suppose there exists $D \geq 1$ such that $a_n = 0$ if $(n, ND) = 1$. Then there exist elements $g_p \in \mathscr{F}_1(N/p, k)$ with $p \mid N$, p prime, such that*

$$f = \sum_{p \mid N} \pi_2(p)_k g_p .$$

In the rest of this section, we give the main applications of Theorem 3.1.

Observe that we have stated Theorem 3.1 for meromorphic modular forms. If f is in $M_1(N, k)$ (resp. $\mathbf{S}_1(N, k)$) then the proof shows that in fact the elements g_p lie in $M_1(N/p, k)$ (resp. $\mathbf{S}_1(N/p, k)$). For the present application, we are of course concerned only with the case of cusp forms, i.e. the case when $f \in \mathbf{S}_1(N, k)$.

Theorem 3.2. *Let $f \in \mathbf{S}_1(N, k)$, $f = \sum a_n q^n$. Assume that f is an eigenfunction of $\mathscr{H}^{(ND)}$. If $a_1 = 0$ then f lies in the space of old forms. Hence if f is primitive, we have $a_1 \neq 0$.*

Proof. Let ε be the Dirichlet character of f. Let p be a prime number, $(p, ND) = 1$. Then

$$a_{np} + \varepsilon(p)p^{k-1}a_{n/p} = \lambda_p a_n$$

where $\lambda_p = \psi_f(T_k(p))$. By induction, we see that $a_{p^\nu} = 0$ for all ν, and then again by induction on the number of primes factors, we see that $a_n = 0$ for all n with $(n, ND) = 1$. We can then apply Theorem 3.1 to conclude the proof.

In view of Theorem 3.2, after multiplying a primitive non-zero eigenfunction of $\mathscr{H}^{(ND)}$ by a scalar, we can achieve $a_1 = 1$, in which case we call f **normalized**.

Theorem 3.3 (i) *Let $f, g \in \mathbf{S}_1(N, k)$ be eigenfunctions of $\mathscr{H}^{(ND)}$ with the same character. If f is primitive, non-zero, then there is some constant c such that $g = cf$. In particular, if g is old, then $g = 0$.*

(ii) *The space $\mathbf{S}_1^+(N, k)$ is the sum of the eigenspaces of $\mathscr{H}^{(ND)}$ whose eigencharacters occur with multiplicity one. The space $\mathbf{S}_1^-(N, k)$ is the sum of the eigenspaces whose eigencharacters occur with multiplicity > 1.*

Proof. Write

$$g = g^{\text{prim}} + g^{\text{old}}$$

where g^{prim} and g^{old} are the components of g in the primitive and old space respectively. Then each component has the same character as g (or f), and it suffices to prove the theorem when g is primitive or g is old.

If g is primitive, non-zero, then we may assume without loss of generality that f, g are normalized. It follows that $f - g$ is a primitive eigenfunction of the Hecke algebra $\mathscr{H}^{(ND)}$, whose first coefficient is equal to 0, whence $f - g = 0$ by Theorem 3.2, thereby concluding the proof in this case.

If g is old, then by Lemma 1 we can express g as a linear combination

$$g = \sum c_i f_i ,$$

where each f_i has the same character. Say f_i occurs in g. Let $h = \pi_1(d_i d_i')g_i$, if g_i is as in Lemma 1, $f_i = \pi_1(d_i)\pi_2(d_i')g_i$. By Theorem 3.2 applied at level $N/d_i d_i'$, we know that $a_1(h) \neq 0$. Hence there exists a constant c such that $a_1(f - ch) = 0$, and $f - ch$ is an eigenfunction of the Hecke algebra $\mathscr{H}^{(ND)}$, so that Theorem 3.2 implies that $f - ch$ is old, and f is old. Since f is primitive, we must have $f = 0$, a contradiction which shows that $g = 0$, and proves the first part of the theorem.

Observe that the character of a non-primitive eigenfunction always has multiplicity > 1, for if g has level N/d, $d > 1$, and character ψ, then

$$\pi_1(d)g \quad \text{and} \quad \pi_2(d)g$$

have the same character, and are not equal. The second part then follows at once.

Theorem 3.3 is sometimes called the **multiplicity 1 theorem**, because it shows that in the space of primitive forms, a character of the Hecke algebra occurs with multiplicity 1. Since the full Hecke algebra $\mathscr{H}$ is commutative, we get:

Corollary. *If f is a primitive eigenfunction of $\mathscr{H}^{(ND)}$, then f is also eigenfunction of $\mathscr{H}$.*

Theorem 3.4. *Let $\{f_1, \ldots, f_r\}$ be a basis of eigenfunctions of $\mathscr{H}^{(ND)}$ for the primitive space $\mathbf{S}_1^+(N, k)$. Let $\psi_1, \ldots, \psi_r$ be the associated characters. Then the map*

$$T \mapsto (\psi_1(T), \ldots, \psi_r(T))$$

is an isomorphism of $\mathscr{H}^{(ND),+}$ with $\mathbf{C}^r$.

Proof. This is the same situation as in level 1, because we now know from multiplicity 1 that the characters $\psi_1, \ldots, \psi_r$ are distinct, whence linearly independent.

In particular, $\mathbf{S}_1^+(N, k)$ is a 1-dimensional module over $\mathscr{H}^{(ND),+}$, with basis element $f_1 + \cdots + f_r$.

§ 4. Proof of the Main Theorem

The proof depends on an analysis of the operator V_p on power series, such that

$$V_p f_\infty(q) = f_\infty(q^p)$$

for various primes p. We shall distinguish two cases, depending on whether $p \mid N$ or $p \nmid N$. We begin by considering a subgroup related to this operator. Then we discuss the two cases, starting with $p \nmid N$.

For any positive integer m, we define $\Gamma_1(N, m)$ to consist of those matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$$

such that

$$a \equiv d \equiv 1 \pmod{N}, \qquad c \equiv 0 \pmod{N}, \qquad b \equiv 0 \pmod{m} .$$

Thus

$$\Gamma_1(N, m) = \Gamma_1(N) \cap \Gamma^0(m) ,$$

if $\Gamma^0(m)$ denotes the subgroup of matrices of $SL_2(\mathbf{Z})$ such that $b \equiv 0 \pmod{m}$.

Lemma 1. *Let p be a prime. Then*

$$(\Gamma_1(N) : \Gamma_1(N, p)) = \begin{cases} p & \text{if } p \mid N \\ p+1 & \text{if } p \nmid N. \end{cases}$$

Let $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. In the first case, the elements

$$\{T^j\} \ (j = 0, \ldots, p-1)$$

form right coset representatives. In the second case $p \nmid N$, these elements together with any matrix

$$Q = \begin{pmatrix} px & 1 \\ Ny & 1 \end{pmatrix} \in SL_2(\mathbf{Z}), \qquad x, y \in \mathbf{Z} ,$$

form right coset representatives. In both cases, $\Gamma_1(N)$ is generated by $\Gamma_1(N, p)$ and T.

Proof. Let

$$\alpha = \begin{pmatrix} a & b \\ Nc & d \end{pmatrix} \in \Gamma_1(N) .$$

Then

$$\alpha T^j = \begin{pmatrix} * & b + ja \\ * & * \end{pmatrix} .$$

If $p \nmid a$ we can pick j such that $b+ja \equiv 0 \pmod{p}$, and if $p \mid N$, then $p \nmid a$, and we can always do this. On the other hand, if $p \nmid N$, and $p \mid a$, then trivially

$$\alpha Q^{-1} \in \Gamma_1(N, p)$$

by straight matrix multiplication, so the coset representatives are as asserted. It is immediately verified that the coset representatives we have given lie in distinct cosets. It also follows at once that $\Gamma_1(N)$ is generated by $\Gamma_1(N, p)$ and T, since T has period p with respect to $\Gamma_1(N, p)$, and the index is either p or $p+1$.

We make use of Chapter VII, § 3, and the operators

$$V_p(\sum a_n q^n) = \sum a_n q^{pn}$$
$$U_p(\sum a_n q^n) = \sum_{p \mid n} a_n q^{n/p}.$$

We put

$$B_p = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$V_p f = p^{-k/2} f \circ [B_p]_k$$

and

$$U_p f = \frac{p^{k/2}}{p} \sum_{b=0}^{p-1} f \circ [B_p]_k^{-1} \circ [T^b]_k$$

We also recall that

$$T_{k,\varepsilon}(p) = U_p + \varepsilon(p) p^{k-1} V_p .$$

Lemma 2. *Let p be a prime. Let f be a meromorphic modular form on $\Gamma_1(N)$. Then:*

(i) *$V_p f$ is on $\Gamma_1(Np)$.*

(ii) *$f \circ [B_p^{-1}]_k$ is on $\Gamma_1(N, p)$ if $p \nmid N$.*

Proof. For (i) it suffices to show that

$$B_p^{-1} \Gamma_1(N) B_p \supset \Gamma_1(Np), \quad \text{that is} \quad B_p \Gamma_1(Np) B_p^{-1} \subset \Gamma_1(N),$$

which follows by straightforward multiplication of matrices.

For (ii), it suffices to prove that

$$B_p \Gamma_1(N) B_p^{-1} \supset \Gamma_1(N, p), \quad \text{that is} \quad B_p^{-1} \Gamma_1(N, p) B_p \subset \Gamma_1(N),$$

which again follows by multiplication of matrices.

The analogue of Lemma 2 (ii) for the case when $p \mid N$ will be given later in Lemma 5. We shall first settle completely the case when $p \nmid N$, which is easier.

Theorem 4.1. *Let p be a prime, $p \nmid N$. Let $f_\alpha(q)$ be a power series such that $V_p f_\alpha = f_\alpha(q^p)$ is in $\mathcal{F}_1(N, k)$. Then $f = 0$.*

Proof. We can write

$$f = f \circ [B_p]_k \circ [B_p^{-1}]_k .$$

By assumption, $f \circ [B_p]_k$ is on $\Gamma_1(N)$. By Lemma 2 (ii), we conclude that f is on $\Gamma_1(N, p)$. Since f has a power series expansion in q, it is invariant under T, and since we saw in Lemma 1 that $\Gamma_1(N, p)$, T generate $\Gamma_1(N)$, we conclude that f also is on $\Gamma_1(N)$.

Following a suggestion of Ihara, we may now conclude the proof as follows. We have seen that f is invariant under

$$\Gamma_1(N) \quad \text{and} \quad B_p \Gamma_1(N) B_p^{-1}.$$

Since these two groups generate a dense subgroup of $SL_2(\mathbf{R})$, we must have $f = 0$.

For the convenience of the reader, we sketch the proof that the subgroup generated by $\Gamma_1(N)$ and $B_p \Gamma_1(N) B_p^{-1}$ is dense in $SL_2(\mathbf{R})$. Let H be its closure, and let $\mathfrak{h}$ be the Lie algebra. Using some obvious elements it is immediate that the group generated by $\Gamma_1(N)$ and $B_p \Gamma_1(N) B_p^{-1}$ contains elements with arbitrary high powers of p in the denominators of the matrix entries, and hence that $\Gamma_1(N)$ is not of finite index in this group, so not of finite index in H. From this it can be seen that H is not discrete, and so $\mathfrak{h} \neq 0$. But $\mathfrak{h}$ must be invariant under inner automorphisms by $\Gamma_1(N)$, and one sees easily that $\mathfrak{h}$ must then be 3-dimensional. Since $SL_2(\mathbf{R})$ is connected, we must have $H = SL_2(\mathbf{R})$.

We recall the theorem we want to prove from the last section.

Let $f = \sum a_n q^n$ be in $\mathcal{F}_1(N, k)$. Suppose there exists $D \geq 1$ such that $a_n = 0$ if $(n, ND) = 1$. Then there exist elements $g_p \in \mathcal{F}_1(N/p, k)$ with $p \mid N$, p prime, such that

$$f = \sum_{p \mid N} \pi_2(p)_k g_p = \sum_{p \mid N} V_p g_p .$$

Without loss of generality, we may assume that $(D, N) = 1$. We shall peel off the primes dividing ND one at a time, starting with those which do not divide N.

From the definition of U_p, V_p we see that:

V_pU_p is the projection on the power series with coefficients a_n such that $p \mid n$.

$(I - V_pU_p)$ is the projection on the power series with coefficients a_n such that $p \nmid n$.

The operators $V_{p_i}U_{p_i}$ obviously commute with each other, and any operator V_p commutes with $U_{p'}$ if p, p' are distinct primes.

Case 1. $p \nmid N$.
Let $p_1, \ldots, p_r$ be the primes dividing ND, and say $p_r \nmid N$. Let

$$g = \prod_{i=1}^{r-1} (I - V_{p_i}U_{p_i})f .$$

Then g is on $\Gamma_1(Np_1^2 \cdots p_{r-1}^2)$ by Lemma 2 (i). Writing

$$g = \sum a_n(g)q^n$$

we see that $p_i \nmid n$ for $i = 1, \ldots, r-1$. By hypothesis, $p_r \mid n$, and $g = V_{p_r}h$ for some h. Since $p_r \nmid Np_1^2 \ldots p_{r-1}^2$, we conclude from Theorem 4.1 that $g = 0$. Hence $a_n(f) = 0$ if n is not divisible by some prime $p_1, \ldots, p_{r-1}$. Inductively, we have reduced the proof of theorem to the case with $D = 1$, i.e. we may assume

$$a_n = 0 \quad \text{if} \quad (n, N) = 1 .$$

To go further, it is convenient to set up the induction more formally.

Let f be on $\Gamma_1(N)$, $f = \sum a_n q^n$. Let $p_1, \ldots, p_r$ be the primes which divide N. Say $p_1, \ldots, p_t$ divide N exactly, and $p_{t+1}^2, \ldots, p_r^2$ divide N. We shall say that f is of **length** $\leqslant s$ if $a_n = 0$ unless $p_i \mid n$ for some $i \leqslant s$. Equivalently, this means that we can write

$$f = \sum_{i=1}^{s} V_{p_i} g_{p_i}$$

where g_{p_i} are power series in q (and we make no a priori requirement that these power series should be modular forms, although we shall see later how to make them such). The induction statement then runs as follows.

Theorem 4.2. *Let f be a meromorphic modular form on $\Gamma_1(N)$. Suppose that f has length $\leqslant s$, and $p = p_s$. Then there exists a meromorphic modular form h_p on $\Gamma_1(N/p)$ such that*

$$f - V_p h_p \quad \text{is on} \quad \Gamma_1(N) ,$$

and $f - V_p h_p$ has length $\leqslant s-1$. If f is holomorphic (resp. cusp) then h_p is holomorphic (resp. cusp).

This statement clearly proves our main theorem. The proof is given first when $p^2 \mid N$, and then when $p \mid N$ exactly, and we shall also need a reduction to the case when f has a Dirichlet character as follows.

Lemma 3. *Let $p \mid N$. If $(a, N) = 1$ then $[a]_k$ and V_pU_p commute on $\mathcal{F}_1(N, k)$.*

Proof. We know that $V_p : \mathcal{F}_1(N, k) \to \mathcal{F}_1(Np, k)$, and also that

$$V_p = \pi_2(p)_k \quad \text{and} \quad [a]_k \quad \text{commute} .$$

On the other hand, on $\mathcal{F}_1(Np, k, \varepsilon)$ for any ε, we have

$$U_p = T_{k,\varepsilon}(p) - \varepsilon(p)p^{k-1}V_p .$$

Hence U_p and $[a]_k$ commute. It follows that V_pU_p and $[a]_k$ commute on $\mathcal{F}_1(N, k, \varepsilon)$, and therefore on $\mathcal{F}_1(N, k)$, thus proving the lemma.

By the lemma, and the fact that for arbitrary f, we have

$$f_\varepsilon = pr_\varepsilon f .$$

is obtained from f by a linear combination of operators $[a]_k$, we conclude that if f satisfies the hypothesis of the theorem, namely

$$\prod_{i=1}^{s} (I - V_{p_i}U_{p_i})f = 0 ,$$

then each component f_ε also satisfies this condition. This reduces the proof of the theorem to the case when $f = f_\varepsilon$, which we now assume.

Lemma 4. *Let $p \mid N$. Let $f \in \mathcal{F}_1(N, k, \varepsilon)$, and assume that ε is not a character mod N/p. If $f = V_p g$ for some g then $f = 0$.*

Proof. Since $f(q) = g(q^p)$ it follows that f is invariant under

$$\begin{pmatrix} 1 & 1/p \\ 0 & 1 \end{pmatrix} .$$

It is also invariant under

$$\begin{pmatrix} 1 & 1 \\ N & N+1 \end{pmatrix} \in \Gamma_1(N) .$$

Hence for any $u, v \in \mathbf{Z}$, f is invariant under

$$\begin{pmatrix} 1 & u/p \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ N & N+1 \end{pmatrix}\begin{pmatrix} 1 & v/p \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+Nu/p & 1+\dfrac{v}{p}+\dfrac{Nuv}{p^2}+(N+1)\dfrac{u}{p} \\ N & 1+N+\dfrac{Nv}{p} \end{pmatrix}$$

Then operating on f with the above matrix leaves f invariant.

If $p^2 \mid N$ select $u \equiv -v \pmod{p}$. Then the above matrix lies in $\Gamma_0(N)$, and the operation also multiplies f by

$$\varepsilon\left(1+\frac{Nv}{p}\right),$$

which is $\neq 1$ for some value of v because ε is not defined mod N/p. This shows that $f=0$.

Suppose that p divides N exactly, so $N=pN'$ with N' prime to p. Then the above matrix has the form

$$\begin{pmatrix} 1+N'u & 1+N'u+\dfrac{u+v+N'uv}{p} \\ N & 1+N'v+N \end{pmatrix}.$$

We can select u, v such that

$$\frac{u+v+N'uv}{p}$$

is integral. Then again the matrix lies in $\Gamma_0(N)$. In fact, we pick any integer $a \not\equiv 1 \pmod{p}$, $a \not\equiv 0 \pmod{p}$, and solve

$$N'v \equiv a-1 \pmod{p}.$$

Then $v \not\equiv 0 \pmod{p}$ and $1+N'v \not\equiv 0 \pmod{p}$. We let

$$u \equiv \frac{-v}{1+N'v}.$$

(If $p=2$, there is no $a \not\equiv 1$, $a \not\equiv 0 \pmod 2$, but then

$$(\mathbf{Z}/N\mathbf{Z})^* = (\mathbf{Z}/N'\mathbf{Z})^*$$

and this case does not arise.) As in the case $p^2 \mid N$, we now see that operating on f with the matrix multiplies f by

$$\varepsilon(1+N'v),$$

which is $\neq 1$ for suitable choice of a, whence again $f=0$. This proves the lemma.

From the lemma, we may assume without loss of generality that ε is defined mod $N/p_1 \cdots p_s$. Indeed,

$$\prod_{i=1}^{s-1} (I - V_{p_i} U_{p_i}) f$$

is annihilated by $I - V_{p_s} U_{p_s}$, and is therefore of type $V_{p_s} g$ for some g. By the lemma, it is either 0 or ε is defined mod N/p_s. Proceeding inductively yields the desired reduction.

Lemmas 5, 6, 7 will now be stated in a self-contained manner.

Lemma 5. *Let $p \mid N$. Let $f \in \mathscr{F}_1(N, k, \varepsilon)$ and assume that ε is defined* mod N/p. *Then $f \circ [B_p^{-1}]_k$ is on $\Gamma_1(N/p, p)$.*

Proof. Let

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N/p, p).$$

Then

$$B_p^{-1} \alpha B_p = \begin{pmatrix} a & b/p \\ pc & d \end{pmatrix} \in \Gamma_0(N).$$

Since $a \equiv d \equiv 1 \pmod{N/p}$ and since ε is defined mod N/p, it follows that

$$f \circ [B_p^{-1} \alpha B_p]_k = f.$$

Hence $f \circ [B_p^{-1}]_k$ is fixed by α, thereby proving the lemma.

Lemma 6. *Let $f \in \mathscr{F}_1(N, k, \varepsilon)$ and assume that ε is defined* mod N/p. *Then $U_p f$ is on $\Gamma_1(N)$ if $p \mid N$ and on $\Gamma_1(N/p)$ if $p^2 \mid N$.*

Proof. By Lemma 5 we know that $f \circ [B_p^{-1}]_k$ is on $\Gamma_1(N/p, p)$, and we note that

$$\Gamma_1(N/p, p) \supset \Gamma_1(N, p).$$

According to Lemma 1, the expression

$$U_p f = p^{k/2-1} \sum_{b=0}^{p-1} f \circ [B_p]_k^{-1} \circ [T^b]_k$$

shows that

$$U_p f = p^{k/2-1} \operatorname{Tr}(f \circ [B_p^{-1}]_k),$$

where the trace is from $\Gamma_1(N, p)$ to $\Gamma_1(N)$ if $p \mid N$, and from $\Gamma_1(N/p, p)$ to $\Gamma_1(N/p)$ if $p^2 \mid N$. This proves the lemma.

Lemma 7. *Let $p \mid N$. Let f be a meromorphic modular form of some level. Assume that $V_p f \in \mathcal{F}_1(N, k, \varepsilon)$, where ε is defined* mod N/p. *Then f is on $\Gamma_1(N/p)$.*

Proof. We can write

$$f = f \circ [B_p]_k \circ [B_p^{-1}]_k ,$$

and apply Lemma 5 to $f \circ [B_p]_k$ to conclude that f is on $\Gamma_1(N/p, p)$. Since f has a power series expansion in q, it also follows that f is invariant under T. By Lemma 1, T and $\Gamma_1(N/p, p)$ generate $\Gamma_1(N/p)$, so the lemma is proved.

Case 2. $p^2 \mid N$.
Let p be a prime such that $p^2 \mid N$. Write

$$f = f_p + f_{\langle p \rangle} ,$$

where

$$f_p = \sum_{p \mid n} a_n q^n .$$

Then by Lemma 6 we see that

$$U_p f \quad \text{is on} \quad \Gamma_1(N/p) ,$$

and by Lemma 2 (i),

$$f_p = V_p U_p f \quad \text{is on} \quad \Gamma_1(N) .$$

Hence

$$f_{\langle p \rangle} = f - f_p \quad \text{is on} \quad \Gamma_1(N) .$$

This concludes the proof in the present case.

Case 3. $p \mid N$ but $p^2 \nmid N$.
We suppose that f is of length $\leqslant s$, with $s \leqslant t$, and put $p = p_s$, assuming that p divides N exactly. We let

$$f^{(2)} = (I - V_{p_1} U_{p_1}) f , \qquad g_{p_1} = U_{p_1} f$$

so that $f^{(2)}$ has length $\leqslant s-1$ with respect to $p_2, \ldots, p_s$, and

$$f = V_{p_1} g_{p_1} + f^{(2)} .$$

We then define g_{p_i} inductively, putting

$$f^{(i+1)} = \prod_{j=1}^{i} (I - V_{p_j} U_{p_j}) f$$

and

$$g_{p_i} = U_{p_i} f^{(i)} .$$

Then

$$f = \sum_{i=1}^{s} V_{p_i} g_{p_i} .$$

By Lemma 7 we conclude that g_{p_i} is on $\Gamma_1(N p_1 \ldots p_{i-1})$ for $i < s$.

By Lemma 2 (i) we conclude that $V_{p_i} g_{p_i}$ is on $\Gamma_1(N p_1 \ldots p_i)$ for $i < s$, and in particular, $V_{p_i} g_{p_i}$ is on $\Gamma_1(N^2/p_s)$.

We then see that $g_p = g_{p_s}$ is on $\Gamma_1(N^2/p^2)$, because $f^{(s)}$ has length $\leqslant 1$ with respect to $p = p_s$, so

$$f^{(s)} = f - \sum_{i=1}^{s-1} V_{p_i} g_{p_i}$$

is on $\Gamma_1(N p_1 \ldots p_{s-1})$, and g_{p_s} is on $\Gamma_1(N^2/p^2)$ by Lemma 7.

We now write

$$f = V_p g_p + \sum_{i=1}^{s-1} V_{p_i} g_{p_i} .$$

Then by Lemma 1,

$$U_p f = p^{k/2-1} \sum f \circ [B_p^{-1}]_k \circ [T^b]_k = p^{k/2-1} \operatorname{Tr}(f \circ [B_p^{-1}]_k) - p^{k/2-1} f \circ [B_p^{-1} Q]_k ,$$

where the trace is taken as the sum over coset representatives of $\Gamma_1(N/p, p)$ in $\Gamma_1(N/p)$, and we may select

$$Q = \begin{pmatrix} px & 1 \\ \dfrac{N}{p} y & 1 \end{pmatrix} \in SL_2(\mathbf{Z}) , \qquad x, y \in \mathbf{Z} ,$$

taking y of the form $\left(\dfrac{N}{p}\right) z$ with an appropriate integer z. In particular,

$$Q \in \Gamma_1((N/p)^2).$$

We define

$$W_p = p^{k/2-1}[B_p^{-1}Q]_k ,$$

and we let

$$h_p = (U_p + W_p)f = p^{k/2-1}\,\mathrm{Tr}\,(f \circ [B_p^{-1}]_k) .$$

Then h_p is on $\Gamma_1(N/p)$. Furthermore,

$$h_p = (U_p + W_p)V_p g_p + \sum_{i=1}^{s-1} (U_p + W_p)V_{p_i} g_{p_i} .$$

Since g_p is on $\Gamma_1((N/p)^2)$, we find

$$W_p V_p g_p = p^{k/2-1} g_p \circ [B_p B_p^{-1} Q]_k = p^{k/2-1} g_p .$$

Let $c_p = 1 + p^{k/2-1}$. Then

$$h_p = c_p g_p + \sum_{i=1}^{s-1} V_{p_i} U_p g_{p_i} + \sum_{i=1}^{s-1} W_p V_{p_i} g_{p_i} .$$

Lemma. *We have $W_p V_{p_i} g_{p_i} = V_{p_i} \varphi_{p_i}$ for some power series φ_{p_i} in q.*

Let us assume the lemma and conclude the proof. Let

$$h_{p_i} = U_p g_{p_i} + \varphi_{p_i} .$$

Then h_{p_i} is a power series in q, and

$$c_p g_p = h_p - \sum_{i=1}^{s-1} V_{p_i} h_{p_i} .$$

From this we get

$$f - V_p(c_p^{-1} h_p) = f - V_p g_p - c_p^{-1} \sum_{i=1}^{s-1} V_{p_i} V_p h_{p_i} .$$

But h_p on $\Gamma_1(N/p)$ implies that $f - V_p(c_p^{-1}h_p)$ is on $\Gamma_1(N)$, and its expression on the right shows that it has length $\leqslant s-1$. Except for the proof of the lemma, this concludes the proof of the inductive step.

Proof of the Lemma. Let

$$W_{p,i} = [B_{p_i} B_p^{-1} Q B_{p_i}^{-1}]_k .$$

Then

$$W_p V_{p_i} = V_{p_i} W_{p,i} .$$

The matrix

$$X = pB_{p_i} B_p^{-1} Q B_{p_i}^{-1} = \begin{pmatrix} px & p_i \\ Ny/p_i & p \end{pmatrix}$$

has determinant p, so $p^2 x - Ny = p$, and X represents $W_{p,i}$. A straightforward matrix multiplication shows that

$$XTX^{-1} \in \Gamma_1(Np_1 \cdots p_{i-1}) ,$$

or in other words, there is some element $\gamma \in \Gamma_1(Np_1 \cdots p_{i-1})$ such that $XT = \gamma X$. Since g_{p_i} is on $\Gamma_1(Np_1 \cdots p_{i-1})$, it follows that $W_{p,i} g_{p_i} = g_{p_i} \circ [X]_k$ is invariant under T, and hence is a power series in q, thus proving the lemma.

The complete proof of the theorem is therefore finished.

Remark. If f was a cusp form to begin with, then the construction shows that g_{p_i}, h_{p_i}, h_p are also cusp forms (on some level), because all constructions are carried out by operating with operators of type $[\alpha]_k$, where $\alpha \in GL_2^+(\mathbf{Q})$. A similar remark applies if $f \in M_1(N, k)$, in which case we end up with holomorphic forms, instead of meromorphic ones.

Chapter IX. The Dedekind Formalism

This chapter pertains both to Part II and Part III. It deals with periods of differentials of third kind. It reproduces with little change the arguments of Dedekind [Ded] for $d\log\eta$. These arguments are typical of those used in deriving the transformation law for more complicated modular forms, and are therefore included here for the convenience of the reader, to give him an early acquaintance with this formalism. We use the transformation constant in the normalization due to Rademacher [Rad] rather than the Dedekind symbol $S(c, d)$ because it is more convenient.

The same arguments as Dedekind can be applied to other modular forms, e.g. Klein forms (cf. [KL 1] for the definition). Goldstein has associated Dedekind type sums to arbitrary Fuchsian groups [G]. For other applications, see for instance Hirzebruch-Zagier [HZ].

§ 1. The Transformation Formalism

As usual, we let

$$\eta(\tau)=q^{1/24}\prod_{n=1}^{\infty}(1-q^n),\qquad q=e^{2\pi i\tau}.$$

Let

$$\alpha=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in SL_2(\mathbf{Z}).$$

We fix a branch $\log\eta$.

For $c\neq 0$ we write

LOG 1. $$\log\eta(\alpha\tau)=\log\eta(\tau)+\frac{1}{2}\log\frac{c\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha)\frac{2\pi i}{24}.$$

Then $\Phi(\alpha)$ is a rational number, and $\Phi(-\alpha)=\Phi(\alpha)$.

Let

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

We have the following properties.

Φ 1.	$\Phi(I)=0$	
Φ 2.	$\Phi(S)=0$	
Φ 3.	$\Phi(\alpha T)=\Phi(\alpha)+1$	
Φ 4.	$\Phi(\alpha S)=\Phi(\alpha)-3$	if $c, d>0$
Φ 5.	$\Phi(\alpha S)=\Phi(\alpha)+3$	if $c>0, d<0$.
Φ 6.	$\Phi(\alpha^{-1})=-\Phi(\alpha)$	
Φ 7.	$\Phi(\alpha')=-\Phi(\alpha)$	if $\alpha'=\begin{pmatrix}-a & b\\ c & -d\end{pmatrix}$

We now prove these properties.

The first is obvious. For **Φ 2**, it is standard from the elementary theory of elliptic functions that

$$\log\eta(S\tau)=\log\eta(\tau)+\frac{1}{2}\log\frac{\tau}{i}.$$

See for instance [L 2], Chapter 18, § 5. This can also be proved via the functional equation of Dirichlet series à la Hecke, cf. Ogg [O 1], p. I-44.

Property **Φ 3** is clear if $c=0$. Suppose $c\neq 0$. Then

$$\alpha T=\begin{pmatrix}a & b+a\\ c & d+c\end{pmatrix}.$$

We get

$$\log\eta(\alpha TT^{-1}\tau)=\log\eta(T^{-1}\tau)+\frac{1}{2}\log\frac{cT^{-1}\tau+d+c}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha T)\frac{2\pi i}{24},$$

so

$$\log\eta(\alpha\tau)=\log\eta(\tau)-\frac{2\pi i}{24}+\frac{1}{2}\log\frac{c\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha T)\frac{2\pi i}{24}.$$

We compare this with **LOG 1** to get **Φ 3**.

Φ 4. $\Phi(\alpha S)=\Phi(\alpha)-3$, assuming $c, d>0$. We have

$$\alpha S=\begin{pmatrix}b & -a\\ d & -c\end{pmatrix}$$

and

$$\log \eta(\alpha S\tau)=\log \eta(\tau)+\frac{1}{2}\log\frac{d\tau-c}{di}+\frac{1}{4}\log d^2+\Phi(\alpha S)\frac{2\pi i}{24}$$

$$=\log \eta(S\tau)+\frac{1}{2}\log\frac{cS\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha)\frac{2\pi i}{24}$$

$$=\log \eta(\tau)+\frac{1}{2}\log\frac{\tau}{i}+\frac{1}{2}\log\frac{cS\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha)\frac{2\pi i}{24}.$$

Having assumed $c, d>0$, the above equalities are logically equivalent with

$$\log\frac{d\tau-c}{i}+2\Phi(\alpha S)\frac{2\pi i}{24}=\log\frac{\tau}{i}+\log\frac{cS\tau}{i}+2\Phi(\alpha)\frac{2\pi i}{24},$$

which holds if and only if

$$2[\Phi(\alpha S)-\Phi(\alpha)]\frac{2\pi i}{24}=\log(cS\tau+d)+\log\tau-\log(d\tau-c)-\frac{\pi i}{2}.$$

Thus to prove **Φ 3** it suffices to prove

$$\log(cS\tau+d)+\log\tau-\log(d\tau-c)=0,$$

and this value is independent of τ, so we put $\tau=$ i, which makes it obviously true, and proves the desired relation.

Φ 5 is proved analogously.

Φ 6. $\Phi(\alpha^{-1})=-\Phi(\alpha)$. Since $\Phi(-\alpha)=\Phi(\alpha)$, we may assume $c>0$. We have

$$\alpha^{-1}=\begin{pmatrix} d & -b\\ -c & a\end{pmatrix}.$$

Also,

$$\log\eta(\alpha^{-1}\alpha\tau)=\log\eta(\alpha\tau)+\frac{1}{2}\log\frac{-c\alpha\tau+a}{-ci}+\frac{1}{4}\log c^2+\Phi(\alpha^{-1})\frac{\pi i}{24}$$

$$=\log\eta(\tau)+\frac{1}{2}\log\frac{c\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha)\frac{2\pi i}{24}$$

$$+\frac{1}{2}\log\frac{-c\alpha\tau+a}{-ci}+\frac{1}{4}\log c^2+\Phi(\alpha^{-1})\frac{2\pi i}{24}$$

$$=\log\eta(\tau),$$

because $\alpha^{-1}\alpha\tau=\tau$. Hence we obtain

$$\frac{1}{2}\log\frac{c\tau+d}{i}+\frac{1}{2}\log\frac{c\alpha\tau-a}{i}=-[\Phi(\alpha^{-1})+\Phi(\alpha)]\frac{2\pi i}{24}.$$

But

$$c\alpha\tau-a=c\frac{a\tau+b}{c\tau+d}-a=-\frac{1}{c\tau+d},$$

and for any $z\in\mathfrak{H}$,

$$\log z+\log(-1/z)=\pi i.$$

Hence the left-hand side in the last relation for $\Phi(\alpha)$ is 0, and the property **Φ 6** is proved.

Φ 7. $\Phi(\alpha')=-\Phi(\alpha)$, where $\alpha'=\begin{pmatrix}-a & b\\ c & -d\end{pmatrix}$. To do this one, we have to take complex conjugates. From the q-product for η, we see that

$$\overline{\log\eta(\tau)}=\log\eta(-\bar\tau).$$

Let $\tau_1=-\bar\tau$, so $\bar\tau_1=-\tau$. Then

$$\alpha\tau=\frac{a\tau+b}{c\tau+d}=-\overline{\left(\frac{-a\tau_1+b}{c\tau_1+d}\right)}=-\overline{\alpha'\tau_1},\quad\text{and}\quad-\alpha\bar\tau=\alpha'\tau_1.$$

Then

$$\log\eta(\alpha\tau)=\log\eta(\tau)+\frac{1}{2}\log\frac{c\tau+d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha)\frac{2\pi i}{24}$$

$$\log\eta(-\alpha\bar\tau)=\log\eta(-\bar\tau)+\frac{1}{2}\log\frac{c\bar\tau+d}{-ci}+\frac{1}{4}\log c^2-\Phi(\alpha)\frac{2\pi i}{24}$$

$$\log\eta(\alpha'\tau_1)=\log\eta(\tau_1)+\frac{1}{2}\log\frac{c\tau_1-d}{ci}+\frac{1}{4}\log c^2+\Phi(\alpha')\frac{2\pi i}{24}.$$

Hence

$$[\Phi(\alpha)+\Phi(\alpha')]\frac{2\pi i}{24}=\frac{1}{2}\log\frac{c\bar\tau+d}{-ci}-\frac{1}{2}\log\frac{c\tau_1-d}{ci}.$$

Put $\tau = i$, so that $\tau_1 = i$. The desired relation amounts to

$$\log \frac{-ci+d}{-ci} - \log \frac{ci-d}{ci} = 0 ,$$

which is true, and proves Φ 7.

Rademacher [Rad] gives other formulas for $\Phi(\alpha)$, especially how it behaves under general composition. We let the reader look into his paper for these, proved by suitable induction from the above.

The Dedekind symbol for relatively prime d, c is defined in terms of $\Phi(\alpha)$ by

$$\Phi(\alpha) = \frac{a}{c} + \frac{d}{c} - 12\mathrm{S}(\alpha) ,$$

and $\mathrm{S}(\alpha) = \mathrm{S}(d, c)$ depends only on d, c. It satisfies the following properties:

S 1.
(i) $\mathrm{S}(-d, -c) = \mathrm{S}(d, c)$
(ii) $\mathrm{S}(-d, c) = -\mathrm{S}(d, c)$
(iii) $\mathrm{S}(d, c) = \mathrm{S}(d, -c)$

S 2. $\mathrm{S}(d, c) = \mathrm{S}(d', c)$ if $d \equiv d' \pmod{c}$

S 3. $\mathrm{S}(0, 1) = \mathrm{S}(d, 1) = 0$

S 4. If $c, d > 0$ then

$$12\mathrm{S}(d, c) + 12\mathrm{S}(c, d) = -3 + \frac{d}{c} + \frac{c}{d} + \frac{1}{cd}$$

S 5. $\mathrm{S}(d, c) = \mathrm{S}(a, c)$ if $ad \equiv 1 \pmod{c}$.

These easily follow from the properties of the Φ-symbol, and are left as an exercise to the reader.

§ 2. Evaluation of the Dedekind Symbol

First we need to make some remarks concerning a limiting argument to be used frequently.

We shall write $\tau = x + iy$, and we shall take limits as τ approaches $-d/c$. For convenience, we shall take

$$\tau = -\frac{d}{c} + iy , \qquad y \to 0 .$$

This will of course be applicable only when $c \neq 0$. If

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$\alpha\tau = \frac{a}{c} - \frac{1}{c^2 iy}$$

and

$$q_{\alpha\tau} = e^{2\pi i \alpha\tau} \to 0 \text{ as } y \to 0,$$

or in other words, as $\tau \to -d/c$ in the prescribed fashion.

Two functions of τ will be said to be **equivalent**, written $\sim$, if their difference approaches 0 as $\tau \to -d/c$ in this fashion. For instance, from the product expression, we get

$$\log \eta(\alpha\tau) = \frac{2\pi i \alpha\tau}{24} + O(q_{\alpha\tau}) ,$$

whence

$$\log \eta(\alpha\tau) \sim \frac{2\pi i \alpha\tau}{24} \sim \frac{2\pi i a}{24\, c} - \frac{2\pi i}{24 c^2 iy} .$$

From **LOG 1** and

$$\alpha\tau = \frac{a}{c} - \frac{1}{c(c\tau + d)}$$

we get

LOG 2 $$0 \sim \log \eta(\tau) + \frac{1}{2} \log \frac{c\tau + d}{ci} + \frac{1}{4} \log c^2 + \frac{2\pi i d}{24\, c} + \frac{2\pi i}{24} \frac{1}{c(c\tau + d)} - \pi i \mathrm{S}(\alpha) .$$

We let

$$\zeta = e^{-2\pi i d/c} \qquad \text{if } c \neq 0 .$$

Then

$$e^{2\pi i \tau} = r\zeta , \qquad \text{where } r = e^{-2\pi y} .$$

We find from the product expansion,

$$\log \eta(\tau) = \frac{2\pi i \tau}{24} - \sum_{n=1}^{\infty} \frac{1}{n} \frac{r^n \zeta^n}{1 - r^n \zeta^n} .$$

We plug this in **LOG 2**, and use the symbol $\sim$ for equivalence to mean that the difference of the two terms has limit 0 as $r \to 1$. We obtain:

$$0 \sim \frac{2\pi i}{24}\left[-\frac{d}{c}+iy\right] - \sum_{n=1}^{\infty}\frac{1}{n}\frac{r^n\zeta^n}{1-r^n\zeta^n}$$

$$+\frac{1}{2}\log\frac{c\tau+d}{ci}+\frac{1}{4}\log c^2$$

$$+\frac{2\pi i d}{24\,c}+\frac{2\pi i}{24}\frac{1}{c(c\tau+d)}-\pi i\mathbf{S}(\alpha)\,,$$

whence

$$0\sim\frac{-2\pi y}{24}-\sum\frac{1}{n}\frac{r^n\zeta^n}{1-r^n\zeta^n}+\frac{1}{2}\log y+\frac{1}{4}\log c^2+\frac{2\pi}{24c^2y}-\pi i\mathbf{S}(\alpha)\,.$$

Taking complex conjugate yields

$$0 \sim \text{real terms} - \sum\frac{1}{n}\frac{r^n\zeta^n}{1-r^n\zeta^n}+\pi i\mathbf{S}(\alpha)$$

so we get

$$2\pi i\mathbf{S}(\alpha) \sim -\sum\frac{1}{n}\left[\frac{r^n\zeta^n}{1-r^n\zeta^n}-\frac{r^n\zeta^n}{1-r^n\zeta^n}\right] \sim -\sum_{n=1}^{\infty}\frac{1}{n}a_n(r) \tag{1}$$

where

$$a_n(r)=\frac{1}{1-r^n\zeta^n}-\frac{1}{1-r^n\zeta^{-n}}\,.$$

The partial sums $\sum_{n\leqslant N} a_n(r)$ *are uniformly bounded for all* N *and* $0\leqslant r\leqslant 1$.

Proof. We combine the terms $a_n(r)$ with

$$n=kc+m, \qquad n=(k+1)c-m, \qquad \text{and} \qquad 0<m<c/2.$$

We note that for $0<m<c/2$ we have:

$$a_{kc+m}(r)+a_{(k+1)c-m}(r)=\frac{1}{1-r^{kc+m}\zeta^m}-\frac{1}{1-r^{kc+m}\zeta^{-m}}$$

$$+\frac{1}{1-r^{(k+1)c-m}\zeta^{-m}}-\frac{1}{1-r^{(k+1)c-m}\zeta^{m}}$$

Combining the two terms with ζ^m gives an estimate

$$\left|\frac{r^{kc}\zeta^m(r^{c-m}+r^m)}{(1-r^{(k+1)c-m}\zeta^m)(1-r^{kc+m}\zeta^m)}\right| \ll r^{kc}\,.$$

We obtain from the preceding estimate an upper bound for the sum of such terms with $k\leqslant K$, namely

$$\sum_{k=1}^{K} r^{kc}(1-r^{c-2m})\leqslant\frac{1-r^{Kc}}{1-r^c}(1-r^{c-2m})\,,$$

which is uniformly bounded. This proves the lemma.

We can take a summation by parts, showing that the limit for $r\to 1$ in the series $\sum a_n(r)/n$ can be taken inside the series sign, and yields

$$2\pi i\mathbf{S}(\alpha)=-\sum\frac{a_n}{n}; \tag{2}$$

where

$$a_n=a_n(1)=\frac{1}{1-\zeta^n}-\frac{1}{1-\zeta^{-n}}\,.$$

Note that

$$a_n=a_{n'} \text{ if } n'\equiv n \pmod c$$

$$a_1+\cdots+a_c=0.$$

Define

$$f(t)=\sum_{n=1}^{c-1}a_nt^n\,.$$

Then $f(t)/(1-t^c)$ is continuous on [0, 1], and

$$2\pi i\mathbf{S}(\alpha)=-\int_0^1\frac{f(t)/t}{1-t^c}dt\,. \tag{3}$$

Proof. We have

$$\frac{f(t)/t}{1-t^c}=\frac{f(t)}{t}\sum_{\nu=0}^{\infty}t^{\nu c}=\sum_{n=1}^{c}\sum_{\nu=0}^{\infty}a_nt^{n+\nu c-1}$$

because

$$\frac{1}{n}=\int_0^1 t^{n-1}dt\,,$$

and the series $\sum a_n/n$ converges, so (3) is clear.

We have a partial fraction decomposition

$$\frac{f(t)/t}{t^c-1}=\frac{1}{c}\sum_{m\not\equiv 0 \pmod c}\frac{f(\zeta^m)}{t-\zeta^m}. \tag{4}$$

We then find

$$\int_0^1\frac{f(t)/t}{1-t^c}dt=-\frac{1}{c}\sum_{m\not\equiv 0}f(\zeta^m)\int_0^1\frac{1}{t-\zeta^m}dt=-\frac{1}{c}\sum_{m\not\equiv 0}f(\zeta^m)[\log(1-\zeta^m)-\log(-\zeta^m)]\,.$$

Lemma 1. *If $0<m<c$ then $f(\zeta^m)=2m-c$.*

Proof. We have:

$$f(\zeta^m)=\sum_{n\not\equiv 0}\left(\frac{1}{1-\zeta^n}-\frac{1}{1-\zeta^{-n}}\right)\zeta^{mn}\,.$$

By cross multiplication, we have for any c-th root of unity $\lambda\neq 1$,

$$\frac{1}{1-\lambda}=-\frac{1}{c}\sum_{k=0}^{c-1}k\lambda^k\,.$$

So

$$a_n=\frac{1}{c}\sum_{k=0}^{c-1}k\zeta^{-nk}-\frac{1}{c}\sum_{k=0}^{c-1}k\zeta^{nk}$$

and

$$f(\zeta^m)=\sum_{n\not\equiv 0}a_n\zeta^{mn}=\frac{1}{c}\sum_{n\equiv 0}\sum_{k=0}^{c-1}k\zeta^{n(m-k)}-\frac{1}{c}\sum_{n\not\equiv 0}\sum_{k=0}^{c-1}k\zeta^{n(m+k)}$$

We interchange the summation over n and k. Then

$$\sum_{n\not\equiv 0}\zeta^{n(m-k)}=\begin{cases}0 & \text{if } m\not\equiv k\\ -1 & \text{if } m\equiv k.\end{cases}$$

Therefore

$$f(\zeta^m)=\frac{1}{c}(mc-\textstyle\sum k-(c-m)c+\sum k)=2m-c\,,$$

as was to be shown.

We now introduce the symbol $((x))$ to mean the unique real number x' such that

$$-1/2\leqslant x'\leqslant 1/2 \quad\text{and}\quad x'\equiv x\ (\mathrm{mod}\ \mathbf{Z}).$$

Thus

$$((x))=x-[x]-1/2\,,$$

where $[x]$ is the largest integer $\leqslant x$.

Since we took $1\leqslant m\leqslant c-1$ in the preceding lemma, we can also write

$$2m-c=2c\left(\left(\frac{m}{c}-\frac{1}{2}\right)\right),$$

Lemma 2. $\log(1-\zeta^m)-\log(-\zeta^m)=\pi i\left(\left(\frac{dm}{c}-\frac{1}{2}\right)\right).$

Proof. Observe that

$$\left(\left(\frac{dm}{c}-\frac{1}{2}\right)\right)=\begin{cases}\left(\left(\frac{dm}{c}\right)\right)+\frac{1}{2} & \text{if } \left(\left(\frac{dm}{c}\right)\right)<0\\ \left(\left(\frac{dm}{c}\right)\right)-\frac{1}{2} & \text{if } \left(\left(\frac{dm}{c}\right)\right)>0\,.\end{cases}$$

To prove the lemma, we distinguish two cases, namely

$$-\frac{dm}{c}\equiv\frac{u}{2}\ (\mathrm{mod}\ \mathbf{Z}) \quad\text{and}\quad -\frac{dm}{c}\equiv-\frac{u}{2}\ (\mathrm{mod}\ \mathbf{Z})\,,$$

where $0<u<1$. The log terms then give the value $\pi i\phi$, where

$$\phi=\frac{1}{2}-\frac{u}{2} \text{ in the first case}$$

$$\phi=\frac{u}{2}-\frac{1}{2} \text{ in the second case.}$$

In both cases the lemma is then clear.

Therefore we finally obtain Dedekind's value:

Theorem 2.1. $\mathbf{S}(\alpha)=\mathbf{S}(d,c)=\sum_{m\not\equiv 0}\left(\left(\frac{m}{c}-\frac{1}{2}\right)\right)\left(\left(\frac{md}{c}-\frac{1}{2}\right)\right).$

Part IV

Congruence Properties and Galois Representations

Chapter X. Congruences and Reduction mod p

The study of modular forms modulo p was originated by Swinnerton-Dyer [Sw D], who determined the structure of the algebra of modular forms mod p. Serre then showed how one can extend this theory in many ways, and in particular obtained results concerning the congruence properties mod higher powers of p for the coefficients of the q-expansions of modular forms. After laying the basic foundations for the q-expansions, we reproduce Swinnerton-Dyer's results, and then some of Serre's basic results, referring to his more extensive papers for the continuation of the theory.

One of the applications to congruence properties of modular forms are to congruence properties of values of zeta functions and L-series at negative integers. Siegel [Si 1], [Si 2] first obtained such results by viewing the values of zeta functions as constant terms of modular forms. Serre greatly expanded these results, cf. for instance [Se 5].

From § 7 onwards, we let $M = M(\mathbf{Z}_{(p)})$ be the algebra of modular forms whose q-expansions have p-integral coefficients, and we let $\overline{M}$ be its reduction mod p, where p is a prime $\geqslant 5$.

§ 1. Kummer Congruences

Nothing about Bernoulli numbers will be used in this chapter other than their definition,

$$\frac{t}{e^t - 1} = \sum B_k \frac{t^k}{k!},$$

from which it is obvious that $B_k = 0$ if k is odd, $k \neq 1$.

Theorem 1.1 *Let p be a prime number, m, n positive integers.*

(i) *If $(p-1) \nmid n$ then B_n/n is p-integral.*

(ii) *If $n \equiv m \not\equiv 0 \pmod{p-1}$, then*

$$\frac{B_n}{n} \equiv \frac{B_m}{m} \pmod{p}.$$

Proof. Let a be a primitive root mod p. Let

$$A(t)=F(at)-F(t)=\frac{at}{e^{at}-1}-\frac{t}{e^t-1},$$

$$=\sum B_k(a^k-1)\frac{t^k}{k!}$$

$$=t\sum\frac{B_k}{k}(a^k-1)\frac{t^{k-1}}{(k-1)!}$$

$$=t\sum A_{k-1}\frac{t^{k-1}}{(k-1)!},\qquad \text{where } A_{k-1}=\frac{B_k}{k}(a^k-1).$$

We have to show:

(i) A_{n-1} is p-integral

(ii) $A_{n-1}\pmod p$ is periodic of period $p-1$.

To do this, let $u=e^t-1$. Then

$$A(t)=\frac{at}{(u+1)^a-1}-\frac{t}{u}=\frac{at}{au+\cdots}-\frac{t}{u}$$

$$=t\left(\sum_{k=0}^{\infty}c_ku^k\right)\quad \text{with } p\text{-integral coefficients } c_k.$$

Let

$$P_1(t)=\sum p_{1k}\frac{t^k}{k!}$$

$$\cdots$$

$$P_r(t)=\sum p_{rk}\frac{t^k}{k!}$$

be power series such that the coefficients p_{in} are p-integral and $p_{in}\bmod p$ are periodic with period $p-1$. Any linear combination of these series with p-integral coefficients also has these properties. Hence it suffices to prove that for each k, the coefficients of the series

$$(e^t-1)^k=\sum a_n\frac{t^n}{n!}$$

have these properties. But $(e^t-1)^k$ is a linear combination of powers e^{rt}, and

$$e^{rt}=\sum r^n\frac{t^n}{n!}.$$

Since $r^n\equiv r^m\pmod p$ if $n\equiv m \bmod p-1$, our theorem is proved.

§ 2. Von Staudt Congruences

Theorem 2.1 *We have*

$$B_n\equiv\sum_{(p-1)\mid n}-\frac{1}{p}\pmod{\mathbf{Z}}.$$

In particular, B_n has a pole or order 1 at p if $n\equiv 0\pmod{p-1}$.

Proof. Since we know from the Kummer congruences that B_n is p-integral if $(p-1)\nmid n$, it will suffice to prove that for each prime p such that $(p-1)\mid n$ we have

$$pB_n\equiv -1\pmod p.$$

Comparing the coefficients of $t^n/n!$ $(n>1)$ in the relation

$$\left(\frac{e^{pt}-1}{t}\right)\sum_{k=0}^{\infty}B_k\frac{t^k}{k!}=\frac{e^{pt}-1}{t}\frac{t}{e^t-1}=\sum_{r=0}^{p-1}e^{rt}$$

we find

$$\sum_{k=0}^{n}\frac{n!}{(n-k+1)!k!}B_kp^{n-k+1}=\sum_{r=1}^{p-1}r^n,$$

or

$$pB_n=-\sum_{k=0}^{n-2}\binom{n}{k}pB_k\frac{p^{n-k}}{n-k+1}+\sum_{r=1}^{p-1}r^n.$$

Since $p^{n-k}/(n-k+1)$ is p-integral, this shows inductively that pB_n is p-integral for all n. Then, since

$$\frac{p^{n-k}}{n-k+1}\equiv 0\pmod p\qquad \text{for}\quad k\leqslant n-2,$$

we deduce

$$pB_n\equiv\sum_{r=1}^{p-1}r^n\pmod p,\qquad \text{for}\quad n\geqslant 2.$$

If $(p-1)\mid n$, then all of the summands r^n are $\equiv 1\pmod p$ by Fermat's theorem, so $pB_n\equiv -1\pmod p$, as was to be shown. (I owe this proof to Zagier.)

§ 3. q-Expansions

We start with the product expansion for the sine,

$$\sin z = z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2\pi^2}\right).$$

Taking the logarithmic derivative yields

$$z\frac{\cos z}{\sin z} = 1 + 2\sum_{n=1}^{\infty} \frac{z^2}{z^2 - n^2\pi^2}$$

$$= 1 - 2\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{z^{2m}}{n^{2m}\pi^{2m}}. \tag{1}$$

On the other hand, from the definition of Bernoulli numbers

$$\frac{t}{e^t - 1} = 1 - \frac{t}{2} + \sum_{k=2}^{\infty} B_k \frac{t^k}{k!}$$

substituting $t = 2iz$, we obtain

$$z\frac{\cos z}{\sin z} = iz\frac{e^{iz} + e^{-iz}}{e^{iz} - e^{-iz}} = 1 + \sum_{k=2}^{\infty} B_k \frac{(2iz)^k}{k!}. \tag{2}$$

Comparing the coefficients in the power series in z in (1) and (2) yields:

Theorem 3.1. *If k is an even integer, then*

$$2\zeta(k) = -\frac{B_k}{k!}(2\pi i)^k.$$

We now start over again using

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z}{n}\right)\left(1 + \frac{z}{n}\right)$$

with logarithmic derivative

$$\pi\frac{\cos \pi z}{\sin \pi z} = \frac{1}{z} + \sum_{n=1}^{\infty}\left[\frac{1}{z-n} + \frac{1}{z+n}\right]. \tag{3}$$

We let

$$q = e^{2\pi i\tau}.$$

Then for τ in the upper half plane $\mathfrak{H}$, we get

$$\pi\frac{\cos \pi\tau}{\sin \pi\tau} = \pi i\frac{q+1}{q-1} = \pi i + \frac{2\pi i}{q-1} = \pi i - 2\pi i\sum_{\nu=0}^{\infty} q^{\nu}. \tag{4}$$

Differentiating the expressions in (3) and (4) repeatedly yields

$$(-1)^{k-1}(k-1)!\sum_{n=-\infty}^{\infty}\frac{1}{(\tau-n)^k} = -\sum_{\nu=1}^{\infty}(2\pi i)^k\nu^{k-1}q^{\nu}. \tag{5}$$

From the definition

$$S_k(\tau) = \sum_{\substack{m,n\\ \neq(0,0)}}\frac{1}{(m\tau+n)^k} \qquad (k \text{ even})$$

we get, summing separately for $m=0$ and $m\neq 0$,

$$S_k = 2\zeta(k) + 2\sum_{m=1}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{(m\tau+n)^k}$$

$$= 2\zeta(k) + 2\sum_{m=1}^{\infty}\sum_{\nu=1}^{\infty}\frac{(2\pi i)^k\nu^{k-1}}{(k-1)!}q^{m\nu}$$

We let

$$\sigma_k(n) = \sum_{d\mid n} d^k.$$

Theorem 3.2. *We have*

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

Since we want power series in q with integer coefficients if possible, and also starting with 1, it is convenient to introduce the constant factor of S_k defined by

$$S_k = 2\zeta(k)E_k.$$

In the light of Theorem 3.1, this yields

$$E_k = 1 - \frac{2k}{B_k}\sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n.$$

§ 4. Modular Forms over $\mathbf{Z}\left[\frac{1}{2}, \frac{1}{3}\right]$.

We define for an even integer k,

$$E_k = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n ,$$

and we abbreviate with Ramanujan,

$$E_4 = Q , \qquad E_6 = R .$$

The functions E_k are the q-expansions of modular forms of weight k. We have

$$Q = E_4 = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n$$

$$R = E_6 = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n$$

In particular, Q and R start with 1 and have integer coefficients, in their q-expansions.

We define

$$\Delta = \frac{1}{1728} (Q^3 - R^2) .$$

Then the q-expansion for Δ starts with q and has integral coefficients. One also has a q-product as follows.

Theorem 4.1. $$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} .$$

For the proof of this, we refer to any text on elliptic functions, e.g. mine, Chapter 18, § 4.

For any field F we can consider the polynomial algebra in two variables

$$F[Q, R]$$

as a graded algebra, namely we define an element to have **weight** k if it can be expressed as a linear combination of monomials

$$Q^a R^b$$

with $4a + 6b = k$. Then

$$F[Q, R] = \bigoplus_k F[Q, R]_k$$

is the direct sum of the graded components of weight k.

When $F = \mathbf{C}$, then M_k consists precisely of the elements of weight k as modular forms.

The elements of a graded component are also said to be **homogeneous**, or **weighted**, or **graded**.

We want to return to the considerations of Chapter I, § 2 concerning the algebra of modular forms, but this time we wish to pay attention to the coefficients. We can therefore refine the preceding result as follows. We let

$$\mathbf{Z}_{2,3} = \mathbf{Z}[\tfrac{1}{2}, \tfrac{1}{3}] .$$

Theorem 4.2. *Let k be even and let*

$$f = \sum a_n q^n$$

be a modular form of weight k. Let $r(k) = \dim M_k$. Let

$$A = \mathbf{Z}_{2,3}\langle a_0, \ldots, a_{r(k)-1} \rangle$$

be the module generated over $\mathbf{Z}_{2,3}$ by the first $r(k)$ coefficients of f. Then all coefficients a_n lie in A, and there exist elements $c_{a,b} \in A$ such that

$$f = \sum_{4a+6b=k} c_{a,b} Q^a R^b ,$$

in other words f is a polynomial of weight k in Q, R with coefficients in A.

Proof. We make the same induction as in the proof of the original theorem, but paying attention to the coefficients. Since

$$\dim M_k = 0 \text{ or } 1$$

for $k < 12$, the assertion is obvious in this case. Let $k \geqslant 12$. Write

$$k = 4r + 6s$$

with some integers r, $s \geqslant 0$. Then $f - a_0 Q^r R^s$ is a modular form of weight k, vanishing at infinity, and therefore divisible by Δ, that is

$$f - a_0 Q^r R^s = \Delta g ,$$

where g has weight $k - 12$. Since the q-expansion for Δ starts with q and has integer coefficients, it is clear that if

$$g = \sum b_n q^n$$

then the coefficients b_n lie in the module generated over $\mathbf{Z}_{2,3}$ by the Fourier coefficients for f. By induction it follows already that f can be written as a polynomial in Q, R with coefficients in the module

$$\mathbf{Z}_{2,3}\langle a_0, a_1, a_2, \dots \rangle .$$

Looking a little closer at the situation yields the more accurate statement concerning the coefficients. Indeed, let

$$f - a_0 Q^r R^s = \sum a'_n q^n = a'_1 q + a'_2 q^2 + \cdots$$

$$\Delta^{-1}(f - a_0 Q^r R^s) = g = b_0 + b_1 q + \cdots$$

We have $\dim M_{k-12} = \dim M_k - 1$, and by induction, it follows that g is a polynomial in Q, R with coefficients in

$$\mathbf{Z}_{2,3}\langle b_0, \dots, b_{r(k)-2} \rangle .$$

However b_v (for $v = 0, \dots, r(k)-2$) lies in

$$\mathbf{Z}_{2,3}\langle a'_1, \dots, a'_{r(k)-1} \rangle$$

and each a'_v lies in A. This proves the theorem.

We needed denominators involving 2, 3 in order to get polynomials in Q and R. However, if we allow Δ, then we can get a basis for M_k over $\mathbf{Z}$.

Theorem 4.3. *Let $M_k(\mathbf{Z})$ be the set of modular forms of weight k,*

$$f = \sum a_n q^n$$

with $a_n \in \mathbf{Z}$. Then $M_k(\mathbf{Z})$ has a $\mathbf{Z}$-basis, which is also a $\mathbf{C}$-basis for $M_k(\mathbf{C})$ as follows:

(i) $k \equiv 0 \pmod 4$: $Q^a \Delta^b$ *with* $4a + 12b = k$

(ii) $k \equiv 2 \pmod 4$: $RQ^a \Delta^b$ *with* $4a + 12b = k - 6$.

Proof. The proof follows the same pattern as above and will be left to the reader.

In fact, I owe the following statement to the first lemma of Victor Miller's thesis. victor@ccr-p.ida.org

Theorem 4.4. *Let $M_k^0(\mathbf{C})$ be the $\mathbf{C}$-vector space of cusp forms of weight k over $\mathbf{C}$, and let r be its dimension. There exists a basis $\{f_1, \dots, f_r\}$ of $M_k^0(\mathbf{C})$ with*

q-expansion coefficients $a_i(f_j)$ such that

$$a_i(f_j) = \delta_{ij}, \qquad 1 \leq i, j \leq r ,$$

and $a_n(f_j) \in \mathbf{Z}$ for all n and $j = 1, \dots, r$.

Proof. Let E_4 and E_6 be as usual. Since

$$B_4 = -1/30 \quad \text{and} \quad B_6 = 1/42 ,$$

we note that E_4 and E_6 have their q-expansion coefficients in $\mathbf{Z}$, and the power series start with 1. Pick non-negative integers a, b such that

$$14 \geq 4a + 6b \equiv k \pmod{12} ,$$

with $a = b = 0$ when $k \equiv 0 \pmod{12}$, and let

$$g_j = \Delta^j E_6^{2(r-j)+a} E_4^b , \qquad \text{for} \quad j = 1, \dots, r .$$

Then clearly

$$a_j(g_j) = 1, \qquad a_i(g_j) = 0 \quad \text{when} \quad i < j .$$

Hence the g_j are linearly independent over $\mathbf{C}$, and thus form a basis for $M_k^0(\mathbf{C})$. Since E_4, E_6 and Δ are all power series in q with coefficients in $\mathbf{Z}$, so are the g_j. The f_i may be constructed from the g_i by a straightforward elimination. The coefficients of f_i are in $\mathbf{Z}$ because the determinant of $r \times r$ matrix of leading coefficients of the g_j is 1. This proves the theorem.

§ 5. Derivatives of Modular Forms

We define the Dedekind **eta function**

$$\eta = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n) .$$

For $q = e^{2\pi i \tau}$, the function η is holomorphic on the upper half plane.

Theorem 5.1. *The eta function satisfies*

$$\eta(\tau + 1) = e^{2\pi i/24} \eta(\tau)$$

$$\eta(-1/\tau) = \sqrt{\tau/i}\, \eta(\tau).$$

Proof. The first relation is trivial from the q-product. As for the second, we know that Δ as a function of lattices, is homogeneous of degree -12, and therefore

$$\Delta(-1/\tau) = \tau^{12} \Delta(\tau) .$$

Taking the 24-th root shows that

$$|\eta(-1/\tau)| = |\sqrt{\tau}|\,|\eta(\tau)|\,.$$

Note that $\sqrt{\tau}$ is holomorphic on the upper half plane. By the maximum modulus principle, the function

$$\frac{\eta(-1/\tau)}{\sqrt{\tau}\,\eta(\tau)}$$

is constant. Putting $\tau = i$ shows that $1 = C\sqrt{i}$, whence $C = \sqrt{1/i}$. This proves the theorem.

We define the logarithmic derivative of η,

$$\frac{2\pi i}{24} P = \eta'/\eta\,.$$

Theorem 5.2. (i) *We have the q-expansion*

$$P = 1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n = 1 - \frac{4}{B_2}\sum_{n=1}^{\infty}\sigma_1(n)q^n\,.$$

(ii) *P satisfies the functional equation*

$$P(\gamma\tau)(c\tau+d)^{-2} = P(\tau) + \frac{12c}{2\pi i}(c\tau+d)^{-1}$$

for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *in* $SL_2(\mathbf{Z})$.

Proof. The first statement comes by taking the logarithmic derivative of the q-product for the eta function. The second statement is verified by induction on the length of a "word" in $SL_2(\mathbf{Z})$, writing an element γ as a product of elements

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

For the particular elements S, T the functional equation follows at once from that of the eta function in Theorem 5.1. The inductive step is trivial and left to the reader.

Symbolically we could write $P = E_2$, but we note that P is not a modular form. There is no modular form of weight 2 other than 0.

The power series for P again has integer coefficients, and begins with 1, so that it can be used in ways similar to Q, R for arithmetic considerations over $\mathbf{Z}$.

We define the differential operators

$$\theta = q\frac{d}{dq} = \frac{1}{2\pi i}\frac{d}{d\tau}$$

and for a given weight k, we let

$$\partial = \partial_k = 12\theta - kP\,.$$

We note that θ and ∂ operate on the power series in q, and as such have integral coefficients.

For an arbitrary matrix

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{R})$$

we recall the operation:

$$f \circ [\alpha]_k(\tau) = f(\alpha\tau)(c\tau+d)^{-k}\,.$$

Lemma. *Let f be holomorphic on the upper half plane. Then*

$$\partial(f\circ[\alpha]_k) = (\partial f)\circ[\alpha]_{k+2}\,.$$

Proof. Immediate from the definitions, and the functional equation for P given in Theorem 5.2, (ii).

Theorem 5.3. (i) *The operator ∂_k maps M_k into M_{k+2} and is a derivation on M_k.*

(ii) *We have*

$$\partial_2 P = -Q - P^2\,, \qquad \partial_4 Q = -4R\,, \qquad \partial_6 R = -6Q^2$$

(iii) $$\theta P = \tfrac{1}{12}(P^2 - Q)\,, \qquad \theta Q = \tfrac{1}{3}(PQ - R)\,, \qquad \theta R = \tfrac{1}{2}(PR - Q^2)\,.$$

Proof. Observe that P is not a modular form, but

$$\partial_2 P + P^2 = (12\theta - P)P = -Q$$

is a modular form. This is immediately implied by the lemma and the transformation law for P in Theorem 5.2 (ii). That the value is $-Q$ comes from the fact that the space of modular forms of weight 4 has dimension 1, whence any modular form of weight 4 is a constant multiple of Q, and this constant is determined

trivially from the q-expansion. The same argument works for the other two relations of (ii). The relations of (iii) are mere rephrasings of (ii) and the definitions.

Suppose that an element $f \in M_k$ is expressed as a product

$$f = gh ,$$

where g has weight m and h has weight n. Then

$$\hat{c}_k = 12\theta - (m+n)P .$$

A direct application of the definitions shows that $\hat{c}_k$ acts like a derivation, i.e.

$$\hat{c}_k(gh) = \hat{c}_m g \cdot h + g \cdot \hat{c}_n h .$$

This proves the theorem.

Most of the time we omit the subscript on $\hat{c}$, viewing $\hat{c}$ as operating on the direct sum of the space M_k, in a graded way.

§ 6. Reduction mod p

Let A be a subring of the complex numbers and let $M_k(A)$ be the subset of those elements $f = \sum a_n q^n$ in M_k having coefficients $a_n \in A$. Let

$$M(A) = \sum M_k(A) .$$

The algebra

$$A[P, Q, R]$$

is embedded in the power series ring $A[[q]]$. We may view the derivation $\theta = q\frac{d}{dq}$ on this power series ring as inducing a derivation on the polynomial algebra $A[P, Q, R]$.

If $\mathfrak{p}$ is a prime ideal of A we have a homomorphism

$$A[[q]] \to \bar{A}[[q]]$$

which is reduction mod $\mathfrak{p}$, namely

$$\sum a_n q^n \mapsto \sum \bar{a}_n q^n$$

where $\bar{a}_n = a_n \bmod \mathfrak{p}$. Then $\hat{c}$ induces an operator $\bar{\partial}$ on the reduced power series.

Similarly, we get a derivation on the polynomial ring

$$\bar{\partial} : \bar{A}[Q, R] \to \bar{A}[Q, R] .$$

We sometimes omit the bar over ∂ when applied to elements of

$$\bar{A}[Q, R]$$

since essentially no confusion can arise, especially if we work throughout an argument only with the reduced ring. If we alternate between the original ring and its reduction, then we keep the notation strict if we feel there is danger of confusion.

If $G(Q, R) \in A[Q, R]$ is a polynomial in the independent variables Q, R, we let

$$\bar{G}(Q, R) = \bar{G}$$

be the polynomial in $\bar{A}[Q, R]$ obtained by reducing the coefficients of G mod $\mathfrak{p}$. So if

$$G = \sum c_{a,b} Q^a R^b$$

then

$$\bar{G} = \sum \bar{c}_{a,b} Q^a R^b .$$

Warning. If $f = G(Q, R)$ we may very well have $\bar{f} = 0$ but $\bar{G} \neq 0$. We shall see important examples of this in a moment.

Also, notationally, we use $\bar{Q}, \bar{R}$ to denote the power series in $\mathbf{F}_p[[q]]$ obtained by reducing the power series for E_4 and E_6 mod p. The slight incompatibility of this notation with that for $\bar{G}$ does not make up for the convenience. The choice of letters (G instead of Q, R) will prevent confusion.

We let $\bar{M}(A)$ be the set of all power series

$$\bar{f} = \sum \bar{a}_n q^n$$

which are reductions of power series $f = \sum a_n q^n$ in $M(A)$. It is a module over $\bar{A}$. If

$$f = F(Q, R)$$

is a polynomial in Q, R with coefficients in A, and F is of weight k, so $f \in M_k(A)$, then

$$\bar{f} = \bar{F}(\bar{Q}, \bar{R}) ,$$

but of course, $\bar{f} \neq \bar{F}(Q, R)$. The result of Swinnerton-Dyer is to determine the structure of the ring $\bar{A}[\bar{Q}, \bar{R}]$ as a factor ring of $\bar{A}[Q, R]$. The cases when the characteristic is 2 or 3 are special, and we give this first.

Theorem 6.1. *Assume that $p=2$ or 3, $A=\mathbf{Z}_{(p)}$. Then*

$$\bar{P}=\bar{Q}=\bar{R}\,,$$

and

$$\overline{M}=\mathbf{F}_p[\bar{A}]\,.$$

Proof. This is immediate from the explicit power series expansions. Cf. also Theorem 4.3.

§ 7. Modular Forms mod p, $p\geqslant 5$

Throughout this section we let A be a local ring $\mathfrak{o}_\mathfrak{p}$ in a number field, with prime ideal $\mathfrak{p}$ dividing p. We let $M_k=M_k(A)$ be the set of modular forms of weight k having q-expansions

$$f=\sum a_n q^n$$

with coefficients $a_n\in A$. We let M be the direct sum of the M_k.

We first give various formulas holding for P, Q, R and their reductions mod p.

Theorem 7.1. *We have in $\mathbf{F}_p[[q]]$:*

(i) $\quad \bar{E}_{p-1}=1 \quad$ *and* $\quad \bar{E}_{p+1}=\bar{P}\,.$

(ii) $\quad \bar{\partial}\bar{E}_{p-1}=\bar{E}_{p+1} \quad$ *and* $\quad \bar{\partial}\bar{E}_{p+1}=-\bar{Q}\bar{E}_{p-1}\,.$

Proof. The Von Staudt congruence shows that

$$\bar{E}_{p-1}=1\,.$$

The Kummer congruence shows that $\bar{E}_{p+1}=\bar{P}$, since clearly $\sigma_p(n)\equiv\sigma_1(n)\pmod p$ for all n. Furthermore, we have

$$\partial E_{p-1}=12\theta E_{p-1}-(p-1)PE_{p-1}\,.$$

Reducing the power series mod p and using the first part of the theorem yields the relation

$$\partial E_{p-1}\equiv E_{p+1}\pmod p$$

as power series. Since $\partial P=-Q-P^2$, the last relation follows at once.

Instead of putting bars, we can also write the relations of Theorem 7.1 as congruences, e.g.

$$\partial E_{p+1}\equiv -QE_{p-1}\pmod p\,,$$

as power series.

On the other hand, by Theorem 4.2, we can write E_{p-1} and E_{p+1} as weighted polynomials in Q, R with p-integral coefficients, so we let

$$E_{p-1}=A(Q,R) \quad \text{and} \quad E_{p+1}=B(Q,R)\,.$$

Theorem 7.2. *We have in $\mathbf{F}_p[Q,R]$:*

$$\partial A(Q,R)\equiv B(Q,R) \quad \textit{and} \quad \partial B(Q,R)\equiv -QA(Q,R)\pmod p\,,$$

or in other words,

$$\bar{\partial}\bar{A}=\bar{B} \quad \textit{and} \quad \bar{\partial}\bar{B}=-\bar{Q}\bar{A}\,.$$

Proof. By definition,

$$\partial A(Q,R)-B(Q,R)=\partial E_{p-1}-E_{p+1}\,.$$

But

$$E_{p+1}=1-\frac{2(p+1)}{B_{p+1}}\sum\sigma_p(n)q^n$$

and by Kummer's congruences, we have

$$\frac{B_{p+1}}{p+1}\equiv\frac{B_2}{2}=\frac{1}{12}\pmod p\,.$$

Since $\bar{\partial}\bar{E}_{p-1}=0$, we find

$$\begin{aligned}E_{p-1}&=12\theta E_{p-1}-(p-1)PE_{p-1}\\&\equiv P\pmod p\\&\equiv E_{p+1}\pmod p\,.\end{aligned}$$

Then

$$\partial E_{p-1}-E_{p+1}$$

is a modular form in M_{p+1}, whose q-expansion coefficients are congruent 0 (mod p). By Theorem 4.2, it follows that

$$\partial A(Q,R)-B(Q,R)\equiv 0\pmod p\,,$$

as was to be shown, for the first relation.

Similarly, $$\partial E_{p+1} + QE_{p-1}$$

is in M_{p+3} and

$$\bar{\partial}\bar{E}_{p+1} = (12\theta - (p+1)\bar{P})\bar{B}(\bar{Q}, \bar{R}) = (12\theta - \bar{P})\bar{P} = -\bar{Q} .$$

Hence in $\mathbf{Z}_{(p)}[[q]]$ we get

$$\partial E_{p+1} + QE_{p-1} \equiv 0 \pmod{p} ,$$

whence by the same q-expansion principle we get in $\mathbf{Z}_{(p)}[Q, R]$

$$\partial B(Q, R) - QA(Q, R) \equiv 0 \pmod{p} ,$$

thereby proving the theorem.

Let $\mathbf{F}$ be the algebraic closure of the prime field $\mathbf{F}_p$. Then:

(i) *Any graded element of* $\mathbf{F}[Q, R]$ *can be written uniquely as a product of irreducible graded elements.*
(ii) *The irreducible graded elements are precisely*

$$Q, R, Q^3 - \alpha R^2 \quad \textit{with} \quad \alpha \in \mathbf{F} .$$

Proof. Write $Q = X^4$ and $R = Y^6$. Then a graded element becomes a function of X, Y, namely

$$G(Q, R) = G^*(X, Y) .$$

Suppose that there is a linear factor

$$G^*(X, Y) = (X - \lambda Y)H(X, Y)$$

for G^* in $\mathbf{F}[X, Y]$. If ζ is a 4-th root of unit and ω is a 3rd root of unity, then there is an automorphism $X \mapsto \zeta X$ and $Y \mapsto \omega Y$ of $\mathbf{F}[X, Y]$ over $\mathbf{F}[Q, R]$. Hence

$$\prod_{\zeta,\omega} (\zeta X - \lambda\omega Y) = X^{12} - \alpha Y^{12} = Q^3 - \alpha R^2$$

must be a factor of $G(Q, R)$ by unique factorization in $\mathbf{F}[X, Y]$. Our assertions are then obvious.

Lemma. (i) *$Q^3 - R^2$ does not divide $\bar{A}$.*

(ii) *If $I = Q^3 - \alpha R^2$ with $\alpha \neq 1$, then $I \nmid \partial I$.*

Proof. We have the q-expansion

$$A(Q, R) = 1 + \text{terms divisible by } p$$
$$Q^3 - R^2 = q + \cdots$$

so the first assertion is clear. As for the second,

$$\partial(Q^3 - \alpha R^2) = 3Q^2(-4R) - \alpha 2R(-6Q^2) = 12Q^2R(\alpha - 1) .$$

Hence if $\alpha \neq 1$ we get $\partial I \neq 0$, and the second assertion follows.

Theorem 7.3. *Let as before $E_{p-1} = A(Q, R)$. Then:*
(i) *$\bar{A}$ has no multiple irreducible factor, and is relatively prime to $\bar{B}$.*
(ii) *$\bar{A} - 1$ is absolutely irreducible.*

Proof. Suppose that

$$\bar{A} = I^m J$$

where I is irreducible, and $I \nmid J$. Note that $m \leq (p-1)/4$, so that $m \not\equiv 0 \pmod{p}$. We must show $m = 1$. We get

$$\partial\bar{A} = mI^{m-1}\partial IJ + I^m\partial J$$

whence

$$\bar{B} = I^{m-1}K_1$$

for some polynomial K_1, and $I \nmid K_1$ because $I \nmid \partial I$ by the lemma. If $m \geq 2$, it then follows that

$$\partial\bar{B} = -Q\bar{A} = I^{m-2}K_2$$

with some polynomial K_2 not divisible by I, which is a contradiction. Hence $m = 1$ and $I \nmid \bar{B}$, thereby proving (i).

For the absolute irreducibility, suppose that we have a factorization

$$\bar{A} - 1 = (G_0 + \cdots + G_m)(H_0 + \cdots + H_n)$$

where G_i, H_i have weight i. Then

$$G_mH_n = \bar{A} \qquad \text{and} \qquad G_0H_0 = -1 .$$

Further,

$$G_mH_{n-1} + G_{m-1}H_n = 0$$
$$G_mH_{n-2} + G_{m-1}H_{n-1} + G_{m-2}H_n = 0$$
$$\cdots$$

Since G_m, H_n are relatively prime, we get $G_m \mid G_{m-1}$ whence $G_{m-1}=0$. Continuing similarly, we end up with $G_0=0$, a contradiction which concludes the proof of the theorem.

Theorem 7.4. *The kernel of the homomorphism*

$$\mathbf{F}[Q, R] \to \mathbf{F}[\bar{Q}, \bar{R}]$$

is the prime ideal $(\bar{A}-1)$.

Proof. Let $\mathfrak{p}$ be the kernel. Since $\mathbf{F}[\bar{Q}, \bar{R}] \neq \mathbf{F}$ it follows that $\mathfrak{p}$ is not maximal. Since $\mathbf{F}[\bar{Q}, \bar{R}]$ is contained in $\mathbf{F}[[q]]$, it is an integral domain and the kernel is prime. Also the kernel contains $\bar{A}-1$. Since $\bar{A}-1$ is prime, we must have the equality by standard dimension theory, as was to be shown.

Let $f \in M_k$. It may happen that there exists some $k' < k$ such that

$$\bar{f}=\bar{g}$$

with $g \in M_{k'}$. We define the **filtration** $w(\bar{f})$ to be the smallest integer $k' \leqslant k$ such that $\bar{f}=\bar{g}$. Thus the function w is defined for elements which are reductions mod p of modular forms.

Theorem 7.5. *Let* $f \in M_k$, *and write* $f=F(Q, R)$ *where* F *is a polynomial of weight* k.
(i) $w(\bar{f})<k$ *if and only if* $\bar{A}$ *divides* $\bar{F}$.
(ii) *If* f *has weight* k *and* g *has weight* k' *and*

$$\bar{f}=\bar{g} \neq 0 ,$$

then $k \equiv k' \pmod{p-1}$.

Proof. Suppose that $w(\bar{f})<k$. Let $g=G(Q, R)$ where G has weight k' and $\bar{f}=\bar{g}$, but $k'<k$. By the preceding theorem, we have the divisibility

$$\bar{F}-\bar{G}=(\bar{A}-1)\bar{H}$$

for some polynomial $\bar{H}$, and

$$\bar{F}=\bar{A}\bar{H}+\bar{G}-\bar{H} .$$

Hence $\bar{F}=\bar{A}\bar{H}_0$, where $\bar{H}_0$ is the graded term of $\bar{H}$ of maximal weight, thus proving the divisibility. Conversely, if $\bar{F}=\bar{A}\bar{H}$, with some graded polynomial $\bar{H}$ of weight $<k$, we let

$$h=H(Q, R) ,$$

and then $\bar{f}=\bar{h}$, so the first assertion is proved. The second follows at once.

In view of the last theorem, for each residue class $\alpha \bmod p-1$ we let

$$\overline{M}^{\alpha}=\sum_{k \equiv \alpha} \overline{M}_k$$

be the sum over all reductions of M_k for k in the given congruence class mod $p-1$. We obtain:

Corollary. *The decomposition*

$$\overline{M}=\sum \overline{M}^{\alpha}$$

is a direct sum decomposition, and $\overline{M}$ *is a graded algebra. i.e.*

$$\overline{M}^{\alpha}\overline{M}^{\beta} \subset \overline{M}^{\alpha+\beta} ,$$

if α, β *are residue classes* mod $p-1$. *In other words,* $\overline{M}$ *is a graded algebra, graded by* $\mathbf{Z}/(p-1)\mathbf{Z}$.

Of course, the decomposition of $\overline{M}^{\alpha}$ as a sum of $\overline{M}_k$ is not direct. Only the decomposition of $\overline{M}$ in terms of the $\overline{M}^{\alpha}$ is direct.

Theorem 7.6. *The ring* $\overline{M}$ *is integrally closed.*

Proof. This is proved by showing with an explicit computation, refining Theorem 7.3, that the affine curve defined by the equation $\overline{A}-1=0$ is non-singular. We omit the details. See [Se 2] and [Se 3] as well as Katz [Ka 1].

§ 8. The Operation of θ on $\overline{M}$

We continue to assume $p \geqslant 5$.

We now analyse the effect of $\theta=\bar{\theta}=q\,d/dq$ on the algebra of modular forms reduced mod p. This entire section is due to Serre [Se 3], [Se 5] and Swinnerton Dyer [Sw D].

From the definition

$$\partial=12\theta-kP$$

and the fact that $\bar{P}=\bar{E}_{p+1}$ (Theorem 7.1) we conclude that

$$\theta=\bar{\theta}\colon \overline{M}^{\alpha} \to \overline{M}^{\alpha+2}$$

acting as a derivation on $\overline{M}$ changes the grading by 2.

Theorem 8.1. *Let $\tilde{f} \in \overline{M}_k$ and suppose that $w(\tilde{f}) = k$.*
(i) We have $w(\theta\tilde{f}) \leqslant w(\tilde{f}) + p + 1$, with equality if and only if $w(\tilde{f}) \not\equiv 0 \pmod{p}$.
(ii) If $w(\tilde{f}) \equiv 0 \pmod{p}$, then $w(\theta\tilde{f}) \leqslant w(\tilde{f}) + 2$.

Proof. We have

$$12\theta\tilde{f} = \tilde{\partial}\tilde{f} + k\tilde{P}\tilde{f} = \overline{\partial f} + k\,\overline{E_{p+1}}\overline{f} = \overline{E_{p-1}\partial f} + k\,\overline{E_{p+1}}\overline{f}\,.$$

Let $F(Q, R)$ be graded, of minimal weight $w(\tilde{f})$ such that $\overline{F}(\overline{Q}, \overline{R}) = \tilde{f}$. This last expression is the reduction of

$$A(Q, R)\partial F(Q, R) + kB(Q, R)F(Q, R)$$

which is graded. If $k \not\equiv 0 \pmod{p}$, then $\overline{A}$ does not divide

$$\overline{A}\partial\overline{F} + k\overline{B}\overline{F}$$

by Theorems 7.3 and 7.5, and hence $w(\theta\tilde{f}) = k + p + 1$.

On the other hand, suppose that $k \equiv 0 \pmod{p}$. Then it is immediately clear that the filtration is $\leqslant k + 2$. This proves the theorem.

In the next theorem, we give a special example of a power series which is *not* a modular form mod p, and which plays an important role in applications.

Theorem 8.2. *Let $k \equiv 0 \pmod{p-1}$. Let*

$$\bar{g} = \sum_{n=1}^{\infty} \bar{\sigma}_{k-1}(n)q^n \in \mathbf{F}_p[[q]]\,.$$

Then:
(i) $\bar{g}^p - \bar{g} = \frac{1}{24}\theta^{p-2}\overline{E}_{p+1}$, and $\bar{g}$ is integral over $\overline{M}$.
(ii) Further, $\bar{g} \neq \bar{\varphi}/\bar{\psi}$, where $\bar{\varphi} \in \overline{M}^{\beta}$, $\bar{\psi} \in \overline{M}^{\lambda}$ are graded elements of $\overline{M}$.

Proof. We have

$$\sigma_{k-1}(n) \equiv \sigma_{p-2}(n) \pmod{p}\,.$$

Then

$$\bar{g}^p - \bar{g} = \sum_{m=1}^{\infty} \bar{\sigma}_{k-1}(m)q^{mp} - \sum_{n=1}^{\infty} \bar{\sigma}_{k-1}(n)q^n\,.$$

If $n = pm$ then

$$\bar{\sigma}_{k-1}(pm) = \bar{\sigma}_{k-1}(m) \qquad \text{and} \qquad d^{k-1} \equiv d^{-1} \pmod{p}$$

if $p \nmid d$. Hence

$$\begin{aligned}\bar{g}^p - \bar{g} &= -\sum_{p|n} \bar{\sigma}_{k-1}(n)q^n \\ &= -\theta^{p-2}\sum \bar{\sigma}_1(n)q^n \\ &= \tfrac{1}{24}\theta^{p-2}\overline{E}_{p+1}\,.\end{aligned}$$

This proves the first assertion.

To prove the second, we compute filtrations. We have $w(\overline{E}_{p+1}) = p+1$ because there is no modular form of weight 2. Also

$$p + 1 + v(p+1) \equiv 1 + v \not\equiv 0 \pmod{p}$$

if $0 \leqslant v \leqslant p-3$. So by Theorem 8.1 (i),

$$w(\theta^{p-2}\overline{E}_{p+1}) = p + 1 + (p-2)(p+1) = p^2 - 1\,. \tag{1}$$

Next we show that $\beta = \lambda$. We have

$$\bar{\varphi}^p - \bar{\varphi}\bar{\psi}^{p-1} = \tfrac{1}{24}\bar{\psi}^p\theta^{p-2}\overline{E}_{p+1}\,.$$

The right hand side lies in a graded component of $\overline{M}$, and each term on the left hand side is in a graded component. Hence these gradings must be equal, so that

$$p\beta \equiv p\lambda + p^2 - 1 \pmod{p-1}\,.$$

This yields $\beta = \lambda$, so we now may assume that $\bar{\varphi}, \bar{\psi} \in \overline{M}^{\alpha}$ for some α.

We write $\bar{g}$ as a quotient

$$\bar{g} = \overline{G}(\overline{Q}, \overline{R})/\overline{H}(\overline{Q}, \overline{R})\,,$$

where $\overline{G}$, $\overline{H}$ are weighted, $\bar{\varphi} = \overline{G}(\overline{Q}, \overline{R})$ and $\bar{\psi} = \overline{H}(\overline{Q}, \overline{R})$. We may assume that $\overline{H}$ is relatively prime to $\overline{A}$. Otherwise, we multiply both $\overline{G}$ and $\overline{H}$ by a factor which makes $\overline{H}$ equal to a power of $\overline{A}$ times a factor prime to $\overline{A}$, and then replace $\overline{H}$ by this factor. We may then assume that $\overline{G}$ is not divisible by $\overline{A}$ by throwing away any power of $\overline{A}$ dividing $\overline{G}$.

With these reduction steps, we then have

$$w(\bar{\varphi}^p) = pw(\bar{\varphi}) \tag{2}$$

because $\overline{G}^p$ is also not divisible by $\overline{A}$ since $\overline{A}$ has no multiple factors. Furthermore, since $\overline{H}$ is prime to $\overline{A}$, we get

$$w(\bar{\psi}^p\tilde{f}) = w(\bar{\psi}^p) + w(\tilde{f}) \tag{3}$$

for any graded element f of $\overline{M}$. We use again the formula

$$\bar{\varphi}^p - \bar{\varphi}\bar{\psi}^{p-1} = \bar{\psi}^p f,$$

with $w(f) = p^2 - 1$, by (1). If $w(\bar{\varphi}) > w(\bar{\psi})$ then using (3),

$$w(\text{left hand side}) = pw(\bar{\varphi})$$
$$w(\text{right hand side}) = pw(\bar{\psi}) + p^2 - 1,$$

and we have a contradiction. If $w(\bar{\varphi}) \leqslant w(\bar{\psi})$ then

$$w(\text{left hand side}) \leqslant pw(\bar{\varphi}),$$

which is strictly smaller than the filtration of the right hand side, again giving a contradiction which concludes the proof of the theorem.

Serre's original proof that $\bar{g}$ is not a quotient of graded elements used the integral closure of $\overline{M}$, which as we have mentioned is a somewhat deeper fact. Ribet showed me how to use the standard filtration arguments to give the proof at the level of the present chapter.

Serre raised the question to give a geometric description of the covering defined by the function $\bar{g}$ over $\overline{M}$. This was answered by Katz [Ka 1], who shows that $\bar{g}$ is the Artin-Schreier generator of the separable part of the modular function field of level p^2, modulo p, over the modular function field of level p. In terms of elliptic curves, this means that $\bar{g}$ generates the separable part of the field of p^2 division points over the field of p division points, for the generic elliptic curve with invariant $\bar{j}$.

The next theorems show how to use the fact that $\bar{g}$ is not a graded element, or a quotient of graded elements of $\overline{M}$.

Theorem 8.3. *Let*

$$f = \sum a_n q_n$$

be a modular form of weight k with coefficients a_n in the quotient field of Λ, and suppose a_n is $\mathfrak{p}$-integral for $n \geqslant 1$.
(i) *If $k \not\equiv 0 \pmod{p-1}$ then a_0 is $\mathfrak{p}$-integral.*
(ii) *If $k \equiv 0 \pmod{p-1}$ then $\operatorname{ord}_{\mathfrak{p}} a_0 \geqslant -\operatorname{ord}_{\mathfrak{p}} k - \operatorname{ord}_{\mathfrak{p}} p$.*

Proof. Suppose that a_0 is not $\mathfrak{p}$-integral. Let

$$-m = \operatorname{ord}_{\mathfrak{p}} a_0 .$$

Let π be a prime element in $\mathfrak{p}$. Then

$$\pi^m f \equiv \pi^m a_0 \not\equiv 0 \pmod{\mathfrak{p}} .$$

We know that $\bar{E}_{p-1} = 1$. Hence by the fact that $\overline{M}$ is graded, i.e. Theorem 7.5 (ii), we conclude that

$$k \equiv 0 \pmod{p-1} .$$

Suppose that $m \geqslant \operatorname{ord}_{\mathfrak{p}} k + \operatorname{ord}_{\mathfrak{p}} p$. Let

$$G_k = -\frac{B_k}{2k} + \sum \sigma_{k-1}(n) q^n .$$

Denoting by $\sim$ the property that the quotient of the two sides is a p-unit, we have by Von Staudt,

$$\frac{B_k}{2k} \sim p^{-\operatorname{ord}_p k - 1},$$

and

$$\frac{B_k}{2ka_0} f \equiv \frac{B_k}{2k} \pmod{\mathfrak{p}}$$

and

$$G_k + \frac{B_k}{2ka_0} f \equiv \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n .$$

This contradicts the preceding theorem concerning $\bar{g}$, which appears on the right hand side.

Theorem 8.4. *Let*

$$f = \sum a_n q^n \quad \text{and} \quad f' = \sum a'_n q^n$$

be modular forms with $\mathfrak{p}$-integral coefficients for $n \geqslant 0$, of weights k, k' respectively. Assume:
(i) $k' \equiv k \not\equiv 0 \pmod{p-1}$
(ii) $a_n \equiv a'_n \pmod{\mathfrak{p}}$ for $n \geqslant 1$.
Then $a_0 \equiv a'_0 \pmod{\mathfrak{p}}$.

Proof. If $k' = k$, consider

$$\pi^{-1}(f' - f) .$$

The preceding theorem applies. If $k' \neq k$, say $k < k'$, and write

$$k' = k + v(p-1) .$$

We can replace f by fE_{p-1}^{v}. Then

$$fE_{p-1}^{v} \equiv f \pmod{\mathfrak{p}} .$$

Again this reduces the theorem to the previous case, and concludes the proof.

Theorem 8.5. *Let f, f' be modular forms of weights k, k' respectively, with* p-*integral coefficients. Assume*

$$f \not\equiv 0 \pmod{\mathfrak{p}} ,$$

and suppose that there is an integer $m \geqslant 1$ such that

$$f' \equiv f \pmod{\mathfrak{p}^m}$$

(i.e. as power series in $A[[q]]$). Then

$$k' \equiv k \bmod (p-1)p^{m-1} .$$

Proof. If $m=1$, the assertion is merely that of Theorem 7.5 which gives us a grading for $\bar{M}$. Suppose $m \geqslant 2$. Let

$$h = k' - k .$$

After replacing f' by $f'E_{(p-1)p^v}$ for large v, we may assume without loss of generality that $k < k'$ and $h \leqslant 4$. Let

$$h = (p-1)p^r h_0 , \quad p \nmid h_0 .$$

Then by Von Staudt,

$$\frac{2h}{B_h} \sim p^{r+1}$$

(we are assuming $p \geqslant 5$). We have to show that $r \geqslant m-1$. Suppose that $r < m-1$. We write

$$fE_h - f' = f - f' + f(E_h - 1)$$

and

$$E_h - 1 = p^{r+1} u \sum_{n=1}^{\infty} \sigma_{h-1}(n) q^n ,$$

where u is a p-unit. Also $f - f' \equiv 0 \pmod{p^m}$. Then

$$u^{-1}p^{-(r+1)}(fE_h - f') \equiv f \sum_{n=1}^{\infty} \sigma_{h-1}(n) q^n \pmod{\mathfrak{p}} .$$

Let

$$\varphi = u^{-1}p^{-(r+1)}(fE_h - f')$$

$$g = \sum_{n=1}^{\infty} \sigma_{h-1}(n) q^n .$$

Then $\bar{g} = \bar{\varphi}/\bar{f}$, and $\bar{\varphi}$, $\bar{f}$ are graded, contradicting Theorem 8.2 concerning $\bar{g}$, and proving our theorem.

Remark. The statement of the theorem remains true for $p=3$, and for $p=2$, Serre shows that

$$k' \equiv k \pmod{2^{m-2}} .$$

Cf. [Se 5].

The preceding theorem is the beginning of Serre's theory of p-adic modular forms. Taking the limit over m implies that p-adic limits of modular forms have a p-adic weight in

$$\lim \mathbf{Z}/(p-1)p^{m-1}\mathbf{Z} .$$

Chapter XI. Galois Representations

Perhaps the most fascinating connection between modular forms and number theory is the way in which they are connected with the existence of non-abelian extensions. Langlands was the first to interpret ordinary class field theory in this vein, pointing out that characters of the Idele class group can be viewed as modular forms on GL_1, and suggesting generalizations to modular forms on GL_2, and higher dimensional groups. Cf. [Lgds 1], [Lgds 2], where the reader will find quite general conjectures phrased in terms of representation theory.

Independently, Shimura [Sh 3] worked out a special case, starting with a different view point, establishing a connection between coefficients of certain modular forms, and the traces of Frobenius elements in extensions K of $\mathbf{Q}$ whose Galois group has a representation in $GL_2(\mathbf{F}_l)$, and K is the field of l-division points of the curve $X_0(11)$, or in the Jacobian of $X_1(N)$, Theorem 7.14 of [Sh 2].

Serre [Se 2] also made conjectures, and pointed out the connection of classical congruence relations for coefficients of modular forms, with respect to certain extensions having GL_2 representations. He also proved the first results known to the effect that extensions coming from division points of elliptic curves without complex multiplication are as big as could be apriori expected [Se 4] and [Se 8]. Serre's conjectures on the existence of certain extensions associated with modular forms were proved by Deligne [De 2].

For modular forms of weight 1, Deligne and Serre [De-Se] proved a basic theorem associating to such forms a finite extension of the rationals. Leaving out more technical considerations of conductor, we state their result as follows.

Let $f=\sum a_n q^n$ be a form of weight 1 on $\Gamma_1(N)$, and assume that f is a normalized eigenfunction of the Hecke algebra, with Dirichlet character ε such that $\varepsilon(-1)=-1$. Then there exists a finite Galois extension K of $\mathbf{Q}$ with group G, and a representation

$$\rho: G \to GL_2(\mathbf{C})$$

such that, for a Frobenius element σ_p with $p \nmid N$, the characteristic polynomial of $\rho(\sigma_p)$ is

$$X^2 - a_p X + \varepsilon(p) .$$

As for the connection between modular forms and elliptic curves, Weil conjectured that any elliptic curve over the rationals is the rational image of a modular curve. Cf. [We 1]. This can be formulated in terms of modular forms of weight 2.

In fact, as Serre pointed out, that the zeta function of an elliptic curve over the rationals should be the Mellin transform of a cusp form of weight 2 was already conjectured by Taniyama, in a paper distributed at the number theory conference in Tokyo, 1955. A substantial insight into this situation was of course provided by Weil's characterization of Dirichlet series with functional equations [W 2].

In this chapter, we give results of Swinnerton Dyer [SwD 1] and Serre [Se 5] showing the connection between congruence properties of modular forms, and Frobenius elements in certain associated Galois extensions, whose existence was proved by Deligne. Although these are special, they give a good introduction to the subject, because they are in a sense typical.

The representations of the Galois group in this chapter occur in $GL_2(\mathbf{F}_l)$. Serre began an extensive theory of those representations in $GL_2(\mathbf{Z}_l)$ [Se 4], [Se 8]. For these l-adic representations, it is a problem to describe "reciprocity laws" concerning the distribution of Frobenius elements. Lang–Trotter [L–T] give a conjecture concerning the asymptotic distribution of such elements satisfying various number theoretic conditions.

Let F be a field. We let P in front of GL or SL denote the group obtained by factoring out the scalar matrices, i.e. the center of these groups. Note that the scalar matrices of SL consists only of ± 1. We are interested in the subgroups of $GL_2(F)$, and their images in $PGL_2(F)$. We begin with the normal subgroups, and prove that $PSL_2(F)$ is simple if F has at least four elements. In § 2, we classify all finite subgroups.

§ 1. Simplicity

A **Borel subgroup** of GL_2 or SL_2 is a subgroup which is conjugate to **the standard Borel subgroup** consisting of all matrices

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix},$$

(with $d=a^{-1}$ in the case of SL_2). We let U be the group of matrices

$$u(b)=\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \qquad b \in F .$$

We let A be the group of diagonal matrices, $\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$. We let

$$s(a)=\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \qquad a \in F^* .$$

We let

$$w=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

For the rest of this section we let $G = SL_2(F)$, or $GL_2(F)$.

Lemma 1. *The matrices*

$$X(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad Y(c) = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

generate $SL_2(F)$.

Proof. Multiplying an arbitrary element of $SL_2(F)$ by matrices of the above type on the right and on the left corresponds to elementary row and column operations (e.g. adding a scalar multiple of a row to the other, etc.). Thus a given matrix can always be brought into a form

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

by such multiplications. Let $W(a) = X(a)Y(-a^{-1})$. We get

$$W(a)W(-1) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix},$$

thereby proving the lemma.

If we let $\bar{U}$ be the group of lower matrices,

$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

then we see that

$$wUw^{-1} = \bar{U}.$$

Also note the commutation relation

$$w\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}w^{-1} = \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix},$$

so that w normalizes A. Similarly,

$$wBw^{-1} = \bar{B}$$

is the group of lower triangular matrices.

We note that

$$B = AU = UA,$$

and also that A normalizes U.

There is a decomposition of G into disjoint subsets,

$$G = B \cup BwB.$$

Indeed, view G as operating on the left of column vectors. The isotropy group of

$$e^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

is obviously U. The orbit Be^1 consists of all column vectors whose second component is 0. On the other hand,

$$we^1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

and therefore the orbit Bwe^1 consists of all vectors whose second component is $\neq 0$, and whose first component is arbitrary. Since these two orbits of B and BwB cover the orbit Ge^1, it follows the union of B and BwB is equal to G (because the isotropy group U is contained in B), and they are obviously disjoint. This decomposition is called the **Bruhat decomposition**.

The Borel subgroup B is a maximal proper subgroup.

Proof. By the Bruhat decomposition, any element not in B lies in BwB, so the assertion follows since B, BwB cover G.

Theorem 1.1. *If F has at least four elements, then $SL_2(F)$ is equal to its own commutator group, and $SL_2(F)/\pm 1$ is simple.*

Proof. We have the commutator relation (by matrix multiplication)

$$s(a)u(b)s(a)^{-1}u(b)^{-1} = u(ba^2 - b) = u(b(a^2 - 1)).$$

We let $G = SL_2(F)$ for this proof. We let G' be the commutator subgroup, and similarly let B' be the commutator subgroup of B. We prove the first assertion that $G = G'$. From the hypothesis that F has at least four elements, we can find an element $a \neq 0$ such that $a^2 \neq 1$, whence the commutator relation shows that $B' = U$. It follows that $G' \supset U$, and since G' is normal we get

$$G' \supset wUw^{-1}.$$

From Lemma 1 we conclude that $G' = G$.

We let Z denote the center of G. It consists of $\pm I$, that is $\pm$ the identity 2×2 matrix, if $G = SL_2(F)$ and is the subgroup of scalar matrices if $G = GL_2(F)$.

Lemma 2. *The intersection of all conjugates of B in G is equal to Z.*

Proof. We leave this to the reader, as a simple fact using conjugation with w.

Lemma 3. *Let $G=SL_2(F)$. If H is normal in G then either $H\subset Z$ or $H\supset G'$.*

Proof. By the maximality of B we must have

$$HB=B \qquad \text{or} \qquad HB=G\,.$$

If $HB=B$ then $H\subset B$. Since H is normal, we conclude that H is contained in every conjugate of B, whence in the center by Lemma 2. On the other hand, suppose that $HB=G$. Write

$$w=hb$$

with $h\in H$ and $b\in B$. Then

$$wUw^{-1}=\bar{U}=hbUb^{-1}h^{-1}=hUh^{-1}\subset HU$$

because H is normal. Since $U\subset HU$ and U, $\bar{U}$ generate $SL_2(F)$, it follows that $HU=G$. Hence

$$G/H=HU/H\approx U/(U\cap H)$$

is abelian, whence $H\supset G'$, as was to be shown.

As we had already proved that $G=G'$, we have also proved the simplicity of G/Z. Observe that Lemma 3 needs no assumption on the cardinality of F.

§ 2. Subgroups of GL_2

Let F be a field. We view $GL_2(F)$ as operating on the 2-dimensional vector space $V=F^2$. We denote by ${}^aF$ the algebraic closure of F, and by ${}^aV$ the extended vector space

$${}^aV=V\otimes {}^aF={}^aF^2\,.$$

By **semi-simple**, we always mean absolutely semisimple, i.e. semisimple over the algebraic closure. For our purposes, it suffices to deal with the separable closure ${}^sF$ and ${}^sV$.

Let K be a separable quadratic extension of F. Let $\{u_1, u_2\}$ be a basis of K. Then we have the regular representation of K with respect to this basis, representing K^* as a subgroup of $GL_2(F)$. The elements of norm 1 correspond precisely to the elements of $SL_2(F)$ in the image of K^*. A different choice of bases of K corresponds to conjugation of this image in $GL_2(F)$. Let C_K denote one of these images.

Then $C_K=C$ is called a **non-split Cartan** subgroup. The subalgebra

$$F[C_K]\subset \mathrm{Mat}_2(F)$$

is isomorphic to K itself, and the units of this algebra are therefore the elements of $C_K\approx K^*$.

The subgroup C_K is a maximal commutative semi-simple subgroup.

Proof. If $\alpha\in GL_2(F)$ commutes with all elements of C_K then α lies in $F[C_K]$, whence in C_K itself as we have just seen.

By the **split Cartan** subgroup we mean the group of diagonal matrices

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

By a **Cartan subgroup** we mean a subgroup conjugate to the split Cartan subgroup or one of the subgroups C_K above.

Every maximal commutative semisimple subgroup of $GL_2(F)$ is a Cartan subgroup, and conversely.

Proof. It is clear that the split Cartan subgroup is maximal commutative semisimple. Suppose that H is a maximal commutative semisimple subgroup of $GL_2(F)$. If H is diagonalizable over F, then H is contained in a conjugate of the split Cartan. On the other hand, suppose H is not diagonalizable over F. It is diagonalizable over the separable closure of F, and the two eigenspaces of dimension 1 give rise to two characters

$$\psi,\ \psi' : H\to {}^sF^*$$

of H in the multiplicative group of the separable closure. For each element $\alpha\in H$, the values, $\psi(\alpha)$, $\psi'(\alpha)$ are the eigenvalues of α, and for some element $\alpha\in H$ these eigenvalues are distinct, otherwise H is diagonalizable over F. Hence the pair of elements

$$\psi(\alpha),\ \psi'(\alpha)$$

are conjugate over F. The image $\psi(H)$ is cyclic, and if $\psi(\alpha)$ generates this image then we see that $\psi(\alpha)$ generates a quadratic extension K of F. The map

$$\alpha\mapsto\psi(\alpha)\,, \qquad\qquad \alpha\in H\,,$$

extends to an F-linear mapping, also denoted by ψ, of the algebra $F[H]$ into K. Since $F[H]$ is semisimple, it follows that

$$\psi : F[H]\to K$$

is an isomorphism. Hence ψ maps H into K^*, and in fact maps H onto K^* because H was taken to be maximal. This proves our assertion.

In the above proof, the two characters ψ, ψ' are called the **characters of the Cartan subgroup**. In the split case, if α has diagonal elements a, d then we get the two characters ψ, ψ' such that

$$\psi(\alpha)=a \qquad \text{and} \qquad \psi'(\alpha)=d\,.$$

In the split case, the values of the characters are in F. In the non-split case, these values are conjugate quadratic over F, and lie in K.

Theorem 2.1. *Let C be a Cartan subgroup of $GL_2(F)$ and N its normalizer. Then C is of index 2 in N.*

Proof. An element of the normalizer of a Cartan subgroup must either fix or interchange the eigenspaces. If it fixes them, then it lies in C by the maximality of C. If it interchanges them, then it does not lie in C, and generates a unique coset of N/C, so that C is of index 2 in N.

In the split case, a representative of N/C which interchanges the eigenspaces is given by

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In the non-split case, let C_K be as above, the Cartan subgroup associated with a separable quadratic extension K of F, and let

$$\sigma : K \to K$$

be the non-trivial automorphism. Let $\{u, u^\sigma\}$ be a normal basis. With respect to this basis, the matrix of σ is precisely the same

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore again in this case we see that there exists a non-trivial element in the normalizer of C. Note that it is trivial to verify the relation

$$M(\sigma)M(x)M(\sigma^{-1})=M(x^\sigma)\,,$$

if $M(x)$ is the matrix associated with an element $x \in K$

Since the order of an element in the multiplicative group of a field is prime to the characteristic we conclude:

If F has characteristic l, then an element of finite order in $GL_2(F)$ is semisimple if and only if its order is prime to l.

Proof. Let $\alpha \in GL_2(F)$. Its Jordan normal form over the algebraic closure is of type

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

The eigenvalue a lies in a finite field, and $a \neq 0$. The assertion is then obvious. We also see:

If the order of α is divisible by l, then the Jordan normal form of α is

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$$

and α can be triangularized over F.

Theorem 2.2. *Let G be a subgroup of $GL_2(\mathbf{F}_l)$. If the order of G is divisible by l, then either G is contained in a Borel subgroup, of $GL_2(\mathbf{F}_l)$, or G contains $SL_2(\mathbf{F}_l)$.*

Proof. Let α be an element of order l in G. We have just seen that α can be triangularized over F.

Consequently α has a one-dimensional eigensubspace W of V. If every element of G has W as eigenspace, then G is contained in the associated Borel subgroup. If not, let $\sigma \in G$ map W into another one-dimensional subspace W', so that $V = W \oplus W'$. Using basis elements of W, W' as a basis for V, we see that α and $\sigma\alpha\sigma^{-1}$ can be represented by the matrices

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}.$$

These generate $SL_2(\mathbf{F}_l)$, and the theorem follows.

Next we consider the case of a subgroup G of $GL_2(F)$, when F has characteristic l (not necessarily a finite field), and the order of G is prime to F. The next lemma shows that eigenspaces can be considered in pairs.

Lemma. *Let G be a subgroup of $GL_2(F)$, where F is a field of characteristic l. Suppose that the order of G is prime to l. If two elements of G have one eigenvector in common, then they have both eigenvectors in common.*

Proof. The two elements can be simultaneously triangularized. Their commutator is then an element with 1 on the diagonal, and must therefore be the identity, otherwise it would have period l.

Let G be a finite subgroup of $GL_2(F)$, of order prime to l. We shall consider the image of G in the projective linear group,

$$G \to GL_2(F) \to GL_2(F)/F^* ,$$

where F^* is identified with the group of scalar matrices, i.e. F^*I. We denote this image by PG, or H.

If G is contained in a Cartan subgroup, then PG is cyclic.

Proof. Let ψ, ψ' be the two characters of the Cartan subgroup (split or non-split), so that

$$\alpha \mapsto (\psi(\alpha), \psi'(\alpha))$$

is an injection of G in $K^* \times K^*$, where $K = F$ or K is quadratic over F. Let D be the diagonal in $K^* \times K^*$. Consider the composite homomorphism

$$G \to K^* \times K^* \to (K^* \times K^*)/D \approx K^* .$$

Its kernel is precisely the group of scalar matrices in G, so PG is embedded in K^*, and every finite subgroup of K^* is cyclic, as was to be shown.

We recall that a finite group G is called **dihedral** if it contains a cyclic subgroup G_0 of index 2, an element $\sigma \in G$ of order 2 not in G_0, such that $\sigma\alpha\sigma^{-1} = \alpha^{-1}$ for all $\alpha \in G_0$.

If G is contained in the normalizer of a Cartan subgroup but not in the Cartan subgroup, then PG is dihedral.

Proof. Let us give the proof in the case of a non-split Cartan. Any $\alpha \in G$ is conjugate over the quadratic extension K to the diagonal matrix of its eigenvalues,

$$\begin{pmatrix} \psi & 0 \\ 0 & \psi^\sigma \end{pmatrix}$$

where σ is the non-trivial automorphism of K over F. Then α^σ is conjugate to

$$\begin{pmatrix} \psi^\sigma & 0 \\ 0 & \psi \end{pmatrix}.$$

In the projectivized group PG, the inverse

$$\begin{pmatrix} \psi^{-1} & 0 \\ 0 & \psi^{-\sigma} \end{pmatrix}$$

yields the same element as $P\alpha^\sigma$, because

$$\begin{pmatrix} \psi\psi^\sigma & 0 \\ 0 & \psi\psi^\sigma \end{pmatrix}$$

is a scalar matrix in F. This proves our assertion.

Theorem 2.3. *Let F be a field of characteristic l. Let G be a finite subgroup of $GL_2(F)$, of order prime to l. Let H be the image of G in $PGL_2(F)$. Then we have the following cases.*

(i) *H is cyclic and G is contained in a Cartan subgroup.*

(ii) *H is dihedral, and G is contained in the normalizer of a Cartan subgroup but not in the Cartan subgroup itself.*

(iii) *H is isomorphic to A_4, S_4, or A_5.*

Proof. The set of eigenspaces of non-trivial elements of H is finite and stable under H. Let

$$E_1, \ldots, E_r$$

be representatives of the orbits under H. Let h be the order of H, and let h_i be the order of the isotropy subgroup of E_i in H. Then the orbit of E_i has h/h_i elements. Counting the number of pairs (α, E) consisting of an element $\alpha \in H$, $\alpha \neq 1$, and an eigenspace E for α in two ways, we find the relation

$$2h - 2 = \frac{h}{h_1}(h_1 - 1) + \cdots + \frac{h}{h_r}(h_r - 1) .$$

We shall determine all solutions of this equation with positive integers h, h_i dividing h. The answer is given in the next lemma.

We analyse the relation, which can be written in the form

$$2\left(1 - \frac{1}{h}\right) = 1 - \frac{1}{h_1} + \cdots + 1 - \frac{1}{h_r}$$

where each h_i divides h, and $h \geqslant 2$.

Lemma. *The only solutions to the above equation are:*

(i) $r = 2$, $h_1 = h_2 = h$.

(ii) $r = 3$, h *even*, $h_1 = h_2$, $h_3 = h/2$.

(iii) $r = 3$, $h = 12$, $h_1 = 2$, $h_2 = 3$, $h_3 = 3$.

(iv) $r = 3$, $h = 24$, $h_1 = 2$, $h_2 = 3$, $h_3 = 4$.

(v) $r = 3$, $h = 60$, $h_1 = 2$, $h_2 = 3$, $h_3 = 5$.

Proof. Suppose some $h_i = h$. Then it is immediate that we are in case (i). We suppose from now on that $h_i < h$ for all i. We must then have $r > 2$, because

$$1 - \frac{1}{h} > 1 - \frac{1}{h_i}.$$

We cannot have $r \geqslant 4$ because otherwise the right-hand side has at least four terms, each of which is at least $1/2$, while the left-hand side is < 2. Therefore $r = 3$, which we assume from now on.

Not all $h_i \geqslant 3$, otherwise the right-hand side is $\geqslant 2$, which is impossible. Hence some $h_i = 2$, and in particular h is even.

Say $h_1 \leqslant h_2 \leqslant h_3$. If $h_3 = h/2$ then it is immediate that we are in case (ii). We therefore assume from now on that $h_3 < h/2$.

We have $h_1 = 2$. We cannot have $h_2 = 2$ also, for otherwise $h_3 = h/2$, which we already have excluded. On the other hand, we cannot have $h_2 \geqslant 4$, otherwise the right-hand side is > 2. Thus $h_2 = 3$, which we assume from now on.

So we have to consider $h_1 = 2$, $h_2 = 3$. If $h_3 \geqslant 6$ then the right-hand side is $\geqslant 2$, which is impossible. Therefore we must have $h_3 = 3$, 4, or 5, which takes care of all of our cases.

We shall now prove that the five cases of the lemma correspond precisely to the five cases of Theorem 2.3. (Note that (iii) in the theorem breaks up into three cases, A_4, S_4 and A_5.)

(i) All elements of H have the same eigenspaces, because in this case there are only two of them. Hence G is contained in the associated Cartan subgroup. Since H is obtained by projectivizing, it follows that H is cyclic.

(ii) In this case, the orbit HE_3 has 2 elements, and the isotropy group H_3 of E_3 in H has index 2 in H, and is normal in H (kernel of the representation as a permutation group of two elements). Let G_3 be the inverse image of H_3 in G. Then G_3 admits E_3 as eigenspace, and by the lemma, we conclude that G_3 is contained in the corresponding Cartan subgroup. It follows that H_3 is cyclic, so H is dihedral. Since

$$G/G_3 \approx H/H_3$$

we see that G permutes the two eigenspaces, and hence cannot be contained in the Cartan subgroup, but is contained in the normalizer.

(iii) $H \approx A_4$.

The orbit of E_3 under H has 4 elements and the isotropy group H_3 has 3 elements. The representation of H as a permutation group of 4 elements is faithful, otherwise an element of the kernel admits 4 distinct eigenspaces, which is impossible. Hence H is isomorphic to a subgroup of the permutation group on the eigenspaces

$$E_1, \ldots, E_4,$$

and must be isomorphic to A_4 since H has order 12.

(iv) $H \approx S_4$.

The orbit of E_2 under H has 8 elements, and the isotropy group H_2 has 3 elements. If E is an eigenspace of H whose isotropy group has order 3 then E is necessarily in the orbit of E_2. Hence we consider the orbit HE_2 as consisting of four pairs of eigenspaces (cf. the lemma again), and we obtain a representation of H as a permutation group of these pairs. The representation is faithful, and hence $H \approx S_4$ (because the order of H is 24). Otherwise, an element $\alpha \in H$ leaves every pair invariant, and $\alpha \neq 1$. Then α has order 2, and interchanges the elements of each pair. This determines α uniquely, and hence α lies in the center of H. This would imply that H has an element of order 6, which is impossible, thus concluding this case.

(v) $H \approx A_5$.

We do not exhibit the isomorphism explicitly, but only prove by group theory that H is simple, whence must be A_5, the unique simple group of order 60.

Every element of H lies in one of the isotropy groups of some eigenspace, and the orders of these isotropy groups are 2, 3, or 5, and in particular are prime. Any two eigenspaces belonging to elements of the same order are in the same orbit of H. Hence any two cyclic subgroups of H are conjugate. So any normal subgroup of H contains all or none of the elements of any given order. Counting pairs of eigenspaces belonging to any given element, we see that H has 15 elements of order 2, 20 elements of order 3 and 24 elements of order 5. Hence H can have no non-trivial normal subgroup, i.e. H is simple, as was to be shown.

§ 3. Applications to Congruences of the Trace of Frobenius

Let $\rho : G_{\mathbf{Q}} \to G$ be a representation of the Galois group of ${}^a\mathbf{Q}$ over $\mathbf{Q}$ into a product of l-adic Lie groups. Let $G_{\mathbf{Q},\rho}$ be the kernel of the representation, and let K_ρ be the fixed field of the kernel. Then ρ induces an embedding of the factor group into the Lie group, and we call K_ρ the field (extension) associated with the representation. We say that ρ is **unramified** at a prime p if p is unramified in its associated field. We also say that K_ρ is cut out by ρ.

Let $\mu^{(l)}$ for a given prime l be the group of all l^ν-th roots of unity, for all positive integers ν. Let $K = \mathbf{Q}(\mu^{(l)})$. The Galois group $\mathrm{Gal}(K/\mathbf{Q})$ is isomorphic in a natural way with $\mathbf{Z}_l^*$. Indeed, for $a \in \mathbf{Z}_l^*$ we have an automorphism σ_a such that

$$\sigma_a \zeta = \zeta^a, \qquad \zeta \in \mu^{(l)}.$$

By ζ^a we mean the following. Let b be some integer congruent to a modulo a high power of l. Then ζ^a is defined to be ζ^b. The homomorphism

$$\chi_l : G_{\mathbf{Q}} \to \mathbf{Z}_l^*$$

obtained by the representation which cuts out the field K above is called the **cyclotomic representation** (or character) at the prime l, or the **l-adic cyclotomic representation**. Elementary algebraic number theory shows that χ_l is unramified outside l.

We note that for odd l, K splits into an extension $\mathbf{Q}(\zeta_l)$ of $\mathbf{Q}$, of degree l-1, and an l-tower, with Galois group isomorphic to $\mathbf{Z}_l$ (additive), actually equal to the subgroup of $\mathbf{Z}_l^*$ consisting of the units congruent to 1 mod l.

Let G_l be the Galois group of the maximal extension of $\mathbf{Q}$ which is ramified only at l. Any representation of G_l into an abelian group factors through the Galois group of the maximal abelian extension of $\mathbf{Q}$ unramified outside l. By Kronecker's theorem, and elementary facts about ramification in cyclotomic fields, it follows that such an abelian representation factors through $\mathbf{Z}_l^*$.

If p is a prime $\neq l$, then p has a Frobenius conjugacy class associated in G_l. We denote by σ_p an element of this class, and call it a **Frobenius element** at p. In $\mathbf{Q}(\mu^{(l)})$, we have

$$\sigma_p(\zeta)=\zeta^p,$$

so that if the representation is abelian, we may view the representation as a Dirichlet character.

The next theorem describes which degeneracies can occur in certain representations of G_l into $GL_2(\mathbf{F}_l)$, and the subsequent theorem then describes certain congruences for the trace of Frobenius in the various cases which can occur.

Theorem 3.1. *Let G_l be the Galois group of the maximal extension of $\mathbf{Q}$ unramified outside l. Let*

$$\rho: G_l \to GL_2(\mathbf{F}_l)$$

be a representation (so ρ is ramified only at l), and assume that there is an even integer $k \geqslant 2$ such that

$$\det \rho = \chi_l^{k-1}.$$

If $G = \operatorname{Im} \rho$ does not contain $SL_2(\mathbf{F}_l)$ then we have one of the following:

(i) *$G \subset$ Borel.*

(ii) *$G \subset$ Normalizer of a Cartan but not in the Cartan.*

(iii) *$PG \approx S_4$.*

In other words, G cannot be contained in a non-split Cartan unless it is also contained in a Borel, and the exceptional cases $PG \approx A_4$ or A_5 cannot occur.

Proof. Suppose that $G \subset C$ where C is a non-split Cartan. Since C is abelian, it follows that ρ factors through the abelianized group

$$G_l^{ab} \approx \mathbf{Z}_l^*.$$

The order of C, whence of G, is prime to l. Hence ρ factors through $\mathbf{Z}/(l-1)\mathbf{Z}$, and therefore the order of G divides $l-1$. But the order of C is l^2-1 and C is cyclic, containing the cyclic subgroup of scalar matrices $\mathbf{F}_l^*$. Hence since $l+1$ and

$l-1$ have g.c.d. $=2$, it follows that G is contained in the group of scalar matrices, that is G is contained in the split Cartan, contained in a Borel.

Next we show that PG cannot be A_4 or A_5, in which case we can assume $l \neq 2$. We have a commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\det \rho} & \mathbf{F}_l^* \\ \downarrow & & \downarrow \\ PG & \longrightarrow & \mathbf{F}_l^*/\mathbf{F}_l^{*2} \end{array}$$

and for $p \neq l$ we have by hypothesis

$$\det \rho(\sigma_p) \equiv p^{k-1} \pmod{l},$$

where σ_p is the Frobenius automorphism at p. By the existence of primes in arithmetic progressions, we see that the image of $\det \rho$ consists of all $(k-1)$-th powers in $\mathbf{F}_l^*$. Since k is even the map

$$G \to \mathbf{F}_l^*/\mathbf{F}_l^{*2}$$

is surjective. Hence the map

$$PG \to \mathbf{F}_l^*/\mathbf{F}_l^{*2}$$

is surjective. Since $(\mathbf{F}_l^* : \mathbf{F}_l^{*2}) = 2$, it follows that PG has a subgroup of index 2. This is not the case for A_4 or A_5, whence we have proved the theorem.

The next theorem shows that when the image of the representation is small, i.e. does not contain $SL_2(\mathbf{F}_l)$, then certain congruences are satisfied by the trace of Frobenius.

Theorem 3.2. *Let*

$$\rho: G_l \to GL_2(\mathbf{F}_l)$$

be a representation as in the preceding theorem. Let

$$a_p = \operatorname{tr} \rho(\sigma_p)$$

where σ_p is the Frobenius automorphism for a prime $p \neq l$. Then corresponding to the three cases of Theorem 3.1, *we have the following three congruences:*

(i) *There is an integer m such that*

$$a_p \equiv p^m + p^{k-1-m} \pmod{l}.$$

(ii) $a_p = 0$ *if* $\left(\frac{p}{l}\right) = -1$.

(iii) $a_p^2/p^{k-1} = 0, 1, 2,$ or $4 \pmod{l}$.

Proof. Consider case (i). If ρ has the characters ψ, ψ',

$$\rho=\begin{pmatrix}\psi & * \\ 0 & \psi'\end{pmatrix}$$

then we can view ψ, ψ' as Dirichlet characters. Since ψ is abelian, it factors through $\mathbf{F}_l^*$, and there is some integer m such that

$$\psi(p)\equiv p^m \pmod{l}\ .$$

Similarly,

$$\psi'(p)\equiv p^{m'} \pmod{l}\ .$$

But $\det\rho=\chi_l^{k-1}$ implies that $m+m'\equiv k-1 \pmod{l-1}$, thereby proving (i).

In case (ii), we have $l\neq 2$. We have a homomorphism

$$G_l\to N\to N/C\approx\{\pm 1\}$$

which is surjective, and whose image is of order 2, hence abelian. It factors through $\mathbf{Z}_l^*$, and in fact through

$$\mathbf{Z}_l^*/\mathbf{Z}_l^{*2}$$

which has order 2. Hence

$$\rho(\sigma_p)\in C\Leftrightarrow\left(\frac{p}{l}\right)=1$$

$$\rho(\sigma_p)\in N,\ \rho(\sigma_p)\notin C\Leftrightarrow\left(\frac{p}{l}\right)=-1\ .$$

If $(p/l)=-1$ then $\rho(\sigma_p)$ is equivalent over the quadratic extension of F_l to a matrix of the form

$$\begin{pmatrix}0 & b\\ c & 0\end{pmatrix}$$

and hence $a_p=\operatorname{tr}\rho(\sigma_p)=0$. This proves (ii).

In case (iii), we note that every element of PG has order 1, 2, 3, or 4. Let d be the period of an element σ. Then σ has eigenvalues α, β and since

$$\sigma^d\sim\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}$$

it follows that

$$\sigma\sim\begin{pmatrix}\alpha & *\\ 0 & \beta\end{pmatrix}$$

where $\beta=\alpha\zeta$ and $\zeta^d=1$. Hence

$$\frac{(\operatorname{tr}\sigma)^2}{\det\sigma}=\zeta^{-1}+2+\zeta\ ,$$

and we have the following table for the desired number.

d	$\dfrac{(\operatorname{tr}\sigma)^2}{\det\sigma}$
1	4
2	0
3	1
4	2

This proves the statement concerning case (iii).

We now introduce the effect of an Euler product on the coefficients. We fix a prime l. A formal Dirichlet series

$$\varphi=\sum_{l\nmid N}a_n n^{-s}$$

with coefficients in $\mathbf{F}_l$ is said to have an **Euler product** of weight k if it can be expressed as a product

$$\varphi=\prod_{p\neq l}\frac{1}{1-a_p p^{-s}+p^{k-1-2s}}\ .$$

The congruence properties of cases (i) and (ii) in the last theorem can be extended to the coefficients a_n for arbitrary n whenever these a_n arise from such an Euler product, as follows.

Lemma. *Assume that φ has an Euler product as above.*

(i) *If $a_p\equiv p^a+p^b \pmod{l}$ for all $p\neq l$, and positive integers $a\leqslant b$, then*

$$a_n\equiv n^a\sigma_{b-a}(n) \pmod{l}$$

where σ_r here denotes the sum of the divisors of n to the r-th power.

(ii) *If $a_p=0$ for $\left(\frac{p}{l}\right)=-1$, then $a_n=0$ for $\left(\frac{n}{l}\right)=-1$.*

Proof. The integers a, b are only determined mod $l-1$, so we can assume

$$0\leqslant a\leqslant b<l-1\ .$$

The p-th factor of the Euler product is then equal (mod l) to

$$\left(\sum p^{va}p^{-sv}\right)\left(\sum p^{jb}p^{-sj}\right).$$

Hence the coefficient of p^{-sn} is equal to

$$\sum_{v+j=n} p^{va+jb} = \sum_{j=0}^{n} p^{(n-j)a+jb} = p^{na} \sum_{j=0}^{n} p^{j(b-a)} = (p^n)^a \sigma_{b-a}(p^n) .$$

By unique factorization, and the multiplicativity of the functions appearing in this last expression, we conclude that

$$a_n = n^a \sigma_{b-a}(n) \pmod{l} ,$$

thereby proving (i).

In case (ii), we consider the p-factor

$$(1 - a_p p^{-s} + p^{k-1} p^{-2s})^{-1} = \frac{1}{1 + p^{k-1} p^{-2s}} ,$$

if $\left(\frac{p}{l}\right) = -1$. The geometric series for this factor has only terms involving even powers of p^{-s}. Hence

$$a_{p^v} = 0 \text{ if } v \text{ is odd} .$$

If $\left(\frac{n}{l}\right) = -1$,

$$a_n = \prod_{p \mid n} a_{p^{n(p)}} ,$$

then some p dividing n must be such that $\left(\frac{p}{l}\right) = -1$, and $n(p)$ is odd. Hence $a_n = 0$, as was to be shown.

So far, modular forms have not played a role. We now introduce them. For the case of forms of weight 2, the existence of a Galois representation having the desired properties, associated with a modular form, was proved by Shimura, as a consequence of what is known as the Eichler–Shimura congruence relation. See also [Sh 7], Theorem 7.14. It was conjectured for arbitrary weight by Serre, and proved by Deligne.

The relevant theorems are stated as Theorems **D 1** and **D 2** below. The proof for the existence of these representations involves a considerable amount of algebraic geometry and cohomology theory. Fortunately, in many applications, one need only know some of the properties of these representations. Thus in this chapter, we are essentially axiomatizing the situation so that the reader needs only minimal knowledge to understand the results as stated.

Theorem D 1. *Let*

$$\bar{f} = \sum \bar{a}_n q^n$$

be a modular form mod l, with coefficients in the algebraic closure of $\mathbf{F}_l$, and such that $\bar{a}_1 = 1$. Let $k = w(\bar{f})$. Assume that the associated Dirichlet series

$$\sum_{l \nmid n} \bar{a}_n n^{-s}$$

has an Euler product. Let F be the field generated over $\mathbf{F}_l$ by the coefficients of $\bar{f}$. Then there exists a unique semisimple representation

$$\rho : G_l \to GL_2(F)$$

such that

$$\operatorname{tr} \rho(\sigma_p) = \bar{a}_p$$
$$\det \rho = \chi_l^{k-1} \pmod{l} ,$$

i.e. $\det \rho(\sigma_p) \equiv p^{k-1} \pmod{l}$ *for all primes $p \neq l$.*

The representation ρ is said to be **associated** with $\bar{f}$, and will be denoted by $\rho_{\bar{f}}$.

We recall that $k = w(\bar{f})$ is the filtration of $\bar{f}$, i.e. the smallest integer $\geqslant 0$ such that $\bar{f}$ is the reduction of a modular form of that weight. Cf. Chapter X, § 7.

The next three theorems give additional information on the three cases (i), (ii), (iii) for such representations. In those theorems, we take $F = \mathbf{F}_l$.

For the next two theorems, *we suppose that $\rho = \rho_{\bar{f}}$ is a representation of G_l into $GL_2(\mathbf{F}_l)$ associated with the modular form $\bar{f} \bmod l$. We let*

$$k = w(\bar{f}) .$$

We assume that $\bar{f} = \sum \bar{a}_n q^n$ is such that the associated formal Dirichlet series

$$\sum_{l \nmid n} \bar{a}_n n^{-s}$$

has an Euler product. We suppose $l \geqslant 5$.

Theorem 3.3 (i). *Suppose that* Im ρ *is contained in a Borel, i.e. suppose that case* (i) *prevails. If $l > k+1$, then:*

(a) *The characters of ρ are* 1 *and* $\chi_l^{k-1} \pmod{l}$.

(b) *We must have $\bar{f} = \bar{G}_k$.*

Proof. Recall that

$$G_m = -\frac{B_m}{2m} + \sum \sigma_{m-1}(n) q^n .$$

We let $0 \leqslant a \leqslant b \leqslant l-2$ as in the lemma, and $a \neq b$ because $a+b$ is odd. Theorem 3.2 and the lemma, (i), show that in the cases other than $a=0$, $b=l-2$, we have

$$(*) \qquad \theta \bar{f} = \theta^{a+1} \bar{G}_{b-a+1} .$$

We have used an extra power of θ to kill those coefficients $\bar{a}_n$ of $\bar{f}$ such that $l \mid n$.

In the case $a=0$, $b=l-2$, the constant term of G_{b-a+1} is not l-integral. In that case, we have instead

$$p a_p \equiv 1 + p \pmod{l} ,$$

whence $n a_n \equiv \sigma_1(n) \bmod l$ for n prime to l, and finally

$$(**) \qquad \theta \bar{f} = \theta^{l-1} \bar{G}_2 = \theta^{l-1} \bar{G}_{l+1} .$$

We cannot have $b-a=1$, i.e. $b-a+1=2$. Otherwise,

$$w(\bar{G}_2) = l+1 ,$$

because $\bar{G}_2 = \bar{G}_{l+1}$, and there is no modular form of weight 2. To get a contradiction for this, we use Serre's filtration Theorem 8.1 of Chapter X. To do this, we note that

$$l+1+v(l+1) \equiv v+1 \not\equiv 0 \pmod{l}$$

for $v=0,\dots,a$, and hence from the right-hand side we get

$$w(\theta^a \bar{G}_2) = l+1+(a+1)(l+1) = (a+2)(l+1) .$$

From the left-hand side, we get

$$w(\theta \bar{f}) \leqslant k+l+1 ,$$

which contradicts $k+1<l$, and concludes the proof that $b-a \neq 1$.

Suppose that $b-a+1>2$. Suppose we are in the case other than $a=0$, $b=l-2$. Then we also obtain

$$0 < b-a+1 < b+1 \leqslant l-1 ,$$

and $b-a+1 \not\equiv 0 \pmod{l-1}$. For $v=0,\dots,a$ we have

$$b-a+1+v(l+1) \equiv b-a+1+v \not\equiv 0 \pmod{l} .$$

Hence by Serre's filtration theorem,

$$w(\theta^{a+1} \bar{G}_{b-a+1}) = (b-a+1)+(a+1)(l+1)$$
$$w(\theta \bar{f}) \leqslant k+l+1 .$$

This is possible only if $a=0$, $b=k-1$. Therefore

$$\theta \bar{f} = \theta \bar{G}_k .$$

If $\bar{f} - \bar{G}_k \neq 0$, then $\theta(f-G_k) \neq 0$, and

$$w(\bar{f} - \bar{G}_k) \equiv k \not\equiv 0 \pmod{l} ,$$

whence

$$w(\theta(\bar{f} - \bar{G}_k)) = k+l-1 ,$$

which is impossible, and concludes the proof of the case (*). The case (**) is handled similarly, cf. [Sw D].

Theorem 3.3 (ii). *If $l>2k$, then case* (ii) *does not occur.*

Proof. Suppose otherwise. By the Lemma and Theorem 3.2 we know that

$$\bar{a}_n = 0 \quad \text{if} \quad \left(\frac{n}{l}\right) = -1 .$$

But

$$n^{(l+1)/2} = n^{(l-1)/2} n \equiv n \quad \text{if} \quad \left(\frac{n}{l}\right) = 1 .$$

Hence

$$\theta \bar{f} = \theta^{(l+1)/2} \bar{f} .$$

We have

$$k + v(l+1) \equiv k + v \not\equiv 0 \pmod{l}$$

for $0 \leqslant v \leqslant (l-1)/2$. Hence

$$w(\theta^{(l+1)/2} \bar{f}) = k + \frac{l+1}{2}(l+1) \equiv k + \frac{l+1}{2} \pmod{l}$$

$$w(\theta \bar{f}) = k+l+1 \equiv k+1 \pmod{l} ,$$

a contradiction which proves the theorem.

Theorem 3.3 (iii). *Case* (iii) *cannot occur if l is sufficiently large.*

Proof. We select some prime $p \neq 2$ such that $a_p \neq 0$. Then there is only a finite number of l dividing

$$a_p^2 - p^{k-1} v, \qquad \text{with} \qquad v = 0, 1, 2, 4$$

(note that this integer is not 0). The assertion follows.

For the applications, one uses the following theorem of Deligne.

Theorem D 2. *Let $f=\sum a_n q^n$ be a modular form with integer coefficients, such that $a_0=0$ and $a_1=1$. Assume that the associated Dirichlet series $\sum a_n n^{-s}$ has an Euler product. Then there exists a representation*

$$\rho_l : G_l \to GL_2(\mathbf{Z}_l)$$

such that $\rho_l(\sigma_p)$ has characteristic polynomial

$$X^2 - a_p X + p^{k-1}$$

for any prime $p \neq l$.

In Deligne's theorem, we let

$$\bar{\rho}_l : G_l \to GL_2(\mathbf{F}_l)$$

be the reduction of $\rho_l \bmod l$, so that this reduction is of the type we have been considering in this section.

Define $GL_2(\mathbf{F}_l)_k$ to consist of those elements α such that $\det \alpha$ is $k-1$ power in $\mathbf{F}_l^*$, and similarly for $GL_2(\mathbf{Z}_l)$.

Theorem 3.4. *For all but a finite number of l, the image of $\bar{\rho}_l$ is equal to $GL_2(\mathbf{F}_l)_k$.*

Proof. The main point is that for all but a finite number of l, the image contains $SL_2(\mathbf{F}_l)$. We use Theorems 3.1, 3.2, 3.3. If $l > k+1$ and if the constant term of G_k is not $\equiv 0 \pmod l$ then case (i) cannot occur. If $l > 2k$, then case (ii) cannot occur and case (iii) cannot occur by Theorem 3.3 (iii).

Using now the condition on the determinant of the representation and its action on the field $\mathbf{Q}(\zeta_l)$ of l-th roots of unity, it is easy to see that the image of ρ_l must be all of $GL_2(\mathbf{F}_l)_k$. This type of argument was originally given by Serre in his study of the fields of division points of elliptic curves without complex multiplication, cf. [Se 4], or [L 2], Chapter 17, § 3, § 4.

Using a lemma of Serre (same references) it is then easy to show that the image of ρ_l is equal to $GL_2(\mathbf{Z}_l)_k$ for almost all l.

Serre has also shown that for the exceptional l such that the image of ρ_l is not all of $GL_2(\mathbf{Z}_l)_k$, it is still true that this image is open in $GL_2(\mathbf{Z}_l)_k$. The arguments involve Lie theory, and again are of the same type that Serre used in his treatment of division points of elliptic curves [Se 2], § 5, p. 14–12.

To get examples of modular forms satisfying the conditions under which we have been proving theorems, one can take cusp forms for $SL_2(\mathbf{Z})$ for those weights k such that M_k^0 has dimension 1, in which case, these are automatically eigenfunctions of Hecke operators, whence have the desired Euler product. There are six such cases, for

$$k = 12, 16, 18, 20, 22, 26 .$$

In Swinnerton-Dyer's paper, the reader will find an explicit determination of the exceptional primes l for these cases, except for $l=59$ and $k=16$.

Although these do not constitute many examples, nevertheless the techniques used generalize to other cases, as in Ribet [Ri 1], where it becomes important to have modular forms with coefficients in fairly general rings, e.g. rings of Hecke operators. In the preceding chapter, we took an intermediate position, using for A a local ring in a number field.

Appendix by Walter Feit. Exceptional Subgroups of GL_2

In Chapter XI, Theorem 2.3 it was shown that if F is a field of characteristic l and G is a finite group contained in $PGL_2(F)$ such that l does not divide the order of G then G is cyclic, dihedral or isomorphic to one of A_4, S_4 or A_5. The purpose of this appendix is to state precisely when A_4, S_4 or A_5 can occur in case F is a finite field.

The following general results from representation theory will be used.

Proposition 1. *Let G be a finite group and let l be a prime which does not divide the order of G. Let F be a field of characteristic l. Then an absolutely irreducible $F[G]$ module is determined up to equivalence by its trace function.*

Proposition 2. *Let G be a finite group and let l be a prime. Let K be an algebraic number field which is a splitting field for all irreducible $K[G]$ modules. Let R be the ring of integers in K and let $\mathfrak{L}$ be a prime divisor of l in R. For a in R let $\bar{a}$ denote the image of a in $\bar{R}=R/\mathfrak{L}$.*

(i) *If χ is a character of G then $\bar{\chi}$ is the trace function of an $\bar{R}[G]$ module of dimension $\chi(1)$.*

(ii) *If l does not divide the order of G then the map which sends χ to $\bar{\chi}$ defines a bijection from the set of all irreducible characters of G onto the set of all trace functions of irreducible $\bar{R}[G]$ modules. Furthermore the irreducible $\bar{R}[G]$ module with trace function $\bar{\chi}$ has dimension $\chi(1)$ over $\bar{R}$.*

Proposition 3. *Let G be a finite group and let F be a field of characteristic $l\neq 0$. Let θ be the trace function of an absolutely irreducible $F[G]$ module. If F_0 is a subfield of F such that $\theta(x)\in F_0$ for all x in G then θ is the trace function of an absolutely irreducible $F_0[G]$ module. (This is essentially Wedderburn's Theorem which asserts that finite division rings are commutative.)*

For any group G let $\mathfrak{Z}(G)$ denote the center of G and let G' denote the commutator subgroup of G.

Suppose that $\mathfrak{Z}(G)\subset G'$. A *covering group* of G is a group $\hat{G}$ such that $\mathfrak{Z}(\hat{G})\subset\hat{G}'$ and $\hat{G}/Z\approx G$ for some subgroup Z of $\mathfrak{Z}(\hat{G})$.

Throughout the rest of this appendix the following notations will be used.

$$H=A_4,\ S_4,\quad \text{or}\quad A_5\,.$$

l is a prime which does not divide $|H|$. Thus $l\neq 2, 3$. In case $H=A_5$, $l\neq 5$. $F=\mathbf{F}_q$ where q is a power of l.

$$\hat{A}_4=SL_2(3),\qquad \hat{A}_5=SL_2(5)\,.$$

Let a be an element of order 8 in $GL_2(3)$. Then $\hat{S}_4$ is the group generated by the following matrices whose entries are elements of $GL_2(3)$.

$$\begin{pmatrix} x & 0\\ 0 & a^{-1}xa\end{pmatrix}\quad x\in SL_2(3),\qquad \begin{pmatrix} 0 & a^2\\ -1 & 0\end{pmatrix}.$$

The following properties of $\hat{H}$ are easily verified

(i) *$\hat{H}$ contains a unique involution. This involution generates $\mathfrak{Z}(\hat{H})$.*

(ii) *$\mathfrak{Z}(\hat{H})$ is contained in every nontrivial normal subgroup of $\hat{H}$.*

(iii) $\mathfrak{Z}(\hat{H})\subseteq\hat{H}'$ *and* $\hat{H}/\mathfrak{Z}(\hat{H})\approx H$.

$\tilde{H}$ is a group which contains a unique involution, $|\mathfrak{Z}(\tilde{H})|=2$, and $\tilde{H}/\mathfrak{Z}(\tilde{H})\approx H$. It is easily seen that $\mathfrak{Z}(\tilde{H})$ is contained in every proper normal subgroup of $\tilde{H}$. Thus in particular $\tilde{H}$ is a covering group of H. The existence of $\hat{H}$ assures the existence of $\tilde{H}$.

It is known that $\tilde{H}$ is determined up to isomorphism by H and so is isomorphic to $\hat{H}$. This will only be needed in case $H=A_5$ and is proved below.

Lemma 1. *Let K be a field of characteristic different from 2. Let G be a subgroup of $PSL_2(K)$ with $G\approx H$. Then the inverse image of G in $SL_2(K)$ is isomorphic to $\tilde{H}$.*

Proof. This is an immediate consequence of the fact that $SL_2(K)$ contains a unique involution and $|\mathfrak{Z}(SL_2(K))|=2$.

Lemma 2. (i) *$\tilde{A}_4$ has exactly three linear characters, μ, μ^2, $\mu^3=1$. There exists a faithful irreducible character χ of degree 2 with $\mathbf{Q}(\chi)=\mathbf{Q}$ such that χ, $\chi\mu$, $\chi\mu^2$ are all the irreducible characters of $\tilde{A}_4$ of degree 2. Furthermore $\mathbf{Q}(\chi\mu)=\mathbf{Q}(\chi\mu^2)=\mathbf{Q}(\sqrt{-3})$.*

(ii) *$\hat{S}_4$ has exactly two linear characters λ, $\lambda^2=1$. There exists a faithful irreducible character χ of degree 2 with $\mathbf{Q}(\chi)=\mathbf{Q}(\sqrt{2})$ such that χ, $\chi\lambda$ are all the irreducible characters of $\hat{S}_4$ of degree 2 whose kernel is contained in $\mathfrak{Z}(\hat{S}_4)$.*

(iii) *$\tilde{A}_5$ has exactly two irreducible characters χ, χ' of degree 2. They are faithful and $\mathbf{Q}(\chi)=\mathbf{Q}(\chi')=\mathbf{Q}(\sqrt{5})$.*

Proof. The group A_4 has no faithful characters of degree less than 3. Since S_4 and A_5 contain a subgroup isomorphic to A_4 the same holds true for S_4 and A_5. Hence by (ii) any irreducible character of $\tilde{H}$ whose kernel is in $\mathfrak{Z}(\tilde{H})$ of degree 2 is faithful. Let $\theta_1,\cdots,\theta_s$ be all the faithful irreducible characters of $\tilde{H}$. Then $\theta_i(1)$

is even as $\mathscr{Z}(\tilde{H})$ is represented by matrices of determinant 1. Let $\theta_i(1)=2d_i$. Furthermore

$$\sum_{i=1}^{s} \theta_i(1)^2 = |\tilde{H}| - |H| = |H| .$$

(i) $4\sum_{i=1}^{s} d_i^2 = 12$. Thus $s=3$, $d_1=d_2=d_3=1$. The group A_4 has a linear character μ which is faithful on A_4/A_4'. Thus μ has order 3. Since $\theta_1(1)=2$ it follows that $\theta_1(x)\neq 0$ for an element x of order 3 in $\tilde{A}_4$. Thus θ_1, $\theta_1\mu$, $\theta_1\mu^2$ are distinct and so are all the irreducible characters of degree 2. $[\mathbf{Q}(\theta_1\mu^i):\mathbf{Q}]\leqslant 3$ as $\theta_1\mu^i$ has at most 3 conjugates. If x has order 3 then for some i, $\sqrt{-3}\in\mathbf{Q}(\theta_1\mu^i(x))$. Thus two of θ_i, θ_j are algebraically conjugate and the other has values in $\mathbf{Q}$.

(ii) $\tilde{S}_4$ contains a subgroup isomorphic to $\tilde{A}_4$. The two nontrivial linear characters of A_4 are conjugate in S_4. Thus if χ is defined as in (i) χ induces a sum of two irreducible characters of degree 2 and $\chi\mu$, $\chi\mu^2$ induce the same irreducible character of degree 4. Let χ denote an extension of χ to $\tilde{S}_4$. Then χ, $\chi\lambda$ are precisely the faithful irreducible characters of degree 2. Hence $[\mathbf{Q}(\chi):\mathbf{Q}]\leqslant 2$. Since $\tilde{S}_4$ contains a unique involution, a Sylow 2-group of $\tilde{S}_4$ is a quaternion group of order 16. Thus there exists y of order 8 in $\tilde{S}_4$ with y conjugate to y^{-1}. Hence if V is the module which affords χ then the characteristic values of y are ω and ω^{-1} where ω is a primitive 8^{th} root of 1. Hence $\chi(y)=\omega+\omega^{-1}=\sqrt{2}$ and so $\mathbf{Q}(\chi)=\mathbf{Q}(\sqrt{2})$.

(iii) $\sum_{i=1}^{s} d_i^2 = 15$. Thus at least one $d_i=1$. Suppose that $\theta_i(1)=2$ for $i=1,\cdots,t$ and $\theta_i(1)>2$ for $i>t$. Hence $t\geqslant 1$. Let x be an element of order 3 in $\tilde{A}_5$. Then x is conjugate to its inverse and so $\theta(x)$ is a rational integer for all irreducible characters θ. Furthermore the centralizer of x has order 6 and the centralizer of the image of x in $\tilde{A}_5/\mathscr{Z}(\tilde{A}_5)\approx A_5$ has order 3. Thus $\sum_{i=1}^{s}\theta_i^2(x)=3$. If $\theta_i(x)=2$ then $\theta_i(x)\neq 0$ and so $t\leqslant 3$. Let y be an element of order 5 in $\tilde{A}_5$. Then y is conjugate to its inverse. Thus if V is a module which affords θ then the characteristic values of y on V are ω, ω^{-1} where ω is a primitive 5^{th} root of 1. Hence $\sqrt{5}\in\mathbf{Q}(\theta_1(y))$. Therefore $[\mathbf{Q}(\theta_1):\mathbf{Q}]$ is even and at most 3. Thus $\mathbf{Q}(\theta_1)=\mathbf{Q}(\sqrt{5})$ and $t=2$ as required.

Lemma 3. *$\tilde{A}_5\approx SL_2(5)$. There exists an outer automorphism which interchanges the two characters of $\tilde{A}_5$ of degree 2.*

Proof. By Lemma 2 and Proposition 2 (i) $\tilde{A}_5$ is isomorphic to a subgroup of $GL_2(5)$. As $\tilde{A}_5=\tilde{A}_5'$ this implies that $\tilde{A}_5$ is isomorphic to a subgroup of $SL_2(5)$. Thus $\tilde{A}_5\approx SL_2(5)$ since $|\tilde{A}_5|=120=|SL_2(5)|$.

The element $\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$ in $GL_2(5)$ induces an outer automorphism of $SL_2(5)$ which interchanges the two conjugate classes of elements of order 5 in $SL_2(5)$. Thus this outer automorphism must interchange the two characters of degree 2.

Lemma 4. *Let G be a subgroup of $PGL_2(F)$ which is isomorphic to H or abelian of type $(2,2)$. Let N be the normalizer of G in $PGL_2(F)$. Then l does not divide $|N|$. If furthermore $G\approx H$ then either $N=G$ or $G\approx A_4$ and $N\approx S_4$.*

Proof. Let G_0, N_0 be the inverse image respectively of G, N in $GL_2(F)$. The only elements of order prime to l which commute with an element of order l are scalars. Thus l does not divide the order of the centralizer of G_0 in $GL_2(F)$. As G_0 admits no outer automorphism of order l ($l>3$) it follows that $l\nmid|N|$. The last statement now follows from Theorem 2.3 in Chapter XI.

Theorem 1. (i) *$PSL_2(F)$ contains a subgroup isomorphic to A_4.*

(ii) *The following are equivalent.*

(a) *$PSL_2(F)$ contains a subgroup isomorphic to S_4.*

(b) $24\,||\,|PSL_2(F)|$.

(c) $q\equiv\pm 1 \pmod 8$.

(d) $\sqrt{2}\in F$.

(iii) *$PGL_2(F)$ contains a subgroup isomorphic to S_4.*

(iv) *The following are equivalent.*

(a) *$PGL_2(F)$ contains a subgroup isomorphic to A_5.*

(b) *$PSL_2(F)$ contains a subgroup isomorphic to A_5.*

(c) $60\,||\,|PSL_2(F)|$.

(d) $q\equiv\pm 1 \pmod 5$.

(e) $\sqrt{5}\in F$.

Proof. (i) By Lemma 2 (i), $\tilde{A}_4$ has a faithful irreducible character χ of degree 2 with $\mathbf{Q}(\chi)=\mathbf{Q}$. Thus by Propositions 2 and 3, $\tilde{A}_4$ is isomorphic to a subgroup of $GL_2(F)$. Hence A_4 is isomorphic to a subgroup of $PGL_2(F)$. This subgroup must be in $PSL_2(F)$ since $|PGL_2(F):PSL_2(F)|=2$ and A_4 has no subgroup of index 2.

(ii) (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are obvious. By the quadratic reciprocity theorem (c) $\Rightarrow$ (d). It remains to show that (d) $\Rightarrow$ (a). Suppose that $\sqrt{2}\in F$. By Lemma 2 (ii) and Propositions 2 and 3 $GL_2(F)$ contains a subgroup G isomorphic to $\tilde{S}_4$. The Sylow 2-group of G is quaternion of order 16. Thus G contains an element y of order 8 which is conjugate to its inverse. Thus the characteristic values of y are ω and ω^{-1}. Hence $y\in SL_2(F)$. $G'\approx\tilde{A}_4$ and $G'\subseteq SL_2(F)$. $\tilde{A}_4$ contains no element of order 8. Therefore $G=\langle y, G'\rangle\subset SL_2(F)$ and so $G/\mathscr{Z}(G)\approx S_4$ is in $PSL_2(F)$.

(iii) If $8\,||\,|PSL_2(F)|$ the result follows from (ii). Suppose that $8\nmid|PSL_2(F)|$. By (i) $PSL_2(F)$ contains a subgroup $G\approx A_4$. Let S be a Sylow 2-subgroup of G. Thus $|S|=4$ and S is a Sylow 2-subgroup of $PSL_2(F)$. Let T be a Sylow 2-subgroup of $PGL_2(F)$ with $S\subset T$. Thus $|T:S|=2$. Consequently $\langle A_4, T\rangle\subset N$ where N is the normalizer of S in $PGL_2(F)$. Thus $24\,||N|$ and $l\nmid|N|$ by Lemma 4. Since $A_4\subset N$, Theorem 2.3 of Chapter XI implies that $N\approx S_4$.

(iv) By the quadratic reciprocity theorem (d) $\Rightarrow$ (e).

Suppose that $\sqrt{5}\in F$. By Lemma 2 (iii) and Propositions 2 and 3, $GL_2(F)$ contains a subgroup isomorphic to $\tilde{A}_5$. Thus $PGL_2(F)$ contains a subgroup isomorphic to A_5. This subgroup is in $PSL_2(F)$ as $A_5'=A_5$.

For any subgroup G of $PGL_2(F)$ let $\mathcal{N}(G)$ denote the normalizer of G in $PGL_2(F)$. If $G \subset PSL_2(F)$ let $\mathcal{N}_1(G)$ denote the normalizer of G in $PSL_2(F)$. Then $\mathcal{N}_1(G) \subset \mathcal{N}(G)$ and $|\mathcal{N}(G):\mathcal{N}_1(G)| \leq 2$. If $|\mathcal{N}(G):\mathcal{N}_1(G)| = 2$ then G has the same number of conjugates in $PGL_2(F)$ and $PSL_2(F)$. Thus any subgroup of $PSL_2(F)$ which is conjugate to G in $PGL_2(F)$ is already conjugate to G in $PSL_2(F)$. If $\mathcal{N}(G) = \mathcal{N}_1(G)$ then the class of subgroups of $PSL_2(F)$ which are conjugate to G in $PGL_2(F)$ breaks up into 2 conjugate classes of subgroups of $PSL_2(F)$. These remarks will be used repeatedly in the next result.

Theorem 2. (i) *There is one conjugate class of subgroups of $PGL_2(F)$ which are isomorphic to A_4.*

(ii) *Suppose that $q \equiv \pm 1 \pmod 8$.*

(a) *There are two conjugate classes of subgroups of $PSL_2(F)$ which are isomorphic to A_4.*

(b) *There are two conjugate classes of subgroups of $PSL_2(F)$ which are isomorphic to S_4.*

Any two subgroups of $PGL_2(F)$ which are isomorphic to S_4 are conjugate in $PGL_2(F)$.

(iii) *Suppose that $q \not\equiv \pm 1 \pmod 8$.*

(a) *Any two subgroups of $PSL_2(F)$ which are isomorphic to A_4 are conjugate in $PSL_2(F)$.*

(b) *Any two subgroups of $PGL_2(F)$ which are isomorphic to S_4 are conjugate in $PGL_2(F)$.*

(iv) *Suppose that $q \equiv \pm 1 \pmod 5$.*

(a) *There is one conjugate class of subgroups of $PGL_2(F)$ which are isomorphic to A_5.*

(b) *There are two conjugate classes of subgroups of $PSL_2(F)$ which are isomorphic to A_5.*

Proof. Let G_1 and G_2 be subgroups of $PGL_2(F)$ which are isomorphic to A_4. Since A_4 has no subgroup of index 2, $G_1, G_2 \subseteq PSL_2(F)$. Let G_1^0, G_2^0 be the inverse images of G_1, G_2 respectively in $SL_2(F)$. Let S_i^0 be a Sylow 2-group of G_i^0. Then S_i^0 is a quaternion group of order 8. Thus S_i^0 has a unique irreducible character of degree 2. Thus by Propositions 1, 2 and 3 it may be assumed that $S_1^0 = S_2^0$ after G_2^0 is replaced by a conjugate in $GL_2(F)$. Let S be the image of S_1^0 in $PGL_2(F)$. Thus S is noncyclic of order 4 and $\langle G_1, G_2 \rangle \subset N(S)$. By Lemma 4 and Theorem 2.3 of Chapter XI, $N(S) \approx A_4$ or S_4. Thus $G_1 = G_2$.

(ii) By Theorem 1 there exists a subgroup G of $PSL_2(F)$ with $G \approx S_4$. Thus $G' \approx A_4$. By Lemma 4 $\mathcal{N}(G') = \mathcal{N}_1(G') = G$. Thus by (i) there are two conjugate classes of subgroups of $PSL_2(F)$ which are isomorphic to A_4. For each such subgroup G_1, $\mathcal{N}_1(G_1') \approx S_4$. Hence there are also two conjugate classes of subgroups

of $PSL_2(F)$ which are isomorphic to S_4. Furthermore any subgroup of $PGL_2(F)$ which is isomorphic to S_4 is in $PSL_2(F)$ and is conjugate to G.

(iii) By Theorem 1 there exists a subgroup G of $PGL_2(F)$ with $G \approx S_4$. Thus $G' \approx A_4$. By Lemma 4 and Theorem 1 $\mathcal{N}(G') = G$ and $\mathcal{N}_1(G') = G'$. Thus $|\mathcal{N}(G'):\mathcal{N}_1(G')| = 2$. Hence by (i) any subgroup of $PGL_2(F)$ which is isomorphic to A_4 is conjugate to G. Since any subgroup of $PGL_2(F)$ which is isomorphic to S_4 is the normalizer a subgroup isomorphic to A_4 the last statement follows.

(iv) By Theorem 1 there exists a subgroup G of $PSL_2(F)$ with $G \approx A_5$. By Lemma 3 the inverse image $\hat{G}$ of G in $SL_2(F)$ is isomorphic to $SL_2(5)$. Let $\hat{G}_1$ and $\hat{G}_2$ be subgroups of $SL_2(F)$ which are isomorphic to $SL_2(5)$. Let θ be the trace function of $\hat{G}_1$. Propositions 1, 2, 3, and Lemmas 2, 3 imply that the trace function of $\hat{G}_2$ is either θ or θ^σ where σ is defined as in Lemma 3. If the trace function of $\hat{G}_2$ is θ then by Proposition 1, $\hat{G}_1$ is conjugate to $\hat{G}_2$ in $GL_2(F)$. If the trace function of $\hat{G}_2$ is θ^σ than $\hat{G}_2^\sigma$ has the trace function θ. Hence by Proposition 1 $\hat{G}_2^\sigma = \hat{G}_2$ is conjugate to $\hat{G}_1$ in $GL_2(F)$. Thus in any case $\hat{G}_1$ is conjugate to $\hat{G}_2$. Therefore any subgroup of $PGL_2(F)$ which is isomorphic to A_5 is conjugate to G in $PGL_2(F)$.

By Lemma 4 $\mathcal{N}(G) = \mathcal{N}_1(G) = G$. Thus there are two conjugate classes of subgroups of $PSL_2(F)$ which are isomorphic to A_5.

Part V

p-Adic Distributions

Chapter XII. General Distributions

We consider functions on a projective system which satisfy a compatibility relation. At each step, the sum of the values in a given fiber are equal to the values at the base point. Mazur isolated this notion [Maz 1], [Maz 2], [Ma-SwD], which turns out to be very prevalent in number theory. This followed the work of Iwasawa, working with group rings formed with a projective system of finite groups, so that the compatibility relation is merely a formulation independent of the group for the basic property of the natural homomorphism of group rings induced by a group homomorphism. Iwasawa's work dealt with projective limits of ideal class groups, a topic pursued especially in papers of Coates with Sinnott, Lichtenbaum and Wiles, [CS 1], [CS 2], [C-Li], [CW]. For projective limits of divisor class groups in the modular function field, cf. the Kubert-Lang series [KL].

The theory also applies to eigenfunctions of Hecke operators.

The reader will immediately see that this eigenfunction property amounts precisely to the compatibility property of a distribution. Cf. Manin [Ma 4], who thus obtains p-adic L-series associated with cusp forms, and also Mazur loc. cit.

Distributions on $\mathbf{Z}_p$ will be identified with power series in § 3, and the Weierstrass Preparation Theorem proved in § 4 gives rise to polynomials which are analogues of zeta functions. The study of their roots in cases when they arise from modular forms and Hecke operators as in Theorem 3.1 of Chapter IV constitute one of the main areas of current research. This chapter lays the general algebraic foundations behind this situation.

§ 1. Definitions

Let $\{X_n\}$ be a sequence of finite sets, and suppose given a sequence of surjective maps

$$\pi_{n+1}: X_{n+1} \to X_n ,$$

so that we can consider the projective limit

$$X \to \cdots \to X_{n+1} \to X_n \to \cdots \to X_1 .$$

For convenience, we took our family of sets indexed by the positive integers. In applications, it often occurs that the sets are ordered by the positive integers

ordered by divisibility. For instance, the family of sets $\mathbf{Z}/N\mathbf{Z}$ arises in the sequel. We shall also consider the projective family

$$\{\mathbf{Z}/p^n\mathbf{Z}\},$$

with a fixed prime number p, and $n=0, 1, 2, \ldots$. In each case, the connecting homomorphism

$$\mathbf{r}_M : \mathbf{Z}/N\mathbf{Z} \to \mathbf{Z}/M\mathbf{Z}$$

for $M|N$ is reduction mod M, denoted by $\mathbf{r}_M$.

This type of projective family will also arise in isomorphic form as follows. We have an isomorphism

$$\frac{1}{N}\mathbf{Z}/\mathbf{Z} \to \mathbf{Z}/N\mathbf{Z}$$

given by multiplication with N. We then have a commutative diagram

$$\begin{array}{ccc} \frac{1}{N}\mathbf{Z}/\mathbf{Z} & \longrightarrow & \mathbf{Z}/N\mathbf{Z} \\ {\scriptstyle N/M}\downarrow & & \downarrow{\scriptstyle \mathbf{r}_M} \\ \frac{1}{M}\mathbf{Z}/\mathbf{Z} & \longrightarrow & \mathbf{Z}/M\mathbf{Z} \end{array}$$

where the left vertical arrow is multiplication with N/M, and the right arrow is reduction mod M. Thus the system

$$\left\{\frac{1}{N}\mathbf{Z}/\mathbf{Z}\right\}$$

is also a projective system, ordered by divisibility.

Let us now return to the general projective system $\{X_n\}$. For each n suppose given a function φ_n of X_n into an abelian group V. We say that the family $\{\varphi_n\}$ is **compatible** if for each n and $x \in X_n$ we have

$$\varphi_n(x) = \sum_{\pi_{n+1} y = x} \varphi_{n+1}(y).$$

The sum is taken over all the elements of X_{n+1} lying above x. In what follows, we often omit the subscripts, and write $\pi y = x$, for instance.

Let K be a ring of operators on V.

Let f be a function on X_m for some integer m, with values in K. If $n \geqslant m$, then we

view f as defined on X_n through the natural projection on X_m. We conclude at once from the compatibility relation that

$$\sum_{x \in X_n} f(x)\varphi_n(x) = \sum_{x \in X_m} f(x)\varphi_m(x).$$

Let X be the projective limit

$$X = \varprojlim X_n,$$

with the limit topology, so that X is a compact space. For each n we have a surjective map

$$\mathbf{r}_n : X \to X_n.$$

For each $x \in X_n$ the inverse image $\mathbf{r}_n^{-1}(x)$ is an open set in X, and the totality of such open sets for all n, x is a basis for the topology of X.

A function f on X is called **locally constant** if and only if there exists n such that f factors through X_n. Such functions are also called **step functions**, and their group is denoted by $\mathbf{St(X, K)}$. For each such function, we can define its integral

$$\int f\, d\varphi = \sum_{x \in X_n} f(x)\varphi_n(x),$$

independent of the choice of n such that f factors through X_n. We then call the family $\{\varphi_n\}$, or the functional $d\varphi$, a **distribution** on X. It is an additive map

$$d\varphi : St(X, K) \to V.$$

Examples of such maps will be given later with Bernoulli numbers.

Let K be a complete field with respect to a non-archimedean valuation, and suppose that V is a non-archimedean Banach space over K, i.e. V is a complete vector space, with a norm

$$|\ |_V : V \to \mathbf{R}^+$$

satisfying

$$|v+w| \leqslant \max\{|v|, |w|\} \qquad v, w \in V$$

$$|cv|_V = |c|_K\, |v|_V \qquad c \in K, v \in V.$$

If φ is bounded, i.e. $|\varphi_n(x)|$ is bounded for all n, $x \in X_n$, then we say that φ is **bounded**, or **quasi-integral** for the valuation. For any $f \in St(X, K)$ we have

$$\left|\int f\, d\varphi\right| \leqslant \|f\|\, \|\varphi\|,$$

where $\|f\|$ is the sup norm of f, and $\|\varphi\|$ is the sup norm of the values $|\varphi_n(x)|$.

Indeed, if f factors through X_n, then

$$\left|\int f\,d\varphi\right| = \left|\sum_{x\in X_n} f(x)\varphi_n(x)\right| \leq \max_{x\in X_n} |f(x)|\,|\varphi_n(x)|$$

by the non-archimedean property, so our assertion is clear.

In particular, if $f\in C(X)$ is a continuous function on X, then we can approximate f uniformly by a sequence $\{f_n\}$ of step functions, and since $\|f-f_n\|\to 0$, we get

$$\|f_n - f_m\| \to 0$$

for $m, n\to\infty$. Hence the integrals

$$\int f_n\,d\varphi$$

converge, and define the integral

$$\int f\,d\varphi$$

for such a continuous function, provided that φ is bounded. This will be the case in important examples, and bounded distributions are also called **measures**.

§ 2. Averaging Operators

Let f be a function defined on $\mathbf{Q}/\mathbf{Z}$, with values in an abelian group V. For $r\in\mathbf{Q}/\mathbf{Z}$, define the **average** $A_N f(r)$ by

$$A_N f(r) = \sum_{Nt=r} f(t)\,.$$

Then we have the obvious formulas

$$A_M \circ A_N = A_{MN}\,,$$

and if $D|N$, then

$$A_N(f\circ D) = DA_{N/D}f\,,$$

where $(f\circ D)(x) = f(Dx)$.

We shall apply these when $N=p^n$ is a prime power, so that f need be defined only for a subgroup of $\mathbf{Q}/\mathbf{Z}$ which for each element r also contains the inverse image $(p\cdot\mathrm{id})^{-1}(r)$. This is satisfied for instance by the subgroup

$$\frac{1}{m_0 p^\infty}\mathbf{Z}/\mathbf{Z}$$

with m_0 fixed. In that case, we use logarithmic notation for the indices of the averaging operators, so that we put

$$A_k f(r) = \sum_{p^k t = r} f(t)\,.$$

The two formulas then read

$$A_{k+l} = A_k \circ A_l$$
$$A_k(f\circ p) = pA_{k-1}f\,.$$

Theorem 2.1. *Suppose that f has values in a vector space V over K. Assume that there are elements $a, b, c\in K$ such that*

$$A_1 f = af - bf\circ p - c\,.$$

Let $\rho, \bar{\rho}$ be the roots of the equation

$$X^2 - aX + bp = 0\,.$$

Define

$$\psi_n(x) = \rho^{-n}[\rho f(x) - bf(px) + u]$$

where

$$u = \frac{c}{(p/\rho)-1}\,.$$

Then $\{\psi_n\}$ defines a distribution on the projective system

$$\left\{\frac{1}{m_0 p^n}\mathbf{Z}/\mathbf{Z}\right\}$$

with values in $K(\rho)\otimes V$.

Proof. This is immediate by using the hypothesis, and computing the average

$$\sum_{i=0}^{p-1} \psi_{n+1}\left(\frac{x+i}{p}\right).$$

We use $a\rho - bp = \rho^2$, and the value for u is chosen so that the distribution relation comes out.

Distributions of the above type occur in the work of Manin [Man 4], [Man 5]. They apply directly to Theorem 3.1 of Chapter IV. They also occur in Mazur-Swinnerton-Dyer [Ma-SwD], see Example 2 below. The distribution value can also be obtained as a limit in the following manner.

For the rest of this section, we assume that V is a Banach space over the complete valuated field K, and that the valuation is p-adic, that is p lies in the maximal ideal of the valuation. We let ord *denote the order of an element at the valuation, so that if $|a|=|p|^{\lambda}$ then* ord $a=\lambda$.

Theorem 2.2. *Assume that there are elements $a, b, c \in K$ such that*

$$A_1 f = af - bf \circ p - c .$$

Let $\rho, \bar{\rho}$ be the roots of the equation

$$X^2 - aX + bp = 0 .$$

Assume that

$$\text{ord } \rho < \text{ord } \bar{\rho} \quad \text{and} \quad \text{ord } \rho \leqslant 0 ,$$

(in other words, ρ has no zero and has a bigger pole than $\bar{\rho}$). Let $\alpha_k = \rho^k + \bar{\rho}^k$. Then:

(i) *The sequence*

$$\left\{\frac{A_k f(x)}{\alpha_k}\right\}$$

converges for each x.

(ii) *If we define for each n the function*

$$\varphi_n(x) = \rho^{-n} \lim A_k f(x)/\alpha_k$$

then the family $\{\varphi_n\}$ defines a distribution on the projective system

$$\left\{\frac{1}{m_0 p^n}\mathbf{Z}/\mathbf{Z}\right\}$$

Proof. Under the assumption of the theorem, we have

$$A_{k+1} f = A_k(A_1 f) = aA_k f - bpA_{k-1} f - p^k c$$

For simplicity of notation, we abbreviate $A_k f$ by $\bar{a}_k$ for the rest of the proof, so we can write symbolically

$$a_{k+1} = a\bar{a}_k - bp\bar{a}_{k-1} - p^k c .$$

From the quadratic equation for α_k we also obtain trivially the difference equation

$$\alpha_{k+1} = a\alpha_k - bp\alpha_{k-1} .$$

Then

$$\begin{aligned}
\alpha_{k+1}\left(\frac{a_{k+1}}{\alpha_{k+1}} - \frac{a_k}{\alpha_k}\right) &= a_{k+1} - \frac{\alpha_{k+1}}{\alpha_k} a_k \\
&= a_{k+1} - \frac{a\alpha_k - bp\alpha_{k-1}}{\alpha_k} a_k \\
&= -bp\bar{a}_{k-1} + bp\frac{\alpha_{k-1}}{\alpha_k}\bar{a}_k + O(p^k c) \\
&= bp\left(\frac{\alpha_{k-1}}{\alpha_k}\bar{a}_k - \bar{a}_{k-1}\right) + O(p^k c) \\
&= bp\alpha_{k-1}\left(\frac{a_k}{\alpha_k} - \frac{a_{k-1}}{\alpha_{k-1}}\right) + O(p^k c) .
\end{aligned}$$

We divide both sides by α_{k+1}. Since ρ has no zero, this only improves the error term $O(p^k c)$. Furthermore, we have

$$\rho\bar{\rho} = bp .$$

Consequently

$$bp\frac{\alpha_{k-1}}{\alpha_{k+1}} = pb\frac{\rho^{k-1} + \bar{\rho}^{k-1}}{\rho^{k+1} + \bar{\rho}^{k+1}}$$

behaves p-adically like $\rho\bar{\rho}/\rho^2$, which has a zero at p. This implies that the successive differences of terms in our sequence approach zero by a constant factor, and hence that the sequence converges. This proves (i).

To verify (ii) is trivial, because the power ρ^{-n} was chosen to make the answer come out right, namely:

$$\sum_{pt=r} \varphi_{n+1}(t) = A_1\left(\rho^{-n-1} \lim \frac{A_k f}{\alpha_k}\right)(r) = \lim \frac{a_{k+1}}{\alpha_{k+1}} \frac{\alpha_{k+1}}{\alpha_k} \frac{1}{\rho^{n+1}} = \varphi_n(r) .$$

This proves the theorem.

In special cases, we wish to compute the limiting value. This amounts to solving explicitly the difference equation satisfied by the A_k, so we make general comments on such equations which are classical. We first mention the homogeneous case.

Suppose given numbers $u_1, \ldots, u_d$. For $n \geqslant d$ we want to solve the system

$$a_n = u_1 a_{n-1} + \cdots + u_d a_{n-d} .$$

We consider the characteristic equation of the system, namely

$$X^d = u_1 X^{d-1} + \cdots + u_d .$$

Let ρ be a root of this equation. Then it is clear that putting

$$a_n=\rho^n$$

gives a solution of the difference equations. Since the solutions form a vector space, we see that if $\rho_1,\cdots,\rho_d$ are distinct roots of the polynomial, then the most general solution of the equation is given by

$$a_n=B_1\rho_1^n+\cdots+B_d\rho_d^n,$$

with some constants $B_1,\cdots,B_d$.

We also want to solve the inhomogeneous system as in Theorem 2.1. We can use an alternate formalism, say with $d=2$ to avoid too many indices. I am indebted to Sommese for pointing out that the technique below used to be drilled into people by books like Hardy's *Pure Mathematics*.

Thus we suppose given a_0, a_1 (initial conditions), and we wish to find the solutions of the system

$$a_k=aa_{k-1}-bpa_{k-2}-p^kc, \qquad k\geqslant 2\,.$$

We consider the formal power series

$$F(T)=\sum_{k=0}^{\infty}a_kT^k\,.$$

We may rewrite this series as

$$F(T)=\sum_{k=2}^{\infty}(aa_{k-1}-bpa_{k-2})T^k+a_0+a_1T-\sum_{k=2}^{\infty}p^kcT^k\,.$$

After shifting indices, this yields

$$F(T)(1-aT+bpT^2)=a_0+(a_1-aa_0)T-\frac{c}{1-pT}\cdot$$

Assume $\rho\neq\bar\rho$, $\rho\neq p$, $\bar\rho\neq p$. Since

$$1-aT+bpT^2=(1-\rho T)(1-\bar\rho T)$$

we obtain the partial fraction decomposition

$$F(T)=\frac{A}{1-\rho T}+\frac{B}{1-\bar\rho T}+\frac{C}{1-pT}$$

with constants A, B, C. Consequently the solution of the difference equation is given by

$$a_k=A\rho^k+B\bar\rho^k+Cp^k\,, \qquad \text{for} \quad k\geqslant 0\,.$$

Since the constants A, B, C can be determined explicitly from the initial conditions, the limit can be determined explicitly in Theorem 2.1. We give examples.

Example 1. Suppose that $c=0$ (homogeneous system). Then $C=0$ and

$$a_k=A\rho^k+B\bar\rho^k.$$

We have

$$a_0=A+B$$
$$a_1=A\rho+B\bar\rho$$

from which we solve:

$$A=\frac{1}{\bar\rho-\rho}(\bar\rho a_0-a_1)$$

$$B=\frac{1}{\rho-\bar\rho}(\rho a_0-a_1)\,.$$

Specialize this to the specific case of the theorem. We have

$$a_0=f(x) \qquad \text{and} \qquad a_1=af(x)-bf(px)\,.$$

Also

$$a=\rho+\bar\rho\,.$$

Therefore

$$A=\frac{1}{\bar\rho-\rho}[-\rho f(x)+bf(px)]$$

$$B=\frac{1}{\rho-\bar\rho}[-\bar\rho f(x)+bf(px)]\,.$$

Since we assume that ρ has a bigger pole than $\bar\rho$, the terms $B\bar\rho^k/\alpha_k$ approach 0 as $k\to\infty$. Furthermore, $\lim\rho^k/\alpha^k=1$. Hence we find the value

$$\boxed{\lim\frac{A_kf(x)}{\alpha_k}=A=\frac{1}{\rho-\bar\rho}[\rho f(x)-bf(px)]\,,}$$

whence the value for the measure at level n:

Corollary. *In Theorem* 2.2, *we have when* $c=0$,

$$\varphi_n(x)=\rho^{-n}\frac{1}{\rho-\bar{\rho}}[\rho f(x)-bf(px)]\,.$$

This is the measure which came into Manin's work [Man 4], [Man 5], although Manin phrases the measure as existing only on the multiplicative group.

Example 2. This example occurs in the work of Mazur and Swinnerton-Dyer [Ma-SwD]. We suppose that

$$A_1 f = af - f\circ p - A_1 f(0)\,,$$

so that $b=1$, and $c=A_1 f(0)$. Suppose also that a is a p-adic unit, so that the root ρ is a p-adic unit. Let

$$S_k = A_k f(0) - A_{k-1} f(0) = \sum_{x\ \mathrm{prim}} f(x)$$

where the sum is taken over the primitive x of period p^k. Thus

$$S_1 = A_1 f(0) - f(0)\,.$$

Under the assumption that $f(0)=0$, so $S_0=0$, we have

$$\boxed{\lim \rho^{-k}S_k = \lim \alpha_k^{-1}S_k = \frac{-N_p S_1}{(\rho^2-p)(1-\bar{\rho})^2}}$$

where $N_p = 1+p-a$.

Proof. Since $\bar{\rho}$ is divisible by p, it is clear that the limits

$$\lim \rho^{-k}S_k \qquad \text{and} \qquad \lim \alpha_k^{-1}S_k$$

are equal. The numbers S_k satisfy the difference equation

$$S_{k+1} = aS_k - pS_{k-1} + (p-1)p^{k-1}S_1\,.$$

One can then solve for S_k of the form

$$S_k = (A\rho^k + B\bar{\rho}^k + Cp^k)S_1\,.$$

One needs

$$A+B+C=0$$
$$A\rho+B\bar{\rho}+C=1\,.$$

One finds from the recursion relation that

$$C=\frac{p-1}{pN_p}\,,$$

and then one solves for A, B, e.g.

$$A=\frac{p-1}{(\rho^2-p)(1-\bar{\rho})}\,.$$

The value for the limit then falls out trivially as before.

In this example, where we took the sum over primitive elements, we can interpret the limit as the measure of the multiplicative group, that is:

$$\boxed{\lim \rho^{-k}S_k = \int_{\mathbf{Z}_p^*} d\varphi\,.}$$

§ 3. The Iwasawa Algebra

Distributions on projective systems usually arise in the context of a projective limit of finite groups, so we make additional remarks pertaining to this situation.

Let G be a compact group, projective limit

$$G=\lim G/G_n\,,$$

where G_n is an open subgroup of finite index. For example,

$$\mathbf{Z}_p = \lim \mathbf{Z}/p^n\mathbf{Z}$$
$$GL_2(\mathbf{Z}_p) = \lim GL_2(\mathbf{Z}/p^n\mathbf{Z})\,.$$

Let $\mathfrak{o}$ be the ring of integers in some finite extension of $\mathbf{Q}_p$. Let $\{\varphi_n\}$ be a distribution on the system $\{G/G_n\}$ with values in $\mathfrak{o}$. Then let θ_n be the element of the group algebra $\mathfrak{o}[G/G_n]$ given by

$$\theta_n = \sum_{x\in G/G_n} \varphi_n(x)x\,.$$

The canonical homomorphism

$$G/G_{n+1} \to G/G_n$$

induces an algebra homomorphism

$$\mathfrak{o}[G/G_{n+1}] \to \mathfrak{o}[G/G_n] ,$$

and we get a commutative diagram

$$\begin{array}{ccc} \varphi_{n+1} & \mapsto & \theta_{n+1} \\ \downarrow & & \downarrow \\ \varphi_n & \mapsto & \theta_n \end{array}$$

or in terms of sets,

$$\begin{array}{ccc} \text{Distr}\,(G/G_{n+1}, \mathfrak{o}) & \to & \mathfrak{o}[G/G_{n+1}] \\ \downarrow & & \downarrow \\ \text{Distr}\,(G/G_n, \mathfrak{o}) & \to & \mathfrak{o}[G/G_n] \end{array}$$

where the horizontal arrows are isomorphisms, and the vertical arrows are: On the left, the averaging map, and on the right the map obtained by reduction, i.e. the canonical map. Thus the algebra of distribution on $\{G/G_n\}$ is isomorphic to the projective limit

$$\text{Distr}\,(G, \mathfrak{o}) \to \lim \mathfrak{o}\,[G/G_n] ,$$

which we denote by $\mathfrak{o}[[G]]$. This notation is justified by the following example with $G = \mathbf{Z}_p$.

Example. Let $\mathbf{Z}_p/p^n\mathbf{Z}_p \approx C_{p^n}$, where C_{p^n} is a multiplicative cyclic group of order p^n. Then we have a commutative diagram

$$\begin{array}{ccc} \mathfrak{o}[C_{p^{n+1}}] & \to & \mathfrak{o}[X]/(X^{p^{n+1}} - 1) \\ \downarrow & & \downarrow \\ \mathfrak{o}[C_{p^n}] & \to & \mathfrak{o}[X]/(X^{p^n} - 1) \end{array}$$

Write $X = T+1$ for another variable T. Then $\mathfrak{o}[X] = \mathfrak{o}[T]$, and

$$\mathfrak{o}[X]/(X^{p^n} - 1) \approx \mathfrak{o}[T]/((T+1)^{p^n} - 1) .$$

A trivial induction shows that

$$(T+1)^{p^n} - 1 \in (p, T)^{n+1}$$

(the right-hand side is the ideal generated by p, T, raised to the $(n+1)$-th power). But we have

$$\lim \mathfrak{o}[T]/(p, T)^n = \mathfrak{o}[[T]] ,$$

in other words, the projective limit of $\mathfrak{o}[T]/(p, T)^n$ is the power series ring in the variable T over $\mathfrak{o}$. This gives us an isomorphism

$$\boxed{\text{Distr}\,(\mathbf{Z}_p, \mathfrak{o}) \approx \mathfrak{o}[[T]]}$$

of the distributions on $\mathbf{Z}_p$ with the power series ring. A given distribution can be written as a power series. We shall see in the next section that such a power series has an associated Weierstrass polynomial. In cases arising naturally in number theory, it is then interesting to study the roots of such polynomials, which are p-adic analogues of zeta functions.

Of course, instead of $\mathbf{Z}_p$ we could take a finite product $\mathbf{Z}_p^d$ for some positive integer d, in which case exactly the same analysis as above shows that we have an isomorphism

$$\boxed{\text{Distr}\,(\mathbf{Z}_p^d, \mathfrak{o}) \approx \mathfrak{o}[[T_1, \ldots, T_d]]}$$

with the power series in d variables.

§ 4. Weierstrass Preparation Theorem

The proof of the Weierstrass theorem in this section is due to Manin [Man 1]. We start with the **Euclidean algorithm**.

Theorem 4.1. *Let $\mathfrak{o}$ be a complete local ring with maximal ideal $\mathfrak{m}$. Let*

$$f(T) = \sum_{i=0}^{\infty} a_i T^i$$

be a power series in $\mathfrak{o}[[T]]$, such that not all a_i lie in $\mathfrak{m}$. Say $a_0, \ldots, a_{n-1} \in \mathfrak{m}$, and $a_n \in \mathfrak{o}^*$ is a unit. Given $g \in \mathfrak{o}[[T]]$ we can solve the equation

$$g = qf + r ,$$

with $q \in \mathfrak{o}[[T]]$, $r \in \mathfrak{o}[T]$, and $\deg r \leqslant n-1$.

Proof. Let α and τ be the projections on the beginning and tail end of the series, given by

$$\alpha: \sum b_i T^i \mapsto \sum_{i=0}^{n-1} b_i T^i$$

$$\tau: \sum b_i T^i \mapsto \sum_{i=n}^{\infty} b_i T^{i-n} .$$

The existence of q, r is equivalent with the condition that there exists q such that

$$\tau(g) = \tau(qf) .$$

But

$$f = \alpha f + T^n \tau(f) .$$

Hence our problem is equivalent with solving

$$\tau(g) = \tau(q\alpha(f)) + \tau(qT^n\tau(f)) = \tau(q\alpha(f)) + q\tau(f) .$$

Note that $\tau(f)$ is invertible. Put $Z = q\tau(f)$. Then the above equation is equivalent with

$$\tau(g) = \tau\left(Z\frac{\alpha(f)}{\tau(f)}\right) + Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)Z .$$

Note that

$$\tau \circ \frac{\alpha(f)}{\tau(f)} : \mathfrak{o}[[T]] \to \mathfrak{m}\mathfrak{o}[[T]] ,$$

because $\alpha(f)/\tau(f) \in \mathfrak{m}\mathfrak{o}[[T]]$. We can therefore invert to find Z, namely

$$Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)^{-1} \tau(g) ,$$

which proves both existence and uniqueness and concludes the proof.

Theorem 4.2. (*Weierstrass Preparation*) *The power series f in the previous theorem can be written in the form*

$$f(T) = (T^n + b_{n-1}T^{n-1} + \cdots + b_0)u ,$$

where $b_i \in \mathfrak{m}$, and u is a unit in $\mathfrak{o}[[T]]$.

Proof. Write

$$T^n = qf + r ,$$

by the Euclidean algorithm. Then q is invertible because

$$q = c_0 + c_1 T + \cdots$$
$$f = \cdots + a_n T^n + \cdots$$

so that

$$1 \equiv c_0 a_n \pmod{\mathfrak{m}} ,$$

and c_0 is a unit in $\mathfrak{o}$. We obtain $qf = T^n - r$, and

$$f = q^{-1}(T^n - r) ,$$

with $r \equiv 0 \pmod{\mathfrak{m}}$. This proves the theorem.

The integer n in Theorems 4.1 and 4.2 is called the **Weierstrass degree** of f. We see that a power series not all of whose coefficients lie in $\mathfrak{m}$ can be expressed as a product of a polynomial having the given Weierstrass degree, times a unit in the power series ring. Furthermore, all the coefficients of the polynomial except the leading one lie in the maximal ideal. Such a polynomial is called **distinguished**.

§ 5. Modules over $\mathbf{Z}_p[[T]]$

The structure of finitely generated modules over $\mathbf{Z}_p[[T]]$ was first determined by Iwasawa [Iw 1], [Iw 2], in connection with towers of cyclotomic fields and the behavior of the p-primary part of the ideal class groups. Serre [Se 1] gave a proof by commutative algebra. In this section we give an elementary proof along the standard lines of row and column operations, due to Paul Cohen.

We let $\Lambda = \mathfrak{o}[[T]]$, where $\mathfrak{o}$ is a complete discrete valuation ring. We denote by p a prime element of $\mathfrak{o}$. By a **finite module** over $\mathfrak{o}$ we mean a finitely generated module annihilated by some power p^k and some distinguished element λ. If $\mathfrak{o} = \mathbf{Z}_p$, then "finite" has its usual meaning.

Theorem 5.1. *Let M be a finitely generated module over Λ. There exists a homomorphism*

$$M \to M'$$

with finite kernel and cokernel, such that

$$M' \approx \Lambda^{(r)} \oplus \prod \Lambda/p^{n_i} \oplus \prod \Lambda/(f_j^{m_j})$$

where each f_j is a distinguished polynomial, irreducible in $o[T]$, i, j range over finite sets of indices, and $\Lambda^{(r)}$ is the product of Λ taken r times, for some integer r.

The rest of this section is devoted to the proof.

Suppose that M has generators $u_1, \ldots, u_n$. Relative to such generators we can form the matrix of relations, whose rows are vectors

$$(\lambda_1, \ldots, \lambda_n)$$

such that

$$\lambda_1 u_1 + \cdots + \lambda_n u_n = 0 .$$

Since Λ is Noetherian, a finite number of the rows generate all of them.

Performing the usual row and column operations on the matrix amounts to changing the generators of the module. We shall describe other operations, corresponding to embedding the module in a bigger one with finite cokernel.

An element $\lambda \in \Lambda$ is called p-**free** if λ does not lie in $p\Lambda$, in other words, if we can apply the Weierstrass preparation theorem to it.

Suppose that there is a relation of the form

$$\lambda_1 u_1 + p(\lambda_2 u_2 + \cdots + \lambda_n u_n) = 0 ,$$

where λ_1 is p-free. We can form the new module M' obtained by adjoining a new generator v with the relations

$$pv = u_1 , \qquad \lambda_1 v = -(\lambda_2 u_2 + \cdots + \lambda_n u_n) .$$

This can be formalized by considering a direct sum

$$M \oplus (v)$$

modulo the desired relations, i.e. modulo the submodule generated by the elements

$$(0, pv) - (u_1, 0) \qquad \text{and} \qquad (0, \lambda_1 v) - (\lambda_2 u_2 + \cdots + \lambda_n u_n, 0) .$$

It is then immediately verified that the canonical map of M into the factor module is injective. The factor module M'/M is annihilated by p and λ_1, whence is finite. Furthermore, the elements $v, u_2, \ldots, u_n$ generate M', and have the relation

$$\lambda_1 v + \lambda_2 u_2 + \cdots + \lambda_n u_n = 0 .$$

In terms of the relation matrix, this means that we shall allow the following operation, replacing the matrix R by a matrix R'.

O 1. *If R contains a row $(\lambda_1, p\lambda_2, \ldots, p\lambda_n)$ with λ_1 not divisible by p, then we let R' be the matrix whose rows are generated by*

$$(\lambda_1, \ldots, \lambda_n)$$

and the rows of R with first element multiplied by p.

Observe that in this first operation, we may have $\lambda_2 = \ldots = \lambda_n = 0$.

Next suppose that some power p^k $(k \geqslant 1)$ divides all elements of R, but that there exists one relation

$$p^k(\lambda_1, \ldots, \lambda_n)$$

such that λ_1 is distinguished (or equivalently, λ_1 is not divisible by p). We may then form the module M' obtained by adjoining a new element v with the relations

$$p^k v = p^k u_1 \qquad \text{and} \qquad \lambda_1 v = -(\lambda_2 u_2 + \cdots + \lambda_n u_n) .$$

Again, it is easily verified that M is embedded in M' and that M'/M is finite. Note that $p^k(v - u_1) = 0$. The relations of the submodule

$$(v, u_2, \ldots, u_n)$$

are generated by R and the additional relation

$$(\lambda_1, \ldots, \lambda_n) .$$

We have a direct sum decomposition

$$M' = (v, u_2, \ldots, u_n) \oplus (v - u_1) ,$$

and the relations of $v - u_1$ are generated by p^k. To prove the theorem, it suffices to consider the first component of M'. Thus our second operation is described as follows.

O 2. *If all elements of R are divisible by p^k, and if there exists one relation $(p^k\lambda_1, \ldots, p^k\lambda_n)$ such that λ_1 is not divisible by p, then we let R' be generated by R and the new row*

$$(\lambda_1, \ldots, \lambda_n) .$$

In order to prove the theorem, it will now suffice to prove that making ordinary row and column operations, and **O 1, O 2** (these being called **admissible operations**) we can obtain a matrix which is essentially diagonalized, in a manner exhibiting the structure of the module as a direct sum as stated in the theorem. This is achieved by the next lemmas.

Lemma 1. *By admissible operations, the matrix R can be transformed into a matrix R' of the form*

$$\begin{pmatrix} \lambda_1 & 0 \dots 0 & 0 \dots 0 \\ \vdots & \dots & \vdots \\ 0 & 0 \dots \lambda_r & 0 \dots 0 \\ * & * \dots * & 0 \dots 0 \\ * & * \dots * & 0 \dots 0 \end{pmatrix}$$

where $\lambda_1, \dots, \lambda_r$ are distinguished.

Proof. By **O 2** we may assume that there is some distinguished element in R. Among all admissible transformations of R, we select one having as some component a distinguished polynomial of minimal degree. By row and column operations, we may assume that this element is in the upper left hand corner, and we denote it by λ_1, so that the matrix looks like this.

$$\begin{pmatrix} \lambda_1 & \lambda_2 \dots \lambda_n \\ * & * \dots * \\ . & . \dots . \\ . & . \dots . \\ . & . \dots . \end{pmatrix}$$

For $j \geqslant 2$ we apply the Euclidean algorithm $\lambda_j = q\lambda_1 + r_j$. If r_j is not divisible by p, we contradict the minimality of the degree of λ_1 unless $r_j = 0$. Without loss of generality, we may therefore assume that $\lambda_2, \dots, \lambda_n$ are either 0 or divisible by p. Using **O 1** repeatedly, we may divide $\lambda_2, \dots, \lambda_n$ by the maximal power of p dividing them all. Thus without loss of generality, we may assume that in fact $\lambda_j = 0$ for $j \geqslant 2$. Our matrix now looks like this.

$$\begin{pmatrix} \lambda_1 & 0 & \dots 0 \\ * & \lambda_{22} \dots \lambda_{2n} \\ * & \lambda_{32} \dots \lambda_{3n} \\ * & . & \dots . \end{pmatrix}$$

Consider the submatrix

$$\begin{pmatrix} \lambda_{22} \dots \lambda_{2n} \\ \lambda_{32} \dots \lambda_{3n} \\ . \quad \dots . \end{pmatrix}$$

If all entries are 0 we are done. If not, suppose that p^k divides all components, and p^k divides one component exactly. After row-column operations, we may assume that p^k divides λ_{22} exactly. By repeated use of **O 1** (with respect to the row

$$(\lambda_1, 0, \dots, 0)),$$

we may assume without loss of generality that λ_{21} is divisible by p^k. Using **O 2**, we we may assume that $k = 0$, i.e. that λ_{22} is distinguished, i.e. the second row is

$$(\lambda_{21}, \lambda_{22}, \dots, \lambda_{2n}),$$

and λ_{22} is not divisible by p. Using the Euclidean algorithm, we may also assume that λ_{2j} is divisible by p for $j > 2$. Furthermore, by the Euclidean algorithm, we may assume that

$$\deg \lambda_{21} < \deg \lambda_{22}.$$

By **O 1**, we may assume that p divides λ_{21}, so $p \mid \lambda_{2j}$ except for $j = 2$. We now apply **O 1** with respect to the second column, i.e. with λ_{22} as our distinguished element, so that we can divide λ_{2j} by p for $j \neq 2$. Iterating this procedure, we may assume that

$$\lambda_{23} = \dots = \lambda_{2n} = 0.$$

By using the Euclidean algorithm $\lambda_{21} = q\lambda_1 + r_1$, and the minimality of the degree of λ_1, we may assume that λ_{21} is divisible by p. We can use **O 1** to replace the second row by

$$(\lambda_{21}/p, \lambda_{22}, 0, \dots, 0).$$

Iterating this procedure, we see that we can assume that $\lambda_{21} = 0$. Thus our matrix of relations may be assumed of the form

$$\begin{pmatrix} \lambda_1 & 0 & \dots 0 \\ 0 & \lambda_2 \dots 0 \\ * & * & * & * & * & * & * \end{pmatrix}$$

Iterating this whole procedure ultimately leads to a matrix as stated in the lemma, thus concluding the proof.

If we let $\lambda = \lambda_1 \dots \lambda_r$, then we see that we may use the same element λ in the diagonal of the top part of the relation matrix. In terms of the module, this means that $\lambda M = 0$, and λ may be assumed distinguished.

Lemma 2. *Suppose that R has the form*

$$\begin{pmatrix} \lambda & 0 \dots 0 \\ 0 & \lambda \dots 0 \\ \dots & \dots \\ 0 & 0 \dots \lambda \\ * & * \dots * \end{pmatrix}$$

where λ is distinguished. Then by admissible transformations we may change R to a matrix of the form

$$\begin{pmatrix} \lambda & 0 \dots 0 \\ \vdots & \vdots \\ 0 & 0 \dots \lambda \\ \lambda_{11} & 0 \dots 0 \\ \vdots & \vdots \\ 0 & 0 \dots \lambda_{rr} \\ * & * \dots * \end{pmatrix}$$

where $\lambda_{ii} = p^{k_i}\lambda'_{ii}$ and λ'_{ii} is distinguished. Furthermore, any relation $(\varphi_1, \dots, \varphi_r)$ in the new matrix is such that

$$\lambda'_{ii} \text{ divides } \varphi_i .$$

Proof. Among all relations in R, select one having a component of minimal degree (not necessarily distinguished). After interchanging rows and columns, we may assume that this relation is

$$(\lambda_{11}, \dots, \lambda_{1n}) ,$$

and

$$\deg \lambda_{11} \leqslant \deg \varphi$$

for any component φ of R. Let p^k divide λ_{11} exactly. Using **O 1** with respect to the relations

$$(0, \dots, \lambda, \dots, 0)$$

we may assume that p^k divides λ_{1j} for $j \geqslant 2$. We may now apply the Euclidean algorithm

$$p^{-k}\lambda_{1j} = qp^{-k}\lambda_{11} + r$$

and use the minimality of the degree of λ_{11} to conclude that $\lambda_{1j} = 0$ if $j \geqslant 2$. The same argument can again be applied to the other columns to get a matrix having the shape stated in the lemma.

If $(\varphi_1, \dots, \varphi_n)$ is any relation, then so is $p^{k_1}(\varphi_1, \dots, \varphi_n)$. Using the Euclidean algorithm

$$p^{k_1}\varphi_1 = qp^{k_1}\lambda'_{11} + r ,$$

we conclude that $r = 0$ and λ'_{11} divides φ_1. Similarly, $\lambda'_{ii} \mid \varphi_i$, thus proving the lemma.

We return to the module interpretation to see that Lemma 2 implies the theorem. Let M be the module defined by the relations of Lemma 2. Let N be the submodule generated by

$$\lambda'_{11}u_1, \dots, \lambda'_{rr}u_r .$$

Then N is finite because $p^k N = 0$ (take $k = \max k_i$), and also $\lambda N = 0$. Furthermore

$$M/N \approx \prod_{i=1}^{r} \Lambda/(\lambda'_{ii}) .$$

Finally, if f, g are distinguished and relatively prime, the map

$$\Lambda/(fg) \to \Lambda/f \oplus \Lambda/g$$

is an embedding with finite cokernel. This allows us to decompose the factors Λ/λ'_{ii} into a direct sum of factors

$$\Lambda/(f_j^{m_j})$$

where f_j is distinguished and irreducible, thereby concluding the proof.

Chapter XIII. Bernoulli Numbers and Polynomials

The oldest distribution is that defined by the Bernoulli polynomials, although of course their classical recurrence property was not called by that name.

Iwasawa pointed out that certain results of Leopoldt [Le 1], [Le 2], [Le 3], and Kubota [Le–Ku 1] could be formulated in terms of distributions. The idea of Leopoldt was that the values of the zeta function taken at negative integers were also the values of a p-adic analytic function, which could be interpreted as the p-adic zeta function, and similarly for L-series. Mazur showed how this p-adic zeta function could be written as a p-adic Mellin transform with respect to a p-adic measure, obtained as a distribution constructed with the Bernoulli numbers, and analogous to the Haar measure on the multiplicative group. This is carried out in the present chapter, and is a reformulation of Iwasawa's work.

The exposition follows Mazur [Maz 2] to a large extent, and also greatly benefited from Serre's treatment in his course at the College de France, 1972.

For direct applications of the topics of this chapter, see also [K–L 2].

§ 1. Bernoulli Numbers and Polynomials

Let

B1. $$F(t)=\frac{t}{e^t-1}=\sum_{k=0}^{\infty}B_k\frac{t^k}{k!}$$

define the **Bernoulli numbers** B_k for $k\geqslant 0$. Then for instance

$$B_0=1,\qquad B_1=-\tfrac{1}{2},\qquad B_2=\tfrac{1}{6}.$$

Observe that

$$F(-t)-F(t)=t,$$

so that F is almost even, and in particular, we have

$$B_k=0 \quad \text{if} \quad k \text{ is odd}, k\neq 1.$$

We shall now define **polynomials** $B_k(X)$ whose constant terms are the Bernoulli numbers. We let

B2. $$F(t,X)=\frac{te^{tX}}{e^t-1}=\sum_{k=0}^{\infty}B_k(X)\frac{t^k}{k!}=F(t)e^{Xt}.$$

It is then clear that

$$B_k(0)=B_k.$$

(There will not be any confusion of meaning, the context will always make it clear whether B_k is the Bernoulli number or polynomial.) We find:

$$B_0(X)=1,\qquad B_1(X)=X-\tfrac{1}{2},\qquad B_2(X)=X^2-X+\tfrac{1}{6}.$$

Let f be a function on $\mathbf{Z}/N\mathbf{Z}$ for some integer N. We find it convenient to abbreviate

$$\mathbf{Z}/N\mathbf{Z}=Z(N).$$

We define

B3. $$F_f^{(N)}(t,X)=\sum_{a=0}^{N-1}f(a)\frac{te^{(a+X)t}}{e^{Nt}-1}$$
$$=\frac{1}{N}\sum_{a=0}^{N-1}f(a)F\left(Nt,\frac{X+a}{N}\right),$$

and we define $B_{k,f}^{(N)}(X)$ by

B4. $$F_f^{(N)}(t,X)=\sum_{k=0}^{\infty}B_{k,f}^{(N)}(X)\frac{t^k}{k!}.$$

Remark. Let $M\mid N$ and suppose that f is factorizable through the canonical homomorphism

$$Z(N)\to Z(M).$$

Then

$$F_f^{(N)}(t,X)=F_f^{(M)}(t,X).$$

Proof. Write each $a=b+qM$ with $0\leqslant b<M$ and $0\leqslant q\leqslant\frac{N}{M}-1$. Let $D=N/M$. Then

$$F_f^{(N)}(t,X)=\sum_{b=0}^{M-1}f(b)\sum_{q=0}^{D-1}te^{(b+X)t}\frac{e^{qMt}}{e^{Nt}-1}=F_f^{(M)}(t,X).$$

Directly from the definition of the polynomials $B_{k,f}(X)$ we find the compatibility relation

B5. $$B_{k,f}^{(N)}(X)=N^{k-1}\sum_{a=0}^{N-1} f(a)B_k\left(\frac{X+a}{N}\right).$$

For $f=1$, this yields in particular

B6. $$B_k(X)=N^{k-1}\sum_{a=0}^{N-1} B_k\left(\frac{X+a}{N}\right)$$

and for $X=0$,

B7. $$B_{k,f}^{(N)}=N^{k-1}\sum_{a \bmod N} f(a)B_k\left(\left\langle\frac{a}{N}\right\rangle\right)$$

where $\langle t\rangle$ denotes the smallest real number $\geqslant 0$ in the residue class of $t \bmod \mathbf{Z}$.

The relation **B6** can also be written in the form

B8. $$N^{k-1}\sum_{x \bmod N} B_k\left(\left\langle y+\frac{x}{N}\right\rangle\right)=B_k(\langle Ny\rangle)$$

for $y\in\mathbf{R}/\mathbf{Z}$. Indeed, the family of elements z in $\mathbf{R}/\mathbf{Z}$ such that

$$\langle Nz\rangle=\langle Ny\rangle$$

is precisely

$$y+\frac{x}{N}\ (\mathrm{mod}\ \mathbf{Z}), \qquad x=0,\ldots,N-1\ .$$

Representatives $\geqslant 0$ for these elements mod $\mathbf{Z}$ are

$$\frac{\langle Ny\rangle+x}{N}, \qquad x=0,\ldots,N-1\ ,$$

and our assertion follows.

The relation **B8** can be formulated as saying:

The function

$$x\mapsto N^{k-1}B_k\left(\left\langle\frac{x}{N}\right\rangle\right)$$

is a distribution on the projective system $\{\mathbf{Z}/N\mathbf{Z}\}$, *denoted by*

$$E_k^{(N)}=N^{k-1}B_k^{(N)}\ .$$

In view of this definition, we may write

B7′. $$B_{k,f}=\int_{Z(N)} f\,dE_k^{(N)}\ .$$

From the present point of view, we obtain a characterization of the Bernoulli polynomials.

Theorem 1.1. *For each integer k, there exists a unique monic polynomial $P_k(X)$, of degree k, with rational coefficients, such that for some prime p,*

$$x\mapsto p^{k-1}P_k\left(\left\langle\frac{x}{p^n}\right\rangle\right)$$

is a distribution on the projective system $\{\mathbf{Z}/p^n\mathbf{Z}\}$, and this polynomial is the Bernoulli polynomial.

The theorem will not be needed, and so will be left as an (easy) exercise.

We abandon for the moment the distribution properties of the Bernoulli polynomials, and derive more formulas.

Take $f=1$. Then we find

$$F(t,X+1)=e^tF(t,X)\ ,$$

so that

$$F(t,X+1)-F(t,X)=te^{Xt}=t\sum X^k\frac{t^k}{k!}\ .$$

From this we have the difference formula

B9. $$B_k(X+1)-B_k(X)=kX^{k-1}\ .$$

In particular,

$$B_k(1)=B_k(0) \quad \text{if} \quad k\geqslant 2\ .$$

This last relation is false for $k=1$.

The difference equation **B9** generalizes as follows. Suppose that f is a function on $Z(N)$. We have

$$F_f(t,X+k)=e^{kt}F_f(t,X)\ .$$

Hence

$$F_f(t,X+N)-F_f(t,X)=(e^{Nt}-1)F_f(t,X)\ ,$$

whence

B10. $$\frac{1}{k}[B_{k,f}(X+N)-B_{k,f}(X)]=\sum_{a=0}^{N-1} f(a)(a+X)^{k-1}.$$

Taking $X=0$ yields

B11. $$\frac{1}{k}[B_{k,f}(N)-B_{k,f}(0)]=\sum_{a=0}^{N-1} f(a)a^{k-1}.$$

We denote sums like the last sum on the right by

$$S_{f,k}(N)=\sum_{a=0}^{N-1} f(a)a^{k}.$$

Then **B11** can also be written

B11. $$\frac{1}{k}[B_{k,f}(N)-B_{k,f}(0)]=S_{f,k-1}(N).$$

We now derive expressions for the coefficients of the Bernoulli polynomials in terms of the Bernoulli numbers.

By definition,

$$F_f(t,X)=F_f(t)e^{Xt}=\sum B_{k,f}(X)\frac{t^k}{k!}$$

$$=\left(\sum B_{i,f}\frac{t^i}{i!}\right)\left(\sum X^j\frac{t^j}{j!}\right).$$

Multiplying out yields

B12. $$B_{k,f}(X)=\sum_{i=0}^{k}\binom{k}{i}B_{i,f}X^{n-i}$$

$$=B_{k,f}+kB_{k-1,f}X+\cdots+kB_{1,f}X^{k-1}+B_{0,f}X^{k}.$$

In particular, for $f=1$, we find

B13. $$B_k(X)=B_k(0)+kB_{k-1}(0)X+\cdots+\left(-\frac{k}{2}\right)X^{k-1}+X^{k}.$$

As an application, we can find a p-adic characterization for the values of Bernoulli numbers in terms of sums $S_{k,f}$.

Theorem 1.2. *Let p be a prime number, let f be a function on $Z(c)$ for some positive integer c, with values in a Banach space over $\mathbf{Q}_p$. Then*

$$B_{k,f}=\lim_{\nu\to\infty}\frac{1}{cp^{\nu}}S_{k,f}(cp^{\nu}),$$

where the limit is p-adic.

Proof. Take $N=cp^{\nu}$. Let D be a common denominator for the coefficients of the polynomial $B_{k,f}(X)$. From **B12** we find

$$\frac{1}{k}[B_{k,f}(cp^{\nu})-B_{k,f}(0)]\equiv B_{k-1,f}cp^{\nu} \bmod \frac{1}{D}p^{2\nu}.$$

We divide by cp^{ν} and use **B11** to conclude the proof.

§ 2. The Multiplicative Distribution

Mazur has shown how to obtain a p-adic measure which plays the role of the measure on the multiplicative group in the real case. We describe his construction in this section, and give his application to the value of L-functions in the next section. It amounts to a measure theoretic formulation of Iwasawa's techniques.

We had defined

$$E_k^{(N)}=N^{k-1}B_k^{(N)}.$$

Let c be a rational number. For N prime to c (i.e. prime to the numerator and denominator of c) we define

$$E_{k,c}^{(N)}(x)=E_k^{(N)}(x)-c^kE_k^{(N)}(c^{-1}x),$$

for $x\in\mathbf{Z}/N\mathbf{Z}$. Multiplication by c or c^{-1} is well defined on $\mathbf{Z}/N\mathbf{Z}$, so our expression makes sense. We can write symbolically

$$E_{k,c}=E_k-c^kE_k\circ c^{-1},$$

without the superscript N.

Alternatively, we could fix a prime p, fix a positive integer m_0, and let

$$N=m_0p^n, \qquad n\to\infty.$$

Then we pick c prime to m_0p, $c\in\mathbf{Z}_p^*$ (a p-adic unit), and define $E_{k,c}^{(N)}$ by the same formula.

The distribution $E_{k,c}^{(N)}$ satisfies the following properties.

E0. $$E_{0,c}=0.$$

Obvious.

E1. $$E_{1,c}^{(N)}(x)=-\left(c\left\langle\frac{c^{-1}x}{N}\right\rangle-\left\langle\frac{x}{N}\right\rangle\right)+\frac{c-1}{2}$$

$$=-h_c(x)+\frac{c-1}{2}$$

where for $x \in \mathbf{Z}/N\mathbf{Z}$ we define

$$h_c(x) = c\left\langle \frac{c^{-1}x}{N} \right\rangle - \left\langle \frac{x}{N} \right\rangle .$$

Proof. We have

$$\begin{aligned} E_{1,c}^{(N)}(x) &= E_1^{(N)}(x) - cE_1^{(N)}(c^{-1}x) \\ &= B_1\left(\left\langle \frac{x}{N} \right\rangle\right) - cB_1\left(\left\langle \frac{c^{-1}x}{N} \right\rangle\right) \\ &= \left\langle \frac{x}{N} \right\rangle - \frac{1}{2} - c\left(\left\langle \frac{c^{-1}x}{N} \right\rangle - \frac{1}{2}\right) \end{aligned}$$

whence the assertion is clear.

E2. $$E_{k,c}^{(N)}(x) \equiv kx^{k-1}E_{1,c}(x) \mod \frac{N}{D(k)}\mathbf{Z}[c, 1/c] ,$$

where $D(k)$ is a least common multiple of the denominators of the coefficients of the polynomial $B_k(X)$.

Proof. Let $r(x)$ be the smallest integer $\geqslant 0$ in the residue class of $x \bmod N$. If $\mathbf{Z}_{(N)}$ denotes $\mathbf{Z}$ localized at the primes dividing N, then we identify

$$\mathbf{Z}_{(N)}/N\mathbf{Z}_{(N)} = \mathbf{Z}/N\mathbf{Z} .$$

We let

$$r(x) = a , \quad \text{so} \quad 0 \leqslant a < N .$$

Let

$$cb \equiv a \pmod N , \qquad 0 \leqslant b < N ,$$

so that for some N-integral y we have

$$cb = a + yN , \qquad y \in \mathbf{Z}[1/c] .$$

Then

$$\frac{b}{N} \equiv \frac{c^{-1}a}{N} \mod \mathbf{Z}[1/c]$$

so that

$$\frac{b}{N} = \frac{c^{-1}x}{N} \qquad \text{and} \qquad r(c^{-1}x) = b .$$

We get

$$cb = a + h_c(x)N ,$$

with

$$h_c(x) = \frac{cb}{N} - \frac{a}{N} = c\left\langle \frac{c^{-1}x}{N} \right\rangle - \left\langle \frac{x}{N} \right\rangle .$$

Using $B_k(X) = X^k - \frac{k}{2}X^{k-1} + \cdots$ from **B13**, we find

$$\begin{aligned} E_k(x) &= N^{k-1}B_k\left(\left\langle \frac{x}{N} \right\rangle\right) \\ &\equiv N^{k-1}\left[\left\langle \frac{x}{N} \right\rangle^k - \frac{k}{2}\left\langle \frac{x}{N} \right\rangle^{k-1}\right] \mod \frac{N}{D(k)} \\ &\equiv N^{k-1}\left[\frac{a^k}{N^k} - \frac{k}{2}\frac{a^{k-1}}{N^{k-1}}\right] \\ &\equiv \frac{a^k}{N} - \frac{k}{2}a^{k-1} . \end{aligned}$$

Then we get the congruences

$$\begin{aligned} E_k(x) - c^k E_k(c^{-1}x) &\equiv \frac{a^k}{N} - \frac{k}{2}a^{k-1} \\ &\quad - \frac{(a+h_c(x)N)^k}{N} + \frac{kc}{2}(a+h_c(x)N)^{k-1} \mod \frac{N}{D(k)}\mathbf{Z}[c, 1/c] \\ &\equiv -kh_c(x)a^{k-1} + k\frac{c-1}{2}a^{k-1} \\ &\equiv kx^{k-1}\left(-h_c(x) + \frac{c-1}{2}\right) \\ &\equiv kx^{k-1}E_{1,c}(x) \end{aligned}$$

as was to be shown.

Theorem 2.1. *The values of $E_{k,c}^{(N)}$ are N-integral.*

Proof. This is immediate from the congruence relation, replacing N by N^v for large v. Then $N^v/D(k)$ is N-integral, and $h_c(x)$ is obviously N-integral. If $2 \mid N$, then $c \equiv 1 \pmod 2$ so that $(c-1)/2$ is N-integral. If $2 \nmid N$, then $(c-1)/2$ is N-integral. This proves the theorem.

We may now pass to the limit. We take the projective system

$$\{\mathbf{Z}/p^n\mathbf{Z}\}$$

whose limit is $\mathbf{Z}_p$. For the applications, it is best to restrict the measure to the units $\mathbf{Z}_p^*$, and as said before, we take c to be a p-adic unit. We let

$$\chi_1 : \mathbf{Z}_p^* \to \mathbf{Z}_p^*$$

be the identity, so that

$$\chi_1^k(x) = x^k, \qquad \text{for} \quad x \in \mathbf{Z}_p^* .$$

Then we may express formula **E2** as follows.

Theorem 2.2. *Let c be a p-adic unit. Then*

$$E_{k,c}(x) = kx^{k-1}E_{1,c}(x) \quad \text{on} \quad \mathbf{Z}_p .$$
$$E_{k,c} = k\chi_1^{k-1}E_{1,c} \quad \text{on} \quad \mathbf{Z}_p^* .$$

We conclude the section with one more property used to define L-functions.

E3. *Let b, c be prime to N. Then*

$$E_{k,c} - b^k E_{k,c} \circ b^{-1} = E_{k,b} - c^k E_{k,b} \circ c^{-1} .$$

Proof. This is trivial by a direct application of the definitions.

§ 3. L-Functions and Bernoulli Numbers

We note that the measure $E_{k,c}$ on $\mathbf{Z}_p^*$ is invariant under multiplicative translations by units in $\mathbf{Z}_p^*$. This is clear by passing to finite levels mod $p^n\mathbf{Z}_p$.

Let ψ be a continuous homomorphism

$$\psi : \mathbf{Z}_p^* \to \mathbf{C}_p^* ,$$

where $\mathbf{C}_p$ is the completion of the algebraic closure of $\mathbf{Q}_p$. If $\psi(c) \neq 1$ we define

$$L(\psi) = L_c(\psi) = \frac{1}{k(\psi(c) - 1)} \int_{\mathbf{Z}_p^*} \psi\chi_1^{-k} \, dE_{k,c} .$$

The expression on the right is independent of k and c.

Proof. The independence from k comes from the relation

$$E_{k,c} = k\chi_1^{k-1}E_{1,c} .$$

To see the independence from c, we multiply **E3** with $\psi\chi_1^{-1}$, and we integrate. We change variables, letting

$$x \mapsto bx ,$$

and use $\chi_1(b) = b$. We then obtain the value

$$k(\psi(c) - 1)(\psi(b) - 1)L_c(\psi) ,$$

from the left hand side of **E3**. By symmetry, using the right hand side of **E3** we obtain the same value with c and b interchanged. This proves that $L_c(\psi)$ is independent of c.

Theorem 3.1. *Assume that the character $\psi : \mathbf{Z}_p^* \to \mathbf{C}_p^*$ has finite order. Then*

$$L(\psi\chi_1^n) = -\frac{B_{n,\psi}}{n} .$$

Proof. Take $k = n$. Then

$$L(\psi\chi_1^n) = \frac{1}{n(\psi(c)c^n - 1)} \int_{\mathbf{Z}_p^*} \psi \, dE_{n,c}$$

$$= \frac{1}{n(\psi(c)c^n - 1)} \int \left[\psi \, dE_n - \psi c^n \, d(E_n \circ c^{-1}) \right]$$

$$= \frac{1}{n(\psi(c)c^n - 1)} \left[\int \psi dE_n - \int \psi(x)c^n \, dE_n(c^{-1}x) \right].$$

In the last integral we let $x \mapsto cx$, so we pick up a factor $\psi(c)$ in front of that integral, while getting rid of c^{-1} inside. Thus

$$L(\psi\chi_1^n) = -\frac{1}{n} \int \psi dE_n = -\frac{1}{n} B_{n,\psi} ,$$

as was to be shown.

It is reasonable to write

$$L(\psi\chi_1^n) = L(1 - n, \psi).$$

With this notation, we see that Theorem 3.1 gives the p-adic analogue to the classical complex theorem giving the values of L-series at integers. This is Mazur's approach to results of Leopoldt [Le 1], and Kubota–Leopoldt [Ku–Le].

We shall now give some applications of the integral formalism to congruence relations between Bernoulli numbers.

Theorem 3.2. *Let $c \in \mathbf{Z}_p^*$. Let $1 \leqslant k$. Then*

$$\int_{\mathbf{Z}_p} x^{k-1}\, dE_{1,c}(x) = (1-c^k)B_k/k\ .$$

Proof. By Theorem 2.2, we have

$$\int_{\mathbf{Z}_p} x^{k-1}\, dE_{1,c}(x) = \frac{1}{k}\int_{\mathbf{Z}_p} dE_{k,c}$$

and

$$E_{k,c} = E_k - c^k E_k \circ c^{-1}\ .$$

Also

$$\int_{\mathbf{Z}_p} dE_k = B_k = \int_{\mathbf{Z}_p} dE_k(c^{-1}x)\ .$$

The desired formula drops out.

Theorem 3.3. *Let $2 \leqslant k \leqslant p-2$. Let*

$$\omega\colon (\mathbf{Z}/p\mathbf{Z})^* \to \mathbf{Z}_p^*$$

be the character such that

$$\omega(a) \equiv a \pmod p\ .$$

Then

$$B_{1,\omega^{k-1}} \equiv B_k/k \pmod p\ .$$

Proof. Let $\psi = \omega^{k-1}$. Choose c to be a primitive root mod p, so that $c^k \not\equiv 1 \bmod p$. Then

$$B_{1,\psi} = \frac{1}{(1-\psi(c)c)}\int_{\mathbf{Z}_p} \omega^{k-1}\, dE_{1,c}\ .$$

On the other hand by Theorem 3.2,

$$B_k/k = \frac{1}{1-c^k}\int_{\mathbf{Z}_p} x^{k-1}\, dE_{1,c}(x)$$

$$= \frac{1}{1-c^k}\int_{\mathbf{Z}_p^*} x^{k-1}\, dE_{1,c}(x) + \frac{1}{1-c^k}\int_{p\mathbf{Z}_p} x^{k-1}\, dE_{1,c}(x)\ .$$

Since $E_{1,c}$ is p-integral valued by Theorem 2.1, and c is a primitive root mod p, the second integral on the right is $\equiv 0 \pmod p$. Furthermore $c^k \not\equiv 1 \pmod p$. Hence it suffices to prove

$$\frac{1-c^k}{1-\psi(c)c}\int_{\mathbf{Z}_p^*} (\omega(x)^{k-1} - x^{k-1})\, dE_{1,c}(x) \equiv 0 \pmod p\ .$$

Again we use the fact that $E_{1,c}(x)$ is p-integral, and

$$\omega(x)^{k-1} - x^{k-1} \equiv 0 \pmod p\ .$$

The factor in front of the integral is a unit, and the theorem is proved.

In some applications it is also useful to know that for $2 \leqslant k \leqslant p-1$, we have

$$B_k \equiv \frac{1}{p}\sum_{a=1}^{p-1} a^k \pmod{p^2}\ .$$

As in Chapter X, § 2 we have

$$pB_k = -\sum_{j=0}^{k-2}\binom{k}{j} pB_j \frac{p^{k-j}}{k-j+1} + \sum_{c=1}^{p-1} c^k\ .$$

By the Kummer congruences, B_j is p-integral for $j=0,\ldots,k-2$, so

$$pB_k \equiv \sum_{c=1}^{p-1} c^k \pmod{p^3}\ ,$$

thereby proving our assertion.

Chapter XIV. The Complex L-Functions

This chapter is inserted for the convenience of the reader, showing how the values of the classical L-series and zeta function at the negative integers are the same as the values of the p-adic functions.

§ 1. The Hurwitz Zeta Function

Let $0<u\leqslant 1$. Define

$$\zeta(s,u)=\sum_{n=0}^{\infty}\frac{1}{(n+u)^s}.$$

The series converges absolutely and defines an analytic function for $\operatorname{Re}s>1$. Recall that

$$\Gamma(s)=\int_0^{\infty}e^{-t}t^s\frac{dt}{t}\qquad\text{for}\quad\operatorname{Re}s>0$$

$$=\int_0^{\infty}e^{-(n+u)t}(n+u)^s t^s\frac{dt}{t}$$

after making the multiplicative translation $t\mapsto(n+u)t$, which leaves the integral invariant. Therefore

$$\Gamma(s)\zeta(s,u)=\sum_{n=0}^{\infty}\frac{\Gamma(s)}{(n+u)^s}=\int_0^{\infty}\sum_{n=0}^{\infty}e^{-(n+u)t}t^s\frac{dt}{t}.$$

But

$$\sum_{n=0}^{\infty}e^{-(n+u)t}=\frac{e^{-ut}}{1-e^{-t}}.$$

Therefore, letting

$$G_u(z)=\frac{e^{-uz}}{1-e^{-z}}$$

we find

$$\Gamma(s)\zeta(s,u)=\int_0^{\infty}G_u(t)t^s\frac{dt}{t},$$

which is usually called the Mellin transform of G_u.

We let

$$F_u(z)=\frac{ze^{uz}}{e^z-1}.$$

Then

$$G_u(-z)=\frac{e^{uz}}{1-e^z}=-\frac{1}{z}F_u(z).$$

Let

$$H_u(s)=\int_C\frac{1}{z}F_u(z)z^s\frac{dz}{z},$$

where the integral is taken over the contour C as shown on Fig. 8, in which K_ε is a small circle of radius ε around the origin.

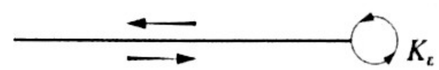

Fig. 8.

As usual, $z^s=e^{s\log z}$, and the log is taken as its principal value on the complex plane from which the negative real axis is deleted. Symbolically we may write

$$\int_C=\int_{-\infty}^{-\varepsilon}+\int_{K_\varepsilon}+\int_{-\varepsilon}^{-\infty},$$

but the integrand in the first and last integral is not the same, corresponding to the two values of z^s which differ by a constant.

The exponential decay of the integrand shows at once that H_u is an entire function of s. We shall see that it gives the analytic continuation for $\zeta(s,u)$.

We change variable, putting $z=-w$. Then writing $G=G_u$, $F=F_u$,

$$H_u(s)=\int_{\infty}^{\varepsilon}\frac{F(-w)}{-w}e^{s\log w}\frac{dw}{w}\cdot e^{-\pi is}$$

$$+\int_{-K_\varepsilon}\frac{F(-w)}{-w}e^{s\log(-w)}\frac{dw}{w}$$

$$+\int_{\varepsilon}^{\infty}\frac{F(-w)}{-w}e^{s\log w}\frac{dw}{w}\cdot e^{\pi is}.$$

Therefore taking the limit as $\varepsilon\to 0$, we find (see below)

$$H_u(s)=e^{-\pi is}\int_0^{\infty}G(t)t^s\frac{dt}{t}-e^{\pi is}\int_0^{\infty}G(t)t^s\frac{dt}{t}$$

$$=-(e^{\pi is}-e^{-\pi is})\int_0^{\infty}G(t)t^s\frac{dt}{t}.$$

This shows that $\zeta(s,u)$ has an analytic continuation to the whole plane, and more precisely:

Theorem 1.1. $H_u(s)=-(e^{\pi is}-e^{-\pi is})\Gamma(s)\zeta(s,u)$.

We needed the

Lemma. *If* $\operatorname{Re}(s)>1$, *then*

$$\int_{-K_\varepsilon}G(w)\,e^{s\log(-w)}\frac{dw}{w}\to 0 \quad as \quad \varepsilon\to 0.$$

Proof. The length of K_ε and $|dw/w|$ have a product which is bounded. But putting $r=|z|$, $\sigma=\operatorname{Re}s$, we have

$$e^{s\log z}\ll e^{s\log r}$$

$$\ll r^{\sigma}$$

and

$$G(z)\ll 1/r \qquad \text{for} \quad r\to 0.$$

This proves the lemma.

Using a standard identity for the gamma function, we also find the expression

$$H_u(s)=-2i\sin\pi s\Gamma(s)\zeta(s,u)=\frac{-2\pi i}{\Gamma(1-s)}\zeta(s,u),$$

so that

Theorem 1.2. $$\zeta(s,u)=-\frac{1}{2\pi i}\Gamma(1-s)H_u(s).$$

Corollary. *If n is an integer $\geqslant 0$ then*

$$\zeta(1-n,u)=-\frac{1}{2\pi i}\Gamma(n)H_u(1-n)$$

and

$$\frac{\zeta(1-n,u)}{\Gamma(n)}=-\textit{residue of } F_u(z)z^{-n-1} \textit{ at } z=0.$$

Next we find an expression for $H_u(s)$ which will lead to the functional equation of L-series.

Theorem 1.3. *For* $\operatorname{Re}s<0$ *we have*

$$H_u(s)=(-2\pi)^s\sum_{n=1}^{\infty}\frac{e^{2\pi iun}e^{\pi is/2}-e^{-2\pi iun}e^{-\pi is/2}}{n^{1-s}}$$

$$=-(2\pi)^s\sum_{n=1}^{\infty}\frac{2i\sin(2\pi un+\pi s/2)}{n^{1-s}}.$$

Proof. Let D_m be the contour indicated on Fig. 9, consisting of the square and the portion of C inside the square, with the given orientation.

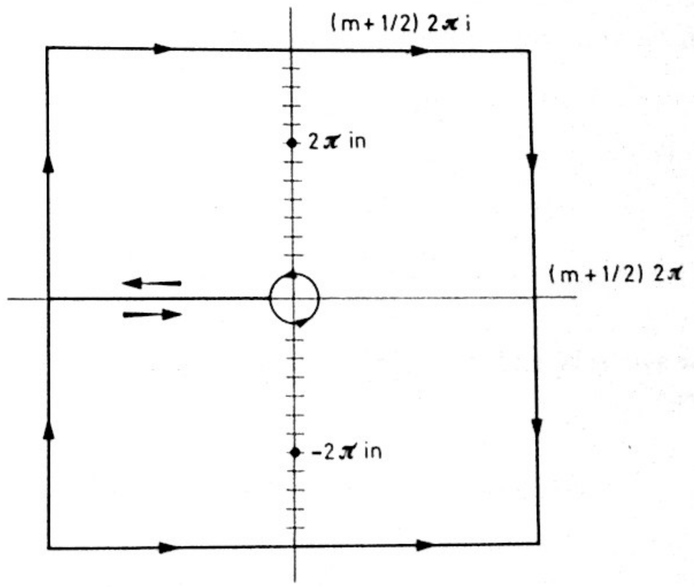

Fig. 9.

Let

$$R_n = \text{residue of } F_u(z)z^{s-2} \text{ at } 2\pi i n,\ n \neq 0,\ -m \leqslant n \leqslant m\ .$$

Then

$$\int_{D_m} \frac{F(z)}{z}\, z^s \frac{dz}{z} = -2\pi i \sum_{\substack{-m \\ n \neq 0}}^{m} R_n\ .$$

For $n \geqslant 1$ we have:

$$R_n = \frac{1}{i}\, e^{2\pi i u n}\, (2\pi n)^{s-1}\, e^{\pi i s/2}$$

$$R_{-n} = \frac{1}{-i}\, e^{-2\pi i u n}\, (2\pi n)^{s-1}\, e^{-\pi i s/2}\ .$$

Note that $F(z)/z$ is bounded on the outside square. Hence if $\operatorname{Re} s < 0$ the integral over the outside square tends to 0 as $m \to \infty$. Hence

$$H_u(s) = \int_C \frac{F(z)}{z}\, z^s \frac{dz}{z} = \lim_{m \to \infty} \int_{D_m} \frac{F(z)}{z}\, z^s \frac{dz}{z}$$

and the theorem follows.

§ 2. Functional Equation

Let $f = f_N$ be a function on $\mathbf{Z}/N\mathbf{Z}$, and let

$$u = a/N\ , \qquad 1 \leqslant a \leqslant N\ , \qquad a \in \mathbf{Z}\ .$$

Then

$$\zeta(s, a/N) = \sum_{n=0}^{\infty} \left(n + \frac{a}{N}\right)^{-s} = N^s \sum_{m \equiv a} \frac{1}{m^s},$$

where the sum is taken for integers $m > 0$, $m \equiv a \pmod N$.

Define

$$\zeta(s, f) = N^{-s} \sum_{a=1}^{N} f(a)\zeta(s, a/N)$$

$$= \sum_{n=1}^{\infty} \frac{f(n)}{n^s}, \qquad \text{for} \quad \operatorname{Re} s > 1\ .$$

Then for $\operatorname{Re}(s) < 0$,

$$H_{a/N}(s) = -(2\pi)^s \sum_{n=1}^{\infty} \frac{e^{2\pi i a n/N}\, e^{\pi i s/2} - e^{-2\pi i a n/N}\, e^{-\pi i s/2}}{n^{1-s}}$$

and

$$\zeta(s, f) = (2\pi/N)^s \sum_{a=1}^{N} f(a)\, \frac{\Gamma(1-s)}{2\pi i} \sum_{n=1}^{\infty} \frac{e^{2\pi i a n/N}\, e^{\pi i s/2} - e^{-2\pi i a n/N}\, e^{-\pi i s/2}}{n^{1-s}}$$

Therefore, if we define

$$\hat{f}(n) = \sum_{a=1}^{N} f(a)\, e^{-2\pi i a n/N}$$

we get the expression:

Theorem 2.1.

$$\zeta(s, f) = \frac{1}{2\pi i}\left(\frac{2\pi}{N}\right)^s \Gamma(1-s) \sum_{n=1}^{\infty} \frac{\hat{f}(-n)e^{\pi i s/2} - \hat{f}(n)e^{-\pi i s/2}}{n^{1-s}}$$

$$= \frac{1}{2\pi i}\left(\frac{2\pi}{N}\right)^s \Gamma(1-s)[\zeta(1-s, \hat{f}^-)\, e^{\pi i s/2} - \zeta(1-s, \hat{f})\, e^{-\pi i s/2}]\ .$$

Consider the special case when $f = \chi$ is a primitive Dirichlet character with conductor N. Then

$$\hat{\chi}(n) = \overline{\chi}(n) S(\chi)\chi(-1)$$

where

$$S(\chi) = \sum \chi(a)\, e^{2\pi i a/N}\ ,$$

and it is standard that $S(\chi) \neq 0$. Hence we obtain for $\zeta(s, \chi) = L(s, \chi)$:

Theorem 2.2.

(i) $L(s, \chi) = \dfrac{1}{2\pi i}\left(\dfrac{2\pi}{N}\right)^s \Gamma(1-s) S(\chi)[e^{\pi i s/2} - \chi(-1)\, e^{-\pi i s/2}]L(1-s, \overline{\chi})\ .$

This is the functional equation. Since

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}\ ,$$

Let

$$R_n = \text{residue of } F_u(z)z^{s-2} \text{ at } 2\pi i n,\ n \neq 0,\ -m \leqslant n \leqslant m\ .$$

Then

$$\int_{D_m} \frac{F(z)}{z}\, z^s \frac{dz}{z} = -2\pi i \sum_{\substack{n=-m \\ n\neq 0}}^{m} R_n\ .$$

For $n \geqslant 1$ we have:

$$R_n = \frac{1}{i}\, e^{2\pi i u n}\, (2\pi n)^{s-1}\, e^{\pi i s/2}$$

$$R_{-n} = \frac{1}{-i} e^{-2\pi i u n}\, (2\pi n)^{s-1}\, e^{-\pi i s/2}\ .$$

Note that $F(z)/z$ is bounded on the outside square. Hence if $\operatorname{Re} s < 0$ the integral over the outside square tends to 0 as $m \to \infty$. Hence

$$H_u(s) = \int_C \frac{F(z)}{z}\, z^s \frac{dz}{z} = \lim_{m\to\infty} \int_{D_m} \frac{F(z)}{z}\, z^s \frac{dz}{z}$$

and the theorem follows.

§ 2. Functional Equation

Let $f = f_N$ be a function on $\mathbf{Z}/N\mathbf{Z}$, and let

$$u = a/N\ , \qquad 1 \leqslant a \leqslant N\ , \qquad a \in \mathbf{Z}\ .$$

Then

$$\zeta(s, a/N) = \sum_{n=0}^{\infty} \left(n + \frac{a}{N}\right)^{-s} = N^s \sum_{m \equiv a} \frac{1}{m^s},$$

where the sum is taken for integers $m > 0$, $m \equiv a \pmod N$.

Define

$$\zeta(s, f) = N^{-s} \sum_{a=1}^{N} f(a)\zeta(s, a/N)$$

$$= \sum_{n=1}^{\infty} \frac{f(n)}{n^s}, \qquad \text{for} \quad \operatorname{Re} s > 1\ .$$

Then for $\operatorname{Re}(s) < 0$,

$$H_{a/N}(s) = -(2\pi)^s \sum_{n=1}^{\infty} \frac{e^{2\pi i a n/N}\, e^{\pi i s/2} - e^{-2\pi i a n/N}\, e^{-\pi i s/2}}{n^{1-s}}$$

and

$$\zeta(s, f) = (2\pi/N)^s \sum_{a=1}^{N} f(a)\, \frac{\Gamma(1-s)}{2\pi i} \sum_{n=1}^{\infty} \frac{e^{2\pi i a n/N}\, e^{\pi i s/2} - e^{-2\pi i a n/N}\, e^{-\pi i s/2}}{n^{1-s}}$$

Therefore, if we define

$$\hat{f}(n) = \sum_{a=1}^{N} f(a)\, e^{-2\pi i a n/N}$$

we get the expression:

Theorem 2.1.

$$\zeta(s, f) = \frac{1}{2\pi i}\left(\frac{2\pi}{N}\right)^s \Gamma(1-s) \sum_{n=1}^{\infty} \frac{\hat{f}(-n)e^{\pi i s/2} - \hat{f}(n)e^{-\pi i s/2}}{n^{1-s}}$$

$$= \frac{1}{2\pi i}\left(\frac{2\pi}{N}\right)^s \Gamma(1-s)[\zeta(1-s, \hat{f}^-)\, e^{\pi i s/2} - \zeta(1-s, \hat{f})\, e^{-\pi i s/2}]\ .$$

Consider the special case when $f = \chi$ is a primitive Dirichlet character with conductor N. Then

$$\hat{\chi}(n) = \overline{\chi}(n) S(\chi)\chi(-1)$$

where

$$S(\chi) = \sum \chi(a)\, e^{2\pi i a/N}\ ,$$

and it is standard that $S(\chi) \neq 0$. Hence we obtain for $\zeta(s, \chi) = L(s, \chi)$:

Theorem 2.2.

(i) $$L(s, \chi) = \frac{1}{2\pi i}\left(\frac{2\pi}{N}\right)^s \Gamma(1-s) S(\chi)[e^{\pi i s/2} - \chi(-1)\, e^{-\pi i s/2}]L(1-s, \overline{\chi})\ .$$

This is the functional equation. Since

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s},$$

we also get another form, namely:

(ii) $$L(s,\chi)=L(1-s,\bar{\chi})\left(\frac{2\pi}{N}\right)^{s} S(\chi)\,\frac{1}{\Gamma(s)}\left(\frac{e^{\pi is/2}-\chi(-1)\,e^{-\pi is/2}}{e^{\pi is}-e^{-\pi is}}\right).$$

Corollary. *Let n be an integer $\geqslant 1$.*

(i) *Suppose that $\chi(-1)=1$. Then $L(1-n,\chi)\neq 0$ if and only if n is even.*

(ii) *Suppose that $\chi(-1)=-1$. Then $L(1-n,\chi)\neq 0$ if and only if n is odd.*

Proof. We are assuming that $L(1,\chi)\neq 0$ in case $n=1$. It is obvious from the Euler product that $L(n,\chi)\neq 0$ for $n\geqslant 1$. We then substitute n for s in the theorem. We use the fact that $L(s,\chi)$ is entire. The quotient of exponential terms will have a non-zero value or have a pole at n according to the various cases under consideration, and in those instances where it has a pole, the L-function $L(1-s,\bar{\chi})$ must have a zero. Considering the four cases separately, the reader will verify at once that they fit the statement of the corollary.

Theorem 2.3. *We have* $L(1-n,\chi)=-\dfrac{B_{n,\chi}}{n}$, *except for* $\zeta(0)=B_1=-\frac{1}{2}$.

Proof. This follows at once from the corollary of Theorem 1.2, and the definition

$$F_\chi(t)=\sum_{a=1}^{c}\chi(a)\,\frac{te^{at}}{e^{ct}-1}=\sum_{n=0}^{\infty}B_{n,\chi}\,\frac{t^n}{n!}.$$

Combining Theorem 2.3 with the corollary, one obtains the conditions under which $B_{n,\chi}\neq 0$.

Chapter XV. The Hecke–Eisenstein and Klein Forms

We have seen in Chapter X, § 3 that the modular form G_k has a q-expansion

$$G_k = 2\zeta(k) + \cdots.$$

So the value of the ordinary zeta function appears as the constant term of a modular form (Eisenstein series, as it is called). This phenomenon, first exploited by Klingen [Kl 3] and Siegel [Si 4], has been highly developed by Serre [Se 5] and others. We give here more examples.

§ 1. Forms of Weight 1

Given a lattice L, we have the usual Weierstrass functions $\sigma(z,L)$ and

$$\zeta(z,L)=\sigma'/\sigma(z,L).$$

There is a function $\eta(z,L)$ which is $\mathbf{R}$-linear in z, such that for $\omega\in L$,

$$\zeta(z+\omega,L)=\zeta(z,L)+\eta(\omega,L)$$

If $L=[\omega_1,\omega_2]$, and we write $w=a_1\omega_1+a_2\omega_2$ with $a_1,a_2\in\mathbf{R}$, not both in $\mathbf{Z}$, then in usual notation,

$$\eta_1=\eta(\omega_1,L),\qquad \eta_2=\eta(\omega_2,L)$$

and

$$\eta(a_1\omega_1+a_2\omega_2)=a_1\eta_1+a_2\eta_2.$$

Let $a=(a_1,a_2)$, $a\cdot\omega=a_1\omega_1+a_2\omega_2$. We define the **Hecke form**

$$H_a(\omega_1,\omega_2)=H_a\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}=\zeta(a_1\omega_1+a_2\omega_2,L)-\eta(a_1\omega_1+a_2\omega_2,L)$$

$$=\zeta(a\cdot\omega,L)-\eta(a\cdot\omega,L).$$

Then for $\lambda \in \mathbf{C}^*$ we get

H 1. $$H_a(\lambda\omega_1, \lambda\omega_2) = \lambda^{-1} H_a(\omega_1, \omega_2),$$

so H_a has weight 1. From the definition of η, it follows at once that

H 2. *H_a depends only on the residue class of a_1, a_2 mod $\mathbf{Z}$.*

Furthermore, if $\alpha \in SL_2(\mathbf{Z})$, then

H 3. $$H_a\left(\alpha\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}\right) = H_{a\alpha}\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}.$$

Now let N be an integer >1, and suppose that a_1, a_2 are rational numbers with denominators dividing N. It is then obvious that H_a is invariant under $\Gamma(N)$, and hence is a modular form of weight 1 on $\Gamma(N)$, if we can show that it has a holomorphic q-expansion. This amounts to quoting classical formulas, given for $L_\tau = [\tau, 1]$. We have

$$\zeta(z, L_\tau) = \eta_2 z + 2\pi i F(q_\tau, q_z)$$

where

$$F(q_\tau, q_z) = -\frac{1}{2} - \frac{q_z}{1-q_z} + \sum_{m=1}^{\infty}\left(\frac{q_\tau^m/q_z}{1-q_\tau^m/q_z} - \frac{q_\tau^m q_z}{1-q^m q_z}\right).$$

See for instance [L 2], Chapter XVIII, § 3.

Put $z = a_1\tau + a_2$, and $h_a(\tau) = H_a(\tau, 1)$. From the Legendre relation $\eta_2\tau - \eta_1 = 2\pi i$, we see at once that

H 4. $$h_a(\tau) = 2\pi i a_1 + 2\pi i F(q_\tau, q_z)\,.$$

Note that $(0, a_2)\alpha = (0, a_2)$ for $\alpha \in \Gamma_1(N)$, whence the form

$$h_{0,a_2} \text{ is on } \Gamma_1(N)\,,$$

Similarly,

$$h_{a_1,0} \text{ is on } \Gamma^1(N)\,,$$

where $\Gamma_1(N)$ consists of those matrices in $SL_2(\mathbf{Z})$ such that

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix} \equiv \begin{pmatrix}1 & 0\\ c & 1\end{pmatrix} \pmod N\,.$$

Furthermore, $\Gamma^1(N)$ is conjugate to $\Gamma_1(N)$ by the matrix

$$w = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},$$

and $(a_1, a_2)w = (a_2, -a_1)$, so

$$H_{(a_1,0)w} = H_{(0,-a_1)}\,.$$

We shall use the value $a_1 = m/N$ where m is an integer prime to N, which we take with $1 \leqslant m \leqslant N-1$. We define

$$G^{1,m} = (2\pi i)^{-1} h_{m/N,0}\,,$$

which is a modular form on $\Gamma^1(N)$. Its q-expansion is given in terms of $q^{1/N}$. In fact, if $f = \sum a_n q^{n/N}$ is a modular form of weight k, let as usual

$$V_N f = \sum a_n q^n$$

be the series obtained by putting q instead of $q^{1/N}$ in the series. Up to a constant factor, $V_N f$ is merely $f \circ [B_N]_k$, where

$$B_N = \begin{pmatrix}N & 0\\ 0 & 1\end{pmatrix}.$$

Lemma. *If $f = \sum a_n q^{n/N}$ is on $\Gamma^1(N)$ then $V_N f$ is on $\Gamma_1(N)$.*

Proof. Note that $f \circ [w]_k$ is on $\Gamma_1(N)$, and $wB_N = w_N$. (See Chapter VII, § 6.) Since $[w_N]_k$ maps $\mathscr{F}_1(N, k)$ into itself, and since we can write

$$f \circ [B_N]_k = f \circ [w]_k \circ [wB_N]_k = f \circ [w]_k \circ [w_N]_k\,,$$

the lemma follows at once.

We put

$$G_{1,m} = V_N G^{1,m},$$

so that $G_{1,m}$ is a modular form of weight 1 on $\Gamma_1(N)$. Using the same identities as in the above lemma, and using also the commutation rule of Lemma 3, Chapter VII, § 6, it is immediate that for an integer $a \neq 0$ and $(a, N) = 1$ we have

H 5. $$[a]_1 G_{1,m} = G_{1,a^{-1}m}.$$

Using **H 4**, the q-expansion of $G_{1,m}$ is given by

H 6. $$G_{1,m} = \frac{m}{N} - \frac{1}{2} - \frac{q^m}{1-q^m} + \sum_{\nu=1}^{\infty}\left[\frac{q^{\nu N-m}}{1-q^{\nu N-m}} - \frac{q^{\nu N+m}}{1-q^{\nu N+m}}\right]$$

Let χ be a non-trivial Dirichlet character on $(\mathbf{Z}/N\mathbf{Z})^*$. We have $\chi(m)=\chi(\nu N+m)$. Define

$$G_{1,\chi} = \sum \chi(m) G_{1,m}.$$

Then we obtain immediately:

Theorem 1.1 *Let χ be an odd character, i.e. $\chi(-1)=-1$. Then*

$$G_{1,\chi} = B_{1,\chi} - 2\sum_{n=1}^{\infty}\sum_{d\mid n}\chi(d)q^n.$$

Furthermore, $G_{1,\chi}$ has Dirichlet character precisely χ.

The proof of this last statement is immediate from **H 5.** Note that the constant term of $G_{1,\chi}$ is $-L(0,\chi)$ according to Chapter XIV.

For p-adic interpolation properties greatly extending such a relation, cf. Katz [K 3], [K 4]. In particular, take N equal to an odd prime number l. Let λ be the prime above l in $\mathbf{Q}(\mu_{l-1})$, and take χ to be the character such that

$$\chi(m)m \equiv 1 \pmod{l}$$

for m prime to l. Then the values of $\chi(m)$ for $m=1,\ldots,l-1$ range over the $(l-1)$th roots of unity, and form a basis over $\mathbf{Z}$ of the ring of integers in the cyclotomic field. Then the sum

$$\sum \chi(m)m$$

is not divisible by the prime λ. Thus

$$\sum \chi(m)\frac{m}{l}$$

has denominator l. Define

$E_{1,\chi}$ = the constant multiple of $G_{1,\chi}$ whose q-expansion has constant term equal to 1.

Theorem 1.2. *We have $E_{1,\chi}\equiv 1 \pmod{l}$.*

Shimura was the first to observe this congruence property, and to suggest that $E_{1,\chi}$ could be used in place of E_{p-1} in the Deligne-Serre theory.

§ 2. The Klein Forms

Let $L=[\omega_1,\omega_2]$ be a lattice, and let $a=(a_1,a_2)$ with real numbers a_1, a_2 not both integers. We write

$$z = a_1\omega_1 + a_2\omega_2 = a\cdot\omega.$$

We define the **Klein forms**

$$\mathfrak{k}_a\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix} = e^{-\eta(a\cdot\omega)(a\cdot\omega)/2}\,\sigma(a\cdot\omega;\omega_1,\omega_2).$$

For $\lambda\in\mathbf{C}^*$ we obtain

K 1. $$\mathfrak{k}_a(\lambda\omega) = \lambda\mathfrak{k}_a(\omega),$$

so $\mathfrak{k}_a$ is homogeneous of degree 1 (weight -1).

Let $b=(b_1,b_2)\in\mathbf{Z}^2$ be a pair of integers. Then

K 2. $$\mathfrak{k}_{a+b}(\omega) = \varepsilon_0(b)\mathfrak{k}_a(\omega),$$

where $\varepsilon_0(b)$ has absolute value 1, and precisely,

$$\varepsilon_0(b) = (-1)^{b_1b_2+b_1+b_2}\,e^{-2\pi i(b_1a_2-b_2a_1)/2}.$$

This follows easily from the Legendre relation.

We now take $\omega_1=\tau$ and $\omega_2=1$, so that

$$z = a_1\tau + a_2.$$

Using the q-expansion for the sigma function, given in most books on elliptic functions (for instance [L 2], Chapter 18, § 2) we can easily derive the q-product for the Klein forms. We let the **Siegel functions** be defined by

$$g_a(\tau) = \mathfrak{k}_a(\tau)\,\Delta(\tau)^{1/12},$$

where $\Delta(\tau)^{1/12}$ is the square of the Dedekind eta function, namely the natural q-product for the 12th root of Δ. Then we find:

K 3. $$g_a(\tau) = -q_\tau^{\frac{1}{2}B_2(a_1)}\,e^{2\pi i a_2(a_1-1)/2}(1-q_z)\prod_{n=1}^{\infty}(1-q_\tau^n q_z)(1-q_\tau^n/q_z),$$

where $q_\tau=e^{2\pi i\tau}$ and $q_z=e^{2\pi iz}$, and $B_2(X)=X^2-X+\frac{1}{6}$.

So far, we needed no further assumption on a_1, a_2. Assume now that they are rational numbers, with denominators dividing an integer $N > 1$, say

$$a_1 = r/N \quad \text{and} \quad a_2 = s/N .$$

Let

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be in $\Gamma(N)$, and write

$$ar + cs = r + \left(\frac{a-1}{N} r + \frac{c}{N} s\right)N , \quad br + ds = s + \left(\frac{b}{N} r + \frac{d-1}{N} s\right)N .$$

Then we find from **K 2**:

K 4. $$\mathfrak{t}_a\left(\alpha\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}\right) = \mathfrak{t}_{a\alpha}\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix} = \varepsilon(\alpha)\mathfrak{t}_a\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$$

where $\varepsilon(\alpha)$ is a $(2N)$th root of unity, given precisely by

$$\varepsilon(\alpha) = -(-1)^{\left(\frac{a-1}{N}r + \frac{c}{N}s + 1\right)\left(\frac{b}{N}r + \frac{d-1}{N}s + 1\right)} e^{2\pi i(br^2 + (d-a)rs - cs^2)/2N^2} .$$

Thus $\mathfrak{t}_a$ is a modular form on $\Gamma(2N^2)$.

Similarly, considering the case with $s = 0$, we see that $\mathfrak{t}_{a_1,0}$ is a form on $\Gamma^1(2N^2)$, and that for $\alpha \in \Gamma^1(N)$, we have

$$\mathfrak{t}_a\left(\alpha\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}\right) = \varepsilon_1(\alpha)\mathfrak{t}_a\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$$

for some root of unity $\varepsilon_1(\alpha)$.

The Klein forms are used to construct units in the modular function field in [KL]. In the next section, we take their logarithmic derivative to construct modular forms of weight 2.

§ 3. Forms of Weight 2

Define

$$h_a(\tau) = (2\pi i)^{-1} g_a'/g_a .$$

Since g_a has weight 0, it follows that h_a is a form of weight 2, by using the differentiation lemma of Chapter X, § 5, namely

$$\partial(f \circ [\alpha]_0) = (\partial f) \circ [\alpha]_2 .$$

The root of unity by which the Siegel function transforms, and a similar root of unity for $\Delta^{1/12}$ (see [L 2], Chapter 18, § 5) disappear in the logarithmic differentiation. From this it follows that if a has denominator N, then

$$h_a \text{ is a modular form of weight 2 on } \Gamma(N) ,$$

$$h_{a_1,0} \text{ is a modular form of weight 2 on } \Gamma^1(N) .$$

The q-expansion for h_a is easily derived from that of g_a. We shall follow the pattern of § 1, and consider only the case for $h_{a_1,0}$, when we obtain

$$h_{a_1,0} = \frac{1}{2}B_2(a_1) - \frac{a_1 q^{a_1}}{1-q^{a_1}} - \sum_{n=1}^{\infty}\left[\frac{(n+a_1)q^{n+a_1}}{1-q^{n+a_1}} + \frac{(n-a_1)q^{n-a_1}}{1-q^{n-a_1}}\right].$$

We use the special value $a_1 = m/N$ for $1 \leqslant m \leqslant N-1$ prime to N. We define

$$G_{2,m} = V_N h_{m/N,0},$$

obtained by replacing $q^{1/N}$ with q in the expansion. Then the same lemma as in § 1 shows that

$$G_{2,m} \text{ is a modular form of weight 2 on } \Gamma_1(N) .$$

As in weight 1, one sees that for $(a, N) = 1$ we have

$$[a]_2 G_{2,m} = G_{2,a^{-1}m} .$$

The q-expansion is given by

$$NG_{2,m} = \frac{1}{2}NB_2\left(\frac{m}{N}\right) - m\frac{q^m}{1-q^m} - \sum_{n=1}^{\infty}\left[\frac{(nN+m)q^{nN+m}}{1-q^{nN+m}} + \frac{(nN-m)q^{nN-m}}{1-q^{nN-m}}\right]$$

Let $\chi \neq 1$ be a Dirichlet character mod N. Define

$$G_{2,\chi} = N\sum_{m=1}^{N-1}\chi(m)G_{2,m} .$$

Then we obtain the q-expansion for $G_{2,\chi}$ from the above, and in particular:

Theorem 3.1. *Suppose that χ is an even character. Then*

$$G_{2,\chi} = \frac{1}{2}B_{2,\chi} - 2\sum_{n=1}^{\infty}\sum_{d|n}\chi(d)dq^n .$$

Furthermore $G_{2,\chi}$ has Dirichlet character χ.

Note that as in the case of weight 1, the constant term is given by $-L(-1, \chi)$.

The weights 1 and 2 are the hardest to deal with. For higher weights, the series $\sum 1/\omega^k$ converge absolutely. They are called **Eisenstein series**, and one can construct much more easily (as Hecke did) modular forms having $B_{k,\chi}/k$ as their constant terms. We leave this as an exercise to the reader. For $k = 1, 2$ one has to use some device to define the corresponding objects since the series don't converge. We have selected a device in the style of the present book, which leads most rapidly to the q-expansions.

Bibliography

The bibliography is very selective. The reader can trace other references in the bibliographies which appear at the end of each of the items listed here. Mostly recent research papers are emphasized here, relevant to the topics chosen for the book.

[A–L] A. Atkin and J. Lehner, Hecke operators on $\Gamma_0(m)$, Math. Ann. **185** (1970) pp. 134–160

[B–M–S] H. Bass, J. Milnor, J. P. Serre, Solution of the congruence subgroup problem for $SL_n(n \geqslant 3)$ and $Sp_{2n}(n \geqslant 2)$, Pub. IHES (1967) pp. 59–137

[B] B. Birch, Elliptic curves, a progress report, AMS conference on number theory, Stonybrook, 1969, pp. 396–400

[C] W. Casselman, On some results of Atkin and Lehner, Math. Ann. **201** (1973) pp. 301–313

[C–Li] J. Coates and S. Lichtenbaum, On l-adic zeta functions, Ann. of Math. **98** No. 3 (1973) pp. 498–550

[CS 1] J. Coates and W. Sinnott, An analogue of Stickelberger's theorem for the higher K-groups, Invent. Math. **24** (1974) pp. 149–161

[CS 2] J. Coates and W. Sinnott, On p-adic L-functions over real quadratic fields, Invent. Math. **25** (1974) pp. 253–279

[CW] J. Coates and A. Wiles, On the conjecture of Birch and Swinnerton-Dyer, to appear, Invent. Math.

[Ded] R. Dedekind, Erläuterungen zu den vorstehenden Fragmenten, Comments on fragments from Riemann's work, in Riemann's Werke (1876), pp. 438–447

[De 1] P. Deligne, Formes modulaires et représentation l-adiques, Séminaire Bourbaki 1968–1969, exp. No. 355

[De 2] ——, Formes modulaires et représentations de $GL(2)$, Lect. Notes Math. **349**, Springer (1973), pp. 55–105

[De–Ra] P. Deligne and M. Rapoport, Les schemas de modules de courbes elliptiques, Modulair forms in one variable, Lect. Notes Math. **349**, Springer (1973) pp. 143–316

[De–Se] P. Deligne and J. P. Serre, Formes modulaires de poids 1, Ann. Sci. ENS **7** (1974) pp. 507–530

[Dr] V. G. Drinfeld, Two theorems on modular curves, Functional Analysis and its applications, Vol. 7 No. 2, translated from the Russian, April–June 1973, pp. 155–156

[E 1] M. EICHLER, Quarternare quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion, Arch. Math. **5** (1954) pp. 355–366

[E 2] ——, Eine Verallgemeinerung der Abelschen Integrale, Math. **67** (1957) pp. 267–298

[E 3] ——, Modular correspondences and their representations, J. Indian Math. Soc. **XX** (1956) pp. 163–206

[G] L. GOLDSTEIN, Dedekind sums for Fuchsian groups, Nagoya Math. J. (1973) pp. 21–47

[H] E. HECKE, Lectures on Dirichlet series, modular functions, and quadratic forms, Institute for Advanced Study, Princeton, 1938

[HZ] F. HIRZEBRUCH and D. ZAGIER, The Atiyah-Singer theorem and elementary number theory, Math. Lecture Series No. 3, Publish or Perish Inc.

[Iw 1] K. IWASAWA, On Γ-extensions of algebraic number fields, Bull. Amer. Math. Soc. **65** (1959) pp. 183–226

[Iw 2] ——, On some invariants of cyclotomic fields, Amer. J. Math. **80** (1958) pp. 773–783

[Iw 3] ——, On some properties of Γ-finite modules, Ann of Math. **70** (1959) pp. 291–312

[Iw 4] ——, On the theory of cyclotomic fields, Ann. of Math. **70** (1959) pp. 530–561

[Iw 5] ——, On p-adic L-functions, Ann. of Math. **89** (1969) pp. 198–205

[Iw 6] ——, Lectures on p-adic L-functions, Ann. of Math. studies No. **74**, Princeton University Press, 1972

[J–L] H. JACQUET and R. LANGLANDS, Automorphic forms on GL_2, Lect. Notes Math. **114**, Springer (1970)

[K 1] N. KATZ, Higher congruences between modular forms, Ann of Math. (1975) pp. 332–367

[K 2] ——, p-adic properties of modular schemes and modular forms, Antwerp Summer Institute, Lect. Notes Math. **350**, Springer (1972) pp. 69–190

[K 3] ——, p-adic interpolation of real analytic Eisenstein series, to appear

[K 4] ——, The Eisenstein measure and p-adic interpolation, to appear, Am. J. Math.

[Kl 1] H. KLINGEN, Über die Werte der Dedekindschen Zetafunktionen, Math. Ann. **145** (1962) pp. 265–272

[Kl 2] ——, Über den arithmetischen Charakter der Fourierkoeffizienten von Modulformen, Math. Ann. **147** (1962) pp. 176–188

[Ku–Le] T. KUBOTA and H. LEOPOLDT, Eine p-adische Theorie der Zetawerte, I. J. Reine Angew. Math. **214–215** (1964) pp. 328–339

[K–L 1] D. KUBERT and S. LANG, Units in the modular function field, I, II, III, Math. Ann. (1975) pp. 67–96, 175–189, 273–285

[K–L 2] ——, Distributions on toroidal groups, Math. Z. **148** (1976) pp. 33–51

[L 1] S. LANG, Algebraic Number Theory, Addison Wesley, 1970

[L 2] ——, Elliptic Functions, Addison Wesley, 1973

[L 3] ——, Introduction to algebraic and abelian functions, Addison Wesley, 1972

[L 4] ——, Introduction to diophantine approximations, Addison Wesley, Reading, 1966

[L–T] S. LANG and H. TROTTER, Distribution of Frobenius elements in GL_2-extensions of the rational numbers, Springer Lecture Notes **504** (1975)

[Lgds 1] R. LANGLANDS, Euler Products, Yale Lecture Notes, 1971

[Lgds 2] ——, Problems in the theory of automorphic forms, Yale Lecture Notes, 1969

[Lgds 3] ——, Modular forms and l-adic representations, Modular Functions in One Variable III (Antwerp Conference), Lect. Notes Math. **349**, Springer (1973)

[Le 1] H. LEOPOLDT, Eine Verallgemeinerung der Bernoullischen Zahlen, Abh. Math. Sem. Univ. Hamburg (1958) pp. 131–140

[Le 2] ——, Über Klassenzahlprimteiler reeller abelscher Zahlkörper als Primteiler verallgemeinerter Bernoullischer Zahlen, Abh. Math. Sem. Univ. Hamburg (1959) pp. 36–47

[Le 3] ——, Über Fermatquotienten von Kreiseinheiten und Klassenzahlformeln modulo p, Rend. Circ. Mat. Palermo (1960) pp. 1–12

[Le 4] ——, Eine p-adische Theorie der Zetawerte II, J. reine angew, Math. **274–275** (1975) pp. 224–239

[Li] W. LI, New forms and functional equations, Math. Ann.

[Maass] H. MAASS, Über eine neue Art von nicht analytischen automorphen Funktionen und die Bestimmung von Dirichlet Reihen durch funktionale Gleichungen, Math. Ann. **121** (1949) pp. 141–183

[Man 1] J. MANIN, Cyclotomic fields and modular curves, Russian Math. Surveys Vol. 26 No. 6, Nov–Dec 1971, pp. 7–78

[Man 2] ——, Parabolic points and zeta functions of modular curves, Izv. Akad. Nauk SSSR, Vol. 6 No. 1 (1972) AMS translation pp. 19–64

[Man 3] ——, Explicit formulas for the eigenvalues of Hecke operators, Acta Arithm. XXIV (1973) pp. 239–249

[Man 4] ——, Periods of parabolic forms and p-adic Hecke series, Math. Sbornik (1973), AMS translation pp. 371–393

[Man 5] ——, The values of p-adic Hecke series at integer points of the critical strip, Math, Sbornik (1974), AMS Translation pp. 631–637

[Maz 1] B. MAZUR, Courbes elliptiques et symboles modulaires, Seminaire Bourbaki, June 1972

[Maz 2] ——, Analyse p-adique, Bourbaki report, 1972

[Maz 3] ——, Rational points of abelian varieties in towers of number fields, Invent. Math. **18** (1972) pp. 183–266

[M–SwD] B. MAZUR and H. SWINNERTON-DYER, Arithmetic of Weil curves, Invent. Math. **25** (1974) pp. 1–61

[Mi] T. MIYAKE, On automorphic forms on GL_2 and Hecke operators, Ann. of Math. **94** (1971) pp. 174–189

[N] M. NEWMAN, Construction and application of a class of modular functions, Proc. London Math. Soc. (3) (1957) pp. 334–350

[O 1] A. OGG, Modular forms and Dirichlet series, Benjamin, 1969

[O 2] ——, On the eigenvalues of Hecke operators, Math. Ann. **179** (1969), pp. 101–108

[O 3] ——, A remark on the Sato–Tate conjecture, Invent. Math. **9** (1970) pp. 198–200

[O 4] ——, On a convolution of L-series, Invent. Math. **7** (1969) pp. 297–312

[O 5] ——, Rational points on certain elliptic curves. AMS conference, St. Louis, 1972, pp. 221–231

[Pe] H. PETERSSON, Konstruktion der sämtlichen lösungen einer Reimannschen Funktionalgleichung durch Dirichlet-Reihen mit Eulerscher Produktentwicklung, I, Math. Ann. **116** (1939) pp. 401–412 and II, III, Math. Ann. **117** (1940–1941), pp. 39–64 and 277–300

[Rad] H. RADEMACHER, Zur Theorie der Modulfunktionen, J. Reine Angew. Math. **167** (1932) pp. 312–336

[Ra 1] R. A. RANKIN, Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions, I, II, Proc. Cambridge Phil. Soc. **35** (1939) pp. 351–372

[Ra 2] ——, An Ω-result for the coefficients of cusp forms, Math. Ann. **203** (1973) pp. 239–250

[Ra 3] ——, The construction of automorphic forms from the derivatives of a given form, Colloquium on zeta functions, Tata Institute of Fundamental Research, (1956) pp. 103–116

[Raz] M. RAZAR, Dirichlet series and Eichler Cohomology, to appear, *Modular forms in one variable V*, (Bonn Conference) Springer Lecture Notes

[Ri 1] K. RIBET, On l-adic representations attached to modular forms, Invent. Math. (1975) pp. 245–275

[Ri 2] ——, Galois action on division points of abelian varieties with real multiplications, Ann. Math. (1976) to appear.

[Sch] B. SCHOENEBERG, Elliptic Modular Functions, Springer Verlag (1974)

[Se 1] J. P. SERRE, Classes des corps cyclotomiques (d'après Iwasawa), Séminaire Bourbaki 1958

[Se 2] ——, Une interprétation des congruences relatives à la fonction τ de Ramanujan, Séminaire Delange–Pisot–Poitou, 1967–1968

[Se 3] ——, Congruences et formes modulaires (d'après Swinnerton-Dyer), Séminaire Bourbaki, 1971–1972

[Se 4] ——, Abelian l-adic representations and elliptic curves, Benjamin, Addison-Wesley, 1968

[Se 5] ——, Formes modulaires et fonctions zeta p-adiques, Summer Institute on Modular Functions, Lect. Notes Math. **350**, Springer (1972)

[Se 6] ——, Divisibilité des coefficients des formes modulaires de poids entier, C. R. Acad. Sci. Paris (1974) pp. 679–682

[Se 7] ——, A course in arithmetic, Springer, New York, 1973

[Se 8] ——, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, Invent. Math. **15** (1972) pp. 259–331

[Sh 1] G. SHIMURA, Sur les intégrales attachées aux formes automorphes, J. Math. Soc. Japan 11 No. 4 (1959) pp. 291–311

[Sh 2] ——, Introduction to the arithmetic theory of automorphic functions, Iwanami Shoten and Princeton University Press, 1971

[Sh 3] ——, A reciprocity law in non-solvable extensions, J. Reine Angew. Math. **221** (1966) pp. 209–220

[Si 1] C. L. SIEGEL, Bernoullische Polynome und quadratische Zahlkörper, Nachr. Akad. Wiss. Göttingen Math.-Phys. K1. II **2** (1968) pp. 7–38

[Si 2] ——, Über die Fourierschen Koeffizienten von Modulformen, Nachr. Akad. Wiss. Göttingen Math.-Phys. K1. II **3** (1970) pp. 15–56

[Si 3] ——, Berechnung von Zetafunktionen an ganzzahligen Stellen, Nachr. Akad. Wiss. Göttingen Math.-Phys. K1. II **10** (1969) pp. 87–102

[SwD] H. SWINNERTON-DYER, On l-adic representations and congruences for coefficients of modular forms, (Antwerp Conference) Lect. Notes Math. **350**, Springer (1973)

[W 1] A. WEIL, Dirichlet series and automorphic forms, Lect. Notes Math. **164**, Springer (1971)

[W 2] ——, Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, Math. Ann. (1966) pp. 149–156

[Z] D. ZAGIER, to appear.

Subject Index