Modular Forms, Hecke Operators, and Modular Abelian Varieties

Kenneth A. Ribet William A. Stein

December 9, 2003

Contents

Preface

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

Kenneth A. Ribet (ribet@math.berkeley.edu)
William A. Stein (was@math.harvard.edu)

1

The Main objects

1.1 Torsion points

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

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

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

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

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

1.1.1 The Tate module

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

Let

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

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

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

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

For any prime ℓ, let

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

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

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

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

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

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

1.2 Galois representations

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

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

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

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

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

Let K be the field cut out by the $\ker(\rho)$, i.e., the fixed field of $\ker(\rho)$. Then K is a finite Galois extension of $\mathbf{Q}$ since $E[n]$ is a finite set and $\ker(\rho)$ is a normal subgroup. Since

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

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

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

1.3 Modular forms

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

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

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

of matrices in $\mathrm{SL}_2(\mathbf{Z})$ that are of the form $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ 1∗01 when reduced modulo N.

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

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

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

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

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

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

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

We will see in that $M_k(N)$ and $S_k(N)$ are finite dimensional. For example, (see) the space $S_{12}(1)$ has dimension one and is spanned by the famous cusp form

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

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

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

1.4 Hecke operators

Mordell defined operators T_n, $n \geq 1$, on $S_k(N)$ which are called *Hecke operators.* These proved very fruitful. The set of such operators forms a commuting family of endomorphisms and is hence "almost" simultaneously diagonalizable. The precise meaning of "almost" and the actual structure of the Hecke algebra $\mathbf{T} = \mathbf{Q}[T_1, T_2, \ldots]$ will be studied in greater detail in.

Often there is a basis $f_1, \ldots, f_r$ of $S_k(N)$ such that each $f = f_i = \sum_{n=1}^{\infty} a_n q^n$ is a simultaneous eigenvector for all the Hecke operators T_n and, moreover, $T_n f = a_n f$. In this situation, the eigenvalues a_n are necessarily algebraic integers and the field $\mathbf{Q}(\ldots, a_n, \ldots)$ generated by all a_n is finite over $\mathbf{Q}$ (see).

The a_n exhibit remarkable properties. For example,

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

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

2
Modular representations and algebraic curves

2.1 Arithmetic of modular forms

Let us give ourselves a cusp form

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

which is an eigenform for all of the Hecke operators T_p. Then the **Mellin transform** of f is the L-function

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

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

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

*E*lliptic curves E_1 and E_2 are *isogenous* if there is a morphism $E_1 \rightarrow E_2$ of algebraic groups, having finite kernel.

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

Conjecture 2.1.2 (Shimura-Taniyama). *Every elliptic curve over* $\mathbf{Q}$ *is modular, that is, isogenous to a curve constructed in the above way.*

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

Then it is well known that

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

One can associate to f a representation

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

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

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

$$\mathrm{tr}(\rho_{\ell,f}(\mathrm{Frob}_p)) = a_p \in K \subset K \otimes \mathbf{Q}_\ell$$

and

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

where χ_ℓ is the ℓth cyclotomic character and ε is the Dirichlet character associated to f. There is an incredible amount of "abuse of notation" packed into this statement. First, the Frobenius Frob_P (note P not p) is only well defined in $\mathrm{Gal}(K/\mathbf{Q})$ (so I think an unstated result is that K must be Galois), and then Frob_p is only well defined up to conjugacy. But this works out since $\rho_{\ell,f}$ is well-defined on $\mathrm{Gal}(K/\mathbf{Q})$ (it kills $\mathrm{Gal}(\overline{\mathbf{Q}}/K)$) and the trace is well-defined on conjugacy classes ($\mathrm{tr}(AB) = \mathrm{tr}(BA)$ so $\mathrm{tr}(ABA^{-1}) = Tr(B)$).

2.2 Characters

Let $f \in S_k(N)$, then for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in 2z$ with $c \equiv 0 \mod N$ we have

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

where $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ is a Dirichlet character mod N. If f is an eigenform for the so called "diamond-bracket operator" $\langle d \rangle$ so that $f|\langle d \rangle = \varepsilon(d) f$ then ε actually takes values in K.

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

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

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

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

Putting these together gives the ℓ-adic cyclotomic character

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

2.3 Parity Conditions

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

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

so $(-1)^k \varepsilon(-1) = 1$ thus

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

Thus the det character is odd so the representation $\rho_{\ell,f}$ is odd.

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

2.4 Conjectures of Serre (mod ℓ version)

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

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

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

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

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

2.5 General remarks on mod p Galois representations

[[This section was written by Joseph Loebach Wetherell.]]

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

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

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

Furthermore ρ is semi-simple if and only if every such sequence splits.

Note that irreducible representations are semi-simple.

One other fact is that semi-simple Galois representations are uniquely determined (up to isomorphism class) by their trace and determinant.

Now, since in the case we are doing, $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact, it follows that the image of any Galois representation ρ into $\mathrm{GL}_2(K_\lambda)$ is compact. Thus we can conjugate it into $\mathrm{GL}_2(\mathcal{O}_\lambda)$. Irreducibility is not needed for this.

Now that we have a representation into $\mathrm{GL}_2(\mathcal{O}_\lambda)$, we can reduce to get a representation $\overline{\rho}$ to $\mathrm{GL}_2(\mathbf{F}_\lambda)$. This reduced representation is not uniquely determined by ρ, since we had a choice of conjugators. However, the trace and determinant are invariant under conjugation, so the trace and determinant of the reduced representation are uniquely determined by ρ.

So we know the trace and determinant of the reduced representation. If we also knew that it was semi-simple, then we would know its isomorphism class, and we would be done. So we would be happy if the reduced representation is irreducible. And in fact, it is easy to see that if the reduced representation is irreducible, then ρ must also be irreducible. Now, it turns out that all ρ of interest to us will be irreducible; unfortunately, we can't go the other way and claim that ρ irreducible implies the reduction is irreducible.

2.6 Serre's Conjecture

Serre has made the following conjecture which is still open at the time of this writing.

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

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

*L*et $\sigma : G \to \mathrm{GL}_2(\mathbf{F})$ ($\mathbf{F}$ is a finite field) be a represenation of the Galois group G. The we say that the *represenions σ is modular* if there is a modular form f, a prime λ, and an embedding $\mathbf{F} \hookrightarrow \overline{\mathbf{F}}_\lambda$ such that $\sigma \cong \overline{\rho}_{\lambda,f}$ over $\overline{\mathbf{F}}_\lambda$.

For more details, see Chapter **??** and [RS01].

2.7 Wiles's Perspective

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

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

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

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

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

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

3
Modular Forms of Level 1

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

3.1 The Definition

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

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

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

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

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

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

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

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

3.2 Some Examples and Conjectures

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

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

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

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

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

Conjecture 3.2.1 (Lehmer). *$\tau(n) \neq 0$ for all $n \geq 1$.*

This conjecture has been verified for $n \leq 22689242781695999$ (see Jordan and Kelly, 1999).

Conjecture 3.2.2 (Edixhoven). *Let p be a prime. There a polynomial time algorithm to compute $\tau(p)$, polynomial in the number of digits of p.*

Edixhoven has proposed an approach to find such an algorithm. His idea is to use ℓ-adic cohomology to find an analogue of the Schoof-Elkies-Atkin algorithm (which counts the number N_q of points on an elliptic curves over a finite field $\mathbf{F}_q$ by computing $N_q \mod \ell$ for many primes ℓ). Here's what Edixhoven has to say about the status of his conjecture (email, October 22, 2003):

> I have made a lot of progress on proving that my method runs in polynomial time, but it is not yet complete. I expect that all should be completed in 2004. For higher weights [...] you need to compute on varying curves such as $X_1(\ell)$ for ℓ up to $\log(p)$ say.
>
> An important by-product of my method is the computation of the mod ℓ Galois representations associated to Δ in time polynomial in ℓ. So, it should be seen as an attempt to make the Langlands correspondence for GL_2 over $\mathbf{Q}$ available computationally.

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

Example 3.2.3. Because E_4 and E_6 generate, it is straightforward to write down a basis for any space $M_k(1)$. For example, the space $M_{36}(1)$ has basis

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

3.3 Modular Forms as Functions on Lattices

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

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

Proposition 3.3.1. *There is a bijection between $\mathcal{R}$ and $\mathcal{E}$ under which $L \in \mathcal{R}$ corresponds to $(\mathbf{C}/L,\ dz) \in \mathcal{E}$.*

Proof. We describe the maps in each direction, but leave the proof that they induce a well-defined bijection as an exercise for the reader. Given $L \in \mathcal{R}$, by Weierstrass theory the quotient $\mathbf{C}/L$ is an elliptic curve, which is equipped with the distinguished differential ω induced by the differential dz on $\mathbf{C}$.

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

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

□

Let

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

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

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

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

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

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

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

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

Lemma 3.3.2. *The lattice function $F : \mathcal{R} \to \mathbf{C}$ associated to $f \in M_k(1)$ is well defined.*

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

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

Multiplying both sides by ω_2^k yields

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

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

Since $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$, we can recover f from F, so the map $f \mapsto F$ is injective. Moreover, it is surjective in the sense that if F is homogeneous of degree $-k$, then F arises from a function $f : \mathfrak{h} \to \mathbf{C}$ that transforms like a modular form. More precisely, if $F : \mathcal{R} \to \mathbf{C}$ satisfies the homogeneity condition $F(\lambda L) = \lambda^{-k}F(L)$, then the function $f : \mathfrak{h} \to \mathbf{C}$ defined by $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$ transforms like a modular form of weight k, as the following computation shows: For any $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ and $\tau \in \mathfrak{h}$, we have

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

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

Proposition 3.3.3. *There is a bijection between $M_k(1)$ and functions $F : \mathcal{R} \to \mathbf{C}$ that are homogeneous of degree $-k$ and holomorphic on $\mathfrak{h} \cup \{\infty\}$. Under this bijection $F : \mathcal{R} \to \mathbf{C}$ corresponds to $f(\tau) = F(\mathbf{Z}\tau + \mathbf{Z})$.*

3.4 Hecke Operators

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

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

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

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

Note that if F is homogeneous of degree $-k$, then $T_n(F)$ is also homogeneous of degree $-k$.

Since $(n,m)=1$ implies $T_nT_m = T_{nm}$ and T_{p^k} is a polynomial in $\mathbf{Z}[T_p]$ (see [Ser73, Cor. 1, pg. 99]), the essential case to consider is n prime.

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

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

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

Recall from Proposition 3.3.1 that there is a bijection between the set $\mathcal{R}$ of lattices in $\mathbf{C}$ and the set $\mathcal{E}$ of isomorphism classes of pairs (E, ω), where ω is a nonzero differential on E.

Suppose $F : \mathcal{R} \to \mathbf{C}$ is homogeneous of degree $-k$, so $F(\lambda L) = \lambda^{-k} F(L)$. Then we may also view T_ℓ as a sum over lattices that contain L with index ℓ, as follows. Suppose $L' \subset L$ is a sublattice of index ℓ and set $L'' = \ell^{-1}L'$. Then we have a chain of inclusions

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

Since $[\ell^{-1}L' : L'] = \ell^2$ and $[L : L'] = \ell$, it follows that $[L'' : L] = \ell$. By homogeneity,

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

3.5 Hecke Operators Directly on q-expansions

Recall that the nth Hecke operator T_n of weight k is

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

Modular forms of weight k correspond to holomorphic functions on lattices of degree $-k$, and T_n extend to an operator on these functions on lattices, so T_n

defines on operator on $M_k(1)$. Recall that Fourier expansion defines an injective map $M_k(1) \subset \mathbf{C}[[q]]$. In this section, we describe $T_n(\sum a_n q^n)$ explicitly as a q-expansion.

3.5.1 Explicit Description of Sublattices

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

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

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

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

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

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

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

Proposition 3.5.1. *Let n be a positive integer. There is a one-to-one correspondence between sublattices $L' \subset L$ of index n and matrices $\left(\begin{smallmatrix}a & b \\ 0 & d\end{smallmatrix}\right)$ with $ad = n$ and $0 \leq b \leq d-1$.*

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

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

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

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

Since $\sigma \in \mathrm{SL}_2(\mathbf{Z})$, we have $n \mid a'd$, and $n \mid ad'$, and $aa'dd' = n^2$. If $a'd > n$, then because $aa'dd' = n^2$, we would have $ad' < n$, which would contradict the fact that $n \mid ad'$; also, $a'd < n$ is impossible since $n \mid a'd$. Thus $a'd = n$ and likewise $ad' = n$. Since $ad = n$ as well, it follows that $a' = a$ and $d' = d$, so $\sigma = \left(\begin{smallmatrix}1 & t \\ 0 & 1\end{smallmatrix}\right)$ for some $t \in \mathbf{Z}$. Then $\sigma\gamma = \left(\begin{smallmatrix}a & b+dt \\ 0 & d\end{smallmatrix}\right)$, which implies that $t = 0$, since $0 \leq b \leq d-1$ and $0 \leq b + dt \leq d-1$. □

Remark 3.5.2. As mentioned earlier, when $n = \ell$ is prime, there are $\ell+1$ sublattices of index ℓ. In general, the number of such sublattices is the sum of the positive divisors of n (exercise).

3.5.2 Hecke operators on q-expansions

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

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

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

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

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

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

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

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

Thus

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

and if $\mu \ge 0$ then the coefficient of q^μ is

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

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

Remark 3.5.3. When $k \geq 1$ the coefficients of q^μ for all μ belong to the $\mathbf{Z}$-module generated by the c_m.

Remark 3.5.4. Setting $\mu = 0$ gives the constant coefficient of $T_n f$ which is

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

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

Remark 3.5.5. Setting $\mu = 1$ shows that the coefficient of q in $T_n f$ is $\sum_{a|1} 1^{k-1} c_n = c_n$. As an immediate corollary we have the following important result.

Corollary 3.5.6. *If f is a cusp form such that $T_n f$ has 0 as coefficient of q for all $n \geq 1$, then $f = 0$.*

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

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

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

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

3.5.3 The Hecke Algebra and Eigenforms

Definition 3.5.7 (Hecke Algebra). The *Hecke algebra* $\mathbf{T}$ associated to $M_k(1)$ is the subring of $\mathrm{End}(M_k(1))$ generated by the operators T_n for all n. Similarly, the *Hecke algebra* associated to $S_k(1)$ is the subring of $\mathrm{End}(S_k(1))$ generated by all Hecke operators T_n.

The Hecke algebra is commutative (e.g., when $(n, m) = 1$ we have $T_n T_m = T_{nm} = T_{mn} = T_m T_n$) of finite rank over $\mathbf{Z}$.

Definition 3.5.8 (Eigenform). An *eigenform* $f \in M_k(1)$ is a nonzero element such that f is an eigenvector for every Hecke operator T_n. If $f \in S_k(1)$ is an eigenform, then f is *normalized* if the coefficient of q in the q-expansion of f is 1. We sometimes called a normalized cuspidal eigenform a *newform*.

If $f = \sum_{n=1}^\infty c_n q^n$ is a normalized eigenform, then Remark 3.5.5 implies that $T_n(f) = c_n f$. Thus the coefficients of a newform are exactly the system of eigenvalues of the Hecke operators acting on the newform.

Remark 3.5.9. It follows from Victor Miller's thesis that $T_1, \ldots, T_n$ generate $\mathbf{T} \subset S_k(1)$, where $n = \dim S_k(1)$.

3.5.4 Examples

```
> M := ModularForms(1,12);
> HeckeOperator(M,2);
[  2049 196560]
[     0    -24]
> S := CuspidalSubspace(M);
> HeckeOperator(S,2);
[-24]
> Factorization(CharacteristicPolynomial(HeckeOperator(M,2)));
[
    <x - 2049, 1>,
    <x + 24, 1>
]
> M := ModularForms(1,40);
> M;
Space of modular forms on Gamma_0(1) of weight 40 and dimension 4
over Integer Ring.
> Basis(M);
[
    1 + 1250172000*q^4 + 7541401190400*q^5 + 9236514405888000*q^6
    + 3770797689077760000*q^7 + O(q^8),
    q + 19291168*q^4 + 37956369150*q^5 + 14446985236992*q^6 +
    1741415886056000*q^7 + O(q^8),
    q^2 + 156024*q^4 + 57085952*q^5 + 1914094476*q^6 -
    27480047616*q^7 + O(q^8),
    q^3 + 168*q^4 - 12636*q^5 + 392832*q^6 - 7335174*q^7 + O(q^8)
]
> HeckeOperator(M,2);
[549755813889 0 1250172000 9236514405888000]
[0 0 549775105056 14446985236992]
[0 1 156024 1914094476]
[0 0 168 392832]
> Factorization(CharacteristicPolynomial(HeckeOperator(M,2)));
[
    <x - 549755813889, 1>,
    <x^3 - 548856*x^2 - 810051757056*x + 213542160549543936, 1>
]
```

3.6 Two Conjectures about Hecke Operators on Level 1 Modular Forms

3.6.1 Maeda's Conjecture

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

There is some numerical evidence for this conjecture. It is true for $k \leq 2000$, according to [FJ02]. Buzzard shows in [Buz96] that for the weights $k \leq 228$ with

$k/12$ a prime, the Galois group of the characteristic polynomial of T_2 is the full symmetric group.

Possible student project: I have computed the characteristic polynomial of T_2 for all weights $k \leq 3000$:

`http://modular.fas.harvard.edu/Tables/charpoly_level1/t2/`

However, I never bothered to try to prove that these are all irreducible, which would establish Maeda's conjecture for $k \leq 3000$. The MathSciNet reviewer of [FJ02] said "In the present paper the authors take a big step forward towards proving Maeda's conjecture in the affirmative by establishing that the Hecke polynomial $T_{p,k}(x)$ is irreducible and has full Galois group over $\mathbb{Q}$ for $k \leq 2000$ and $p < 2000, p$ prime." Thus stepping forward to $k \leq 3000$, at least for $p = 2$, might be worth doing.

3.6.2 The Gouvea-Mazur Conjecture

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

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

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

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

Notice that the above examples, with $p = 5$ and $k_1 = 12$, are consistent with this conjecture. However, the conjecture is false in general. Frank Calegari and Kevin Buzzard recently found the first counterexample, when $p = 59$, $k_1 = 16$, $\alpha = 1$, and $k_2 = 16 + 59 \cdot 58 = 3438$. We have $d(16, 0, 59) = 0d(16, 1, 59) = 1$, $d(16, \alpha, 59) = 0$ for all other α. However, initial computations strongly suggest (but do not prove!) that $d(3438, 1, 59) = 2$. It is a finite, but difficult, computation to decide what $d(3438, 1, 59)$ really is (see Section 3.7). Using a trace formula, Calegari and Buzzard at least showed that either $d(3438, 1, 59) \geq 2$ or there exists $\alpha < 1$ such that $d(3438, \alpha, 59) > 0$, both of which contradict Conjecture 3.6.2.

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

3.7 A Modular Algorithm for Computing Characteristic Polynomials of Hecke Operators

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

In this section, we describe an algorithm to compute the characteristic polynomial of the Hecke operator $T_{p,k}$, which is adapted for the case when $p > 2$. It could be generalized to modular forms for $\Gamma_1(N)$, given a method to compute a basis of q-expansions to "low precision" for the space of modular forms of weight k and level N. By "low precision" we mean to precision $O(q^{dp+1})$, where $T_1, T_2, \ldots, T_d$ generate the Hecke algebra $\mathbf{T}$ as a ring. The algorithm described here uses nothing more than the basics of modular forms and some linear algebra; in particular, no trace formulas or modular symbols are involved.

3.7.1 *Review of Basic Facts About Modular Forms*

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

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

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

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

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

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

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

3.7.2 The Naive Approach

Let k be an even positive integer and p be a prime. Our goal is to compute the characteristic polynomial of the Hecke operator $T_{p,k}$ acting on $S_k(1)$. In practice, when k and p are both reasonably large, e.g., $k = 886$ and $p = 59$, then the coefficients of the characteristic polynomial are huge (the roots of the characteristic polynomial are $O(p^{k/2-1})$). A naive way to compute the characteristic polynomial of $T_{p,k}$ is to use (3.7.1) to compute the matrix $[T_{p,k}]$ of $T_{p,k}$ on the basis of Proposition 3.7.1, where E_4 and E_6 are computed to precision $p \dim M_k(1)$, and to then compute the characteristic polynomial of $[T_{p,k}]$ using, e.g., a modular algorithm (compute the characteristic polynomial modulo many primes, and use the Chinese Remainder Theorem). The difficulty with this approach is that the coefficients of the q-expansions of $E_4^a E_6^b$ to precision $p \dim M_k(1)$ quickly become enormous, so both storing them and computing with them is costly, and the components of $[T_{p,k}]$ are also huge so the characteristic polynomial is difficult to compute. See Example 3.2.3 above, where the coefficients of the q-expansions are already large.

3.7.3 The Eigenform Method

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

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

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

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

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

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

2. Using (3.7.1), we compute the matrix $[T_2]$ of T_2 with respect to Miller's basis $f_1, \ldots, f_d$.

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

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

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

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

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

since $a_p(f_i) = 0$ for all $i \neq p$ and $a_p(f_p) = 1$. Once we have computed b_p, we can compute the characteristic polynomial of T_p, because it is the minimal polynomial of b_p. We spend the rest of this section discussing how to make this step practical.

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

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

1. $f(x)$ has no repeated roots modulo ℓ,

2. ℓ does not divide any denominator involved in our representation of $\mathbf{e}$, and

3. the image of e_1 in $\mathbf{F}_\ell[x]/(f(x))$ is invertible.

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

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

3.7.4 How to Write Down an Eigenvector over an Extension Field

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

Algorithm 3.7.5. Let A be an $n \times n$ matrix over an arbitrary field k and suppose that the characteristic polynomial $f(x) = x^n + \cdots + a_1 x + a_0$ of A is irreducible. Let α be a root of $f(x)$ in an algebraic closure $\overline{k}$ of k. Factor $f(x)$ over $k(\alpha)$ as $f(x) = (x - \alpha)g(x)$. Then for any element $v \in k^n$ the vector $g(A)v$ is either 0 or it is an eigenvector of A with eigenvalue α. The vector $g(A)v$ can be computed by finding Av, $A(Av)$, $A(A(Av))$, and then using that

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

where the coefficients c_i are determined by the recurrence

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

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

Proof. By the Cayley-Hamilton theorem [Lan93, XIV.3] we have that $f(A) = 0$. Consequently, for any $v \in k^n$, we have $(A - \alpha)g(A)v = 0$ so that $Ag(A)v = \alpha v$. Since f is irreducible it is the polynomial of least degree satisfied by A and so $g(A) \neq 0$. Therefore $g(A)v \neq 0$ for all v not in the proper closed subspace $\ker(g(A))$. □

3.7.5 Simple Example: Weight 36, $p = 3$

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

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

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

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

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

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

This matrix has (irreducible) characteristic polynomial

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

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

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

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

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

which has characteristic polynomial

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

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

$$x^3 + 104875308x^2 - 144593891972573904x - 21175292105104984004394432.$$

4
Analytic theory of modular curves

4.1 The Modular group

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

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

4.1.1 The Upper half plane

Let

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

be the open complex upper half plane. The group

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

acts by linear fractional transformations ($z \mapsto (az+b)/(cz+d)$) on $\mathbf{C} \cup \{\infty\}$. By the following lemma, $\mathrm{SL}_2(\mathbf{R})$ also acts on $\mathfrak{h}$:

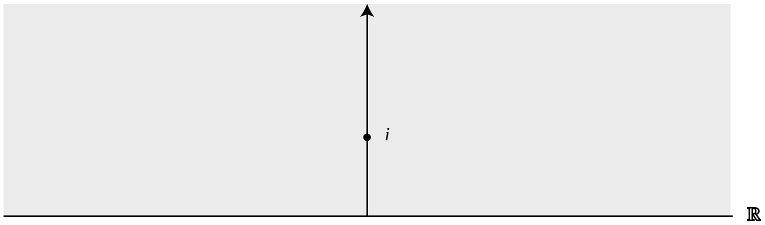


FIGURE 4.1.1. The upper half plane $\mathfrak{h}$

Lemma 4.1.1. *Suppose $g \in \mathrm{SL}_2(\mathbf{R})$ and $z \in \mathfrak{h}$. Then*

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

Proof. Apply the identity $\mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z})$ to both sides of the asserted equality and simplify. □

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

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

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

4.1.2 Fundamental domain for the modular group

4.1.3 Conjugating an element of the upper half plane into the fundamental domain

4.1.4 Writing an element in terms of generators

4.1.5 Generators for the modular group

4.2 Congruence subgroups

4.2.1 Definition

4.2.2 Fundamental domains for congruence subgroups

4.2.3 Coset representatives

4.2.4 Generators for congruence subgroups

Simple method

Sophisticated method

4.3 Modular curves

4.3.1 The upper half plane is a disk

4.3.2 The upper half plane union the cusps

4.3.3 The Poincaré metric

4.3.4 Fuchsian groups and Riemann surfaces

Definition of Fuchsian group. Quotient of upper half plane.

4.3.5 Riemann surfaces attached to congruence subgroups

$X_0(N)$

4.4 Points on modular curves parameterize elliptic curves with extra structure

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

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

Lemma 4.4.1. *Suppose $\varphi : \mathbf{C}/\Lambda_1 \to \mathbf{C}/\Lambda_2$ is nonzero. Then the kernel of φ is isomorphic to $\Lambda_2/T(\Lambda_1)$.*

Lemma 4.4.2. *Two complex tori $\mathbf{C}/\Lambda_1$ and $\mathbf{C}/\Lambda_2$ are isomorphic if and only if there is a complex number α such that $\alpha\Lambda_1 = \Lambda_2$.*

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

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

Proposition 4.4.3. *Suppose $\tau, \tau' \in \mathfrak{h}$. Then $E_\tau \cong E_{\tau'}$ if and only if there exists $g \in \mathrm{SL}_2(\mathbf{Z})$ such that $\tau = g(\tau')$. Thus the set of isomorphism classes of elliptic curves over $\mathbf{C}$ is in natural bijection with the orbit space $\mathrm{SL}_2(\mathbf{Z})\backslash\mathfrak{h}$.*

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

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

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

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

Let $E = \mathbf{C}/\Lambda$ be an elliptic curve over $\mathbf{C}$ and N a positive integer. Using Lemma 4.4.1, we see that

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

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

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

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

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

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

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

Notice that $e(P_\tau, Q_\tau) = -1 \in \mathbf{Z}/N\mathbf{Z}$. Also if $\mathbf{C}/\Lambda \cong \mathbf{C}/\Lambda'$ via multiplication by α, and $P, Q \in (\mathbf{C}/\Lambda)[N]$, then $e(\alpha(P), \alpha(Q)) = e(P, Q)$.

Theorem 4.4.4. *Let N be a positive integer.*

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

2. *The non-cuspidal points on $X_1(N)$ correspond to pairs (E, P) where P is a point on E of exact order N. (Two pairs (E, P) and (E', P') isomorphic if there is an isomorphism $\varphi : E \to E'$ such that $\varphi(P) = P'$.)*

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

This theorem follows from Propositions 4.4.5 and 4.4.7 below.

Proposition 4.4.5. *Let E be an elliptic curve over $\mathbf{C}$. If C is a cyclic subgroup of E of order N, then there exists $\tau \in \mathfrak{h}$ such that (E, C) is isomorphic to (E_τ, C_τ). If P is a point on E of order N, then there exists $\tau \in \mathbf{C}$ such that (E, P) is isomorphic to (E_τ, P_τ). If P, Q is a basis for $E[N]$ and $e(P, Q) = -1 \in \mathbf{Z}/N\mathbf{Z}$, then there exists $\tau \in \mathbf{C}$ such that (E, P, Q) is isomorphic to (E_τ, P_τ, Q_τ).*

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

Suppose $P = a\omega_1/N + b\omega_2/N$ is a point of order N. Then $\gcd(a, b, N) = 1$, otherwise P would have order strictly less than N, a contradiction. Thus we can modify a and b by adding multiples of N to them (this follows from the fact that $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective), so that $P = a\omega_1/N + b\omega_2/N$ and $\gcd(a, b) = 1$. There exists $c, d \in \mathbf{Z}$ such that $ad - bc = 1$, so $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$ form a basis for Λ, and C is generated by $P = \omega_1'/N$. If necessary, replace ω_2' by $-\omega_2'$ so that $\tau = \omega_2'/\omega_1' \in \mathfrak{h}$. Then (E, P) is isomorphic to (E_τ, P_τ). Also, if C is the subgroup generated by P, then (E, C) is isomorphic to (E_τ, C_τ).

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

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

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

Remark 4.4.6. Part 3 of Theorem 2.4 in Chapter 11 of Husemöller's book on elliptic curves is **wrong**, since he neglects the Weil pairing condition. Also the first paragraph of his proof of the theorem is incomplete.

The following proposition completes the proof of Theorem 4.4.4.

Proposition 4.4.7. *Suppose $\tau, \tau' \in \mathfrak{h}$. Then (E_τ, C_τ) is isomorphic $(E_{\tau'}, C_{\tau'})$ if and only if there exists $g \in \Gamma_0(N)$ such that $g(\tau) = \tau'$. Also, (E_τ, P_τ) is isomorphic $(E_{\tau'}, P_{\tau'})$ if and only if there exists $g \in \Gamma_1(N)$ such that $g(\tau) = \tau'$. Finally, (E_τ, P_τ, Q_τ) is isomorphic $(E_{\tau'}, P_{\tau'}, Q_{\tau'})$ if and only if there exists $g \in \Gamma(N)$ such that $g(\tau) = \tau'$.*

Proof. We prove only the first assertion, since the others are proved in a similar way. Suppose (E_τ, C_τ) is isomorphic to (E_τ', C_τ'). Then there is $\lambda \in \mathbf{C}$ such that $\lambda\Lambda_\tau = \Lambda_{\tau'}$. Thus $\lambda\tau = a\tau' + b$ and $\lambda 1 = c\tau' + d$ with $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ (as we saw in the proof of Proposition 4.4.3). Dividing the second equation by N we get $\lambda \frac{1}{N} = \frac{c}{N}\tau' + \frac{d}{N}$, which lies in $\Lambda_{\tau'} = \mathbf{Z}\tau' + \frac{1}{N}\mathbf{Z}$, by hypothesis. Thus $c \equiv 0 \pmod{N}$, so $g \in \Gamma_0(N)$, as claimed. For the converse, note that if $N \mid c$, then $\frac{c}{N}\tau' + \frac{d}{N} \in \Lambda_{\tau'}$. □

4.5 The Genus of $X(N)$

Let N be a positive integer. The aim of this section is to establish some facts about modular curves associated to congruence subgroups and compute the genus of $X(N)$. Similar methods can be used to compute the genus of $X_0(N)$ and $X_1(N)$ (for $X_0(N)$ see [Shi94, §1.6] and for $X_1(N)$ see [DI95, §9.1]).

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

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

Theorem 4.5.1 (Hurwitz Formula). *Let $f : X \to Y$ be as above. Then*

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

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

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

Let X be one of the modular curves $X_0(N)$, $X_1(N)$, or $X(N)$ corresponding to a congruence subgroup Γ, and let $Y = X(1) = \mathbf{P}^1$. There is a natural map $f : X \to Y$ got by sending the equivalence class of τ modulo the congruence subgroup Γ to the equivalence class of τ modulo $\mathrm{SL}_2(\mathbf{Z})$. This is "the" map $X \to \mathbf{P}^1$ that we mean everywhere below.

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

Proposition 4.5.2. *Suppose $N > 2$. The degree of the map $X_0(N) \to \mathbf{P}^1$ is $N\prod_{p|N}(1 + 1/p)$. The degree of the map $X_1(N) \to \mathbf{P}^1$ is $\frac{1}{2}N^2\prod_{p|N}(1 - 1/p^2)$. The degree of the map from $X(N) \to \mathbf{P}^1$ is $\frac{1}{2}N^3\prod_{p|N}(1 - 1/p^2)$. If $N = 2$, then the degrees are 3, 3, and 6, respectively.*

Proof. This follows from the discussion above, Exercise X about indices of congruence subgroups in $\mathrm{SL}_2(\mathbf{Z})$, and the observation that for $N > 2$ the groups $\Gamma(N)$ and $\Gamma_1(N)$ do not contain -1 and the group $\Gamma_0(N)$ does. □

Proposition 4.5.3. *Let X be $X_0(N)$, $X_1(N)$ or $X(N)$. Then the map $X \to \mathbf{P}^1$ is ramified at most over ∞ and the two points corresponding to elliptic curves with extra automorphisms (i.e., the two elliptic curves with j-invariants 0 and 1728).*

Proof. Since we have a tower $X(N) \to X_1(N) \to X_0(N) \to \mathbf{P}^1$, it suffices to prove the assertion for $X = X(N)$. Since we do not claim that there is no ramification over ∞, we may restrict to $Y(N)$. By Theorem 4.4.4, the points on $Y(N)$ correspond to isomorphism classes of triples (E, P, Q), where E is an elliptic curve over $\mathbf{C}$ and P, Q are a basis for $E[N]$. The map from $Y(N)$ to $\mathbf{P}^1$ sends the isomorphism class of (E, P, Q) to the isomorphism class of E. The equivalence class of (E, P, Q) also contains $(E, -P, -Q)$, since $-1 : E \to E$ is an isomorphism. The only way the fiber over E can have cardinality smaller than the degree is if there is an extra equivalence $(E, P, Q) \to (E, \varphi(P), \varphi(Q))$ with φ an automorphism of E not equal to ± 1. The theory of CM elliptic curves shows that there are only two isomorphism classes of elliptic curves E with automorphisms other than ± 1, and these are the ones with j-invariant 0 and 1728. This proves the proposition. □

Theorem 4.5.4. *For $N > 2$, the genus of $X(N)$ is*

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

For $N = 1, 2$, the genus is 0.

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

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

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

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

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

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

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

so

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

Substituting the formula for d from Proposition 4.5.2 yields the claimed formula. □

5
Modular Symbols

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

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

Some standard references for modular symbols are [Man72] [Lan95, IV], [Cre97], and [Mer94]. Sections 5.1–5.2 below very closely follow Section 1 of Manin's paper [Man72].

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

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

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

5.1 Modular symbols

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

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

hence an isomorphism

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

For more details, see [Lan95, §IV.1].

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

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

Definition 5.1.1 (Modular symbol). The homology class $\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{R})$ associated to $\alpha, \beta \in \mathfrak{h}^*$ is called the *modular symbol* attached to α and β.

Proposition 5.1.2. *The symbols $\{\alpha, \beta\}$ have the following properties:*

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

Remark 5.1.3. We only have $\{\alpha, \beta\} = \{\beta, \alpha\}$ if $\{\alpha, \beta\} = 0$, so the modular symbols notation, which suggests "unordered pairs", is actively misleading.

Proposition 5.1.4. *For any $\alpha \in \mathfrak{h}^*$, the map $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ that sends g to $\{\alpha, g\alpha\}$ is a surjective group homomorphism that does not depend on the choice of α.*

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

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

so the map is a group homomorphism. To see that the map does not depend on the choice of α, suppose $\beta \in \mathfrak{h}^*$. By Proposition 5.1.2, we have $\{\alpha, \beta\} = \{g\alpha, g\beta\}$. Thus

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

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

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

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

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

5.2 Manin symbols

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

5.2.1 Using continued fractions to obtain surjectivity

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

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

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

Proposition 5.2.1. *Any element of* $\mathrm{H}_1(X(G), \mathbf{Z})$ *is a sum of elements of the form* $[r]$*, and the representation* $\sum n_r\{\alpha_r, \beta_r\}$ *of* $h \in \mathrm{H}_1(X(G), \mathbf{Z})$ *can be chosen so that* $\sum n_r(\pi(\beta_r) - \pi(\alpha_r)) = 0 \in \mathrm{Div}(X(G))$.

Proof. By Proposition 5.1.4, every element h of $\mathrm{H}_1(X(G), \mathbf{Z})$ is of the form $\{0, g(0)\}$ for some $g \in \mathbf{G}$. If $g(0) = \infty$, then $h = [G]$ and $\pi(\infty) = \pi(0)$, so we may assume $g(0) = a/b \neq \infty$, with a/b in lowest terms and $b > 0$. Also assume $a > 0$, since the case $a < 0$ is treated in the same way. Let

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

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

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

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

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

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

Lemma 5.2.2. *If* $x = \sum_{j=1}^{t} n_j\{\alpha_j, \beta_j\}$ *is a* $\mathbf{Z}$*-linear combination of modular symbols for* G *and* $\sum n_j(\pi(\beta_j) - \pi(\alpha_j)) = 0 \in \mathrm{Div}(X(G))$*, then* $x \in \mathrm{H}_1(X(G), \mathbf{Z})$.

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

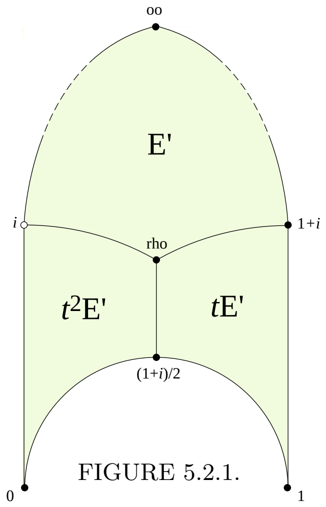


FIGURE 5.2.1.

5.2.2 Triangulating $X(G)$ to obtain injectivity

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

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

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

Let B be the subgroup of C generated by symbols (r), for all $r \in G\backslash\Gamma$ that satisfy $r = rt$, and by $(r) + (rt) + (rt^2)$ for all other r. If $r = rt$, then $rt(0) = r(0)$, so $r(\infty) = r(0)$, so $(r) \in Z$. Also, using (5.2.1) below, we see that for any r, the element $(r) + (rt) + (rt^2)$ lies in Z.

The map $G\backslash\Gamma \to \mathrm{H}_1(X(G), \mathbf{R})$ that sends (r) to $[r]$ induces a homomorphism $C \to \mathrm{H}_1(X(G), \mathbf{R})$, so by Proposition 5.2.1 we obtain a surjective homomorphism

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

Theorem 5.2.3 (Manin). *The map $\psi : Z/B \to \mathrm{H}_1(X(G), \mathbf{Z})$ is an isomorphism.*

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

Let E denote the *interior* of the triangle with vertices 0, 1, and ∞, as illustrated in Figure 5.2.1. Let E' denote the union of the interior of the region bounded by the path from i to $\rho = e^{\pi i/3}$ to $1+i$ to ∞ with the indicated path from i to ρ, not including the vertex i.

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

Figure 5.2.1:

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

(5.2.1)

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

If $(0,\infty)$ is the path from 0 to ∞, then $t(0,\infty) = (\infty,1)$ and $t^2(0,\infty) = (1,0)$. Also, $s(0,\infty) = (\infty,0)$. Thus each half side of E is Γ-conjugate to the side from i to ∞. Also, each 1-simplex in Figure 5.2.1, i.e., the sides that connected two adjacent labeled vertices such as i and ρ, maps homeomorphically into $X(\Gamma)$. This is clear for the half sides, since they are conjugate to a path in the interior of the standard fundamental domain for Γ, and for the medians (lines from midpoints to ρ) since the path from i to ρ is on an edge of the standard fundamental domain with no self identifications.

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

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

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

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

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

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

$$e_1(r) = e_1(r') \qquad \text{implies} \qquad r' = rs. \tag{5.2.2}$$

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

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

2-sided: The 2-sided 2-cells $e_2(r)$ are indexed by the cosets $r = Gg$ such that $rt = r$. Note that for such an r, we have $\pi(rE') = \pi(rtE') = \pi(rt^2E')$. The 2-cell $e_2(r)$ is $\pi(gE')$. The image $g(\rho, i)$ of the half median maps to a line from the center of $e_2(r)$ to the edge $\pi(g(i)) = \pi(g(1+i))$. Orient $e_2(r)$ in a way compatible with the e_1. Since $Ggt = Gg$,

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

so

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

Thus

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

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

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

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

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

Let C, Z, B be as defined before the statement of the Theorem 5.2.3. We have $\mathrm{H}_1(X(G), \mathbf{Z}) \cong Z_1/B_1$, and would like to prove that $Z/B \cong Z_1/B_1$.

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

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

If $n_r \neq 0$ then $r \neq rs$, so (5.2.2) implies that $e_1(r) \neq e_1(rs)$. If $n_r \neq 0$ and $n_{r'} \neq 0$ with $r' \neq r$, then $r \neq rs$ and $r' \neq r's$, so $e_1(r), e_1(rs), e_1(r'), e_1(r's)$ are all distinct. We conclude that $\sum n_r(e_1(rs) - e_1(r)) \neq 0$, which proves that φ is injective.

Suppose $(r) \in C$. Then

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

since

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

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

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

□

5.3 Hecke Operators

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

There is a commutative ring

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

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

Suppose

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

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

Example 5.3.1. Using MAGMA we see that the homology $\mathrm{H}_1(X_0(11), \mathbf{R})$ is generated by $\{-1/7, 0\}$ and $\{-1/5, 0\}$.

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

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

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

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

```
> x := 5*pi(M!<1,[Cusps()|0,Infinity()]>);
> y := pi(M!<1,[-2/5,0]>);
> x;
{-1/5, 0}
> y;
-1*{-1/7, 0} + -1*{-1/5, 0}
> x eq y;
false
```

Definition 5.3.2 (Hecke operators). We define the *Hecke operator* T_p on $\mathrm{H}_1(X_0(N), \mathbf{R})$ as follows. When p is a prime with $p \nmid N$, we have

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

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

Example 5.3.3. We continue with Example 5.3.1. If we apply the Hecke operator T_2 to both $5\{0, \infty\}$ and $\{-1/5, 0\}$, the "non-well-definedness" cancels out.

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

Examples 5.3.1 shows that it is not clear that the definition of T_p given above makes sense. For example, if $\{\alpha, \beta\}$ is replaced by an equivalent modular symbol $\{\alpha', \beta'\}$, why does the formula for T_p give the same answer? We will not address this question further here, but will revisit it later when we have a more natural and intrinsic definition of Hecke operators. We only remark that T_p is induced by a "correspondence" from $X_0(N)$ to $X_0(N)$, so T_p preserve $\mathrm{H}_1(X_0(N), \mathbf{Z})$.

5.4 Modular Symbols and Rational Homology

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

Theorem 5.4.1 (Manin). *For any $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, we have*

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

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

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

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

so

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

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

There are general theorems about the denominator of $\{\alpha, \beta\}$ in some cases. Example 5.3.1 above demonstrated the following theorem in the case $N = 11$.

Theorem 5.4.2 (Ogg [Ogg71]). *Let N be a prime. Then the image*

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

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

5.5 Special Values of L-functions

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

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

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

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

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

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

so, by a theorem of Hecke which we will prove [later], $L(f, s)$ is holomorphic on all $\mathbf{C}$. Note that $L(f, s)$ is the Mellin transform of the modular form f:

$$L(f, s) = (2\pi)^s \Gamma(s)^{-1} \int_0^{i\infty} (-iz)^s f(z) \frac{dz}{z}. \qquad (5.5.1)$$

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

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

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

Theorem 5.5.1. *Let E be an elliptic curve over $\mathbf{Q}$. Then $L(E,1)/\Omega_E \in \mathbf{Q}$.*

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

By Equation 5.5.1, we have

$$L(E,1) = 2\pi \int_0^{i\infty} -izf(z)dz/z. \tag{5.5.2}$$

If $q = e^{2\pi iz}$, then $dq = 2\pi iqdz$, so $2\pi if(z)dz = dq/q$, and (5.5.2) becomes

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

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

A standard fact is that if

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

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

The next crucial observation is that by Theorem 5.4.1, there is an integer n such that $n\{0,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$. This is relevant because if

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

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

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

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

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

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

6

Modular Forms of Higher Level

6.1 Modular Forms on $\Gamma_1(N)$

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

Definition 6.1.1 (Modular Forms). The space of *modular forms* of level N and weight k is

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

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

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

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

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

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

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

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

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

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

Lemma 6.1.2. *Let $\alpha \in \mathrm{SL}_2(\mathbf{Z})$. Then there exists a positive integer h such that $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \alpha^{-1}\Gamma_1(N)\alpha$.*

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

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

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

on the upper half plane. We say that f vanishes at x if $b_{n/h} = 0$, so a cusp form is a form that vanishes at every cusp.

6.2 The Diamond Bracket and Hecke Operators

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

6.2.1 Diamond Bracket Operators

The group $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$, and the quotient $\Gamma_0(N)/\Gamma_1(N)$ is isomorphic to $(\mathbf{Z}/N\mathbf{Z})^*$. From this structure we obtain an action of $(\mathbf{Z}/N\mathbf{Z})^*$ on $S_k(\Gamma_1(N))$, and use it to decompose $S_k(\Gamma_1(N))$ as a direct sum of more manageable chunks $S_k(\Gamma_1(N), \varepsilon)$.

Definition 6.2.1 (Dirichlet character). A *Dirichlet character* ε modulo N is a homomorphism

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

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

Remarks 6.2.2.

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

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

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

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

Definition 6.2.3. Let $\overline{d} \in (\mathbf{Z}/N\mathbf{Z})^*$ and $f \in S_k(\Gamma_1(N))$. The map $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ is surjective, so there exists a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ such that $d \equiv \overline{d} \pmod{N}$. The *diamond bracket d operator* is then

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

Remark 6.2.4. Fred Diamond was named after diamond bracket operators.

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

For each Dirichlet character ε mod N let

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

where d_γ is the lower-left entry of γ.

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

Lemma 6.2.5. *The operator $\langle d\rangle$ on the finite-dimensional vector space $S_k(\Gamma_1(N))$ is diagonalizable.*

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

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

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

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

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

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

6.2.2 Hecke Operators on q-expansions

Suppose

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

and let p be a prime. Then

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

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

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

We record the relations

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

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

6.3 Old and New Subspaces

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

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

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

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

Definition 6.3.1 (Old Subspace). The *old subspace* of $S_k(\Gamma_1(N))$ is the subspace generated by the images of the φ_M for all $M \mid N$ with $M \neq N$.

Definition 6.3.2 (New Subspace). The *new subspace* of $S_k(\Gamma_1(N))$ is the complement of the old subspace with respect to the Petersson inner product.

Since I haven't introduced the Petersson inner product yet, note that the new subspace of $S_k(\Gamma_1(N))$ is the largest subspace of $S_k(\Gamma_1(N))$ that is stable under the Hecke operators and has trivial intersection with the old subspace of $S_k(\Gamma_1(N))$.

Definition 6.3.3 (Newform). A *newform* is an element f of the new subspace of $S_k(\Gamma_1(N))$ that is an eigenvector for every Hecke operator, which is normalized so that the coefficient of q in f is 1.

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

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

In the converse direction, it is a deep theorem of Breuil, Conrad, Diamond, Taylor, and Wiles that if E is any elliptic curve over $\mathbf{Q}$, then E is isogenous to A_f for some f of level equal to the conductor N of E.

When f has weight greater than 2, Scholl constructs, in an analogous way, a Grothendieck motive (=compatible collection of cohomology groups) $\mathcal{M}_f$ attached to f.

7
Newforms and Euler Products

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

7.1 Atkin, Lehner, Li Theory

The results of [Li75] about newforms are proved using many linear transformations that do not necessarily preserve $S_k(\Gamma_1(N), \varepsilon)$. Thus we introduce more general spaces of cusp forms, which these transformations preserve. These spaces are also useful because they make precise how the space of cusp forms for the full congruence subgroup $\Gamma(N)$ can be understood in terms of spaces $S_k(\Gamma_1(M), \varepsilon)$ for various M and ε, which justifies our usual focus on these latter spaces. This section follows [Li75] closely.

Let M and N be positive integers and define

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

and

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

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

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

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

We have

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

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

Let

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

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

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

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

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

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

and

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

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

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

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

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

and

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

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

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

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

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

$$B_N : S_k(M, N, \varepsilon) \to S_k(MN, 1, \varepsilon) = S_k(\Gamma_1(MN), \varepsilon). \tag{7.1.1}$$

Putting these maps together allows us to completely understand the cusp forms $S_k(\Gamma(N))$ in terms of spaces $S_k(\Gamma_1(N^2), \varepsilon)$, for all Dirichlet characters ε that arise from characters modulo N. (Recall that $\Gamma(N)$ is the principal congruence subgroup $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}))$. This is because $S_k(\Gamma(N))$ is isomorphic to the direct sum of $S_k(N, N, \varepsilon)$, as ε various over all Dirichlet characters modulo N.

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

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

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

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

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

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

Remark 7.1.1. Because of these relations, (7.1.1) describe $S_k(\Gamma(N))$ as a module over the ring generated by T_p for $p \nmid N$.

Definition 7.1.2 (Old Subspace). The *old subspace* $S_k(M, N, \varepsilon)_{\mathrm{old}}$ is the subspace of $S_k(M, N, \varepsilon)$ generated by all $f | B_d$ and $g | C_e$ where $f \in S_k(M', N)$, $g \in S_k(M, N')$, and M', N' are proper factors of M, N, respectively, and $d \mid M/M'$, $e \mid N/N'$.

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

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

Definition 7.1.3 (Petersson Inner Product). If $f, g \in S_k(\Gamma(N))$, the *Petersson inner product* of f and g is

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

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

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

Proposition 7.1.4 (Petersson). *If $p \nmid N$ and $f \in S_k(\Gamma_1(N), \varepsilon)$, then $\langle f|T_p, g\rangle = \varepsilon(p)\langle f, g|T_p\rangle$.*

Remark 7.1.5. The proposition implies that the T_p, for $p \nmid N$, are diagonalizable. Be careful, because the T_p, with $p \mid N$, need not be diagonalizable.

Definition 7.1.6 (New Subspace). The *new subspace* $S_k(M, N, \varepsilon)_{\text{new}}$ is the orthogonal complement of $S_k(M, N, \varepsilon)_{old}$ in $S_k(M, N, \varepsilon)$ with respect to the Petersson inner product.

Both the old and new subspaces of $S_k(M, N, \varepsilon)$ are preserved by the Hecke operators T_p with $(p, NM) = 1$.

Remark 7.1.7. Li [Li75] also gives a purely algebraic definition of the new subspace as the intersection of the kernels of various trace maps from $S_k(M, N, \varepsilon)$, which are obtained by averaging over coset representatives.

Definition 7.1.8 (Newform). A *newform* $f = \sum a_n q_N^n \in S_k(M, N, \varepsilon)$ is an element of $S_k(M, N, \varepsilon)_{\text{new}}$ that is an eigenform for all T_p, for $p \nmid NM$, and is normalized so that $a_1 = 1$.

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

After proving many technical but elementary lemmas about the operators B_d, C_d, U_p^N, T_p, and W_q^M, Li uses the lemmas to deduce the following theorems. The proofs are all elementary, but there is little I can say about them, except that you just have to read them.

Theorem 7.1.9. *Suppose $f = \sum a_n q_N^n \in S_k(M, N, \varepsilon)$ and $a_n = 0$ for all n with $(n, K) = 1$, where K is a fixed positive integer. Then $f \in S_k(M, N, \varepsilon)_{\text{old}}$.*

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

Theorem 7.1.10. Let $f = \sum a_n q_N^n$ be a newform in $S_k(M, N, \varepsilon)$, p a prime with $(p, MN) = 1$, and $q \mid MN$ a prime. Then

1. $f|T_p = a_p f$, $f|U_q^N = a_q f$, and for all $n \geq 1$,

$$\begin{aligned} a_p a_n &= a_{np} + \varepsilon(p)p^{k-1}a_{n/p}, \\ a_q a_n &= a_{nq}. \end{aligned}$$

If $L(f, s) = \sum_{n \geq 1} a_n n^{-s}$ is the Dirichlet series associated to f, then $L(f, s)$ has an Euler product

$$L(f, s) = \prod_{q|MN} (1 - a_q q^{-s})^{-1} \prod_{p \nmid MN} (1 - a_p p^{-s} + \varepsilon(p)p^{k-1}p^{-2s})^{-1}.$$

2. (a) If ε is not a character mod MN/q, then $|a_q| = q^{(k-1)/2}$.
   (b) If ε is a character mod MN/q, then $a_q = 0$ if $q^2 \mid MN$, and $a_q^2 = \varepsilon(q)q^{k-2}$ if $q^2 \nmid MN$.

7.2 The U_p Operator

Let N be a positive integer and M a divisor of N. For each divisor d of N/M we define a map

$$\alpha_d : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)) : \quad f(\tau) \mapsto f(d\tau).$$

We verify that $f(d\tau) \in S_k(\Gamma_1(N))$ as follows. Recall that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we write

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

The transformation condition for f to be in $S_k(\Gamma_1(N))$ is that $f|[\gamma]_k(\tau) = f(\tau)$. Let $f(\tau) \in S_k(\Gamma_1(M))$ and let $\iota_d = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $f|[\iota_d]_k(\tau) = d^{k-1}f(d\tau)$ is a modular form on $\Gamma_1(N)$ since $\iota_d^{-1}\Gamma_1(M)\iota_d$ contains $\Gamma_1(N)$. Moreover, if f is a cusp form then so is $f|[\iota_d]_k$.

Proposition 7.2.1. *If $f \in S_k(\Gamma_1(M))$ is nonzero, then*

$$\left\{ \alpha_d(f) \; : \; d \mid \frac{N}{M} \right\}$$

is linearly independent.

Proof. If the q-expansion of f is $\sum a_n q^n$, then the q-expansion of $\alpha_d(f)$ is $\sum a_n q^{dn}$. The matrix of coefficients of the q-expansions of $\alpha_d(f)$, for $d \mid (N/M)$, is upper triangular. Thus the q-expansions of the $\alpha_d(f)$ are linearly independent, hence the $\alpha_d(f)$ are linearly independent, since the map that sends a cusp form to its q-expansion is linear. □

When $p \mid N$, we denote by U_p the Hecke operator T_p acting on the image space $S_k(\Gamma_1(N))$. For clarity, in this section we will denote by $T_{p,M}$, the Hecke operator $T_p \in \mathrm{End}(S_k(\Gamma_1(M)))$. For $f = \sum a_n q^n \in S_k(\Gamma_1(N))$, we have

$$f|U_p = \sum a_{np} q^n.$$

Suppose $f = \sum a_n q^n \in S_k(\Gamma_1(M))$ is a normalized eigenform for all of the Hecke operators T_n and $\langle n \rangle$, and p is a prime that does not divide M. Then

$$f|T_{p,M} = a_p f \quad \text{and} \quad f|\langle p \rangle = \varepsilon(p) f.$$

Assume $N = p^r M$, where $r \geq 1$ is an integer. Let

$$f_i(\tau) = f(p^i \tau),$$

so $f_0, \ldots, f_r$ are the images of f under the maps $\alpha_{p^0}, \ldots, \alpha_{p^r}$, respectively, and $f = f_0$. We have

$$\begin{aligned} f|T_{p,M} &= \sum_{n \geq 1} a_{np} q^n + \varepsilon(p) p^{k-1} \sum a_n q^{pn} \\ &= f_0|U_p + \varepsilon(p) p^{k-1} f_1, \end{aligned}$$

so

$$f_0|U_p = f|T_{p,M} - \varepsilon(p)p^{k-1}f_1 = a_p f_0 - \varepsilon(p)p^{k-1}f_1.$$

Also

$$f_1|U_p = \left(\sum a_n q^{pn}\right)|U_p = \sum a_n q^n = f_0.$$

More generally, for any $i \geq 1$, we have $f_i|U_p = f_{i-1}$.

The operator U_p preserves the two dimensional vector space spanned by f_0 and f_1, and the matrix of U_p with respect to the basis f_0, f_1 is

$$A = \begin{pmatrix} a_p & 1 \\ -\varepsilon(p)p^{k-1} & 0 \end{pmatrix},$$

which has characteristic polynomial

$$X^2 - a_p X + p^{k-1}\varepsilon(p). \tag{7.2.1}$$

7.2.1 A Connection with Galois Representations

This leads to a striking connection with Galois representations. Let f be a newform and let $K = K_f$ be the field generated over $\mathbf{Q}$ by the Fourier coefficients of f. Let ℓ be a prime and λ a prime lying over ℓ. Then Deligne (and Serre, when $k = 1$) constructed a representation

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, K_\lambda).$$

If $p \nmid N\ell$, then ρ_λ is unramified at p, so if $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ if a Frobenius element, then $\rho_\lambda(\mathrm{Frob}_p)$ is well defined, up to conjugation. Moreover, one can show that

$$\det(\rho_\lambda(\mathrm{Frob}_p)) = p^{k-1}\varepsilon(p), \quad \text{and}$$
$$\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_p)) = a_p.$$

(We will discuss the proof of these relations further in the case $k = 2$.) Thus the characteristic polynomial of $\rho_\lambda(\mathrm{Frob}_p) \in \mathrm{GL}_2(E_\lambda)$ is

$$X^2 - a_p X + p^{k-1}\varepsilon(p),$$

which is the same as (7.2.1).

7.2.2 When is U_p Semisimple?

Question 7.2.2. Is U_p semisimple on the span of f_0 and f_1?

If the eigenvalues of U_p are distinct, then the answer is yes. If the eigenvalues are the same, then $X^2 - a_p X + p^{k-1}\varepsilon(p)$ has discriminant 0, so $a_p^2 = 4p^{k-1}\varepsilon(p)$, hence

$$a_p = 2p^{\frac{k-1}{2}}\sqrt{\varepsilon(p)}.$$

Open Problem 7.2.3. Does there exist an eigenform $f = \sum a_n q^n \in S_k(\Gamma_1(N))$ such that $a_p = 2p^{\frac{k-1}{2}}\sqrt{\varepsilon(p)}$?

It is a curious fact that the Ramanujan conjectures, which were proved by Deligne in 1973, imply that $|a_p| \leq 2p^{(k-1)/2}$, so the above equality remains taunting. When $k = 2$, Coleman and Edixhoven proved that $|a_p| < 2p^{(k-1)/2}$.

7.2.3 An Example of Non-semisimple U_p

Suppose $f = f_0$ is a normalized eigenform. Let W be the space spanned by f_0, f_1 and let V be the space spanned by f_0, f_1, f_2, f_3. Then U_p acts on V/W by $\overline{f}_2 \mapsto 0$ and $\overline{f}_3 \mapsto \overline{f}_2$. Thus the matrix of the action of U_p on V/W is $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, which is nonzero and nilpotent, hence not semisimple. Since W is invariant under U_p this shows that U_p is not semisimple on V, i.e., U_p is not diagonalizable.

7.3 The Cusp Forms are Free of Rank One over $\mathbf{T}_{\mathbf{C}}$

7.3.1 Level 1

Suppose $N = 1$, so $\Gamma_1(N) = \mathrm{SL}_2(\mathbf{Z})$. Using the Petersson inner product, we see that all the T_n are diagonalizable, so $S_k = S_k(\Gamma_1(1))$ has a basis

$$f_1, \ldots, f_d$$

of normalized eigenforms where $d = \dim S_k$. This basis is canonical up to ordering. Let $\mathbf{T}_{\mathbf{C}} = \mathbf{T} \otimes \mathbf{C}$ be the ring generated over $\mathbf{C}$ by the Hecke operator T_p. Then, having fixed the basis above, there is a canonical map

$$\mathbf{T}_{\mathbf{C}} \hookrightarrow \mathbf{C}^d : \quad T \mapsto (\lambda_1, \ldots, \lambda_d),$$

where $f_i | T = \lambda_i f_i$. This map is injective and $\dim \mathbf{T}_{\mathbf{C}} = d$, so the map is an isomorphism of $\mathbf{C}$-vector spaces.

The form

$$v = f_1 + \cdots + f_n$$

generates S_k as a $\mathbf{T}$-module. Note that v is canonical since it does not depend on the ordering of the f_i. Since v corresponds to the vector $(1, \ldots, 1)$ and $\mathbf{T} \cong \mathbf{C}^d$ acts on $S_k \cong \mathbf{C}^d$ componentwise, this is just the statement that $\mathbf{C}^d$ is generated by $(1, \ldots, 1)$ as a $\mathbf{C}^d$-module.

There is a perfect pairing $S_k \times \mathbf{T}_{\mathbf{C}} \to \mathbf{C}$ given by

$$\left\langle \sum f, \, T_n \right\rangle = a_1(f | T_n) = a_n(f),$$

where $a_n(f)$ denotes the nth Fourier coefficient of f. Thus we have simultaneously:

1. S_k is free of rank 1 over $\mathbf{T}_{\mathbf{C}}$, and
2. $S_k \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C})$ as $\mathbf{T}$-modules.

Combining these two facts yields an isomorphism

$$\mathbf{T}_{\mathbf{C}} \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C}). \tag{7.3.1}$$

This isomorphism sends an element $T \in \mathbf{T}$ to the homomorphism

$$X \mapsto \langle v | T, X \rangle = a_1(v | T | X).$$

Since the identification $S_k = \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}_{\mathbf{C}}, \mathbf{C})$ is canonical and since the vector v is canonical, we see that the isomorphism (7.3.1) is canonical.

Recall that M_k has as basis the set of products $E_4^a E_6^b$, where $4a + 6b = k$, and S_k is the subspace of forms where the constant coefficient of their q-expansion is 0. Thus there is a basis of S_k consisting of forms whose q-expansions have coefficients in $\mathbf{Q}$. Let $S_k(\mathbf{Z}) = S_k \cap \mathbf{Z}[[q]]$, be the submodule of S_k generated by cusp forms with Fourier coefficients in $\mathbf{Z}$, and note that $S_k(\mathbf{Z}) \otimes \mathbf{Q} \cong S_k(\mathbf{Q})$. Also, the explicit formula $(\sum a_n q^n)|T_p = \sum a_{np} q^n + p^{k-1} \sum a_n q^{np}$ implies that the Hecke algebra $\mathbf{T}$ preserves $S_k(\mathbf{Z})$.

Proposition 7.3.1. *The Fourier coefficients of each f_i are totally real algebraic integers.*

Proof. The coefficient $a_n(f_i)$ is the eigenvalue of T_n acting on f_i. As observed above, the Hecke operator T_n preserves $S_k(\mathbf{Z})$, so the matrix $[T_n]$ of T_n with respect to a basis for $S_k(\mathbf{Z})$ has integer entries. The eigenvalues of T_n are algebraic integers, since the characteristic polynomial of $[T_n]$ is monic and has integer coefficients.

The eigenvalues are real since the Hecke operators are self-adjoint with respect to the Petersson inner product. □

Remark 7.3.2. A *CM field* is a quadratic imaginary extension of a totally real field. For example, when $n > 2$, the field $\mathbf{Q}(\zeta_n)$ is a CM field, with totally real subfield $\mathbf{Q}(\zeta_n)^+ = \mathbf{Q}(\zeta_n + 1/\zeta_n)$. More generally, one shows that the eigenvalues of any newform $f \in S_k(\Gamma_1(N))$ generate a totally real or CM field.

Proposition 7.3.3. *We have $v \in S_k(\mathbf{Z})$.*

Proof. This is because $v = \sum \mathrm{Tr}(T_n) q^n$, and, as we observed above, there is a basis so that the matrices T_n have integer coefficients. □

Example 7.3.4. When $k = 36$, we have

$$\begin{aligned} v = 3q &+ 139656q^2 - 104875308q^3 + 34841262144q^4 + 892652054010q^5 \\ &- 4786530564384q^6 + 878422149346056q^7 + \cdots . \end{aligned}$$

The normalized newforms f_1, f_2, f_3 are

$$\begin{aligned} f_i = q &+ aq^2 + (-1/72a^2 + 2697a + 478011548)q^3 + (a^2 - 34359738368)q^4 \\ &(a^2 - 34359738368)q^4 + (-69/2a^2 + 14141780a + 1225308030462)q^5 + \cdots , \end{aligned}$$

for a each of the three roots of $X^3 - 139656X^2 - 59208339456X - 1467625047588864$.

7.3.2 General Level

Now we consider the case for general level N. Recall that there are maps

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

for all M dividing N and all divisor d of N/M.

The *old subspace* of $S_k(\Gamma_1(N))$ is the space generated by all images of these maps with $M|N$ but $M \neq N$. The *new subspace* is the orthogonal complement of the old subspace with respect to the Petersson inner product.

There is an algebraic definition of the new subspace. One defines trace maps

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

for all $M < N$, $M \mid N$ which are adjoint to the above maps (with respect to the Petersson inner product). Then f is in the new part of $S_k(\Gamma_1(N))$ if and only if f is in the kernels of all of the trace maps.

It follows from Atkin-Lehner-Li theory that the T_n acts semisimply on the new subspace $S_k(\Gamma_1(M))_{\mathrm{new}}$ for all $M \geq 1$, since the common eigenspaces for all T_n each have dimension 1. Thus $S_k(\Gamma_1(M))_{\mathrm{new}}$ has a basis of normalized eigenforms. We have a natural map

$$\bigoplus_{M|N} S_k(\Gamma_1(M))_{\mathrm{new}} \hookrightarrow S_k(\Gamma_1(N)).$$

The image in $S_k(\Gamma_1(N))$ of an eigenform f for some $S_k(\Gamma_1(M))_{\mathrm{new}}$ is called a *newform* of level $M_f = M$. Note that a newform of level less than N is not necessarily an eigenform for all of the Hecke operators acting on $S_k(\Gamma_1(N))$; in particular, it can fail to be an eigenform for the T_p, for $p \mid N$.

Let

$$v = \sum_f f(q^{\frac{N}{M_f}}) \in S_k(\Gamma_1(N)),$$

where the sum is taken over all newforms f of weight k and some level $M \mid N$. This generalizes the v constructed above when $N = 1$ and has many of the same good properties. For example, $S_k(\Gamma_1(N))$ is free of rank 1 over $\mathbf{T}$ with basis element v. Moreover, the coefficients of v lie in $\mathbf{Z}$, but to show this we need to know that $S_k(\Gamma_1(N))$ has a basis whose q-expansions lie in $\mathbf{Q}[[q]]$. This is true, but we will not prove it here. One way to proceed is to use the Tate curve to construct a q-expansion map $\mathrm{H}^0(X_1(N), \Omega_{X_1(N)/\mathbf{Q}}) \to \mathbf{Q}[[q]]$, which is compatible with the usual Fourier expansion map.

Example 7.3.5. The space $S_2(\Gamma_1(22))$ has dimension 6. There is a single newform of level 11,

$$f = q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 + \cdots .$$

There are four newforms of level 22, the four $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of

$$\begin{aligned} g = q - \zeta q^2 &+ (-\zeta^3 + \zeta - 1)q^3 + \zeta^2 q^4 + (2\zeta^3 - 2)q^5 \\ &+ (\zeta^3 - 2\zeta^2 + 2\ \zeta - 1)q^6 - 2\zeta^2 q^7 + ... \end{aligned}$$

where ζ is a primitive 10th root of unity.

Warning 7.3.6. Let $S = S_2(\Gamma_0(88))$, and let $v = \sum \mathrm{Tr}(T_n)q^n$. Then S has dimension 9, but the Hecke span of v only has dimension 7. Thus the more "canonical looking" element $\sum \mathrm{Tr}(T_n)q^n$ is not a generator for S.

7.4 Decomposing the Anemic Hecke Algebra

We first observe that it make no difference whether or not we include the Diamond bracket operators in the Hecke algebra. Then we note that the $\mathbf{Q}$-algebra generated by the Hecke operators of index coprime to the level is isomorphic to a product of fields corresponding to the Galois conjugacy classes of newforms.

Proposition 7.4.1. *The operators $\langle d \rangle$ on $S_k(\Gamma_1(N))$ lie in $\mathbf{Z}[\ldots, T_n, \ldots]$.*

Proof. It is enough to show $\langle p\rangle \in \mathbf{Z}[\ldots, T_n, \ldots]$ for primes p, since each $\langle d\rangle$ can be written in terms of the $\langle p\rangle$. Since $p \nmid N$, we have that

$$T_{p^2} = T_p^2 - \langle p\rangle p^{k-1},$$

so $\langle p\rangle p^{k-1} = T_p^2 - T_{p^2}$. By Dirichlet's theorem on primes in arithmetic progression [Lan94, VIII.4], there is another prime q congruent to p mod N. Since p^{k-1} and q^{k-1} are relatively prime, there exist integers a and b such that $ap^{k-1} + bq^{k-1} = 1$. Then

$$\langle p\rangle = \langle p\rangle(ap^{k-1} + bq^{k-1}) = a({T_p}^2 - T_{p^2}) + b({T_q}^2 - T_{q^2}) \in \mathbf{Z}[\ldots, T_n, \ldots].$$

□

Let S be a space of cusp forms, such as $S_k(\Gamma_1(N))$ or $S_k(\Gamma_1(N), \varepsilon)$. Let

$$f_1, \ldots, f_d \in S$$

be representatives for the Galois conjugacy classes of newforms in S of level N_{f_i} dividing N. For each i, let $K_i = \mathbf{Q}(\ldots, a_n(f_i), \ldots)$ be the field generated by the Fourier coefficients of f_i.

Definition 7.4.2 (Anemic Hecke Algebra). The *anemic Hecke algebra* is the subalgebra

$$\mathbf{T}_0 = \mathbf{Z}[\ldots, T_n, \ldots : (n, N) = 1] \subset \mathbf{T}$$

of $\mathbf{T}$ obtained by adjoining to $\mathbf{Z}$ only those Hecke operators T_n with n relatively prime to N.

Proposition 7.4.3. *We have* $\mathbf{T}_0 \otimes \mathbf{Q} \cong \prod_{i=1}^{d} K_i$.

The map sends T_n to $(a_n(f_1), \ldots, a_n(f_d))$. The proposition can be proved using the discussion above and Atkin-Lehner-Li theory, but we will not give a proof here.

Example 7.4.4.
When $S = S_2(\Gamma_1(22))$, then $\mathbf{T}_0 \otimes \mathbf{Q} \cong \mathbf{Q} \times \mathbf{Q}(\zeta_{10})$ (see Example 7.3.5). When $S = S_2(\Gamma_0(37))$, then $\mathbf{T}_0 \otimes \mathbf{Q} \cong \mathbf{Q} \times \mathbf{Q}$.

8
Hecke operators as correspondences

Our goal is to view the Hecke operators T_n and $\langle d\rangle$ as objects defined over $\mathbf{Q}$ that act in a compatible way on modular forms, modular Jacobians, and homology. In order to do this, we will define the Hecke operators as correspondences.

8.1 The Definition

Definition 8.1.1 (Correspondence). Let C_1 and C_2 be curves. A *correspondence* $C_1 \rightsquigarrow C_2$ is a curve C together with nonconstant morphisms $\alpha : C \to C_1$ and $\beta : C \to C_2$. We represent a correspondence by a diagram

$$\begin{tikzcd} & C \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ C_1 & & C_2 \end{tikzcd}$$

Given a correspondence $C_1 \rightsquigarrow C_2$ the *dual correspondence* $C_2 \rightsquigarrow C_1$ is obtained by looking at the diagram in a mirror

$$\begin{tikzcd} & C \arrow[dl, "\beta"'] \arrow[dr, "\alpha"] & \\ C_2 & & C_1 \end{tikzcd}$$

In defining Hecke operators, we will focus on the simple case when the modular curve is $X_0(N)$ and Hecke operator is T_p, where $p \nmid N$. We will view T_p as a correspondence $X_0(N) \rightsquigarrow X_0(N)$, so there is a curve $C = X_0(pN)$ and maps α and β fitting into a diagram

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N). \end{tikzcd}$$

The maps α and β are degeneracy maps which forget data. To define them, we view $X_0(N)$ as classifying isomorphism classes of pairs (E, C), where E is an elliptic curve and C is a cyclic subgroup of order N (we will not worry about what happens at the cusps, since any rational map of nonsingular curves extends uniquely to a morphism). Similarly, $X_0(pN)$ classifies isomorphism classes of pairs (E, G) where $G = C \oplus D$, C is cyclic of order N and D is cyclic of order p. Note that since $(p, N) = 1$, the group G is cyclic of order pN and the subgroups C and D are uniquely determined by G. The map α forgets the subgroup D of order p, and β quotients out by D:

$$\alpha : (E, G) \mapsto (E, C) \tag{8.1.1}$$

$$\beta : (E, G) \mapsto (E/D, (C + D)/D) \tag{8.1.2}$$

We translate this into the language of complex analysis by thinking of $X_0(N)$ and $X_0(pN)$ as quotients of the upper half plane. The first map α corresponds to the map

$$\Gamma_0(pN)\backslash\mathfrak{h} \to \Gamma_0(N)\backslash\mathfrak{h}$$

induced by the inclusion $\Gamma_0(pN) \hookrightarrow \Gamma_0(N)$. The second map β is constructed by composing the isomorphism

$$\Gamma_0(pN)\backslash\mathfrak{h} \xrightarrow{\sim} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1} \backslash\mathfrak{h} \tag{8.1.3}$$

with the map to $\Gamma_0(N)\backslash\mathfrak{h}$ induced by the inclusion

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1} \subset \Gamma_0(N).$$

The isomorphism (8.1.3) is induced by $z \mapsto \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) z = pz$; explicitly, it is

$$\Gamma_0(pN)z \mapsto \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) \Gamma_0(pN) \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) z.$$

(Note that this is well-defined.)

The maps α and β induce pullback maps on differentials

$$\alpha^*, \beta^* : \mathrm{H}^0(X_0(N), \Omega^1) \to \mathrm{H}^0(X_0(pN), \Omega^1).$$

We can identify $S_2(\Gamma_0(N))$ with $H^0(X_0(N), \Omega^1)$ by sending the cusp form $f(z)$ to the holomorphic differential $f(z)dz$. Doing so, we obtain two maps

$$\alpha^*, \beta^* : S_2(\Gamma_0(N)) \to S_2(\Gamma_0(pN)).$$

Since α is induced by the identity map on the upper half plane, we have $\alpha^*(f) = f$, where we view $f = \sum a_n q^n$ as a cusp form with respect to the smaller group $\Gamma_0(pN)$. Also, since β^* is induced by $z \mapsto pz$, we have

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

The factor of p is because

$$\beta^*(f(z)dz) = f(pz)d(pz) = pf(pz)dz.$$

Let X, Y, and C be curves, and α and β be nonconstant holomorphic maps, so we have a correspondence

$$\begin{tikzcd} & C \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X & & Y. \end{tikzcd}$$

By first pulling back, then pushing forward, we obtain induced maps on differentials

$$H^0(X,\Omega^1) \xrightarrow{\alpha^*} H^0(C,\Omega^1) \xrightarrow{\beta_*} H^0(Y,\Omega^1).$$

The composition $\beta_* \circ \alpha^*$ is a map $H^0(X,\Omega^1) \to H^0(Y,\Omega^1)$. If we consider the dual correspondence, which is obtained by switching the roles of X and Y, we obtain a map $H^0(Y,\Omega^1) \to H^0(X,\Omega^1)$.

Now let α and β be as in (8.1.1). Then we can recover the action of T_p on modular forms by considering the induced map

$$\beta_* \circ \alpha^* : H^0(X_0(N),\Omega^1) \to H^0(X_0(N),\Omega^1)$$

and using that $S_2(\Gamma_0(N)) \cong H^0(X_0(N),\Omega^1)$.

8.2 Maps induced by correspondences

In this section we will see how correspondences induce maps on divisor groups, which in turn induce maps on Jacobians.

Suppose $\varphi : X \to Y$ is a morphism of curves. Let $\Gamma \subset X \times Y$ be the graph of φ. This gives a correspondence

$$\begin{tikzcd} & \Gamma \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X & & Y \end{tikzcd}$$

We can reconstruct φ from the correspondence by using that $\varphi(x) = \beta(\alpha^{-1}(x))$. [draw picture here]

More generally, suppose Γ is a curve and that $\alpha : \Gamma \to X$ has degree $d \geq 1$. View $\alpha^{-1}(x)$ as a divisor on Γ (it is the formal sum of the points lying over x, counted with appropriate multiplicities). Then $\beta(\alpha^{-1}(x))$ is a divisor on Y. We thus obtain a map

$$\mathrm{Div}^n(X) \xrightarrow{\beta\circ\alpha^{-1}} \mathrm{Div}^{dn}(Y),$$

where $\mathrm{Div}^n(X)$ is the group of divisors of degree n on X. In particular, setting $d = 0$, we obtain a map $\mathrm{Div}^0(X) \to \mathrm{Div}^0(Y)$.

We now apply the above construction to T_p. Recall that T_p is the correspondence

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N), \end{tikzcd}$$

where α and β are as in Section 8.1 and the induced map is

$$(E,C) \overset{\alpha^*}{\mapsto} \sum_{D\in E[p]} (E, C\oplus D) \overset{\beta_*}{\mapsto} \sum_{D\in E[p]} (E/D, (C+D)/D).$$

Thus we have a map $\mathrm{Div}(X_0(N)) \to \mathrm{Div}(X_0(N))$. This strongly resembles the first definition we gave of T_p on level 1 forms, where T_p was a correspondence of lattices.

8.3 Induced maps on Jacobians of curves

Let X be a curve of genus g over a field k. Recall that there is an important association

$$\left\{ \text{ curves } X/k \right\} \longrightarrow \left\{ \text{ Jacobians } \mathrm{Jac}(X) = J(X) \text{ of curves } \right\}$$

between curves and their Jacobians.

Definition 8.3.1 (Jacobian). Let X be a curve of genus g over a field k. Then the *Jacobian* of X is an abelian variety of dimension g over k whose underlying group is functorially isomorphic to the group of divisors of degree 0 on X modulo linear equivalence. (For a more precise definition, see Section **??** (Jacobians section).)

There are many constructions of the Jacobian of a curve. We first consider the Albanese construction. Recall that over $\mathbf{C}$, any abelian variety is isomorphic to $\mathbf{C}^g/L$, where L is a lattice (and hence a free $\mathbf{Z}$-module of rank $2g$). There is an embedding

$$\begin{aligned} \iota : \mathrm{H}_1(X,\mathbf{Z}) &\hookrightarrow \mathrm{H}^0(X,\Omega^1)^* \\ \gamma &\mapsto \int_\gamma \bullet \end{aligned}$$

Then we realize $\mathrm{Jac}(X)$ as a quotient

$$\mathrm{Jac}(X) = \mathrm{H}^0(X,\Omega^1)^*/\iota(\mathrm{H}_1(X,\mathbf{Z})).$$

In this construction, $\mathrm{Jac}(X)$ is most naturally viewed as covariantly associated to X, in the sense that if $X \to Y$ is a morphism of curves, then the resulting map $\mathrm{H}^0(X,\Omega^1)^* \to \mathrm{H}^0(Y,\Omega^1)^*$ on tangent spaces induces a map $\mathrm{Jac}(X) \to \mathrm{Jac}(Y)$.

There are other constructions in which $\mathrm{Jac}(X)$ is contravariantly associated to X. For example, if we view $\mathrm{Jac}(X)$ as $\mathrm{Pic}^0(X)$, and $X \to Y$ is a morphism, then pullback of divisor classes induces a map $\mathrm{Jac}(Y) = \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X) = \mathrm{Jac}(X)$.

If $F : X \rightsquigarrow Y$ is a correspondence, then F induces an a map $\mathrm{Jac}(X) \to \mathrm{Jac}(Y)$ and also a map $\mathrm{Jac}(Y) \to \mathrm{Jac}(X)$. If $X = Y$, so that X and Y are the same, it can often be confusing to decide which duality to use. Fortunately, for T_p, with p prime to N, it does not matter which choice we make. But it matters a lot if $p \mid N$ since then we have non-commuting confusable operators and this has resulted in mistakes in the literature.

8.4 More on Hecke operators

Our goal is to move things down to $\mathbf{Q}$ from $\mathbf{C}$ or $\overline{\mathbf{Q}}$. In doing this we want to understand T_n (or T_p), that is, how they act on the associated Jacobians and how they can be viewed as correspondences. In characteristic p the formulas of Eichler-Shimura will play an important role.

We consider T_p as a correspondence on $X_1(N)$ or $X_0(N)$. To avoid confusion we will mainly consider T_p on $X_0(N)$ with $p \nmid N$. Thus assume, unless otherwise stated, that $p \nmid N$. Remember that T_p was defined to be the correspondence

$$\begin{tikzcd}
& X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N) & & X_0(N)
\end{tikzcd}$$

Think of $X_0(pN)$ as consisting of pairs $(\underline{E}, D)$ where D is a cyclic subgroup of E of order p and $\underline{E}$ is the *enhanced* elliptic curve consisting of an elliptic curve E along with a cyclic subgroup of order N. The degeneracy map α forgets the subgroup D and the degeneracy map β divides by it. By contravariant functoriality we have a commutative diagram

$$\begin{tikzcd}
H^0(X_0(N),\Omega^1) \arrow[r, "T_p^*=\alpha_*\circ\beta^*"] \arrow[d, equal] & H^0(X_0(N),\Omega^1) \arrow[d, equal] \\
S_2(\Gamma_0(N)) \arrow[r, "T_p"] & S_2(\Gamma_0(N))
\end{tikzcd}$$

Our convention to define T_p^* as $\alpha_* \circ \beta^*$ instead of $\beta_* \circ \alpha^*$ was completely psychological because there is a canonical duality relating the two. We chose the way we did because of the analogy with the case of a morphism $\varphi : Y \to X$ with graph Γ which induces a correspondence

$$\begin{tikzcd}
& \Gamma \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\
Y & & X
\end{tikzcd}$$

Since the morphism φ induces a map on global sections in the other direction

$$H^0(X,\Omega^1) = \Gamma(X) \xrightarrow{\varphi^*} \Gamma(Y) = H^0(Y,\Omega^1)$$

it is psychologically natural for more general correspondence such as T_p to map from the right to the left.

The morphisms α and β in the definition of T_p are defined over $\mathbf{Q}$. This can be seen using the general theory of representable functors. Thus since T_p is defined over $\mathbf{Q}$ most of the algebraic geometric objects we will construct related to T_p will be defined over $\mathbf{Q}$.

8.5 Hecke operators acting on Jacobians

The Jacobian $J(X_0(N)) = J_0(N)$ is an abelian variety defined over $\mathbf{Q}$. There are both covariant and contravariant ways to construct $J_0(N)$. Thus a map α :

$X_0(pN) \to X_0(N)$ induces maps

$$\begin{tikzcd}
J_0(pN) \arrow[r, equal] & J_0(pN) \arrow[d, "\alpha_*"] \\
J_0(N) \arrow[u, "\alpha^*"] \arrow[r, "p+1"'] & J_0(N)
\end{tikzcd}$$

Note that $\alpha_* \circ \alpha^* : J_0(N) \to J_0(N)$ is just multiplication by $\deg(\alpha) = p+1$, since there are $p+1$ subgroups of order p in $\underline{E}$. (At least when $p \nmid N$, when $p|N$ there are only p subgroups.)

There are two possible ways to define T_p as an endomorphism of $J_0(N)$. We could either define T_p as $\beta_* \circ \alpha^*$ or equivalently as $\alpha_* \circ \beta^*$ (assuming still that $p \nmid N$).

8.5.1 The Albanese Map

There is a way to map the curve $X_0(N)$ into its Jacobian since the underlying group structure of $J_0(N)$ is

$$J_0(N) = \frac{\left\{ \text{divisors of degree 0 on } X_0(N) \right\}}{\left\{ \text{principal divisors} \right\}}$$

Once we have chosen a rational point, say ∞, on $X_0(N)$ we obtain the Albanese map

$$\theta : X_0(N) \to J_0(N) : x \mapsto x - \infty$$

which sends a point x to the divisor $x - \infty$. The map θ gives us a way to pullback differentials on $J_0(N)$. Let $\operatorname{Cot} J_0(N)$ denote the cotangent space of $J_0(N)$ (or the space of regular differentials). The diagram

$$\begin{tikzcd}
\operatorname{Cot} J_0(N) \arrow[d, "\theta^*"', "\wr"] & \operatorname{Cot} J_0(N) \arrow[l, "\xi_p^*"'] \arrow[d, "\theta^*", "\wr"'] \\
H^0(X_0(N), \Omega^1) & H^0(X_0(N), \Omega^1) \arrow[l, "T_p^*"']
\end{tikzcd}$$

may be taken to give a definition of ξ_p since there is a unique endomorphism $\xi_p : J_0(N) \to J_0(N)$ inducing a map ξ_p^* which makes the diagram commute.

Now suppose Γ is a correspondence $X \rightsquigarrow Y$ so we have a diagram

$$\begin{tikzcd}
& \Gamma \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X & & Y
\end{tikzcd}$$

For example, think of Γ as the graph of a morphism $\varphi : X \to Y$. Then Γ should induce a natural map

$$H^0(Y, \Omega^1) \longrightarrow H^0(X, \Omega^1).$$

Taking Jacobians we see that the composition

$$J(X) \xrightarrow{\alpha^*} J(\Gamma) \xrightarrow{\beta_*} J(Y)$$

gives a map $\beta_* \circ \alpha^* : J(X) \to J(Y)$. On cotangent spaces this induces a map

$$\alpha^* \circ \beta_* : H^0(Y, \Omega^1) \to H^0(X, \Omega^1).$$

Now, after choice of a rational point, the map $X \to J(X)$ induces a map $\operatorname{Cot} J(X) \to H^0(X, \Omega^1)$. This is in fact independent of the choice of rational point since differentials on $J(X)$ are invariant under translation.

The map $J(X) \to J(Y)$ is preferred in the literature. It is said to be induced by the Albanese functoriality of the Jacobian. We could have just as easily defined a map from $J(Y) \to J(X)$. To see this let

$$\psi = \beta_* \circ \alpha^* : J(X) \to J(Y).$$

Dualizing induces a map $\psi^\vee = \alpha_* \circ \beta^*$:

$$\begin{tikzcd}
J(X)^\vee \arrow[d, "\cong"] & J(Y)^\vee \arrow[l, "\psi^\vee"'] \\
J(X) & J(Y) \arrow[u, "\cong"]
\end{tikzcd}$$

Here we have used autoduality of Jacobians. This canonical duality is discussed in [MFK94] and [Mum70] and in Milne's article in [Sch65].

8.5.2 The Hecke Algebra

We now have $\xi_p = T_p \in \operatorname{End} J_0(N)$ for every prime p. If $p|N$, then we must decide between $\alpha_* \circ \beta^*$ and $\beta_* \circ \alpha^*$. The usual choice is the one which induces the usual T_p on cusp forms. If you don't like your choice you can get out of it with Atkin-Lehner operators.

Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_p, \ldots] \subset \operatorname{End} J_0(N)$$

then $\mathbf{T}$ is the same as $\mathbf{T}_{\mathbf{Z}} \subset \operatorname{End}(S_2(\Gamma_0(N)))$. To see this first note that there is a map $\mathbf{T} \to \mathbf{T}_{\mathbf{Z}}$ which is not a prior injective, but which is injective because elements of $\operatorname{End} J_0(N)$ are completely determined by their action on $\operatorname{Cot} J_0(N)$.

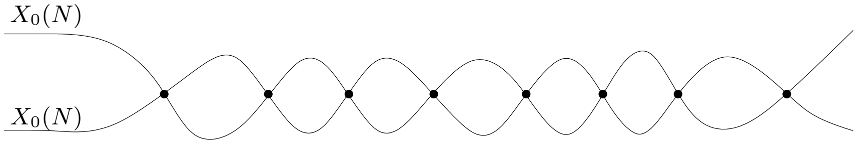


FIGURE 8.6.1. The reduction mod p of the Deligne-Rapoport model of $X_0(Np)$

8.6 The Eichler-Shimura Relation

Suppose $p \nmid N$ is a prime. The Hecke operator T_p and the Frobenius automorphism Frob_p induce, by functoriality, elements of $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$, which we also denote T_p and Frob_p. The Eichler-Shimura relation asserts that the relation

$$T_p = \mathrm{Frob}_p + p\,\mathrm{Frob}_p^{-1} \tag{8.6.1}$$

holds in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$. In this section we sketch the main idea behind why (8.6.1) holds. For more details and a proof of the analogous statement for $J_1(N)$, see [Con01].

Since $J_0(N)$ is an abelian variety defined over $\mathbf{Q}$, it is natural to ask for the primes p such that $J_0(N)$ have good reduction. In the 1950s Igusa showed that $J_0(N)$ has good reduction for all $p \nmid N$. He viewed $J_0(N)$ as a scheme over $\mathrm{Spec}(\mathbf{Q})$, then "spread things out" to make an abelian scheme over $\mathrm{Spec}(\mathbf{Z}[1/N])$. He did this by taking the Jacobian of the normalization of $X_0(N)$ (which is defined over $\mathbf{Z}[1/N]$) in $\mathbf{P}^n_{\mathbf{Z}[1/N]}$.

The Eichler-Shimura relation is a formula for T_p in characteristic p, or more precisely, for the corresponding endomorphisms in $\mathrm{End}(J_0(N)_{\mathbf{F}_p}))$ for all p for which $J_0(N)$ has good reduction at p. If $p \nmid N$, then $X_0(N)_{\mathbf{F}_p}$ has many of the same properties as $X_0(N)_{\mathbf{Q}}$. In particular, the noncuspidal points on $X_0(N)_{\mathbf{F}_p}$ classify isomorphism classes of enhanced elliptic curves $\underline{E} = (E, C)$, where E is an elliptic curve over $\mathbf{F}_p$ and C is a cyclic subgroup of E of order N. (Note that two pairs are considered *isomorphic* if they are isomorphic over $\overline{\mathbf{F}}_p$.)

Next we ask what happens to the map $T_p : J_0(N) \to J_0(N)$ under reduction modulo p. To this end, consider the correspondence

$$\begin{tikzcd} & X_0(Np) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N) \end{tikzcd}$$

that defines T_p. The curve $X_0(N)$ has good reduction at p, but $X_0(Np)$ typically does not. Deligne and Rapaport [DR73] showed that $X_0(Np)$ has relatively benign reduction at p. Over $\mathbf{F}_p$, the reduction $X_0(Np)_{\mathbf{F}_p}$ can be viewed as two copies of $X_0(N)$ glued at the supersingular points, as illustrated in Figure 8.6.1.

The set of supersingular points

$$\Sigma \subset X_0(N)(\overline{\mathbf{F}}_p)$$

is the set of points in $X_0(N)$ represented by pairs $\underline{E} = (E, C)$, where E is a supersingular elliptic curve (so $E(\overline{\mathbf{F}}_p)[p] = 0$). There are exactly $g+1$ supersingular points, where g is the genus of $X_0(N)$.

Consider the correspondence $T_p : X_0(N) \rightsquigarrow X_0(N)$ which takes an enhanced elliptic curve $\underline{E}$ to the sum $\sum \underline{E}/D$ of all quotients of $\underline{E}$ by subgroups D of order p. This is the correspondence

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N), \end{tikzcd} \tag{8.6.2}$$

where the map α forgets the subgroup of order p, and β quotients out by it. From this one gets $T_p : J_0(N) \to J_0(N)$ by functoriality.

Remark 8.6.1. There are many ways to think of $J_0(N)$. The cotangent space $\operatorname{Cot} J_0(N)$ of $J_0(N)$ is the space of holomorphic (or translation invariant) differentials on $J_0(N)$, which is isomorphic to $S_2(\Gamma_0(N))$. This gives a connection between our geometric definition of T_p and the definition, presented earlier, of T_p as an operator on a space of cusp forms.

The Eichler-Shimura relation takes place in $\operatorname{End}(J_0(N)_{\mathbf{F}_p})$. Since $X_0(N)$ reduces "nicely" in characteristic p, we can apply the Jacobian construction to $X_0(N)_{\mathbf{F}_p}$.

Lemma 8.6.2. *The natural reduction map*

$$\operatorname{End}(J_0(N)) \hookrightarrow \operatorname{End}(J_0(N)_{\mathbf{F}_p})$$

is injective.

Proof. Let $\ell \nmid Np$ be a prime. By [ST68, Thm. 1, Lem. 2], the reduction to characteristic p map induces an isomorphism

$$J_0(N)(\overline{\mathbf{Q}})[\ell^\infty] \cong J_0(N)(\overline{\mathbf{F}}_p)[\ell^\infty].$$

If $\varphi \in \operatorname{End}(J_0(N))$ reduces to the 0 map in $\operatorname{End}(J_0(N)_{\mathbf{F}_p})$, then $J_0(N)(\overline{\mathbf{Q}})[\ell^\infty]$ must be contained in $\ker(\varphi)$. Thus φ induces the 0 map on $\operatorname{Tate}_\ell(J_0(N))$, so $\varphi = 0$. □

Let $F : X_0(N)_{\mathbf{F}_p} \to X_0(N)_{\mathbf{F}_p}$ be the Frobenius map in characteristic p. Thus, if $K = K(X_0(N))$ is the function field of the nonsingular curve $X_0(N)$, then $F : K \to K$ is induced by the pth power map $a \mapsto a^p$.

Remark 8.6.3. The Frobenius map corresponds to the pth powering map on points. For example, if $X = \operatorname{Spec}(\mathbf{F}_p[t])$, and $z = (\operatorname{Spec}(\overline{\mathbf{F}}_p) \to X)$ is a point defined by a homomorphism $\alpha : \mathbf{F}_p[t] \mapsto \overline{\mathbf{F}}_p$, then $F(z)$ is the composite

$$\mathbf{F}_p[t] \xrightarrow{x \mapsto x^p} \mathbf{F}_p[t] \xrightarrow{\alpha} \overline{\mathbf{F}}_p.$$

If $\alpha(t) = \xi$, then $F(z)(t) = \alpha(t^p) = \xi^p$.

By both functorialities, F induces maps on the Jacobian of $X_0(N)_{\mathbf{F}_p}$:

$$\operatorname{Frob}_p = F_* \quad \text{and} \quad \operatorname{Ver}_p = \operatorname{Frob}_p^\vee = F^*,$$

which we illustrate as follows:

$$\begin{tikzcd} J_0(N)_{\mathbf{F}_p} \arrow[r, bend left, "\operatorname{Ver}_p"] & J_0(N)_{\mathbf{F}_p} \arrow[l, bend left, "\operatorname{Frob}_p"] \end{tikzcd}$$

Note that $\mathrm{Ver}_p \circ \mathrm{Frob}_p = \mathrm{Frob}_p \circ \mathrm{Ver}_p = [p]$ since p is the degree of F (for example, if $K = \mathbf{F}_p(t)$, then $F(K) = \mathbf{F}_p(t^p)$ is a subfield of degree p, so the map induced by F has degree p).

Theorem 8.6.4 (Eichler-Shimura Relation). *Let N be a positive integer and $p \nmid N$ be a prime. Then the following relation holds:*

$$T_p = \mathrm{Frob}_p + \mathrm{Ver}_p \in \mathrm{End}(J_0(N)_{\mathbf{F}_p}).$$

Sketch of Proof. We view $X_0(pN)_{\mathbf{F}_p}$ as two copies of $X_0(N)_{\mathbf{F}_p}$ glued along corresponding supersingular points Σ, as in Figure 8.6.1. This diagram and the correspondence (8.6.2) that defines T_p translate into the following diagram of schemes over $\mathbf{F}_p$:

$$\begin{tikzcd}
& \Sigma \arrow[dl, hook'] \arrow[dr, hook] & \\
X_0(N)_{\mathbf{F}_p} \arrow[dr, hook, "r"] \arrow[dd, "\cong"'] & & X_0(N)_{\mathbf{F}_p} \arrow[dl, hook', "s"'] \arrow[dd, "\cong"] \\
& X_0(pN)_{\mathbf{F}_p} \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p}
\end{tikzcd}$$

The maps r and s are defined as follows. Recall that a point of $X_0(N)_{\mathbf{F}_p}$ is an enhanced elliptic curve $\underline{E} = (E, C)$ consisting of an elliptic curve E (not necessarily defined over $\mathbf{F}_p$) along with a cyclic subgroup C of order N. We view a point on $X_0(Np)$ as a triple $(E, C, E \to E')$, where (E, C) is as above and $E \to E'$ is an isogeny of degree p. We use an isogeny instead of a cyclic subgroup of order p because $E(\overline{\mathbf{F}}_p)[p]$ has order either 1 or p, so the data of a cyclic subgroup of order p holds very little information.

The map r sends $\underline{E}$ to $(\underline{E}, \varphi)$, where φ is the isogeny of degree p,

$$\varphi : E \xrightarrow{\mathrm{Frob}} E^{(p)}.$$

Here $E^{(p)}$ is the curve obtained from E by hitting all defining equations by Frobenious, that is, by pth powering the coefficients of the defining equations for E. We introduce $E^{(p)}$ since if E is not defined over $\mathbf{F}_p$, then Frobenious does not define an endomorphism of E. Thus r is the map

$$r : \quad \underline{E} \mapsto (\underline{E}, E \xrightarrow{\mathrm{Frob}_p} E^{(p)}),$$

and similarly we define s to be the map

$$s : \quad \underline{E} \mapsto (E^{(p)}, C, E \xleftarrow{\mathrm{Ver}_p} E^{(p)})$$

where Ver_p is the dual of Frob_p (so $\mathrm{Ver}_p \circ \mathrm{Frob}_p = \mathrm{Frob}_p \circ \mathrm{Ver}_p = [p]$).

We view α as the map sending $(\underline{E}, E \to E')$ to $\underline{E}$, and similarly we view β as the map sending $(\underline{E}, E \to E')$ to the pair (E', C'), where C' is the image of C in

E' via $E \to E'$. Thus

$$\begin{aligned} \alpha: &\quad (E \to E') \mapsto E \\ \beta: &\quad (E' \to E) \mapsto E' \end{aligned}$$

It now follows immediately that $\alpha \circ r = \mathrm{id}$ and $\beta \circ s = \mathrm{id}$. Note also that $\alpha \circ s = \beta \circ r = F$ is the map $E \mapsto E^{(p)}$.

Away from the finitely many supersingular points, we may view $X_0(pN)_{\mathbf{F}_p}$ as the disjoint union of two copies of $X_0(N)_{\mathbf{F}_p}$. Thus away from the supersingular points, we have the following equality of correspondences:

$$\begin{tikzcd}
& X_0(pN)_{\mathbf{F}_p} \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & & & & X_0(N)_{\mathbf{F}_p} \arrow[dl, "\mathrm{id}=\alpha\circ r"'] \arrow[dr, "F=\beta\circ r"] & & & & X_0(N)_{\mathbf{F}_p} \arrow[dl, "F=\alpha\circ s"'] \arrow[dr, "\mathrm{id}=\beta\circ s"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p} & =' & X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p} & + & X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p},
\end{tikzcd}$$

where $F = \mathrm{Frob}_p$, and the $='$ means equality away from the supersingular points. Note that we are simply "pulling back" the correspondence; in the first summand we use the inclusion r, and in the second we use the inclusion s.

This equality of correspondences implies that the equality

$$T_p = \mathrm{Frob}_p + \mathrm{Ver}_p$$

of endomorphisms holds on a dense subset of $J_0(N)_{\mathbf{F}_p}$, hence on all $J_0(N)_{\mathbf{F}_p}$. □

8.7 Applications of the Eichler-Shimura Relation

8.7.1 The Characteristic Polynomial of Frobenius

How can we apply the relation $T_p = \mathrm{Frob} + \mathrm{Ver}$ in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$? Let $\ell \nmid pN$ be a prime and consider the ℓ-adic Tate module

$$\mathrm{Tate}_\ell(J_0(N)) = \left(\varprojlim J_0(N)[\ell^\nu]\right) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$$

which is a vector space of dimension $2g$ over $\mathbf{Q}_\ell$, where g is the genus of $X_0(N)$ or the dimension of $J_0(N)$. Reduction modulo p induces an isomorphism

$$\mathrm{Tate}_\ell(J_0(N)) \to \mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$$

(see the proof of Lemma 8.6.2). On $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$ we have linear operators Frob_p, Ver_p and T_p which, as we saw in Section 8.6, satisfy

$$\begin{aligned} \mathrm{Frob}_p + \mathrm{Ver}_p &= T_p, \qquad \text{and} \\ \mathrm{Frob}_p \circ \mathrm{Ver}_p &= \mathrm{Ver}_p \circ \mathrm{Frob}_p = [p]. \end{aligned}$$

The endomorphism $[p]$ is invertible on $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$, since p is prime to ℓ, so Ver_p and Frob_p are also invertible and

$$T_p = \mathrm{Frob}_p + [p]\,\mathrm{Frob}_p^{-1}.$$

Multiplying both sides by Frob_p and rearranging, we see that

$$\mathrm{Frob}_p^2 - T_p \mathrm{Frob}_p + [p] = 0 \in \mathrm{End}(\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})).$$

This is a beautiful quadratic relation, so we should be able to get something out of it. We will come back to this shortly, but first we consider the various objects acting on the ℓ-adic Tate module.

The module $\mathrm{Tate}_\ell(J_0(N))$ is acted upon in a natural way by

1. The Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of $\mathbf{Q}$, and
2. $\mathrm{End}_{\mathbf{Q}}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ (which acts by functoriality).

These actions commute with each other since endomorphisms defined over $\mathbf{Q}$ are not affected by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Reducing modulo p, we also have the following commuting actions:

3. The Galois group $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ of $\mathbf{F}_p$, and
4. $\mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$.

Note that a decomposition group group $D_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts, after quotienting out by the corresponding inertia group, in the same way as $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and the action is unramified, so action 3 is a special case of action 1.

The Frobenius elements $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and $\mathrm{Frob}_\in \mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ induce the same operator on $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$. Note that while φ_p naturally lives in a quotient of a decomposition group, one often takes a lift to get an element in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

On $\mathrm{Tate}_\ell(J_0(N)_{\mathbf{F}_p})$ we have a quadratic relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0.$$

This relation plays a role when one separates out pieces of $J_0(N)$ in order to construct Galois representations attached to newforms of weight 2. Let

$$R = \mathbf{Z}[\ldots, T_p, \ldots] \subset \mathrm{End}\, J_0(N),$$

where we only adjoin those T_p with $p \nmid N$. Think of R as a reduced Hecke algebra; in particular, R is a subring of $\mathbf{T}$. Then

$$R \otimes \mathbf{Q} = \prod_{i=1}^{r} E_i,$$

where the E_i are totally real number fields. The factors E_i are in bijection with the Galois conjugacy classes of weight 2 newforms f on $\Gamma_0(M)$ (for some $M|N$). The bijection is the map

$$f \mapsto \mathbf{Q}(\text{coefficients of } f) = E_i$$

Observe that the map is the same if we replace f by one of its conjugates. This decomposition is a decomposition of a subring

$$R \otimes \mathbf{Q} \subset \mathrm{End}(J_0(N)) \otimes \mathbf{Q} \overset{\text{def}}{=} \mathrm{End}(J_0(N) \otimes \mathbf{Q}).$$

Thus it induces a direct product decomposition of $J_0(N)$, so $J_0(N)$ gets divided up into subvarieties which correspond to conjugacy classes of newforms.

The relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0 \tag{8.7.1}$$

suggests that

$$\mathrm{tr}(\varphi_p) = T_p \qquad \text{and} \quad \det \varphi_p = p. \tag{8.7.2}$$

This is true, but (8.7.2) does not follow formally just from the given quadratic relation. It can be proved by combining (8.7.1) with the Weil pairing.

8.7.2 The Cardinality of $J_0(N)(\mathbf{F}_p)$

Proposition 8.7.1. *Let $p \nmid N$ be a prime, and let f be the characteristic polynomial of T_p acting on $S_2(\Gamma_0(N))$. Then*

$$\#J_0(N)(\mathbf{F}_p) = f(p+1).$$

9
Abelian Varieties

This chapter provides foundational background about abelian varieties and Jacobians, with an aim toward what we will need later when we construct abelian varieties attached to modular forms. We will not give complete proofs of very much, but will try to give precise references whenever possible, and many examples.

We will follow the articles by Rosen [Ros86] and Milne [Mil86] on abelian varieties. We will try primarily to explain the statements of the main results about abelian varieties, and prove results when the proofs are not too technical and enhance understanding of the statements.

9.1 Abelian Varieties

Definition 9.1.1 (Variety). A *variety* X over a field k is a finite-type separated scheme over k that is geometrically integral.

The condition that X be geometrically integral means that $X_{\overline{k}}$ is reduced (no nilpotents in the structure sheaf) and irreducible.

Definition 9.1.2 (Group variety). A *group variety* is a group object in the category of varieties. More precisely, a group variety X over a field k is a variety equipped with morphisms

$$m : X \times X \to X \quad \text{and} \quad i : X \to X$$

and a point $1_X \in A(k)$ such that m, i, and 1_X satisfy the axioms of a group; in particular, for every k-algebra R they give $X(R)$ a group structure that depends in a functorial way on R.

Definition 9.1.3 (Abelian Variety). An *abelian variety* A over a field k is a complete group variety.

Theorem 9.1.4. *Suppose A is an abelian variety. Then*

1. *The group law on A is commutative.*

2. *A is projective, i.e., there is an embedding from A into $\mathbf{P}^n$ for some n.*

3. *If $k = \mathbf{C}$, then $A(k)$ is analytically isomorphic to V/L, where V is a finite-dimensional complex vector space and L is a lattice in V. (A lattice is a free $\mathbf{Z}$-module of rank equal to $2 \dim V$ such that $\mathbf{R}L = V$.)*

Proof. Part 1 is not too difficult, and can be proved by showing that every morphism of abelian varieties is the composition of a homomorphism with a translation, then applying this result to the inversion map (see [Mil86, Cor. 2.4]). Part 2 is proved with some effort in [Mil86, §7]. Part 3 is proved in [Mum70, §I.1] using the exponential map from Lie theory from the tangent space at 0 to A. □

9.2 Complex Tori

Let A be an abelian variety over $\mathbf{C}$. By Theorem 9.1.4, there is a complex vector space V and a lattice L in V such that $A(\mathbf{C}) = V/L$, that is to say, $A(\mathbf{C})$ is a complex torus.

More generally, if V is any complex vector space and L is a lattice in V, we call the quotient $T = V/L$ a *complex torus.* In this section, we prove some results about complex tori that will help us to understand the structure of abelian varieties, and will also be useful in designing algorithms for computing with abelian varieties.

The differential 1-forms and first homology of a complex torus are easy to understand in terms of T. If $T = V/L$ is a complex torus, the tangent space to $0 \in T$ is canonically isomorphic to V. The $\mathbf{C}$-linear dual $V^* = \mathrm{Hom}_{\mathbf{C}}(V, \mathbf{C})$ is isomorphic to the $\mathbf{C}$-vector space $\Omega(T)$ of holomorphic differential 1-forms on T. Since $V \to T$ is the universal covering of T, the first homology $\mathrm{H}_1(T, \mathbf{Z})$ of T is canonically isomorphic to L.

9.2.1 Homomorphisms

Suppose $T_1 = V_1/L_1$ and $T_2 = V_2/L_2$ are two complex tori. If $\varphi : T_1 \to T_2$ is a (holomorphic) homomorphism, then φ induces a $\mathbf{C}$-linear map from the tangent space of T_1 at 0 to the tangent space of T_2 at 0. The tangent space of T_i at 0 is canonically isomorphic to V_i, so φ induces a $\mathbf{C}$-linear map $V_1 \to V_2$. This maps sends L_1 into L_2, since $L_i = \mathrm{H}_1(T_i, \mathbf{Z})$. We thus have the following diagram:

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] \\
L_1 \arrow[r, "\rho_{\mathbf{Z}}(\varphi)"] \arrow[d] & L_2 \arrow[d] \\
V_1 \arrow[r, "\rho_{\mathbf{C}}(\varphi)"] \arrow[d] & L_2 \arrow[d] \\
T_1 \arrow[r, "\varphi"] \arrow[d] & T_2 \arrow[d] \\
0 & 0
\end{tikzcd} \tag{9.2.1}$$

We obtain two faithful representations of $\mathrm{Hom}(T_1, T_2)$,

$$\rho_{\mathbf{C}} : \mathrm{Hom}(T_1, T_2) \to \mathrm{Hom}_{\mathbf{C}}(V_1, V_2)$$

$$\rho_{\mathbf{Z}} : \mathrm{Hom}(T_1, T_2) \to \mathrm{Hom}_{\mathbf{Z}}(L_1, L_2).$$

Suppose $\psi \in \mathrm{Hom}_{\mathbf{Z}}(L_1, L_2)$. Then $\psi = \rho_{\mathbf{Z}}(\varphi)$ for some $\varphi \in \mathrm{Hom}(T_1, T_2)$ if and only if there is a complex linear homomorphism $f : V_1 \to V_2$ whose restriction to L_1 is ψ. Note that $f = \psi \otimes \mathbf{R}$ is uniquely determined by ψ, so ψ arises from some φ precisely when f is $\mathbf{C}$-linear. This is the case if and only if $fJ_1 = J_2 f$, where $J_n : V_n \to V_n$ is the $\mathbf{R}$-linear map induced by multiplication by $i = \sqrt{-1} \in \mathbf{C}$.

Example 9.2.1.

1. Suppose $L_1 = \mathbf{Z} + \mathbf{Z}i \subset V_1 = \mathbf{C}$. Then with respect to the basis $1, i$, we have $J_1 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. One finds that $\mathrm{Hom}(T_1, T_1)$ is the free $\mathbf{Z}$-module of rank 2 whose image via $\rho_{\mathbf{Z}}$ is generated by J_1 and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. As a ring $\mathrm{Hom}(T_1, T_1)$ is isomorphic to $\mathbf{Z}[i]$.

2. Suppose $L_1 = \mathbf{Z} + \mathbf{Z}\alpha i \subset V_1 = \mathbf{C}$, with $\alpha^3 = 2$. Then with respect to the basis $1, \alpha i$, we have $J_1 = \left(\begin{smallmatrix} 0 & -\alpha \\ 1/\alpha & 0 \end{smallmatrix}\right)$. Only the scalar integer matrices commute with J_1.

Proposition 9.2.2. *Let T_1 and T_2 be complex tori. Then $\mathrm{Hom}(T_1, T_2)$ is a free $\mathbf{Z}$-module of rank at most $4 \dim T_1 \cdot \dim T_2$.*

Proof. The representation $\rho_{\mathbf{Z}}$ is faithful (injective) because φ is determined by its action on L_1, since L_1 spans V_1. Thus $\mathrm{Hom}(T_1, T_2)$ is isomorphic to a subgroup of $\mathrm{Hom}_{\mathbf{Z}}(L_1, L_2) \cong \mathbf{Z}^d$, where $d = 2 \dim V_1 \cdot 2 \dim V_2$. □

Lemma 9.2.3. *Suppose $\varphi : T_1 \to T_2$ is a homomorphism of complex tori. Then the image of φ is a subtorus of T_2 and the connected component of $\ker(\varphi)$ is a subtorus of T_1 that has finite index in $\ker(\varphi)$.*

Proof. Let $W = \ker(\rho_{\mathbf{C}}(\varphi))$. Then the following diagram, which is induced by φ, has exact rows and columns:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & & \\
0 \arrow[r] & L_1 \cap W \arrow[r] \arrow[d] & L_1 \arrow[r] \arrow[d] & L_2 \arrow[r] \arrow[d] & L_2/\varphi(L_1) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & W \arrow[r] \arrow[d] & V_1 \arrow[r] \arrow[d] & V_2 \arrow[r] \arrow[d] & V_2/\varphi(V_1) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \ker(\varphi) \arrow[r] & T_1 \arrow[r] \arrow[d] & T_2 \arrow[r] \arrow[d] & T_2/\varphi(T_1) \arrow[r] \arrow[d] & 0 \\
 & & 0 & 0 & 0 &
\end{tikzcd}$$

Using the snake lemma, we obtain an exact sequence

$$0 \to L_1 \cap W \to W \to \ker(\varphi) \to L_2/\varphi(L_1) \to V_2/\varphi(V_1) \to T_2/\varphi(T_1) \to 0.$$

Note that $T_2/\varphi(T_1)$ is compact because it is the continuous image of a compact set, so the cokernel of φ is a torus (it is given as a quotient of a complex vector space by a lattice).

The kernel $\ker(\varphi) \subset T_1$ is a closed subset of the compact set T_1, so is compact. Thus $L_1 \cap W$ is a lattice in W. The map $L_2/\varphi(L_1) \to V_2/\varphi(V_1)$ has kernel generated by the saturation of $\varphi(L_1)$ in L_2, so it is finite, so the torus $W/(L_1 \cap W)$ has finite index in $\ker(\varphi)$. □

Remark 9.2.4. The category of complex tori is not an abelian category because kernels need not be in the category.

9.2.2 Isogenies

Definition 9.2.5 (Isogeny). An *isogeny* $\varphi : T_1 \to T_2$ of complex tori is a surjective morphism with finite kernel. The *degree* $\deg(\varphi)$ of φ is the order of the kernel of φ.

Note that $\deg(\varphi \circ \varphi') = \deg(\varphi)\deg(\varphi')$.

Lemma 9.2.6. *Suppose that φ is an isogeny. Then the kernel of φ is isomorphic to the cokernel of $\rho_{\mathbf{Z}}(\varphi)$.*

Proof. (This is essentially a special case of Lemma 9.2.3.) Apply the snake lemma to the morphism (9.2.1) of short exact sequences, to obtain a six-term exact sequence

$$0 \to K_L \to K_V \to K_T \to C_L \to C_V \to C_T \to 0,$$

where K_X and C_X are the kernel and cokernel of $X_1 \to X_2$, for $X = L, V, T$, respectively. Since φ is an isogeny, the induced map $V_1 \to V_2$ must be an isomorphism, since otherwise the kernel would contain a nonzero subspace (modulo a lattice), which would be infinite. Thus $K_V = C_V = 0$. It follows that $K_T \cong C_L$, as claimed. □

One consequence of the lemma is that if φ is an isogeny, then

$$\deg(\varphi) = [L_1 : \rho_{\mathbf{Z}}(\varphi)(L_1)] = |\det(\rho_{\mathbf{Z}}(\varphi))|.$$

Proposition 9.2.7. *Let T be a complex torus of dimension d, and let n be a positive integer. Then multiplication by n, denoted $[n]$, is an isogeny $T \to T$ with kernel $T[n] \cong (\mathbf{Z}/n\mathbf{Z})^{2d}$ and degree n^{2d}.*

Proof. By Lemma 9.2.6, $T[n]$ is isomorphic to L/nL, where $T = V/L$. Since $L \approx \mathbf{Z}^{2d}$, the proposition follows. □

We can now prove that isogeny is an equivalence relation.

Proposition 9.2.8. *Suppose $\varphi : T_1 \to T_2$ is a degree m isogeny of complex tori of dimension d. Then there is a unique isogeny $\hat{\varphi} : T_2 \to T_1$ of degree m^{2d-1} such that $\hat{\varphi} \circ \varphi = \varphi \circ \hat{\varphi} = [m]$.*

Proof. Since $\ker(\varphi) \subset \ker([m])$, the map $[m]$ factors through φ, so there is a morphism $\hat{\varphi}$ such that $\hat{\varphi} \circ \varphi = [m]$:

$$\begin{tikzcd}
T_1 \arrow[r, "\varphi"] \arrow[dr, "{[m]}"'] & T_2 \arrow[d, "\hat{\varphi}"] \\
 & T_1
\end{tikzcd}$$

We have

$$(\varphi \circ \hat{\varphi} - [m]) \circ \varphi = \varphi \circ \hat{\varphi} \circ \phi - [m] \circ \varphi = \varphi \circ \hat{\varphi} \circ \phi - \varphi \circ [m] = \varphi \circ (\hat{\varphi} \circ \phi - [m]) = 0.$$

This implies that $\varphi \circ \hat{\varphi} = [m]$, since φ is surjective. Uniqueness is clear since the difference of two such morphisms would vanish on the image of φ. To see that $\hat{\varphi}$ has degree m^{2d-1}, we take degrees on both sides of the equation $\hat{\varphi} \circ \varphi = [m]$. □

9.2.3 Endomorphisms

The ring $\mathrm{End}(T) = \mathrm{Hom}(T, T)$ is called the *endomorphism ring* of the complex torus T. The *endomorphism algebra* of T is $\mathrm{End}_0(T) = \mathrm{End}(T) \otimes_{\mathbf{Z}} \mathbf{Q}$.

Definition 9.2.9 (Characteristic polynomial). The *characteristic polynomial* of $\varphi \in \mathrm{End}(T)$ is the characteristic polynomial of the $\rho_{\mathbf{Z}}(\varphi)$. Thus the characteristic polynomial is a monic polynomial of degree $2 \dim T$.

9.3 Abelian Varieties as Complex Tori

In this section we introduce extra structure on a complex torus $T = V/L$ that will enable us to understand whether or not T is isomorphic to $A(\mathbf{C})$, for some abelian variety A over $\mathbf{C}$. When $\dim T = 1$, the theory of the Weierstrass $\wp$ function implies that T is always $E(\mathbf{C})$ for some elliptic curve. In contrast, the generic torus of dimension > 1 does not arise from an abelian variety.

In this section we introduce the basic structures on complex tori that are needed to understand which tori arise from abelian varieties, to construct the dual of an abelian variety, to see that $\mathrm{End}_0(A)$ is a semisimple $\mathbf{Q}$-algebra, and to understand the polarizations on an abelian variety. For proofs, including extensive motivation from the one-dimensional case, read the beautifully written book [SD74] by Swinnerton-Dyer, and for another survey that strongly influenced the discussion below, see Rosen's [Ros86].

9.3.1 Hermitian and Riemann Forms

Let V be a finite-dimensional complex vector space.

Definition 9.3.1 (Hermitian form). A *Hermitian form* is a conjugate-symmetric pairing

$$H : V \times V \to \mathbf{C}$$

that is $\mathbf{C}$-linear in the first variable and $\mathbf{C}$-antilinear in the second. Thus H is $\mathbf{R}$-bilinear, $H(iu, v) = iH(u, v) = H(u, \bar{i}v)$, and $H(u, v) = \overline{H(v, u)}$.

Write $H = S + iE$, where $S, E : V \times V \to \mathbf{R}$ are real bilinear pairings.

Proposition 9.3.2. *Let H, S, and E be as above.*

1. *We have that S is symmetric, E is antisymmetric, and*
$$S(u, v) = E(iu, v), \quad S(iu, iv) = S(u, v), \quad E(iu, iv) = E(u, v).$$
2. *Conversely, if E is a real-valued antisymmetric bilinear pairing on V such that $E(iu, iv) = E(u, v)$, then $H(u, v) = E(iu, v) + iE(u, v)$ is a Hermitian form on V. Thus there is a bijection between the Hermitian forms on V and the real, antisymmetric bilinear forms E on V such that $E(iu, iv) = E(u, v)$.*

Proof. To see that S is symmetric, note that $2S = H + \overline{H}$ and $H + \overline{H}$ is symmetric because H is conjugate symmetric. Likewise, $E = (H - \overline{H})/(2i)$, so

$$E(v, u) = \frac{1}{2i}\left(H(v, u) - \overline{H(v, u)}\right) = \frac{1}{2i}\left(\overline{H(u, v)} - H(u, v)\right) = -E(u, v),$$

which implies that E is antisymmetric. To see that $S(u, v) = E(iu, v)$, rewrite both $S(u, v)$ and $E(iu, v)$ in terms of H and simplify to get an identity. The other two identities follow since

$$H(iu, iv) = iH(u, iv) = i\bar{i}H(u, v) = H(u, v).$$

Suppose $E : V \times V \to \mathbf{R}$ is as in the second part of the proposition. Then

$$H(iu, v) = E(i^2u, v) + iE(iu, v) = -E(u, v) + iE(iu, v) = iH(u, v),$$

and the other verifications of linearity and antilinearity are similar. For conjugate symmetry, note that

$$\begin{aligned} H(v,u) &= E(iv,u) + iE(v,u) = -E(u,iv) - iE(u,v) \\ &= -E(iu,-v) - iE(u,v) = H(u,v). \end{aligned}$$

□

Note that the set of Hermitian forms is a group under addition.

Definition 9.3.3 (Riemann form). A *Riemann form* on a complex torus $T = V/L$ is a Hermitian form H on V such that the restriction of $E = \mathrm{Im}(H)$ to L is integer valued. If $H(u,u) \geq 0$ for all $u \in V$ then H is *positive semi-definite* and if H is positive and $H(u,u) = 0$ if and only if $u = 0$, then H is *nondegenerate.*

Theorem 9.3.4. *Let T be a complex torus. Then T is isomorphic to $A(\mathbf{C})$, for some abelian variety A, if and only if there is a nondegenerate Riemann form on T.*

This is a nontrivial theorem, which we will not prove here. It is proved in [SD74, Ch.2] by defining an injective map from positive divisors on $T = V/L$ to positive semi-definite Riemann forms, then constructing positive divisors associated to theta functions on V. If H is a nondegenerate Riemann form on T, one computes the dimension of a space of theta functions that corresponds to H in terms of the determinant of $E = \mathrm{Im}(H)$. Since H is nondegenerate, this space of theta functions is nonzero, so there is a corresponding nondegenerate positive divisor D. Then a basis for

$$L(3D) = \{f \,:\, (f) + 3D \text{ is positive }\} \cup \{0\}$$

determines an embedding of T in a projective space.

Why the divisor $3D$ instead of D above? For an elliptic curve $y^2 = x^3 + ax + b$, we could take D to be the point at infinity. Then $L(3D)$ consists of the functions with a pole of order at most 3 at infinity, which contains 1, x, and y, which have poles of order 0, 2, and 3, respectively.

Remark 9.3.5. (Copied from page 39 of [SD74].) When $n = \dim V > 1$, however, a general lattice L will admit no nonzero Riemann forms. For if $\lambda_1, \ldots, \lambda_{2n}$ is a base for L then E as an $\mathbf{R}$-bilinear alternating form is uniquely determined by the $E(\lambda_i, \lambda_j)$, which are integers; and the condition $E(z,w) = E(iz,iw)$ induces linear relations with real coefficients between $E(\lambda_i, \lambda_j)$, which for general L have no nontrivial integer solutions.

9.3.2 Complements, Quotients, and Semisimplicity of the Endomorphism Algebra

Lemma 9.3.6. *If T possesses a nondegenerate Riemann form and $T' \subset T$ is a subtorus, then T' also possesses a nondegenerate Riemann form.*

Proof. If H is a nondegenerate Riemann form on a torus T and T' is a subtorus of T, then the restriction of H to T' is a nondegenerate Riemann form on T' (the restriction is still nondegenerate because H is positive definite). □

Lemma 9.3.6 and Lemma 9.2.3 together imply that the kernel of a homomorphism of abelian varieties is an extension of an abelian variety by a finite group.

Lemma 9.3.7. *If T possesses a nondegenerate Riemann form and $T \to T'$ is an isogeny, then T' also possesses a nondegenerate Riemann form.*

Proof. Suppose $T = V/L$ and $T' = V'/L'$. Since the isogeny is induced by an isomorphism $V \to V'$ that sends L into L', we may assume for simplicity that $V = V'$ and $L \subset L'$. If H is a nondegenerate Riemann form on V/L, then $E = \mathrm{Re}(H)$ need not be integer valued on L'. However, since L has finite index in L', there is some integer d so that dE is integer valued on L'. Then dH is a nondegenerate Riemann form on V/L'. □

Note that Lemma 9.3.7 implies that the quotient of an abelian variety by a finite subgroup is again an abelian variety.

Theorem 9.3.8 (Poincare Reducibility). *Let A be an abelian variety and suppose $A' \subset A$ is an abelian subvariety. Then there is an abelian variety $A'' \subset A$ such that $A = A' + A''$ and $A' \cap A''$ is finite. (Thus A is isogenous to $A' \times A''$.)*

Proof. We have $A(\mathbf{C}) \approx V/L$ and there is a nondegenerate Riemann form H on V/L. The subvariety A' is isomorphic to V'/L', where V' is a subspace of V and $L' = V' \cap L$. Let V'' be the orthogonal complement of V' with respect to H, and let $L'' = L \cap V''$. To see that L'' is a lattice in V'', it suffices to show that L'' is the orthogonal complement of L' in L with respect to $E = \mathrm{Im}(H)$, which, because E is integer valued, will imply that L'' has the correct rank. First, suppose that $v \in L''$; then, by definition, v is in the orthogonal complement of L' with respect to H, so for any $u \in L'$, we have $0 = H(u, v) = S(u, v) + iE(u, v)$, so $E(u, v) = 0$. Next, suppose that $v \in L$ satisfies $E(u, v) = 0$ for all $u \in L'$. Since $V' = \mathbf{R}L'$ and E is $\mathbf{R}$-bilinear, this implies $E(u, v) = 0$ for any $u \in V'$. In particular, since V' is a complex vector space, if $u \in L'$, then $S(u, v) = E(iu, v) = 0$, so $H(u, v) = 0$.

We have shown that L'' is a lattice in V'', so $A'' = V''/L''$ is an abelian subvariety of A. Also $L' + L''$ has finite index in L, so there is an isogeny $V'/L' \oplus V''/L'' \to V/L$ induced by the natural inclusions. □

Proposition 9.3.9. *Suppose $A' \subset A$ is an inclusion of abelian varieties. Then the quotient A/A' is an abelian variety.*

Proof. Suppose $A = V/L$ and $A' = V'/L'$, where V' is a subspace of V. Let $W = V/V'$ and $M = L/(L \cap V')$. Then, W/M is isogenous to the complex torus V''/L'' of Theorem 9.3.8 via the natural map $V'' \to W$. Applying Lemma 9.3.7 completes the proof. □

Definition 9.3.10. An abelian variety A is *simple* if it has no nonzero proper abelian subvarieties.

Proposition 9.3.11. *The algebra* $\mathrm{End}_0(A)$ *is semisimple.*

Proof. Using Theorem 9.3.8 and induction, we can find an isogeny

$$A \simeq A_1^{n_1} \times A_2^{n_2} \times \cdots \times A_r^{n_r}$$

with each A_i simple. Since $\mathrm{End}_0(A) = \mathrm{End}(A) \otimes \mathbf{Q}$ is unchanged by isogeny, and $\mathrm{Hom}(A_i, A_j) = 0$ when $i \neq j$, we have

$$\mathrm{End}_0(A) = \mathrm{End}_0(A_1^{n_1}) \times \mathrm{End}_0(A_2^{n_2}) \times \cdots \times \mathrm{End}_0(A_r^{n_r})$$

Each of $\mathrm{End}_0(A_i^{n_i})$ is isomorphic to $M_{n_i}(D_i)$, where $D_i = \mathrm{End}_0(A_i)$. By Schur's Lemma, $D_i = \mathrm{End}_0(A_i)$ is a division algebra over $\mathbf{Q}$ (proof: any nonzero endomorphism has trivial kernel, and any injective linear transformation of a $\mathbf{Q}$-vector space is invertible), so $\mathrm{End}_0(A)$ is a product of matrix algebras over division algebras over $\mathbf{Q}$, which proves the proposition. □

9.3.3 Theta Functions

Suppose $T = V/L$ is a complex torus.

Definition 9.3.12 (Theta function). Let $M : V \times L \to C$ and $J : L \to \mathbf{C}$ be set-theoretic maps such that for each $\lambda \in L$ the map $z \mapsto M(z, \lambda)$ is $\mathbf{C}$-linear. A *theta function* of type (M, J) is a function $\theta : V \to \mathbf{C}$ such that for all $z \in V$ and $\lambda \in L$, we have

$$\theta(z + \lambda) = \theta(z) \cdot \exp(2\pi i(M(z, \lambda) + J(\lambda))).$$

Suppose that $\theta(z)$ is a nonzero holomorphic theta function of type (M, J). The $M(z, \lambda)$, for various λ, cannot be unconnected. Let $F(z, \lambda) = 2\pi i(M(z, \lambda) + J(\lambda))$.

Lemma 9.3.13. *For any $\lambda, \lambda' \in L$, we have*

$$F(z, \lambda + \lambda') = F(z + \lambda, \lambda') + F(z, \lambda) \pmod{2\pi i}.$$

Thus

$$M(z, \lambda + \lambda') = M(z, \lambda) + M(z, \lambda'), \tag{9.3.1}$$

and

$$J(\lambda + \lambda') - J(\lambda) - J(\lambda') \equiv M(\lambda, \lambda') \pmod{\mathbf{Z}}.$$

Proof. Page 37 of [SD74]. □

Using (9.3.1) we see that M extends uniquely to a function $\tilde{M} : V \times V \to \mathbf{C}$ which is $\mathbf{C}$-linear in the first argument and $\mathbf{R}$-linear in the second. Let

$$E(z, w) = \tilde{M}(z, w) - M(w, z),$$

$$H(z, w) = E(iz, w) + iE(z, w).$$

Proposition 9.3.14. *The pairing H is Riemann form on T with real part E.*

We call H the Riemann form associated to θ.

9.4 A Summary of Duality and Polarizations

Suppose A is an abelian variety over an arbitrary field k. In this section we summarize the most important properties of the dual abelian variety $A^\vee$ of A. First we review the language of sheaves on a scheme X, and define the Picard group of X as the group of invertible sheaves on X. The dual of A is then a variety whose points correspond to elements of the Picard group that are algebraically equivalent to 0. Next, when the ground field is $\mathbf{C}$, we describe how to view $A^\vee$ as a complex torus in terms of a description of A as a complex torus. We then define the Néron-Severi group of A and relate it to polarizations of A, which are certain homomorphisms $A \to A^\vee$. Finally we observe that the dual is functorial.

9.4.1 Sheaves

We will use the language of sheaves, as in [Har77], which we now quickly recall. A *pre-sheaf of abelian groups* $\mathcal{F}$ on a scheme X is a contravariant functor from the category of open sets on X (morphisms are inclusions) to the category of abelian groups. Thus for every open set $U \subset X$ there is an abelian group $\mathcal{F}(U)$, and if $U \subset V$, then there is a restriction map $\mathcal{F}(V) \to \mathcal{F}(U)$. (We also require that $\mathcal{F}(\emptyset) = 0$, and the map $\mathcal{F}(U) \to \mathcal{F}(U)$ is the identity map.) A *sheaf* is a pre-sheaf whose sections are determined locally (for details, see [Har77, §II.1]).

Every scheme X is equipped with its structure sheaf $\mathcal{O}_X$, which has the property that if $U = \mathrm{Spec}(R)$ is an affine open subset of X, then $\mathcal{O}_X(U) = R$. A *sheaf of $\mathcal{O}_X$-modules* is a sheaf $\mathcal{M}$ of abelian groups on X such that each abelian group has the structure of $\mathcal{O}_X$-module, such that the restriction maps are module morphisms. A *locally-free sheaf* of $\mathcal{O}_X$-modules is a sheaf $\mathcal{M}$ of $\mathcal{O}_X$-modules, such that X can be covered by open sets U so that $\mathcal{M}|_U$ is a free $\mathcal{O}_X$-module, for each U.

9.4.2 The Picard Group

An *invertible sheaf* is a sheaf $\mathcal{L}$ of $\mathcal{O}_X$-modules that is locally free of rank 1. If $\mathcal{L}$ is an invertible sheaf, then the sheaf-theoretic Hom, $\mathcal{L}^\vee = \mathcal{H}\mathrm{om}(\mathcal{L}, \mathcal{O}_X)$ has the property that $\mathcal{L}^\vee \otimes \mathcal{L} = \mathcal{O}_X$. The group $\mathrm{Pic}(X)$ of invertible sheaves on a scheme X is called *the Picard group* of X. See [Har77, §II.6] for more details.

Let A be an abelian variety over a field k. An invertible sheaf $\mathcal{L}$ on A is *algebraically equivalent to* 0 if there is a connected variety T over k, an invertible sheaf $\mathcal{M}$ on $A \times_k T$, and $t_0, t_1 \in T(k)$ such that $\mathcal{M}_{t_0} \cong \mathcal{L}$ and $\mathcal{M}_{t_1} \cong \mathcal{O}_A$. Let $\mathrm{Pic}^0(A)$ be the subgroup of elements of $\mathrm{Pic}(A)$ that are algebraically equivalent to 0.

The *dual* $A^\vee$ of A is a (unique up to isomorphism) abelian variety such that for every field F that contains the base field k, we have $A^\vee(F) = \mathrm{Pic}^0(A_F)$. For the precise definition of $A^\vee$ and a proof that $A^\vee$ exists, see [Mil86, §9–10].

9.4.3 The Dual as a Complex Torus

When A is defined over the complex numbers, so $A(\mathbf{C}) = V/L$ for some vector space V and some lattice L, [Ros86, §4] describes a construction of $A^\vee$ as a complex torus, which we now describe. Let

$$V^* = \{f \in \mathrm{Hom}_{\mathbf{R}}(V, \mathbf{C}) \,:\, f(\alpha t) = \overline{\alpha} f(t), \text{ all } \alpha \in \mathbf{C},\, t \in V\}.$$

Then V^* is a complex vector space of the same dimension as V and the map $\langle f, v\rangle = \mathrm{Im} f(t)$ is an $\mathbf{R}$-linear pairing $V^* \times V \to \mathbf{R}$. Let

$$L^* = \{f \in V^* \,:\, \langle f, \lambda\rangle \in \mathbf{Z}, \text{ all } \lambda \in L\}.$$

Since A is an abelian variety, there is a nondegenerate Riemann form H on A. The map $\lambda : V \to V^*$ defined by $\lambda(v) = H(v, \cdot)$ is an isomorphism of complex vector spaces. If $v \in L$, then $\lambda(v) = H(v, \cdot)$ is integer valued on L, so $\lambda(L) \subset L^*$. Thus λ induces an isogeny of complex tori $V/L \to V^*/L^*$, so by Lemma 9.3.7 the torus V^*/L^* possesses a nondegenerate Riemann form (it's a multiple of H). In [Ros86, §4], Rosen describes an explicit isomorphism between V^*/L^* and $A^\vee(\mathbf{C})$.

9.4.4 The Néron-Several Group and Polarizations

Let A be an abelian variety over a field k. Recall that $\mathrm{Pic}(A)$ is the group of invertible sheaves on A, and $\mathrm{Pic}^0(A)$ is the subgroup of invertible sheaves that are algebraically equivalent to 0. The *Néron-Severi group* of A is the quotient $\mathrm{Pic}(A)/\mathrm{Pic}^0(A)$, so by definition we have an exact sequence

$$0 \to \mathrm{Pic}^0(A) \to \mathrm{Pic}(A) \to \mathrm{NS}(A) \to 0.$$

Suppose $\mathcal{L}$ is an invertible sheaf on A. One can show that the map $A(k) \to \mathrm{Pic}^0(A)$ defined by $a \mapsto t_a^*\mathcal{L} \otimes \mathcal{L}^{-1}$ is induced by homomorphism $\varphi_{\mathcal{L}} : A \to A^\vee$. (Here $t_a^*\mathcal{L}$ is the pullback of the sheaf $\mathcal{L}$ by translation by a.) Moreover, the map $\mathcal{L} \mapsto \varphi_{\mathcal{L}}$ induces a homomorphism from $\mathrm{Pic}(A) \to \mathrm{Hom}(A, A^\vee)$ with kernel $\mathrm{Pic}^0(A)$. The group $\mathrm{Hom}(A, A^\vee)$ is free of finite rank, so $\mathrm{NS}(A)$ is a free abelian group of finite rank. Thus $\mathrm{Pic}^0(A)$ is saturated in $\mathrm{Pic}(A)$ (i.e., the cokernel of the inclusion $\mathrm{Pic}^0(A) \to \mathrm{Pic}(A)$ is torsion free).

Definition 9.4.1 (Polarization). A *polarization* on A is a homomorphism $\lambda : A \to A^\vee$ such that $\lambda_{\overline{k}} = \varphi_{\mathcal{L}}$ for some $\mathcal{L} \in \mathrm{Pic}(A_{\overline{k}})$. A polarization is *principal* if it is an isomorphism.

When the base field k is algebraically closed, the polarizations are in bijection with the elements of $\mathrm{NS}(A)$. For example, when $\dim A = 1$, we have $\mathrm{NS}(A) = \mathbf{Z}$, and the polarizations on A are multiplication by n, for each integer n.

9.4.5 The Dual is Functorial

The association $A \mapsto A^\vee$ extends to a contravariant functor on the category of abelian varieties. Thus if $\varphi : A \to B$ is a homomorphism, there is a natural choice of homomorphism $\varphi^\vee : B^\vee \to A^\vee$. Also, $(A^\vee)^\vee = A$ and $(\varphi^\vee)^\vee = \varphi$.

Theorem 9.4.2 below describes the kernel of $\varphi^\vee$ in terms of the kernel of φ. If G is a finite group scheme, the *Cartier dual* of G is $\mathrm{Hom}(G, \mathbf{G}_m)$. For example, the Cartier dual of $\mathbf{Z}/m\mathbf{Z}$ is μ_m and the Cartier dual of μ_m is $\mathbf{Z}/m\mathbf{Z}$. (If k is algebraically closed, then the Cartier dual of G is just G again.)

Theorem 9.4.2. *If $\varphi : A \to B$ is a surjective homomorphism of abelian varieties with kernel G, so we have an exact sequence $0 \to G \to A \to B \to 0$, then the kernel of $\varphi^\vee$ is the Cartier dual of G, so we have an exact sequence $0 \to G^\vee \to B^\vee \to A^\vee \to 0$.*

9.5 Jacobians of Curves

We begin this lecture about Jacobians with an inspiring quote of David Mumford:

> "The Jacobian has always been a corner-stone in the analysis of algebraic curves and compact Riemann surfaces. [...] Weil's construction [of the Jacobian] was the basis of his epoch-making proof of the Riemann Hypothesis for curves over finite fields, which really put characteristic p algebraic geometry on its feet." – Mumford, *Curves and Their Jacobians*, page 49.

9.5.1 Divisors on Curves and Linear Equivalence

Let X be a projective nonsingular algebraic curve over an algebraically field k. A *divisor* on X is a formal finite $\mathbf{Z}$-linear combination $\sum_{i=1}^{m} n_i P_i$ of closed points in X. Let $\mathrm{Div}(X)$ be the group of all divisors on X. The *degree* of a divisor $\sum_{i=1}^{m} n_i P_i$ is the integer $\sum_{i=1}^{m} n_i$. Let $\mathrm{Div}^0(X)$ denote the subgroup of divisors of degree 0.

Suppose k is a perfect field (for example, k has characteristic 0 or k is finite), but do not require that k be algebraically closed. Let the group of divisors on X over k be the subgroup

$$\mathrm{Div}(X) = \mathrm{Div}(X/k) = \mathrm{H}^0(\mathrm{Gal}(\overline{k}/k), \mathrm{Div}(X/\overline{k}))$$

of elements of $\mathrm{Div}(X/\overline{k})$ that are fixed by every automorphism of $\overline{k}/k$. Likewise, let $\mathrm{Div}^0(X/k)$ be the elements of $\mathrm{Div}(X/k)$ of degree 0.

A *rational function* on an algebraic curve X is a function $X \to \mathbf{P}^1$, defined by polynomials, which has only a finite number of poles. For example, if X is the elliptic curve over k defined by $y^2 = x^3 + ax + b$, then the field of rational functions on X is the fraction field of the integral domain $k[x, y]/(y^2 - (x^3 + ax + b))$. Let $K(X)$ denote the field of all rational functions on X defined over k.

There is a natural homomorphism $K(X)^* \to \mathrm{Div}(X)$ that associates to a rational function f its divisor

$$(f) = \sum \mathrm{ord}_P(f) \cdot P$$

where $\mathrm{ord}_P(f)$ is the order of vanishing of f at P. Since X is nonsingular, the local ring of X at a point P is isomorphic to $k[[t]]$. Thus we can write $f = t^r g(t)$ for some unit $g(t) \in k[[t]]$. Then $R = \mathrm{ord}_P(f)$.

Example 9.5.1. If $X = \mathbf{P}^1$, then the function $f = x$ has divisor $(0) - (\infty)$. If X is the elliptic curve defined by $y^2 = x^3 + ax + b$, then

$$(x) = (0, \sqrt{b}) + (0, -\sqrt{b}) - 2\infty,$$

and

$$(y) = (x_1, 0) + (x_2, 0) + (x_3, 0) - 3\infty,$$

where x_1, x_2, and x_3 are the roots of $x^3 + ax + b = 0$. A uniformizing parameter t at the point ∞ is x/y. An equation for the elliptic curve in an affine neighborhood of ∞ is $Z = X^3 + aXZ^2 + bZ^3$ (where $\infty = (0, 0)$ with respect to these coordinates) and $x/y = X$ in these new coordinates. By repeatedly substituting Z into this equation we see that Z can be written in terms of X.

It is a standard fact in the theory of algebraic curves that if f is a nonzero rational function, then $(f) \in \mathrm{Div}^0(X)$, i.e., the number of poles of f equals the number of zeros of f. For example, if X is the Riemann sphere and f is a polynomial, then the number of zeros of f (counted with multiplicity) equals the degree of f, which equals the order of the pole of f at infinity.

The *Picard group* $\mathrm{Pic}(X)$ of X is the group of divisors on X modulo linear equivalence. Since divisors of functions have degree 0, the subgroup $\mathrm{Pic}^0(X)$ of divisors on X of degree 0, modulo linear equivalence, is well defined. Moreover, we have an exact sequence of abelian groups

$$0 \to K(X)^* \to \mathrm{Div}^0(X) \to \mathrm{Pic}^0(X) \to 0.$$

Thus for any algebraic curve X we have associated to it an abelian group $\mathrm{Pic}^0(X)$. Suppose $\pi : X \to Y$ is a morphism of algebraic curves. If D is a divisor on Y, the pullback $\pi^*(D)$ is a divisor on X, which is defined as follows. If $P \in \mathrm{Div}(Y/\overline{k})$ is a point, let $\pi^*(P)$ be the sum $\sum e_{Q/P} Q$ where $\pi(Q) = P$ and $e_{Q/P}$ is the ramification degree of Q/P. (Remark: If t is a uniformizer at P then $e_{Q/P} = \mathrm{ord}_Q(\phi^* t_P)$.) One can show that $\pi^* : \mathrm{Div}(Y) \to \mathrm{Div}(X)$ induces a homomorphism $\mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X)$. Furthermore, we obtain the contravariant *Picard functor* from the category of algebraic curves over a fixed base field to the category of abelian groups, which sends X to $\mathrm{Pic}^0(X)$ and $\pi : X \to Y$ to $\pi^* : \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X)$.

Alternatively, instead of defining morphisms by pullback of divisors, we could consider the push forward. Suppose $\pi : X \to Y$ is a morphism of algebraic curves and D is a divisor on X. If $P \in \mathrm{Div}(X/\overline{k})$ is a point, let $\pi_*(P) = \pi(P)$. Then π_* induces a morphism $\mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y)$. We again obtain a functor, called the covariant *Albanese functor* from the category of algebraic curves to the category of abelian groups, which sends X to $\mathrm{Pic}^0(X)$ and $\pi : X \to Y$ to $\pi_* : \mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y)$.

9.5.2 Algebraic Definition of the Jacobian

First we describe some universal properties of the Jacobian under the hypothesis that $X(k) \neq \emptyset$. Thus suppose X is an algebraic curve over a field k and that $X(k) \neq \emptyset$. The Jacobian variety of X is an abelian variety J such that for an extension k'/k, there is a (functorial) isomorphism $J(k') \to \mathrm{Pic}^0(X/k')$. (I don't know whether this condition uniquely characterizes the Jacobian.)

Fix a point $P \in X(k)$. Then we obtain a map $f : X(k) \to \mathrm{Pic}^0(X/k)$ by sending $Q \in X(k)$ to the divisor class of $Q-P$. One can show that this map is induced by an injective morphism of algebraic varieties $X \hookrightarrow J$. This morphism has the following universal property: if A is an abelian variety and $g : X \to A$ is a morphism that sends P to $0 \in A$, then there is a unique homomorphism $\psi : J \to A$ of abelian varieties such that $g = \psi \circ f$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g"'] & J \arrow[d, "\psi"] \\ & A \end{tikzcd}$$

This condition uniquely characterizes J, since if $f' : X \to J'$ and J' has the universal property, then there are unique maps $J \to J'$ and $J' \to J$ whose composition in both directions must be the identity (use the universal property with $A = J$ and $f = g$).

If X is an arbitrary curve over an arbitrary field, the Jacobian is an abelian variety that represents the "sheafification" of the "relative Picard functor". Look in Milne's article or Bosch-Lüktebohmert-Raynaud *Neron Models* for more details. Knowing this totally general definition won't be important for this course, since we will only consider Jacobians of modular curves, and these curves always have a rational point, so the above properties will be sufficient.

A useful property of Jacobians is that they are canonically principally polarized, by a polarization that arises from the "θ divisor" on J. In particular, there is always an isomorphism $J \to J^\vee = \mathrm{Pic}^0(J)$.

9.5.3 The Abel-Jacobi Theorem

Over the complex numbers, the construction of the Jacobian is classical. It was first considered in the 19th century in order to obtain relations between integrals of rational functions over algebraic curves (see Mumford's book, *Curves and Their Jacobians*, Ch. III, for a nice discussion).

Let X be a Riemann surface, so X is a one-dimensional complex manifold. Thus there is a system of coordinate charts (U_α, t_α), where $t_\alpha : U_\alpha \to \mathbf{C}$ is a homeomorphism of U_α onto an open subset of $\mathbf{C}$, such that the change of coordinate maps are analytic isomorphisms. A *differential* 1*-form* on X is a choice of two continuous functions f and g to each local coordinate $z = x + iy$ on $U_\alpha \subset X$ such that $f\,dx + g\,dy$ is invariant under change of coordinates (i.e., if another local coordinate patch U'_α intersects U_α, then the differential is unchanged by the change of coordinate map on the overlap). If $\gamma : [0,1] \to X$ is a path and $\omega = f\,dx + g\,dy$ is a 1-form, then

$$\int_\gamma \omega := \int_0^1 \left(f(x(t), y(t)) \frac{dx}{dt} + g(x(t), y(t)) \frac{dy}{dt} \right) dt \in \mathbf{C}.$$

From complex analysis one sees that if γ is homologous to γ', then $\int_\gamma \omega = \int_{\gamma'} \omega$. In fact, there is a nondegenerate pairing

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

If X has genus g, then it is a standard fact that the complex vector space $\mathrm{H}^0(X, \Omega^1_X)$ of holomorphic differentials on X is of dimension g. The integration pairing defined above induces a homomorphism from integral homology to the dual V of the differentials:

$$\Phi : \mathrm{H}_1(X, \mathbf{Z}) \to V = \mathrm{Hom}(H^0(X, \Omega^1_X), \mathbf{C}).$$

This homomorphism is called the *period mapping*.

Theorem 9.5.2 (Abel-Jacobi). *The image of* Φ *is a lattice in* V*.*

The proof involves repeated clever application of the residue theorem.

The intersection pairing

$$\mathrm{H}_1(X, \mathbf{Z}) \times \mathrm{H}_1(X, \mathbf{Z}) \to \mathbf{Z}$$

defines a nondegenerate alternating pairing on $L = \Phi(\mathrm{H}_1(X, \mathbf{Z}))$. This pairing satisfies the conditions to induce a nondegenerate Riemann form on V, which gives $J = V/L$ to structure of abelian variety. The abelian variety J is the Jacobian of X, and if $P \in X$, then the functional $\omega \mapsto \int_P^Q \omega$ defines an embedding of X into J. Also, since the intersection pairing is perfect, it induces an isomorphism from J to $J^\vee$.

Example 9.5.3. For example, suppose $X = X_0(23)$ is the modular curve attached to the subgroup $\Gamma_0(23)$ of matrices in $\mathrm{SL}_2(\mathbf{Z})$ that are upper triangular modulo 24. Then $g = 2$, and a basis for $\mathrm{H}_1(X_0(23), \mathbf{Z})$ in terms of modular symbols is

$$\{-1/19, 0\}, \quad \{-1/17, 0\}, \quad \{-1/15, 0\}, \quad \{-1/11, 0\}.$$

The matrix for the intersection pairing on this basis is

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

With respect to a reduced integral basis for

$$\mathrm{H}^0(X, \Omega^1_X) \cong S_2(\Gamma_0(23)),$$

the lattice $\Phi(\mathrm{H}_1(X, \mathbf{Z}))$ of periods is (approximately) spanned by

```
[
    (0.59153223605591049412844857432 - 1.68745927346801253993135357636*i
        0.762806324458047168681080323846571478727 - 0.60368764497868211035115379488*i),
    (-0.59153223605591049412844857432 - 1.68745927346801253993135357636*i
        -0.762806324458047168681080323846571478727 - 0.60368764497868211035115379488*i),
    (-1.354338560513957662809528899804 - 1.0837716284893304295801997808568748714097*i
        -0.59153223605591049412844857401 + 0.480083983510648319229045987467*i),
    (-1.52561264891609433736216065099 0.342548176804273349105263499648)
]
```

9.5.4 Every abelian variety is a quotient of a Jacobian

Over an infinite field, *every* abelin variety can be obtained as a quotient of a Jacobian variety. The modular abelian varieties that we will encounter later are, by definition, exactly the quotients of the Jacobian $J_1(N)$ of $X_1(N)$ for some N. In this section we see that merely being a quotient of a Jacobian does not endow an abelian variety with any special properties.

Theorem 9.5.4 (Matsusaka). *Let A be an abelian variety over an algebraically closed field. Then there is a Jacobian J and a surjective map $J \to A$.*

This was originally proved in *On a generating curve of an abelian variety*, Nat. Sc. Rep. Ochanomizu Univ. **3** (1952), 1–4. Here is the Math Review by P. Samuel:

> An abelian variety A is said to be generated by a variety V (and a mapping f of V into A) if A is the group generated by $f(V)$. It is proved that every abelian variety A may be generated by a curve defined over the algebraic closure of def(A). A first lemma shows that, if a variety V is the carrier of an algebraic system $(X(M))_{M \in U}$ of curves ($X(M)$ being defined, non-singular and disjoint from the singular bunch of V for almost all M in the parametrizing variety U) if this system has a simple base point on V, and if a mapping f of V into an abelian variety is constant on some $X(M_0)$, then f is a constant; this is proved by specializing on M_0 a generic point M of U and by using specializations of cycles [Matsusaka, Mem. Coll. Sci. Kyoto Univ. Ser. A. Math. 26, 167–173 (1951); these Rev. 13, 379]. Another lemma notices that, for a normal projective variety V, a suitable linear family of plane sections of V may be taken as a family $(X(M))$. Then the main result follows from the complete reducibility theorem. This result is said to be the basic tool for generalizing Chow's theorem ("the Jacobian variety of a curve defined over k is an abelian projective variety defined over k").

Milne [Mil86, §10] proves the theorem under the weaker hypothesis that the base field is infinite. We briefly sketch his proof now. If $\dim A = 1$, then A is the Jacobian of itself, so we may assume $\dim A > 1$. Embed A into $\mathbf{P}^n$, then, using the Bertini theorem, cut $A \subset \mathbf{P}^n$ by hyperplane sections $\dim(A) - 1$ times to obtain a nonsingular curve C on A of the form $A \cap V$, where V is a linear subspace of $\mathbf{P}^n$. Using standard arguments from Hartshorne [Har77], Milne shows (Lemma 10.3) that if W is a nonsingular variety and $\pi : W \to A$ is a finite morphism, then $\pi^{-1}(C)$ is geometrically connected (the main point is that the pullback of an ample invertible sheaf by a finite morphism is ample). (A morphism $f : X \to Y$ is *finite* if for every open affine subset $U = \mathrm{Spec}(R) \subset Y$, the inverse image $f^{-1}(U) \subset X$ is an affine open subset $\mathrm{Spec}(B)$ with B a finitely generated R-**module**. Finite morphisms have finite fibers, but not conversely.) We assume this lemma and deduce the theorem.

Let J be the Jacobian of C; by the universal property of Jacobians there is a unique homomorphism $f : J \to A$ coming from the inclusion $C \hookrightarrow A$. The image $A_1 = f(J)$ is an abelian subvariety since images of homomorphisms of abelian varieties are abelian varieties. By the Poincare reducibility theorem (we only proved this over $\mathbf{C}$, but it is true in general), there is an abelian subvariety $A_2 \subset A$ such that $A_1 + A_2 = A$ and $A_1 \cap A_2$ is finite. The isogeny $g : A_1 \times A_2 \to A$ given by $g(x, y) = x + y \in A$ is a finite morphism (any isogeny of abelian varieties is finite, flat, and surjective by Section 8 of [Mil86]). The inverse image $g^{-1}(A_1)$ is a union of $\#(A_1 \cap A_2)$ irreducible components; if this intersection is nontrivial, then likewise $g^{-1}(C)$ is reducible, which is a contradiction. This does not complete the proof, since it is possible that g is an isomorphism, so we use one additional trick. Suppose n is a positive integer coprime to the residue characteristic, and let

$$h = 1 \times [n] : A_1 \times A_2 \to A_1 \times A_2$$

be the identity map on the first factor and multiplication by n on the second. Then h is finite and $(h \circ g)^{-1}(A_1)$ is a union of $n^{2\dim A_2} = \deg(h)$ irreducible components, hence $(h \circ g)^{-1}(C)$ is reducible, a contradiction.

Question 9.5.5. Is Theorem 9.5.4 false for some abelian variety A over some finite field k?

Question 9.5.6 (Milne). Using the theorem we can obtain a sequence of Jacobian varieties $J_1, J_2, \ldots$ that form a complex

$$\cdots \to J_2 \to J_1 \to A \to 0.$$

(In each case the image of J_{i+1} is the connected component of the kernel of $J_i \to J_{i-1}$.) Is it possible to make this construction in such a way that the sequence terminates in 0?

Question 9.5.7 (Yau). Let A be an abelian variety. What can be said about the minimum of the dimensions of all Jacobians J such that there is a surjective morphism $J \to A$?

Remark 9.5.8. Brian Conrad has explained to the author that if A is an abelian variety over an infinite field, then A can be embedded in a Jacobian J. This does not follow directly from Theorem 9.5.4 above, since if $J \twoheadrightarrow A^\vee$, then the dual map $A \to J$ need not be injective.

9.6 Néron Models

The main references for Néron models are as follows:

1. [AEC2]: Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves.* Chapter IV of this book contains an extremely well written and motivated discussion of Néron models of elliptic curves over Dedekind domains with perfect residue field. In particular, Silverman gives an almost complete construction of Néron models of elliptic curves. Silverman very clearly really wants his reader to understand the construction. Highly recommended.

2. [BLR]: Bosch, Lütkebohmert, Raynaud, *Néron Models.* This is an excellent and accessible book that contains a complete construction of Néron models and some of their generalizations, a discussion of their functorial properties, and a sketch of the construction of Jacobians of families of curves. The goal of this book was to redo in scheme-theoretic language Néron original paper, which is written in a language that was ill-adapted to the subtleties of Néron models.

3. Artin, *Néron Models*, in Cornell-Silverman. This is the first-ever exposition of Néron's original paper in the language of schemes.

9.6.1 What are Néron Models?

Suppose E is an elliptic curve over $\mathbf{Q}$. If Δ is the minimal discriminant of E, then E has good reduction at p for all $p \nmid \Delta$, in the sense that E extends to an abelian scheme $\mathcal{E}$ over $\mathbf{Z}_p$ (i.e., a "smooth" and "proper" group scheme). One can not ask for E to extend to an abelian scheme over $\mathbf{Z}_p$ for all $p \mid \Delta$. One can, however, ask whether there is a notion of "good" model for E at these bad primes. To quote [BLR, page 1], "It came as a surprise for arithmeticians and algebraic geometers when A. Néron, relaxing the condition of properness and concentrating on the group structure and the smoothness, discovered in the years 1961–1963 that such models exist in a canonical way."

Before formally defining Néron models, we describe what it means for a morphism $f : X \to Y$ of schemes to be smooth. A morphism $f : X \to Y$ is finite type if for every open affine $U = \mathrm{Spec}(R) \subset Y$ there is a finite covering of $f^{-1}(U)$ by open affines $\mathrm{Spec}(S)$, such that each S is a finitely generated R-algebra.

Definition 9.6.1. A morphism $f : X \to Y$ is *smooth at $x \in X$* if it is of finite type and there are open affine neighborhoods $\mathrm{Spec}(A) \subset X$ of x and $\mathrm{Spec}(R) \subset Y$ of $f(x)$ such that

$$A \cong R[t_1, \ldots, t_{n+r}]/(f_1, \ldots, f_n)$$

for elements $f_1, \ldots, f_n \in R[t_1, \ldots, t_{n+r}]$ and all $n \times n$ minors of the Jacobian matrix $(\partial f_i/\partial t_j)$ generate the unit ideal of A. The morphism f is *étale* at x if, in addition, $r = 0$. A morphism is *smooth of relative dimension d* if it is smooth at x for every $x \in X$ and $r = d$ in the isomorphism above.

Smooth morphisms behave well. For example, if f and g are smooth and $f \circ g$ is defined, then $f \circ g$ is automatically smooth. Also, smooth morphisms are closed under base extension: if $f : X \to Y$ is a smooth morphism over S, and S' is a scheme over S, then the induced map $X \times_S S' \to Y \times_S S'$ is smooth. (If you've never seen products of schemes, it might be helpful to know that $\mathrm{Spec}(A) \times \mathrm{Spec}(B) = \mathrm{Spec}(A \otimes B)$. Read [Har77, §II.3] for more information about fiber products, which provide a geometric way to think about tensor products. Also, we often write $X_{S'}$ as shorthand for $X \times_S S'$.)

We are now ready for the definition. Suppose R is a Dedekind domain with field of fractions K (e.g., $R = \mathbf{Z}$ and $K = \mathbf{Q}$).

Definition 9.6.2 (Néron model). Let A be an abelian variety over K. The *Néron model* $\mathcal{A}$ of A is a smooth commutative group scheme over R such that for any smooth morphism $S \to R$ the natural map of abelian groups

$$\mathrm{Hom}_R(S, \mathcal{A}) \to \mathrm{Hom}_K(S \times_R K, A)$$

is a bijection. This is called the Néron mapping property: In more compact notation, it says that there is an isomorphism $\mathcal{A}(S) \cong A(S_K)$.

Taking $S = \mathcal{A}$ in the definition we see that $\mathcal{A}$ is unique, up to a unique isomorphism.

It is a deep theorem that Néron models exist. Fortunately, Bosch, Lütkebohmert, and Raynaud devoted much time to create a carefully written book [BLR90] that explains the construction in modern language. Also, in the case of elliptic curves, Silverman's second book [Sil94] is extremely helpful.

The basic idea of the construction is to first observe that if we can construct a Néron model at each localization $R_{\mathfrak{p}}$ at a nonzero prime ideal of R, then each of these local models can be glued to obtain a global Néron model (this uses that there are only finitely many primes of bad reduction). Thus we may assume that R is a discrete valuation ring.

The next step is to pass to the "strict henselization" R' of R. A local ring R with maximal ideal $\wp$ is henselian if "every simple root lifts uniquely"; more precisely, if whenever $f(x) \in R[x]$ and $\alpha \in R$ is such that $f(\alpha) \equiv 0 \pmod{\wp}$ and $f'(\alpha) \not\equiv 0 \pmod{\wp}$, there is a unique element $\tilde{\alpha} \in R$ such that $\tilde{\alpha} \equiv \alpha \pmod{\wp}$ and $f(\tilde{\alpha}) = 0$. The strict henselization of a discrete valuation ring R is an extension of R that is henselian and for which the residue field of R' is the separable closure of the residue field of R (when the residue field is finite, the separable close is just the algebraic closure). The strict henselization is not too much bigger than R, though it is typically not finitely generated over R. It is, however, much smaller than the completion of R (e.g., $\mathbf{Z}_p$ is uncountable). The main geometric property of a strictly henselian ring R with residue field k is that if X is a smooth scheme over R, then the reduction map $X(R) \to X(k)$ is surjective.

Working over the strict henselization, we first resolve singularities. Then we use a generalization of the theorem that Weil used to construct Jacobians to pass from a birational group law to an actual group law. We thus obtain the Néron model over the strict henselization of R. Finally, we use Grothendieck's faithfully flat descent to obtain a Néron model over R.

When A is the Jacobian of a curve X, there is an alternative approach that involves the "minimal proper regular model" of X. For example, when A is an elliptic curve, it is the Jacobian of itself, and the Néron model can be constructed in terms of the minimal proper regular model $\mathcal{X}$ of A as follows. In general, the model $\mathcal{X} \to R$ is not also smooth. Let $\mathcal{X}'$ be the smooth locus of $\mathcal{X} \to R$, which is obtained by removing from each closed fiber $\mathcal{X}_{\mathbf{F}_p} = \sum n_i C_i$ all irreducible components with multiplicity $n_i \geq 2$ and all singular points on each C_i, and all points where at least two C_i intersect each other. Then the group structure on A extends to a group structure on $\mathcal{X}'$, and $\mathcal{X}'$ equipped with this group structure is the Néron model of A.

Explicit determination of the possibilities for the minimal proper regular model of an elliptic curve was carried out by Kodaira, then Néron, and finally in a very explicit form by Tate. Tate codified a way to find the model in what's called "Tate's Algorithm" (see Antwerp IV, which is available on my web page: `http://modular.fas.harvard.edu/scans/antwerp/`, and look at Silverman, chapter IV, which also has important implementation advice).

9.6.2 The Birch and Swinnerton-Dyer Conjecture and Néron Models

Throughout this section, let A be an abelian variety over $\mathbf{Q}$ and let $\mathcal{A}$ be the corresponding Néron model over $\mathbf{Z}$. We work over $\mathbf{Q}$ for simplicity, but could work over any number field.

Let $L(A, s)$ be the Hasse-Weil L-function of A (see Section [to be written]). Let $r = \mathrm{ord}_{s=1} L(A, s)$ be the analytic rank of A. The Birch and Swinnerton-Dyer Conjecture asserts that $A(\mathbf{Q}) \approx \mathbf{Z}^r \oplus A(\mathbf{Q})_{\mathrm{tor}}$ and

$$\frac{L^{(r)}(A,1)}{r!} = \frac{(\prod c_p) \cdot \Omega_A \cdot \mathrm{Reg}_A \cdot \#\text{Ш}(A)}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^{\vee}(\mathbf{Q})_{\mathrm{tor}}}.$$

We have not defined most of the quantities appearing in this formula. In this section, we will define the Tamagawa numbers c_p, the real volume Ω_A, and the Shafarevich-Tate group $Ш(A)$ in terms of the Néron model $\mathcal{A}$ of A.

We first define the Tamagawa numbers c_p, which are the orders groups of connected components. Let p be a prime and consider the closed fiber $\mathcal{A}_{\mathbf{F}_p}$, which is a smooth commutative group scheme over $\mathbf{F}_p$. Then $\mathcal{A}_{\mathbf{F}_p}$ is a disjoint union of one or more connected components. The connected component $\mathcal{A}^0_{\mathbf{F}_p}$ that contains the identity element is a subgroup of $\mathcal{A}_{\mathbf{F}_p}$ (Intuition: the group law is continuous and the continuous image of a connected set is connected, so the group structure restricts to $\mathcal{A}^0_{\mathbf{F}_p}$).

Definition 9.6.3 (Component Group). The *component group* of A at p is

$$\Phi_{A,p} = \mathcal{A}_{\mathbf{F}_p}/\mathcal{A}^0_{\mathbf{F}_p}.$$

Fact: The component group $\Phi_{A,p}$ is a finite flat group scheme over $\mathbf{F}_p$, and for all but finitely many primes p, we have $\Phi_{A,p} = 0$.

Definition 9.6.4 (Tamagawa Numbers). The *Tamagawa number* of A at a prime p is

$$c_p = \#\Phi_{A,p}(\mathbf{F}_p).$$

Next we define the real volume Ω_A. Choose a basis

$$\omega_1, \ldots, \omega_d \in \mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}})$$

for the global differential 1-forms on $\mathcal{A}$, where $d = \dim A$. The wedge product $w = \omega_1 \wedge \omega_2 \wedge \cdots \wedge \omega_d$ is a global d-form on $\mathcal{A}$. Then w induces a differential d-form on the real Lie group $A(\mathbf{R})$.

Definition 9.6.5 (Real Volume). The *real volume* of A is

$$\Omega_A = \left| \int_{A(\mathbf{R})} w \right| \in \mathbf{R}_{>0}.$$

Finally, we give a definition of the Shafarevich-Tate group in terms of the Néron model. Let $\mathcal{A}_0$ be the scheme obtained from the Néron model $\mathcal{A}$ over A by removing from each closed fiber all nonidentity components. Then $\mathcal{A}_0$ is again a smooth commutative group scheme, but it need not have the Néron mapping property.

Recall that an étale morphism is a morphism that is smooth of relative dimension 0. A sheaf of abelian groups on the étale site $\mathbf{Z}_{\text{ét}}$ is a functor (satisfying certain axioms) from the category of étale morphism $X \to \mathbf{Z}$ to the category of abelian groups. There are enough sheaves on $\mathbf{Z}_{\text{ét}}$ so that there is a cohomology theory for such sheaves, which is called étale cohomology. In particular if $\mathcal{F}$ is a sheaf on $\mathbf{Z}_{\text{ét}}$, then for every integer q there is an abelian group $\mathrm{H}^q(\mathbf{Z}_{\text{ét}}, \mathcal{F})$ associated to $\mathcal{F}$ that has the standard properties of a cohomology functor.

The group schemes $\mathcal{A}_0$ and $\mathcal{A}$ both determine sheaves on the étale site, which we will also denote by $\mathcal{A}_0$ and $\mathcal{A}$.

Definition 9.6.6 (Shafarevich-Tate Group). Suppose $A(\mathbf{R})$ is connected that that $\mathcal{A}_0 = \mathcal{A}$. Then the *Shafarevich-Tate* group of A is $\mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A})$. More generally, suppose only that $A(\mathbf{R})$ is connected. Then the Shafarevich-Tate group is the image of the natural map

$$f : \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}_0) \to \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}).$$

Even more generally, if $A(\mathbf{R})$ is not connected, then there is a natural map $r : \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}) \to \mathrm{H}^1(\mathrm{Gal}(\mathbf{C}/\mathbf{R}), A(\mathbf{C}))$ and $Ш(A) = \mathrm{im}(f) \cap \ker(r)$.

Mazur proves in the appendix to [Maz72] that this definition is equivalent to the usual Galois cohomology definition. To do this, he considers the exact sequence $0 \to \mathcal{A}_0 \to \mathcal{A} \to \Phi_A \to 0$, where Φ_A is a sheaf version of $\oplus_p \Phi_{A,p}$. The main input is Lang's Theorem, which implies that over a local field, unramified Galois cohomology is the same as the cohomology of the corresponding component group.

Conjecture 9.6.7 (Shafarevich-Tate). *The group* $\mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A})$ *is finite.*

When A has rank 0, all component groups $\Phi_{A,p}$ are trivial, $A(\mathbf{R})$ is connected, and $A(\mathbf{Q})_{\mathrm{tor}}$ and $A^\vee(\mathbf{Q})_{\mathrm{tor}}$ are trivial, the Birch and Swinnerton-Dyer conjecture takes the simple form

$$\frac{L(A,1)}{\Omega_A} = \# \mathrm{H}^1(\mathbf{Z}_{\text{ét}}, \mathcal{A}).$$

Later, when A is modular, we will (almost) interpret $L(A,1)/\Omega_A$ as the order of a certain group that involves modular symbols. Thus the BSD conjecture asserts that two groups have the same order; however, they are not isomorphic, since, e.g., when $\dim A = 1$ the modular symbols group is always cyclic, but the Shafarevich-Tate group is never cyclic (unless it is trivial).

9.6.3 Functorial Properties of Neron Models

The definition of Néron model is functorial, so one might expect the formation of Néron models to have good functorial properties. Unfortunately, it doesn't.

Proposition 9.6.8. *Let A and B be abelian varieties. If $\mathcal{A}$ and $\mathcal{B}$ are the Néron models of A and B, respectively, then the Néron model of $A \times B$ is $\mathcal{A} \times \mathcal{B}$.*

Suppose $R \subset R'$ is a finite extension of discrete valuation rings with fields of fractions $K \subset K'$. Sometimes, given an abelian variety A over a field K, it is easier to understand properties of the abelian variety, such as reduction, over K'. For example, you might have extra information that implies that $A_{K'}$ decomposes as a product of well-understood abelian varieties. It would thus be useful if the Néron model of $A_{K'}$ were simply the base extension $\mathcal{A}_{R'}$ of the Néron model of A over R. This is, however, frequently not the case.

Distinguishing various types of ramification will be useful in explaining how Néron models behave with respect to base change, so we now recall the notions of tame and wild ramification. If π generates the maximal ideal of R and v' is the valuation on R', then the extension is *unramified* if $v'(\pi) = 1$. It is *tamely ramified* if $v'(\pi)$ is not divisible by the residue characteristic of R, and it is *wildly ramified* if $v'(\pi)$ is divisible by the residue characteristic of R. For example, the extension $\mathbf{Q}_p(p^{1/p})$ of $\mathbf{Q}_p$ is wildly ramified.

Example 9.6.9. If R is the ring of integers of a p-adic field, then for every integer n there is a unique unramified extension of R of degree n. See [Cp86, §I.7], where Fröhlich uses Hensel's lemma to show that the unramified extensions of $K =$

Frac(R) are in bijection with the finite (separable) extensions of the residue class field.

The Néron model does not behave well with respect to base change, except in some special cases. For example, suppose A is an abelian variety over the field of fractions K of a discrete valuation ring R. If K' is the field of fractions of a finite unramified extension R' of R, then the Néron model of $A_{K'}$ is $\mathcal{A}_{R'}$, where $\mathcal{A}$ is the Néron model of A over R. Thus the Néron model over an unramified extension is obtained by base extending the Néron model over the base. This is not too surprising because in the construction of Néron model we first passed to the strict henselization of R, which is a limit of unramified extensions.

Continuing with the above notation, if K' is tamely ramified over K, then in general $\mathcal{A}_{R'}$ need *not* be the Néron model of $A_{K'}$. Assume that K' is Galois over K. In [Edi92], Bas Edixhoven describes the Néron model of A_K in terms of $\mathcal{A}_{R'}$. To describe his main theorem, we introduce the restriction of scalars of a scheme.

Definition 9.6.10 (Restriction of Scalars). Let $S' \to S$ be a morphism of schemes and let X' be a scheme over S'. Consider the functor

$$\mathcal{R}(T) = \mathrm{Hom}_{S'}(T \times_S S', X')$$

on the category of all schemes T over S. If this functor is representable, the representing object $X = \mathrm{Res}_{S'/S}(X')$ is called the *restriction of scalars* of X' to S.

Edixhoven's main theorem is that if G is the Galois group of K' over K and $X = \mathrm{Res}_{R'/R}(\mathcal{A}_{R'})$ is the restriction of scalars of $\mathcal{A}_{R'}$ down to R, then there is a natural map $\mathcal{A} \to X$ whose image is the closed subscheme X^G of fixed elements.

We finish this section with some cautious remarks about exactness properties of Néron models. If $0 \to A \to B \to C \to 0$ is an exact sequence of abelian varieties, then the functorial definition of Néron models produces a complex of Néron models

$$0 \to \mathcal{A} \to \mathcal{B} \to \mathcal{C} \to 0,$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are the Néron models of A, B, and C, respectively. This complex can fail to be exact at every point. For an in-depth discussion of conditions when we have exactness, along with examples that violate exactness, see [BLR90, Ch. 7], which says: "we will see that, except for quite special cases, there will be a defect of exactness, the defect of right exactness being much more serious than the one of left exactness."

To give examples in which right exactness fails, it suffices to give an optimal quotient $B \to C$ such that for some p the induced map $\Phi_{B,p} \to \Phi_{C,p}$ on component groups is not surjective (recall that optimal means $A = \ker(B \to C)$ is an abelian variety). Such quotients, with B and C modular, arise naturally in the context of Ribet's level optimization. For example, the elliptic curve E given by $y^2 + xy = x^3 + x^2 - 11x$ is the optimal new quotient of the Jacobian $J_0(33)$ of $X_0(33)$. The component group of E at 3 has order 6, since E has semistable reduction at 3 (since $3 \,||\, 33$) and $\mathrm{ord}_3(j(E)) = -6$. The image of the component group of $J_0(33)$ in the component group of E has order 2:

```
> OrderOfImageOfComponentGroupOfJON(ModularSymbols("33A"),3);
2
```

Note that the modular form associated to E is congruent modulo 3 to the form corresponding to $J_0(11)$, which illustrates the connection with level optimization.

10

Abelian Varieties Attached to Modular Forms

In this chapter we describe how to decompose $J_1(N)$, up to isogeny, as a product of abelian subvarieties A_f corresponding to Galois conjugacy classes of cusp forms f of weight 2. This was first accomplished by Shimura (see [Shi94, Theorem 7.14]). We also discuss properties of the Galois representation attached to f.

In this chapter we will work almost exclusively with $J_1(N)$. However, everything goes through exactly as below with $J_1(N)$ replaced by $J_0(N)$ and $S_2(\Gamma_1(N))$ replaced by $S_2(\Gamma_0(N))$. Since, $J_1(N)$ has dimension much larger than $J_0(N)$, so for computational investigations it is frequently better to work with $J_0(N)$.

See Brian Conrad's appendix to [ribet-stein: Lectures on Serre's Conjectures] for a much more extensive exposition of the construction discussed below, which is geared toward preparing the reader for Deligne's more general construction of Galois representations associated to newforms of weight $k \geq 2$ (for that, see Conrad's book ...).

10.1 Decomposition of the Hecke Algebra

Let N be a positive integer and let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(J_1(N))$$

be the algebra of all Hecke operators acting on $J_1(N)$. Recall from Section 7.4 that the anemic Hecke algebra is the subalgebra

$$\mathbf{T}_0 = \mathbf{Z}[\ldots, T_n, \ldots : (n, N) = 1] \subset \mathbf{T}$$

of $\mathbf{T}$ obtained by adjoining to $\mathbf{Z}$ only those Hecke operators T_n with n relatively prime to N.

Remark 10.1.1. Viewed as $\mathbf{Z}$-modules, $\mathbf{T}_0$ need not be saturated in $\mathbf{T}$, i.e., $\mathbf{T}/\mathbf{T}_0$ need not be torsion free. For example, if $\mathbf{T}$ is the Hecke algebra associated to

$S_2(\Gamma_1(24))$ then $\mathbf{T}/\mathbf{T}_0 \cong \mathbf{Z}/2\mathbf{Z}$. Also, if $\mathbf{T}$ is the Hecke algebra associated to $S_2(\Gamma_0(54))$, then $\mathbf{T}/\mathbf{T}_0 \cong \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}$.

If $f = \sum a_n q^n$ is a newform, then the field $K_f = \mathbf{Q}(a_1, a_2, \ldots)$ has finite degree over $\mathbf{Q}$, since the a_n are the eigenvalues of a family of commuting operators with integral characteristic polynomials. The *Galois conjugates* of f are the newforms $\sigma(f) = \sum \sigma(a_n) q^n$, for $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. There are $[K_f : \mathbf{Q}]$ Galois conjugates of f.

As in Section 7.4, we have a canonical decomposition

$$\mathbf{T}_0 \otimes \mathbf{Q} \cong \prod_f K_f, \tag{10.1.1}$$

where f varies over a set of representatives for the Galois conjugacy classes of newforms in $S_2(\Gamma_1(N))$ of level dividing N. For each f, let

$$\pi_f = (0, \ldots, 0, 1, 0, \ldots, 0) \in \prod K_f$$

be projection onto the factor K_f of the product (10.1.1). Since $\mathbf{T}_0 \subset \mathbf{T}$, and $\mathbf{T}$ has no additive torsion, we have $\mathbf{T}_0 \otimes \mathbf{Q} \subset \mathbf{T} \otimes \mathbf{Q}$, so these projectors π_f lie in $\mathbf{T}_\mathbf{Q} = \mathbf{T} \otimes \mathbf{Q}$. Since $\mathbf{T}_\mathbf{Q}$ is commutative and the π_f are mutually orthogonal idempotents whose sum is $(1, 1, \ldots, 1)$, we see that $\mathbf{T}_\mathbf{Q}$ breaks up as a product of algebras

$$\mathbf{T}_\mathbf{Q} \cong \prod_f L_f, \qquad t \mapsto \sum_f t\pi_f.$$

10.1.1 The Dimension of L_f

Proposition 10.1.2. *If f, L_f and K_f are as above, then $\dim_{K_f} L_f$ is the number of divisors of N/N_f where N_f is the level of the newform f.*

Proof. Let V_f be the complex vector space spanned by all images of Galois conjugates of f via all maps α_d with $d \mid N/N_f$. It follows from [Atkin-Lehner-Li theory – multiplicity one] that the images via α_d of the Galois conjugates of f are linearly independent. (Details: More generally, if f and g are newforms of level M, then by Proposition 7.2.1, $B(f) = \{\alpha_d(f) : d \mid N/N_f\}$ is a linearly independent set and likewise for $B(g)$. Suppose some nonzero element f' of the span of $B(f)$ equals some element g' of the span of $B(g)$. Since T_p, for $p \nmid N$, commutes with α_d, we have $T_p(f') = a_p(f)f'$ and $T_p(g') = a_p(g)g'$, so $0 = T_p(0) = T_p(f' - g') = a_p(f)f' - a_p(g)g'$. Since $f' = g'$, this implies that $a_p(f) = a_p(g)$. Because a newform is determined by the eigenvalues of T_p for $p \nmid N$, it follows that $f = g$.) Thus the $\mathbf{C}$-dimension of V_f is the number of divisors of N/N_f times $\dim_\mathbf{Q} K_f$.

The factor L_f is isomorphic to the image of $\mathbf{T}_\mathbf{Q} \subset \mathrm{End}(S_k(\Gamma_1(N)))$ in $\mathrm{End}(V_f)$. As in Section **??**, there is a single element $v \in V_f$ so that $V_f = \mathbf{T}_\mathbf{C} \cdot v$. Thus the image of $\mathbf{T}_\mathbf{Q}$ in $\mathrm{End}(V_f)$ has dimension $\dim_\mathbf{C} V_f$, and the result follows. □

Let's examine a particular case of this proposition. Suppose p is a prime and $f = \sum a_n q^n$ is a newform of level N_f coprime to p, and let $N = p \cdot N_f$. We will show that

$$L_f = K_f[U]/(U^2 - a_p U + p), \tag{10.1.2}$$

hence $\dim_{K_f} L_f = 2$ which, as expected, is the number of divisors of $N/N_f = p$. The first step is to view L_f as the space of operators generated by the Hecke operators T_n acting on the span V of the images $f(dz) = f(q^d)$ for $d \mid (N/N_f) = p$. If $n \neq p$, then T_n acts on V as the scalar a_n, and when $n = p$, the Hecke operator T_p acts on $S_k(\Gamma_1(p \cdot N_f))$ as the operator also denoted U_p. By Section 7.2, we know that U_p corresponds to the matrix $\left(\begin{smallmatrix} a_p & 1 \\ -p & 0 \end{smallmatrix}\right)$ with respect to the basis $f(q), f(q^p)$ of V. Thus U_p satisfies the relation $U_p^2 - a_p U + p$. Since U_p is not a scalar matrix, this minimal polynomial of U_p is quadratic, which proves (10.1.2).

More generally, see [DDT94, Lem. 4.4] (Diamond-Darmon-Taylor) for an explicit presentation of L_f as a quotient

$$L_f \cong K_f[\ldots, U_p, \ldots]/I$$

where I is an ideal and the U_p correspond to the prime divisors of N/N_f.

10.2 Decomposition of $J_1(N)$

Let f be a newform in $S_2(\Gamma_1(N))$ of level a divisor M of N, so $f \in S_2(\Gamma_1(M))_{\text{new}}$ is a normalized eigenform for all the Hecke operators of level M. We associate to f an abelian subvariety A_f of $J_1(N)$, of dimension $[L_f : \mathbf{Q}]$, as follows. Recall that π_f is the fth projector in $\mathbf{T}_0 \otimes \mathbf{Q} = \prod_g K_g$. We can not define A_f to be the image of $J_1(N)$ under π_f, since π_f is only, a priori, an element of $\mathrm{End}(J_1(N)) \otimes \mathbf{Q}$. Fortunately, there exists a positive integer n such that $n\pi_f \in \mathrm{End}(J_1(N))$, and we let

$$A_f = n\pi_f(J_1(N)).$$

This is independent of the choice of n, since the choices for n are all multiples of the "denominator" n_0 of π_f, and if A is any abelian variety and n is a positive integer, then $nA = A$.

The natural map $\prod_f A_f \to J_1(N)$, which is induced by summing the inclusion maps, is an isogeny. Also A_f is simple if f is of level N, and otherwise A_f is isogenous to a power of $A'_f \subset J_1(N_f)$. Thus we obtain an isogeny decomposition of $J_1(N)$ as a product of $\mathbf{Q}$-simple abelian varieties.

Remark 10.2.1. The abelian varieties A_f frequently decompose further over $\overline{\mathbf{Q}}$, i.e., they are not absolutely simple, and it is an interesting problem to determine an isogeny decomposition of $J_1(N)_{\overline{\mathbf{Q}}}$ as a product of simple abelian varieties. It is still not known precisely how to do this computationally for any particular N.

This decomposition can be viewed in another way over the complex numbers. As a complex torus, $J_1(N)(\mathbf{C})$ has the following model:

$$J_1(N)(\mathbf{C}) = \mathrm{Hom}(S_2(\Gamma_1(N)), \mathbf{C})/H_1(X_1(N), \mathbf{Z}).$$

The action of the Hecke algebra $\mathbf{T}$ on $J_1(N)(\mathbf{C})$ is compatible with its action on the cotangent space $S_2(\Gamma_1(N))$. This construction presents $J_1(N)(\mathbf{C})$ naturally as $V/\mathcal{L}$ with V a complex vector space and $\mathcal{L}$ a lattice in V. The anemic Hecke algebra $\mathbf{T}_0$ then decomposes V as a direct sum $V = \bigoplus_f V_f$. The Hecke operators act on V_f and $\mathcal{L}$ in a compatible way, so $\mathbf{T}_0$ decomposes $\mathcal{L} \otimes \mathbf{Q}$ in a compatible way. Thus $\mathcal{L}_f = V_f \cap \mathcal{L}$ is a lattice in V_f, so we may $A_f(\mathbf{C})$ view as the complex torus $V_f/\mathcal{L}_f$.

Lemma 10.2.2. *Let $f \in S_2(\Gamma_1(N))$ be a newform of level dividing N and $A_f = n\pi_f(J_1(N))$ be the corresponding abelian subvariety of $J_1(N)$. Then the Hecke algebra $\mathbf{T} \subset \mathrm{End}(J_1(N))$ leaves A_f invariant.*

Proof. The Hecke algebra $\mathbf{T}$ is commutative, so if $t \in \mathbf{T}$, then

$$tA_f = tn\pi_f(J_1(N)) = n\pi_f(tJ_1(N)) \subset n\pi_f(J_1(N)) = A_f.$$

□

Remark 10.2.3. Viewing $A_f(\mathbf{C})$ as $V_f/\mathcal{L}_f$ is extremely useful computationally, since $\mathcal{L}$ can be computed using modular symbols, and $\mathcal{L}_f$ can be cut out using the Hecke operators. For example, if f and g are nonconjugate newforms of level dividing N, we can explicitly compute the group structure of $A_f \cap A_g \subset J_1(N)$ by doing a computation with modular symbols in $\mathcal{L}$. More precisely, we have

$$A_f \cap A_g \cong (\mathcal{L}/(\mathcal{L}_f + \mathcal{L}_g))_{\mathrm{tor}}.$$

Note that A_f depends on viewing f as an element of $S_2(\Gamma_1(N))$ for some N. Thus it would be more accurate to denote A_f by $A_{f,N}$, where N is any multiple of the level of f, and to reserve the notation A_f for the case $N = 1$. Then $\dim A_{f,N}$ is $\dim A_f$ times the number of divisors of N/N_f.

10.2.1 Aside: Intersections and Congruences

Suppose f and g are not Galois conjugate. Then the intersection $\Psi = A_f \cap A_g$ is finite, since $V_f \cap V_g = 0$, and the integer $\#\Psi$ is of interest. This cardinality is related to congruence between f and g, but the exact relation is unclear. For example, one might expect that $p \mid \#\Psi$ if and only if there is a prime $\wp$ of the compositum $K_f.K_g$ of residue characteristic p such that $a_q(f) \equiv a_q(g) \pmod{\wp}$ for all $q \nmid N$. If $p \mid \#\Psi$, then such a prime $\wp$ exists (take $\wp$ to be induced by a maximal ideal in the support of the nonzero $\mathbf{T}$-module $\Psi[p]$). The converse is frequently true, but is sometimes false. For example, if N is the prime 431 and

$$\begin{aligned} f &= q - q^2 + q^3 - q^4 + q^5 - q^6 - 2q^7 + \cdots \\ g &= q - q^2 + 3q^3 - q^4 - 3q^5 - 3q^6 + 2q^7 + \cdots , \end{aligned}$$

then $f \equiv g \pmod 2$, but $A_f \cap A_g = 0$. This example implies that “multiplicity one fails” for level 431 and $p = 2$, so the Hecke algebra associated to $J_0(431)$ is not Gorenstein (see [Lloyd Kilford paper] for more details).

10.3 Galois Representations Attached to A_f

It is important to emphasize the case when f is a newform of level N, since then A_f is $\mathbf{Q}$-simple and there is a compatible family of 2-dimensional ℓ-adic representations attached to f, which arise from torsion points on A_f.

Proposition 10.1.2 implies that $L_f = K_f$. Fix such an f, let $A = A_f$, let $K = K_f$, and let

$$d = \dim A = \dim_{\mathbf{Q}} K = [K : \mathbf{Q}].$$

Let ℓ be a prime and consider the $\mathbf{Q}_\ell$-adic Tate module $\mathrm{Tate}_\ell(A)$ of A:

$$\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes \varprojlim_{\nu>0} A[\ell^\nu].$$

Note that as a $\mathbf{Q}_\ell$-vector space $\mathrm{Tate}_\ell(A) \cong \mathbf{Q}_\ell^{2d}$, since $A[n] \cong (\mathbf{Z}/n\mathbf{Z})^{2d}$, as groups.

There is a natural action of the ring $K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$ on $\mathrm{Tate}_\ell(A)$. By algebraic number theory

$$K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell = \prod_{\lambda \mid \ell} K_\lambda,$$

where λ runs through the primes of the ring $\mathcal{O}_K$ of integers of K lying over ℓ and K_λ denotes the completion of K with respect to the absolute value induced by λ. Thus $\mathrm{Tate}_\ell(A)$ decomposes as a product

$$\mathrm{Tate}_\ell(A) = \prod_{\lambda \mid \ell} \mathrm{Tate}_\lambda(A)$$

where $\mathrm{Tate}_\lambda(A)$ is a K_λ vector space.

Lemma 10.3.1. *Let the notation be as above. Then for all λ lying over ℓ,*

$$\dim_{K_\lambda} \mathrm{Tate}_\lambda(A) = 2.$$

Proof. Write $A = V/\mathcal{L}$, with $V = V_f$ a complex vector space and $\mathcal{L}$ a lattice. Then $\mathrm{Tate}_\lambda(A) \cong \mathcal{L} \otimes \mathbf{Q}_\ell$ as K_λ-modules (not as $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-modules!), since $A[\ell^n] \cong \mathcal{L}/\ell^n\mathcal{L}$, and $\varprojlim_n \mathcal{L}/\ell^n\mathcal{L} \cong \mathbf{Z}_\ell \otimes \mathcal{L}$. Also, $\mathcal{L} \otimes \mathbf{Q}$ is a vector space over K, which must have dimension 2, since $\mathcal{L} \otimes \mathbf{Q}$ has dimension $2d = 2 \dim A$ and K has degree d. Thus

$$\mathrm{Tate}_\lambda(A) \cong \mathcal{L} \otimes K_\lambda \approx (K \oplus K) \otimes_K K_\lambda \cong K_\lambda \oplus K_\lambda$$

has dimension 2 over K_λ. □

Now consider $\mathrm{Tate}_\lambda(A)$, which is a K_λ-vector space of dimension 2. The Hecke operators are defined over $\mathbf{Q}$, so $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on $\mathrm{Tate}_\ell(A)$ in a way compatible with the action of $K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$. We thus obtain a homomorphism

$$\rho_\ell = \rho_{f,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{K \otimes \mathbf{Q}_\ell} \mathrm{Tate}_\ell(A) \approx \mathrm{GL}_2(K \otimes \mathbf{Q}_\ell) \cong \prod_\lambda \mathrm{GL}_2(K_\lambda).$$

Thus ρ_ℓ is the direct sum of ℓ-adic Galois representations ρ_λ where

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{End}_{K_\lambda}(\mathrm{Tate}_\lambda(A))$$

gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathrm{Tate}_\lambda(A)$.

If $p \nmid \ell N$, then ρ_λ is unramified at p (see [ST68, Thm. 1]). In this case it makes sense to consider $\rho_\lambda(\varphi_p)$, where $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a Frobenius element at p. Then $\rho_\lambda(\varphi_p)$ has a well-defined trace and determinant, or equivalently, a well-defined characteristic polynomial $\Phi(X) \in K_\lambda[X]$.

Theorem 10.3.2. *Let $f \in S_2(\Gamma_1(N), \varepsilon)$ be a newform of level N with Dirichlet character ε. Suppose $p \nmid \ell N$, and let $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be a Frobenius element at p. Let $\Phi(X)$ be the characteristic polynomial of $\rho_\lambda(\varphi_p)$. Then*

$$\Phi(X) = X^2 - a_p X + p \cdot \varepsilon(p),$$

where a_p is the pth coefficient of the modular form f (thus a_p is the image of T_p in E_f and $\varepsilon(p)$ is the image of $\langle p \rangle$).

Let $\varphi = \varphi_p$. By the Cayley-Hamilton theorem

$$\rho_\lambda(\varphi)^2 - \mathrm{tr}(\rho_\lambda(\varphi))\rho_\lambda(\varphi) + \det(\rho_\lambda(\varphi)) = 0.$$

Using the Eichler-Shimura congruence relation (see) we will show that $\mathrm{tr}(\rho_\lambda(\varphi)) = a_p$, but we defer the proof of this until ....

We will prove that $\det(\rho_\lambda(\varphi)) = p$ in the special case when $\varepsilon = 1$. This will follow from the equality

$$\det(\rho_\lambda) = \chi_\ell, \tag{10.3.1}$$

where χ_ℓ is the ℓth cyclotomic character

$$\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^* \subset K_\lambda^*,$$

which gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on μ_{ℓ^∞}. We have $\chi_\ell(\varphi) = p$ because φ induces induces pth powering map on μ_{ℓ^∞}.

It remains to establish (10.3.1). The simplest case is when A is an elliptic curve. In [Sil92,], Silverman shows that $\det(\rho_\ell) = \chi_\ell$ using the Weil pairing. We will consider the Weil pairing in more generality in the next section, and use it to establish (10.3.1).

10.3.1 The Weil Pairing

Let $T_\ell(A) = \varprojlim_{n\geq 1} A[\ell^n]$, so $\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes T_\ell(A)$. The Weil pairing is a non-degenerate perfect pairing

$$e_\ell : T_\ell(A) \times T_\ell(A^\vee) \to \mathbf{Z}_\ell(1).$$

(See e.g., [Mil86, §16] for a summary of some of its main properties.)

Remark 10.3.3. Identify $\mathbf{Z}/\ell^n\mathbf{Z}$ with μ_{ℓ^n} by $1 \mapsto e^{-2\pi i/\ell^n}$, and extend to a map $\mathbf{Z}_\ell \to \mathbf{Z}_\ell(1)$. If $J = \mathrm{Jac}(X)$ is a Jacobian, then the Weil pairing on J is induced by the canonical isomorphism

$$T_\ell(J) \cong \mathrm{H}^1(X, \mathbf{Z}_\ell) = \mathrm{H}^1(X, \mathbf{Z}) \otimes \mathbf{Z}_\ell,$$

and the cup product pairing

$$\mathrm{H}^1(X, \mathbf{Z}_\ell) \otimes_{\mathbf{Z}_\ell} \mathrm{H}^1(X, \mathbf{Z}_\ell) \xrightarrow{\cup} \mathbf{Z}_\ell.$$

For more details see the discussion on pages 210–211 of Conrad's appendix to [RS01], and the references therein. In particular, note that $\mathrm{H}^1(X, \mathbf{Z}_\ell)$ is isomorphic to $\mathrm{H}_1(X, \mathbf{Z}_\ell)$, because $\mathrm{H}_1(X, \mathbf{Z}_\ell)$ is self-dual because of the intersection pairing. It is easy to see that $\mathrm{H}_1(X, \mathbf{Z}_\ell) \cong T_\ell(J)$ since by Abel-Jacobi $J \cong T_0(J)/\mathrm{H}_1(X, \mathbf{Z})$, where $T_0(J)$ is the tangent space at J at 0 (see Lemma 10.3.1).

Here $\mathbf{Z}_\ell(1) \cong \varprojlim \mu_{\ell^n}$ is isomorphic to $\mathbf{Z}_\ell$ as a ring, but has the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ induced by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\varprojlim \mu_{\ell^n}$. Given $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, there is an element $\chi_\ell(\sigma) \in \mathbf{Z}_\ell^*$ such that $\sigma(\zeta) = \zeta^{\chi_\ell(\sigma)}$, for every ℓ^nth root of unity ζ. If we view $\mathbf{Z}_\ell(1)$ as just $\mathbf{Z}_\ell$ with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then the action of $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathbf{Z}_\ell(1)$ is left multiplication by $\chi_\ell(\sigma) \in \mathbf{Z}_\ell^*$.

Definition 10.3.4 (Cyclotomic Character). The homomorphism

$$\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^*$$

is called the *ℓ-adic cyclotomic character.*

If $\varphi : A \to A^\vee$ is a polarization (so it is an isogeny defined by translation of an ample invertible sheaf), we define a pairing

$$e_\ell^\varphi : T_\ell(A) \times T_\ell(A) \to \mathbf{Z}_\ell(1) \tag{10.3.2}$$

by $e_\ell^\varphi(a,b) = e_\ell(a, \varphi(b))$. The pairing (10.3.2) is a skew-symmetric, nondegenerate, bilinear pairing that is $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-equivariant, in the sense that if $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then

$$e_\ell^\varphi(\sigma(a), \sigma(b)) = \sigma \cdot e_\ell^\varphi(a,b) = \chi_\ell(\sigma) e_\ell^\varphi(a,b).$$

We now apply the Weil pairing in the special case $A = A_f \subset J_1(N)$. Abelian varieties attached to modular forms are equipped with a canonical polarization called the *modular polarization.* The canonical principal polarization of $J_1(N)$ is an isomorphism $J_1(N) \xrightarrow{\sim} J_1(N)^\vee$, so we obtain the modular polarization $\varphi = \varphi_A : A \to A^\vee$ of A, as illustrated in the following diagram:

$$\begin{tikzcd}
J_1(N) \arrow[r, "\text{autoduality}\,\cong"] & J_1(N)^\vee \arrow[d] \\
A \arrow[u] \arrow[r, "\text{polarization}\,\varphi_A"] & A^\vee
\end{tikzcd}$$

Consider (10.3.2) with $\varphi = \varphi_A$ the modular polarization. Tensoring over $\mathbf{Q}$ and restricting to $\mathrm{Tate}_\lambda(A)$, we obtain a nondegenerate skew-symmetric bilinear pairing

$$e : \mathrm{Tate}_\lambda(A) \times \mathrm{Tate}_\lambda(A) \to \mathbf{Q}_\ell(1). \tag{10.3.3}$$

The nondegeneracy follows from the nondegeneracy of e_ℓ^φ and the observation that

$$e_\ell^\varphi(\mathrm{Tate}_\lambda(A), \mathrm{Tate}_{\lambda'}(A)) = 0$$

when $\lambda \neq \lambda'$. This uses the Galois equivariance of e_ℓ^ϕ carries over to Galois equivariance of e, in the following sense. If $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and $x, y \in \mathrm{Tate}_\lambda(A)$, then

$$e(\sigma x, \sigma y) = \sigma e(x,y) = \chi_\ell(\sigma) e(x,y).$$

Note that σ acts on $\mathbf{Q}_\ell(1)$ as multiplication by $\chi_\ell(\sigma)$.

10.3.2 The Determinant

There are two proofs of the theorem, a fancy proof and a concrete proof. We first present the fancy proof. The pairing e of (10.3.3) is a skew-symmetric and bilinear form so it determines a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-equivarient homomorphism

$$\bigwedge_{K_\lambda}^{2} \mathrm{Tate}_\lambda(A) \to \mathbf{Q}_\ell(1). \tag{10.3.4}$$

It is not *a priori* true that we can take the wedge product over K_λ instead of $\mathbf{Q}_\ell$, but we can because $e(tx, y) = e(x, ty)$ for any $t \in K_\lambda$. This is where we use that A is attached to a newform with trivial character, since when the character is nontrivial, the relation between $e(T_p x, y)$ and $e(x, T_p y)$ will involve $\langle p \rangle$. Let $D = \bigwedge^2 \mathrm{Tate}_\lambda(A)$ and note that $\dim_{K_\lambda} D = 1$, since $\mathrm{Tate}_\lambda(A)$ has dimension 2 over K_λ.

There is a canonical isomorphism

$$\mathrm{Hom}_{\mathbf{Q}_\ell}(D, \mathbf{Q}_\ell(1)) \cong \mathrm{Hom}_{K_\lambda}(D, K_\lambda(1)),$$

and the map of (10.3.4) maps to an isomorphism $D \cong K_\lambda(1)$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-modules. Since the representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on D is the determinant, and the representation on $K_\lambda(1)$ is the cyclotomic character χ_ℓ, it follows that $\det \rho_\lambda = \chi_\ell$.

Next we consider a concrete proof. If $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then we must show that $\det(\sigma) = \chi_\ell(\sigma)$. Choose a basis $x, y \in \mathrm{Tate}_\lambda(A)$ of $\mathrm{Tate}_\lambda(A)$ as a 2 dimensional K_λ vector space. We have $\sigma(x) = ax + cy$ and $\sigma(y) = bx + dy$, for $a, b, c, d \in K_\lambda$. Then

$$\begin{aligned}
\chi_\ell(\sigma)e(x, y) &= \langle \sigma x, \sigma y\rangle \\
&= e(ax + cy, bx + dy) \\
&= e(ax, bx) + e(ax, dy) + e(cy, bx) + e(cy, dy) \\
&= e(ax, dy) + e(cy, bx) \\
&= e(adx, y) - e(bcx, y) \\
&= e((ad - bc)x, y) \\
&= (ad - bc)e(x, y)
\end{aligned}$$

To see that $e(ax, bx) = 0$, note that

$$e(ax, bx) = e(abx, x) = -e(x, abx) = -e(ax, bx).$$

Finally, since e is nondegenerate, there exists x, y such that $e(x, y) \neq 0$, so $\chi_\ell(\sigma) = ad - bc = \det(\sigma)$.

10.4 Remarks About the Modular Polarization

Let A and φ be as in Section 10.3.1. The degree $\deg(\varphi)$ of the modular polarization of A is an interesting arithmetic invariant of A. If $B \subset J_1(N)$ is the sum of all modular abelian varieties A_g attached to newforms $g \in S_2(\Gamma_1(N))$, with g not a Galois conjugate of f and of level dividing N, then $\ker(\varphi) \cong A \cap B$, as illustrated in the following diagram:

$$\begin{tikzcd}
& \ker(\varphi_B) \arrow[d, "\cong"] \arrow[dr] & & \\
\ker(\varphi_A i) \arrow[r, "\cong"] \arrow[dr] & A \cap B \arrow[r] \arrow[d] & B \arrow[d] \arrow[dr] & \\
& A \arrow[r] \arrow[dr, "\varphi"] & J_1(N) \arrow[r] \arrow[d] & B^\vee \\
& & A^\vee &
\end{tikzcd}$$

Note that $\ker(\varphi_B)$ is also isomorphic to $A \cap B$, as indicated in the diagram.

In connection with Section **??**, the quantity $\ker(\varphi_A) = A \cap B$ is closely related to congruences between f and eigenforms orthogonal to the Galois conjugates of f.

When A has dimension 1, we may alternatively view A as a quotient of $X_1(N)$ via the map

$$X_1(N) \to J_1(N) \to A^\vee \cong A.$$

Then $\varphi_A : A \to A$ is pullback of divisors to $X_1(N)$ followed by push forward, which is multiplication by the degree. Thus $\varphi_A = [n]$, where n is the degree of the morphism $X_1(N) \to A$ of algebraic curves. The *modular degree* is

$$\deg(X_1(N) \to A) = \sqrt{\deg(\varphi_A)}.$$

More generally, if A has dimension greater than 1, then $\deg(\varphi_A)$ has order a perfect square (for references, see [Mil86, Thm. 13.3]), and we define the *modular degree* to be $\sqrt{\deg(\varphi_A)}$.

Let f be a newform of level N. In the spirit of Section 10.2.1 we use congruences to define a number related to the modular degree, called the congruence number. For a subspace $V \subset S_2(\Gamma_1(N))$, let $V(\mathbf{Z}) = V \cap \mathbf{Z}[[q]]$ be the elements with integral q-expansion at ∞ and $V^\perp$ denotes the orthogonal complement of V with respect to the Petersson inner product. The *congruence number* of f is

$$r_f = \# \frac{S_2(\Gamma_1(N))(\mathbf{Z})}{V_f(\mathbf{Z}) + V_f^\perp(\mathbf{Z})},$$

where V_f is the complex vector space spanned by the Galois conjugates of f. We thus have two positive associated to f, the congruence number r_f and the modular degree m_f of of A_f.

Theorem 10.4.1. $m_f \mid r_f$

Ribet mentions this in the case of elliptic curves in [ZAGIER, 1985] [Zag85a], but the statement is given incorrectly in that paper (the paper says that $r_f \mid m_f$, which is wrong). The proof for dimension greater than one is in [AGASHE-STEIN, Manin constant...]. Ribet also subsequently proved that if $p^2 \nmid N$, then $\mathrm{ord}_p(m_f) = \mathrm{ord}_p(r_f)$.

We can make the same definitions with $J_1(N)$ replaced by $J_0(N)$, so if $f \in S_2(\Gamma_0(N))$ is a newform, $A_f \subset J_0(N)$, and the congruence number measures congruences between f and other forms in $S_2(\Gamma_0(N))$. In [FM99, Ques. 4.4], they ask whether it is always the case that $m_f = r_f$ when A_f is an elliptic curve, and m_f and r_f are defined relative to $\Gamma_0(N)$. I implemented an algorithm in MAGMA to compute r_f, and found the first few counterexamples, which occur when

$$N = 54, 64, 72, 80, 88, 92, 96, 99, 108, 120, 124, 126, 128, 135, 144.$$

For example, the elliptic curve A labeled 54B1 in [Cre97] has $r_A = 6$ and $m_A = 2$. To see directly that $3 \mid r_A$, observe that if f is the newform corresponding to E and g is the newform corresponding to $X_0(27)$, then $g(q) + g(q^2)$ is congruent to f modulo 3. This is consistent with Ribet's theorem that if $p \mid r_A/m_A$ then $p^2 \mid N$. There seems to be no absolute bound on the p that occur.

It would be interesting to determine the answer to the analogue of the question of Frey-Mueller for $\Gamma_1(N)$. For example, if $A \subset J_1(54)$ is the curve isogeneous to 54B1, then $m_A = 18$ is divisible by 3. However, I do not know r_A in this case, because I haven't written a program to compute it for $\Gamma_1(N)$.

11
Modularity of Abelian Varieties

11.1 Modularity Over $\mathbf{Q}$

Definition 11.1.1 (Modular Abelian Variety). Let A be an abelian variety over $\mathbf{Q}$. Then A is *modular* if there exists a positive integer N and a surjective map $J_1(N) \to A$ defined over $\mathbf{Q}$.

The following theorem is the culmination of a huge amount of work, which started with Wiles's successful attack [Wil95] on Fermat's Last Theorem, and finished with [BCDT01].

Theorem 11.1.2 (Breuil, Conrad, Diamond, Taylor, Wiles). *Let E be an elliptic curve over $\mathbf{Q}$. Then E is modular.*

We will say nothing about the proof here.

If A is an abelian variety over $\mathbf{Q}$, let $\mathrm{End}_{\mathbf{Q}}(A)$ denote the ring of endomorphisms of A that are defined over $\mathbf{Q}$.

Definition 11.1.3 (GL_2-type). An abelian variety A over $\mathbf{Q}$ is of GL_2-*type* if the endomorphism algebra $\mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$ contains a number field of degree equal to the dimension of A.

For example, every elliptic curve E over $\mathbf{Q}$ is trivially of GL_2-type, since $\mathbf{Q} \subset \mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(E)$.

Proposition 11.1.4. *If A is an abelian variety over $\mathbf{Q}$, and $K \subset \mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$ is a field, then $[K : \mathbf{Q}]$ divides $\dim A$.*

Proof. As discussed in [Rib92, §2], K acts faithfully on the tangent space $\mathrm{Tan}_0(A/\mathbf{Q})$ over $\mathbf{Q}$ to A at 0, which is a $\mathbf{Q}$ vector space of dimension $\dim(A)$. Thus $\mathrm{Tan}_0(A/\mathbf{Q})$ is a vector space over K, hence has $\mathbf{Q}$-dimension a multiple of $[K : \mathbf{Q}]$. □

Proposition 11.1.4 implies, in particular, that if E is an elliptic curve over $\mathbf{Q}$, then $\mathrm{End}_{\mathbf{Q}}(E) = \mathbf{Q}$. Recall that E *has CM* or is a *complex multiplication* elliptic curve if $\mathrm{End}_{\overline{\mathbf{Q}}}(E) \neq \mathbf{Z}$). Proposition 11.1.4 implies that if E is a CM elliptic curve, the extra endomorphisms are *never* defined over $\mathbf{Q}$.

Proposition 11.1.5. *Suppose $A = A_f \subset J_1(N)$ is an abelian variety attached to a newform of level N. Then A is of GL_2-type.*

Proof. The endomorphism ring of A_f contains $\mathcal{O}_f = \mathbf{Z}[\ldots, a_n(f), \ldots]$, hence the field $K_f = \mathbf{Q}(\ldots, a_n(f), \ldots)$ is contained in $\mathbf{Q} \otimes \mathrm{End}_{\mathbf{Q}}(A)$. Since $A_f = n\pi J_1(N)$, where π is a projector onto the factor K_f of the anemic Hecke algebra $\mathbf{T}_0 \otimes_{\mathbf{Z}} \mathbf{Q}$, we have $\dim A_f = [K_f : \mathbf{Q}]$. (One way to see this is to recall that the tangent space $T = \mathrm{Hom}(S_2(\Gamma_1(N)), \mathbf{C})$ to $J_1(N)$ at 0 is free of rank 1 over $\mathbf{T}_0 \otimes_{\mathbf{Z}} \mathbf{C}$.) □

Conjecture 11.1.6 (Ribet). *Every abelian variety over $\mathbf{Q}$ of GL_2-type is modular.*

Suppose

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

is an odd irreducible continuous Galois representation, where odd means that

$$\det(\rho(c)) = -1,$$

where c is complex conjugation. We say that ρ is *modular* if there is a newform $f \in S_k(\Gamma_1(N))$, and a prime ideal $\wp \subset \mathcal{O}_f$ such that for all $\ell \nmid Np$, we have

$$\begin{aligned} \mathrm{Tr}(\rho(\mathrm{Frob}_\ell)) &\equiv a_\ell \pmod{\wp}, \\ \mathrm{Det}(\rho(\mathrm{Frob}_\ell)) &\equiv \ell^{k-1} \cdot \varepsilon(\ell) \pmod{\wp}. \end{aligned}$$

Here χ_p is the p-adic cyclotomic character, and ε is the (Nebentypus) character of the newform f.

Conjecture 11.1.7 (Serre). *Every odd irreducible continuous representation*

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$$

is modular. Moreover, there is a formula for the "optimal" weight $k(\rho)$ and level $N(\rho)$ of a newform that gives rise to ρ.

In [Ser87], Serre describes the formula for the weight and level. Also, it is now known due to work of Ribet, Edixhoven, Coleman, Voloch, Gross, and others that if ρ is modular, then ρ arises from a form of the conjectured weight and level, except in some cases when $p = 2$. (For more details see the survey paper [RS01].) However, the full Conjecture 11.1.7 is known in very few cases.

Remark 11.1.8. There is interesting recent work of Richard Taylor which connects Conjecture 11.1.7 with the open question of whether every variety of a certain type has a point over a solvable extension of $\mathbf{Q}$. The question of the existence of solvable points ("solvability of varieties in radicals") seems very difficult. For example, we don't even know the answer for genus one curves, or have a good reason to make a conjecture either way (as far as I know). There's a book of Mike Fried that discusses this solvability question.

Serre's conjecture is very strong. For example, it would imply modularity of all abelian varieties over $\mathbf{Q}$ that could possibly be modular, and the proof of this implication does not rely on Theorem 11.1.2.

Theorem 11.1.9 (Ribet). *Serre's conjectures on modularity of all odd irreducible mod p Galois representations implies Conjecture 11.1.6.*

To give the reader a sense of the connection between Serre's conjecture and modularity, we sketch some of the key ideas of the proof of Theorem 11.1.9; for more details the reader may consult Sections 1–4 of [Rib92].

Without loss, we may assume that A is $\mathbf{Q}$-simple. As explained in the not trivial [Rib92, Thm. 2.1], this hypothesis implies that

$$K = \mathbf{Q} \otimes_{\mathbf{Z}} \mathrm{End}_{\mathbf{Q}}(A)$$

is a number field of degree $\dim(A)$. The Tate modules

$$\mathrm{Tate}_\ell(A) = \mathbf{Q}_\ell \otimes \varprojlim_{n\geq 1} A[\ell^n]$$

are free of rank two over $K \otimes \mathbf{Q}_\ell$, so the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $\mathrm{Tate}_\ell(A)$ defines a representation

$$\rho_{A,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(K \otimes \mathbf{Q}_\ell).$$

Remarks 11.1.10. That these representations take values in GL_2 is why such A are said to be "of GL_2-type". Also, note that the above applies to $A = A_f \subset J_1(N)$, and the ℓ-adic representations attached to f are just the factors of $\rho_{A,\ell}$ coming from the fact that $K \otimes \mathbf{Q}_\ell \cong \prod_{\lambda|\ell} K_\lambda$.

The deepest input to Ribet's proof is Faltings's isogeny theorem, which Faltings proved in order to prove Mordell's conjecture (there are only a finite number of L-rational points on any curve over L of genus at least 2).

If B is an abelian variety over $\mathbf{Q}$, let

$$L(B,s) = \prod_{\text{all primes } p} \frac{1}{\det\left(1 - p^{-s}\cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(A)\right)} = \prod_p L_p(B,s),$$

where ℓ is a prime of good reduction (it makes no difference which one).

Theorem 11.1.11 (Faltings). *Let A and B be abelian varieties. Then A is isogenous to B if and only if $L_p(A,s) = L_p(B,s)$ for almost all p.*

Using an analysis of Galois representations and properties of conductors and applying results of Faltings, Ribet finds an infinite set Λ of primes of K such that all $\rho_{A,\lambda}$ are irredudible and there only finitely many Serre invariants $N(\rho_{A,\lambda})$ and $k(\rho_{A,\lambda})$. For each of these λ, by Conjecture 11.1.7 there is a newform f_λ of level $N(\rho_{A,\lambda}))$ and weight $k(\rho_{A,\lambda})$ that gives rise to the mod ℓ representation $\rho_{A,\lambda}$. Since Λ is infinite, but there are only finitely many Serre invariants $N(\rho_{A,\lambda}))$, $k(\rho_{A,\lambda})$, there must be a single newform f and an infinite subset Λ' of Λ so that for every $\lambda \in \Lambda'$ the newform f gives rise to $\rho_{A,\lambda}$.

Let $B = A_f \subset J_1(N)$ be the abelian variety attached to f. Fix any prime p of good reduction. There are infinitely many primes $\lambda \in \Lambda'$ such that $\rho_{A,\lambda} \cong \rho_{B,\tilde{\lambda}}$ for some $\tilde{\lambda}$, and for these λ,

$$\det\left(1 - p^{-s}\cdot \mathrm{Frob}_p \,| A[\lambda]\right) = \det\left(1 - p^{-s}\cdot \mathrm{Frob}_p \,| B[\tilde{\lambda}]\right).$$

This means that the degree two polynomials in p^{-s} (over the appropriate fields, e.g., $K \otimes \mathbf{Q}_\ell$ for A)

$$\det\left(1 - p^{-s}\cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(A)\right)$$

and

$$\det\left(1 - p^{-s} \cdot \mathrm{Frob}_p \,|\, \mathrm{Tate}_\ell(B)\right)$$

are congruent modulo infinitely many primes. Therefore they are equal. By Theorem 11.1.11, it follows that A is isogenous to $B = A_f$, so A is modular.

11.2 Modularity of Elliptic Curves over $\overline{\mathbf{Q}}$

Definition 11.2.1 (Modular Elliptic Curve). An elliptic curve E over $\overline{\mathbf{Q}}$ is *modular* if there is a surjective morphism $X_1(N) \to E$ for some N.

Definition 11.2.2 (Q-curve). An elliptic curve E over **Q**-bar is a **Q**-*curve* if for every $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ there is an isogeny $E^\sigma \to E$ (over $\overline{\mathbf{Q}}$).

Theorem 11.2.3 (Ribet). *Let E be an elliptic curve over $\overline{\mathbf{Q}}$. If E is modular, then E is a* **Q**-*curve, or E has CM.*

This theorem is proved in [Rib92, §5].

Conjecture 11.2.4 (Ribet). *Let E be an elliptic curve over $\overline{\mathbf{Q}}$. If E is a* **Q**-*curve, then E is modular.*

In [Rib92, §6], Ribet proves that Conjecture 11.1.7 implies Conjecture 11.2.4. He does this by showing that if a **Q**-curve E does not have CM then there is a **Q**-simple abelian variety A over **Q** of GL_2-type such that E is a simple factor of A over $\overline{\mathbf{Q}}$. This is accomplished finding a model for E over a Galois extension K of **Q**, restricting scalars down to **Q** to obtain an abelian variety $B = \mathrm{Res}_{K/\mathbf{Q}}(E)$, and using Galois cohomology computations (mainly in H^2's) to find the required A of GL_2-type inside B. Then Theorem 11.1.9 and our assumption that Conjecture 11.1.7 is true together immediately imply that A is modular.

Ellenberg and Skinner [ES00] have recently used methods similar to those used by Wiles to prove strong theorems toward Conjecture 11.2.4. See also Ellenberg's survey [Ell02], which discusses earlier modularity results of Hasegawa, Hashimoto, Hida, Momose, and Shimura, and gives an example to show that there are infinitely many **Q**-curves whose modularity is not known.

Theorem 11.2.5 (Ellenberg, Skinner). *Let E be a* **Q**-*curve over a number field K with semistable reduction at all primes of K lying over 3, and suppose that K is unramified at 3. Then E is modular.*

12
L-functions

12.1 L-functions Attached to Modular Forms

Let $f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma_1(N))$ be a cusp form.

Definition 12.1.1 (L-series). The *L-series* of f is

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

Definition 12.1.2 (Λ-function). The *completed Λ function* of f is

$$\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s),$$

where

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

is the Γ function (so $\Gamma(n) = (n-1)!$ for positive integers n).

We can view $\Lambda(f,s)$ as a (Mellin) transform of f, in the following sense:

Proposition 12.1.3. *We have*

$$\Lambda(f,s) = N^{s/2}\int_0^\infty f(iy)y^s \frac{dy}{y},$$

and this integral converges for $\mathrm{Re}(s) > \frac{k}{2}+1$.

Proof. We have

$$\begin{aligned}\int_0^\infty f(iy)y^s\frac{dy}{y} &= \int_0^\infty \sum_{n=1}^\infty a_n e^{-2\pi n y} y^s \frac{dy}{y}\\ &= \sum_{n=1}^\infty a_n \int_0^\infty e^{-t}(2\pi n)^{-s} t^s \frac{dt}{t} \qquad (t = 2\pi n y)\\ &= (2\pi)^{-s}\Gamma(s)\sum_{n=1}^\infty \frac{a_n}{n^s}.\end{aligned}$$

To go from the first line to the second line, we reverse the summation and integration and perform the change of variables $t = 2\pi n y$. (We omit discussion of convergence.) □

12.1.1 Analytic Continuation and Functional Equation

We define the *Atkin-Lehner operator* W_N on $S_k(\Gamma_1(N))$ as follows. If $w_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$, then $[w_N^2]_k$ acts as $(-N)^{k-2}$, so if

$$W_N(f) = N^{1-\frac{k}{2}} \cdot f|[w_N]_k,$$

then $W_N^2 = (-1)^k$. (Note that W_N is an involution when k is even.) It is easy to check directly that if $\gamma \in \Gamma_1(N)$, then $w_N \gamma w_N^{-1} \in \Gamma_1(N)$, so W_N preserves $S_k(\Gamma_1(N))$. Note that in general W_N does *not* commute with the Hecke operators T_p, for $p \mid N$.

The following theorem is mainly due to Hecke (and maybe other people, at least in this generality). For a very general version of this theorem, see [Li75].

Theorem 12.1.4. *Suppose $f \in S_k(\Gamma_1(N), \chi)$ is a cusp form with character χ. Then $\Lambda(f, s)$ extends to an entire (holomorphic on all of $\mathbf{C}$) function which satisfies the functional equation*

$$\Lambda(f, s) = i^k \Lambda(W_N(f), k - s).$$

Since $N^{s/2}(2\pi)^{-s}\Gamma(s)$ is everywhere nonzero, Theorem 12.1.4 implies that $L(f, s)$ also extends to an entire function.

It follows from Definition 12.1.2 that $\Lambda(cf, s) = c\Lambda(f, s)$ for any $c \in \mathbf{C}$. Thus if f is a W_N-eigenform, so that $W_N(f) = wf$ for some $w \in \mathbf{C}$, then the functional equation becomes

$$\Lambda(f, s) = i^k w \Lambda(f, k - s).$$

If $k = 2$, then W_N is an involution, so $w = \pm 1$, and the sign in the functional equation is $\varepsilon(f) = i^k w = -w$, which is the negative of the sign of the Atkin-Lehner involution W_N on f. It is straightforward to show that $\varepsilon(f) = 1$ if and only if $\mathrm{ord}_{s=1} L(f, s)$ is even. Parity observations such as this are extremely useful when trying to understand the Birch and Swinnerton-Dyer conjecture.

Sketch of proof of Theorem 12.1.4 when $N = 1$. We follow [Kna92, §VIII.5] closely.

Note that since $w_1 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$, the condition $W_1(f) = f$ is satisfied for any $f \in S_k(1)$. This translates into the equality

$$f\left(-\frac{1}{z}\right) = z^k f(z). \qquad (12.1.1)$$

Write $z = x + iy$ with x and y real. Then (12.1.1) along the positive imaginary axis (so $z = iy$ with y positive real) is

$$f\left(\frac{i}{y}\right) = i^k y^k f(iy). \tag{12.1.2}$$

From Proposition 12.1.3 we have

$$\Lambda(f, s) = \int_0^\infty f(iy) y^{s-1} dy, \tag{12.1.3}$$

and this integral converges for $\mathrm{Re}(s) > \frac{k}{2} + 1$.

Again using growth estimates, one shows that

$$\int_1^\infty f(iy) y^{s-1} dy$$

converges for all $s \in \mathbf{C}$, and defines an entire function. Breaking the path in (12.1.3) at 1, we have for $\mathrm{Re}(s) > \frac{k}{2} + 1$ that

$$\Lambda(f, s) = \int_0^1 f(iy) y^{s-1} dy + \int_1^\infty f(iy) y^{s-1} dy.$$

Apply the change of variables $t = 1/y$ to the first term and use (12.1.2) to get

$$\begin{aligned}\int_0^1 f(iy) y^{s-1} dy &= \int_\infty^1 -f(i/t) t^{1-s} \frac{1}{t^2} dt \\ &= \int_1^\infty f(i/t) t^{-1-s} dt \\ &= \int_1^\infty i^k t^k f(it) t^{-1-s} dt \\ &= i^k \int_1^\infty f(it) t^{k-1-s} dt.\end{aligned}$$

Thus

$$\Lambda(f, s) = i^k \int_1^\infty f(it) t^{k-s-1} dt + \int_1^\infty f(iy) y^{s-1} dy.$$

The first term is just a translation of the second, so the first term extends to an entire function as well. Thus $\Lambda(f, s)$ extends to an entire function.

The proof of the general case for $\Gamma_0(N)$ is almost the same, except the path is broken at $1/\sqrt{N}$, since $i/\sqrt{N}$ is a fixed point for w_N. □

12.1.2 A Conjecture About Nonvanishing of $L(f, k/2)$

Suppose $f \in S_k(1)$ is an eigenform. If $k \equiv 2 \pmod 4$, then $L(f, k/2) = 0$ for reasons related to the discussion after the statement of Theorem 12.1.4. On the other hand, if $k \equiv 0 \pmod 4$, then $\mathrm{ord}_{s=k/2} L(f, k/2)$ is even, so $L(f, k/2)$ may or may not vanish.

Conjecture 12.1.5. *Suppose $k \equiv 0 \pmod 4$. Then $L(f, k/2) \neq 0$.*

According to [CF99], Conjecture 12.1.5 is true for weight k if there is some n such that the characteristic polynomial of T_n on $S_k(1)$ is irreducible. Thus Maeda's conjecture implies Conjecture 12.1.5. Put another way, if you find an f of level 1 and weight $k \equiv 0 \pmod 4$ such that $L(f, k/2) = 0$, then Maeda's conjecture is false for weight k.

Oddly enough, I personally find Conjecture 12.1.5 less convincing that Maeda's conjecture, despite it being a weaker conjecture.

12.1.3 Euler Products

Euler products make very clear how L-functions of eigenforms encode deep arithmetic information about representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Given a "compatible family" of ℓ-adic representations ρ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, one can define an Euler product $L(\rho, s)$, but in general it is very hard to say anything about the analytic properties of $L(\rho, s)$. However, as we saw above, when ρ is attached to a modular form, we know that $L(\rho, s)$ is entire.

Theorem 12.1.6. *Let $f = \sum a_n q^n$ be a newform in $S_k(\Gamma_1(N), \varepsilon)$, and let $L(f, s) = \sum_{n\geq 1} a_n n^{-s}$ be the associated Dirichlet series. Then $L(f, s)$ has an Euler product*

$$L(f,s) = \prod_{p\mid N} \frac{1}{1 - a_p p^{-s}} \cdot \prod_{p\nmid N} \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{k-1} p^{-2s}}.$$

Note that it is not really necessary to separate out the factors with $p \mid N$ as we have done, since $\varepsilon(p) = 0$ whenever $p \mid N$. Also, note that the denominators are of the form $F(p^{-s})$, where

$$F(X) = 1 - a_p X + \varepsilon(p) p^{k-1} X^2$$

is the reverse of the characteristic polynomial of Frob_p acting on any of the ℓ-adic representations attached to f, with $p \neq \ell$.

Recall that if p is a prime, then for every $r \geq 2$ the Hecke operators satisfy the relationship

$$T_{p^r} = T_{p^{r-1}} T_p - p^{k-1} \varepsilon(p) T_{p^{r-2}}. \qquad (12.1.4)$$

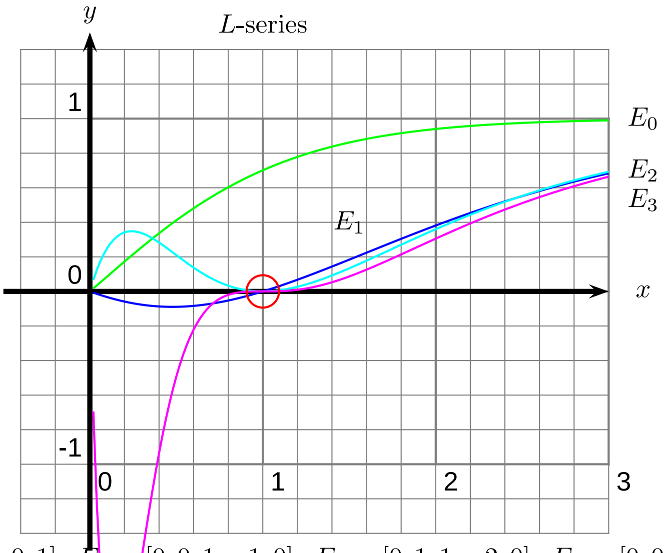
Lemma 12.1.7. *For every prime p we have the formal equality*

$$\sum_{r\geq 0} T_{p^r} X^r = \frac{1}{1 - T_p X + \varepsilon(p) p^{k-1} X^2}. \qquad (12.1.5)$$

Proof. Multiply both sides of (12.1.5) by $1 - T_p X + \varepsilon(p) p^{k-1} X^2$ to obtain the equation

$$\sum_{r\geq 0} T_{p^r} X^r - \sum_{r\geq 0} (T_{p^r} T_p) X^{r+1} + \sum_{r\geq 0} (\varepsilon(p) p^{k-1} T_{p^r}) X^{r+2} = 1.$$

This equation is true if and only if the lemma is true. Equality follows by checking the first few terms and shifting the index down by 1 for the second sum and down by 2 for the third sum, then using (12.1.4). □



$E_0 = [0,0,0,0,1]$, $E_1 = [0,0,1,-1,0]$, $E_2 = [0,1,1,-2,0]$, $E_3 = [0,0,1,-7,6]$
FIGURE 12.1.1. Graph of $L(E,s)$ for s real, for curves of ranks 0 to 3.

Note that $\varepsilon(p) = 0$ when $p \mid N$, so when $p \mid N$

$$\sum_{r\geq 0} T_{p^r} X^r = \frac{1}{1 - T_p X}.$$

Since the eigenvalues a_n of f also satisfy (12.1.4), we obtain each factor of the Euler product of Theorem 12.1.6 by substituting the a_n for the T_n and p^{-s} for X into (12.1.4). For $(n,m) = 1$, we have $a_{nm} = a_n a_m$, so

$$\sum_{n\geq 1} \frac{a_n}{n^s} = \prod_p \left(\sum_{r\geq 0} \frac{a_{p^r}}{p^{rs}} \right),$$

which gives the full Euler product for $L(f,s) = \sum a_n n^{-s}$.

12.1.4 Visualizing L-function

A. Shwayder did his Harvard junior project with me on visualizing L-functions of elliptic curves (or equivalently, of newforms $f = \sum a_n q^n \in S_2(\Gamma_0(N))$ with $a_n \in \mathbf{Z}$ for all n. The graphs in Figures 12.1.1–**??** of $L(E,s)$, for s real, and $|L(E,s)|$, for s complex, are from his paper.

13
The Birch and Swinnerton-Dyer Conjecture

This chapter is about the conjecture of Birch and Swinnerton-Dyer on the arithmetic of abelian varieties. We focus primarily on abelian varieties attached to modular forms.

In the 1960s, Sir Peter Swinnerton-Dyer worked with the EDSAC computer lab at Cambridge University, and developed an operating system that ran on that computer (so he told me once). He and Bryan Birch programmed EDSAC to compute various quantities associated to elliptic curves. They then formulated the conjectures in this chapter in the case of dimension 1 (see [Bir65, Bir71, SD67]). Tate formulated the conjectures in a functorial way for abelian varieties of arbitrary dimension over global fields in [Tat66], and proved that if the conjecture is true for an abelian variety A, then it is also true for each abelian variety isogenous to A.

Suitably interpreted, the conjectures may by viewed as generalizing the analytic class number formula, and Bloch and Kato generalized the conjectures to Grothendieck motives in [BK90].

13.1 The Rank Conjecture

Let A be an abelian variety over a number field K.

Definition 13.1.1 (Mordell-Weil Group). The *Mordell-Weil group* of A is the abelian group $AK)$ of all K-rational points on A.

Theorem 13.1.2 (Mordell-Weil). *The Mordell-Weil group $A(K)$ of A is finitely generated.*

The proof is nontrivial and combines two ideas. First, one proves the "weak Mordell-Weil theorem": for any integer m the quotient $A(K)/mA(K)$ is finite. This is proved by combining Galois cohomology techniques with standard finiteness theorems from algebraic number theory. The second idea is to introduce the Néron-

Tate canonical height $h : A(K) \to \mathbf{R}_{\geq 0}$ and use properties of h to deduce, from finiteness of $A(K)/mA(K)$, that $A(K)$ itself is finitely generated.

Definition 13.1.3 (Rank). By the structure theorem $A(K) \cong \mathbf{Z}^r \oplus G_{\text{tor}}$, where r is a nonnegative integer and G_{tor} is the torsion subgroup of G. The *rank* of A is r.

Let $f \in S_2(\Gamma_1(N))$ be a newform of level N, and let $A = A_f \subset J_1(N)$ be the corresponding abelian variety. Let $f_1, \ldots, f_d$ denote the $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates of f, so if $f = \sum a_n q^n$, then $f_i = \sum \sigma(a_n) q^n$, for some $\sigma \in \text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

Definition 13.1.4 (L-function of A). We define the L-function of $A = A_f$ (or any abelian variety isogenous to A) to be

$$L(A, s) = \prod_{i=1}^{d} L(f_i, s).$$

By Theorem 12.1.4, each $L(f_i, s)$ is an entire function on $\mathbf{C}$, so $L(A, s)$ is entire. In Section 13.4 we will discuss an intrinsic way to define $L(A, s)$ that does not require that A be attached to a modular form. However, in general we do not know that $L(A, s)$ is entire.

Conjecture 13.1.5 (Birch and Swinnerton-Dyer). *The rank of* $A(\mathbf{Q})$ *is equal to* $\text{ord}_{s=1} L(A, s)$.

One motivation for Conjecture 13.1.5 is the following *formal* observation. Assume for simplicity of notation that $\dim A = 1$. By Theorem 12.1.6, the L-function $L(A, s) = L(f, s)$ has an Euler product representation

$$L(A, s) = \prod_{p \mid N} \frac{1}{1 - a_p p^{-s}} \cdot \prod_{p \nmid N} \frac{1}{1 - a_p p^{-s} + p \cdot p^{-2s}},$$

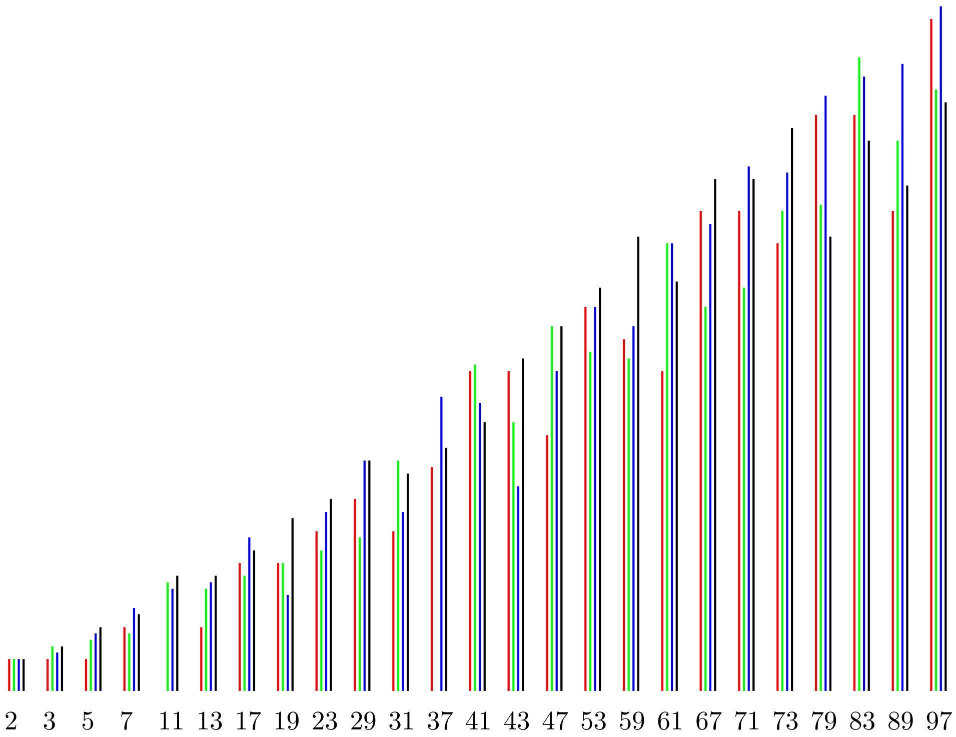
which is valid for $\text{Re}(s)$ sufficiently large. (Note that $\varepsilon = 1$, since A is a modular elliptic curve, hence a quotient of $X_0(N)$.) There is no loss in considering the product $L^*(A, s)$ over only the good primes $p \nmid N$, since $\text{ord}_{s=1} L(A, s) = \text{ord}_{s=1} L^*(A, s)$ (because $\prod_{p \mid N} \frac{1}{1 - a_p p^{-s}}$ is nonzero at $s = 1$). We then have *formally* that

$$\begin{aligned} L^*(A, 1) &= \prod_{p \nmid N} \frac{1}{1 - a_p p^{-1} + p^{-1}} \\ &= \prod_{p \nmid N} \frac{p}{p - a_p + 1} \\ &= \prod_{p \nmid N} \frac{p}{\#A(\mathbf{F}_p)} \end{aligned}$$

The intuition is that if the rank of A is large, i.e., $A(\mathbf{Q})$ is large, then each group $A(\mathbf{F}_p)$ will also be large since it has many points coming from reducing the elements of $A(\mathbf{Q})$ modulo p. It seems likely that if the groups $\#A(\mathbf{F}_p)$ are unusually large, then $L^*(A, 1) = 0$, and computational evidence suggests the more precise Conjecture 13.1.5.

Example 13.1.6. Let A_0 be the elliptic curve $y^2 + y = x^3 - x^2$, which has rank 0 and conductor 11, let A_1 be the elliptic curve $y^2 + y = x^3 - x$, which has rank 1 and conductor 37, let A_2 be the elliptic curve $y^2 + y = x^3 + x^2 - 2x$, which has rank 2 and conductor 389, and finally let A_3 be the elliptic curve $y^2 + y = x^3 - 7x + 6$, which has rank 3 and conductor 5077. By an exhaustive search, these are known to be the smallest-conductor elliptic curves of each rank. Conjecture 13.1.5 is known to be true for them, the most difficult being A_3, which relies on the results of [GZ86].

The following diagram illustrates $|\#A_i(\mathbf{F}_p)|$ for $p < 100$, for each of these curves. The height of the red line (first) above the prime p is $|\#A_0(\mathbf{F}_p)|$, the green line (second) gives the value for A_1, the blue line (third) for A_2, and the black line (fourth) for A_3. The intuition described above suggests that the clumps should look like triangles, with the first line shorter than the second, the second shorter than the third, and the third shorter than the fourth—however, this is visibly not the case. The large Mordell-Weil group over $\mathbf{Q}$ does not increase the size of every $E(\mathbf{F}_p)$ as much as we might at first suspect. Nonetheless, the first line is no longer than the last line for every p except $p = 41, 79, 83, 97$.



Remark 13.1.7. Suppose that $L(A, 1) \neq 0$. Then assuming the Riemann hypothesis for $L(A, s)$ (i.e., that $L(A, s) \neq 0$ for $\mathrm{Re}(s) > 1$), Goldfeld [Gol82] proved that the Euler product for $L(A, s)$, formally evaluated at 1, converges but *does not* converge to $L(A, 1)$. Instead, it converges (very slowly) to $L(A, 1)/\sqrt{2}$. For further details and insight into this strange behavior, see [Con03].

Remark 13.1.8. The Clay Math Institute has offered a one million dollar prize for a proof of Conjecture 13.1.5 for elliptic curves over $\mathbf{Q}$. See [Wil00].

Theorem 13.1.9 (Kolyvagin-Logachev). *Suppose $f \in S_2(\Gamma_0(N))$ is a newform such that $\mathrm{ord}_{s=1} L(f, s) \leq 1$. Then Conjecture 13.1.5 is true for A_f.*

Theorem 13.1.10 (Kato). *Suppose $f \in S_2(\Gamma_1(N))$ and $L(f,1) \neq 0$. Then Conjecture 13.1.5 is true for A_f.*

13.2 Refined Rank Zero Conjecture

Let $f \in S_2(\Gamma_1(N))$ be a newform of level N, and let $A_f \subset J_1(N)$ be the corresponding abelian variety.

The following conjecture refines Conjecture 13.1.5 in the case $L(A, 1) \neq 0$. We recall some of the notation below, where we give a formula for $L(A, 1)/\Omega_A$, which can be computed up to an vinteger, which we call the Manin index. Note that the definitions, results, and proofs in this section are all true exactly as stated with $X_1(N)$ replaced by $X_0(N)$, which is relevant if one wants to do computations.

Conjecture 13.2.1 (Birch and Swinnerton-Dyer). *Suppose* $L(A, 1) \neq 0$. *Then*

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

By Theorem 13.1.10, the group $\text{Ш}(A)$ is finite, so the right hand side makes sense. The right hand side is a rational number, so if Conjecture 13.2.1 is true, then the quotient $L(A, 1)/\Omega_A$ should also be a rational number. In fact, this is true, as we will prove below (see Theorem 13.2.11). Below we will discuss aspects of the proof of rationality in the case that A is an elliptic curve, and at the end of this section we give a proof of the general case.

In to more easily understanding $L(A, 1)/\Omega_A$, it will be easiest to work with $A = A_f^\vee$, where $A_f^\vee$ is the dual of A_f. We view A naturally as a quotient of $J_1(N)$ as follows. Dualizing the map $A_f \hookrightarrow J_1(N)$ we obtain a surjective map $J_1(N) \to A_f^\vee$. Passing to the dual doesn't affect whether or not $L(A, 1)/\Omega_A$ is rational, since changing A by an isogeny does not change $L(A, 1)$, and only changes Ω_A by multiplication by a nonzero rational number.

13.2.1 The Number of Real Components

Definition 13.2.2 (Real Components). Let c_∞ be the number of connected components of $A(\mathbf{R})$.

If A is an elliptic curve, then $c_\infty = 1$ or 2, depending on whether the graph of the affine part of $A(\mathbf{R})$ in the plane $\mathbf{R}^2$ is connected. For example, Figure 13.2.1 shows the real points of the elliptic curve defined by $y^2 = x^3 - x$ in the three affine patches that cover $\mathbf{P}^2$. The completed curve has two real components.

In general, there is a simple formula for c_∞ in terms of the action of complex conjugation on $\mathrm{H}_1(A(\mathbf{R}), \mathbf{Z})$, which can be computed using modular symbols. The formula is

$$\log_2(c_\infty) = \dim_{\mathbf{F}_2} A(\mathbf{R})[2] - \dim(A).$$

13.2.2 The Manin Index

The map $J_1(N) \to A$ induces a map $\mathcal{J} \to \mathcal{A}$ on Néron models. Pullback of differentials defines a map

$$\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}) \to \mathrm{H}^0(\mathcal{J}, \Omega^1_{\mathcal{J}/\mathbf{Z}}). \tag{13.2.1}$$

One can show that there is a q-expansion map

$$\mathrm{H}^0(\mathcal{J}, \Omega^1_{\mathcal{J}/\mathbf{Z}}) \to \mathbf{Z}[[q]] \tag{13.2.2}$$

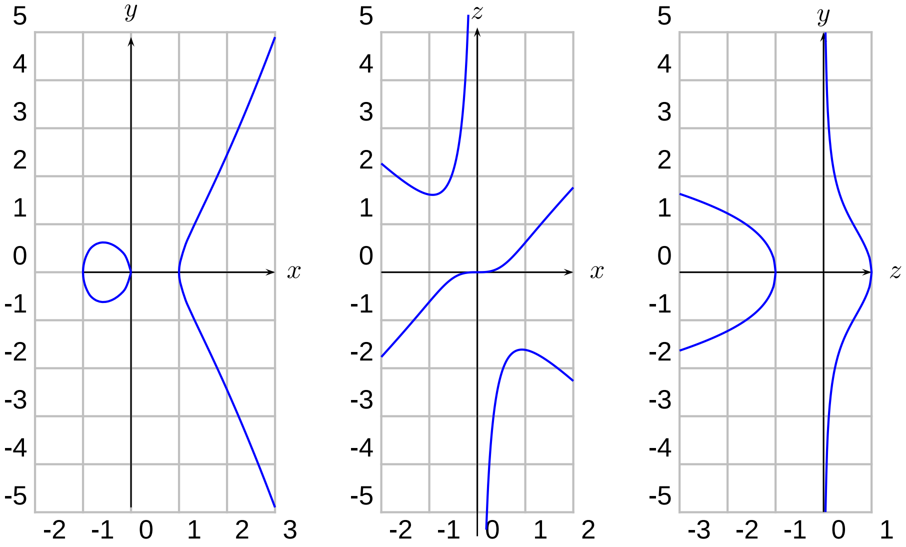


FIGURE 13.2.1. Graphs of real solutions to $y^2z = x^3 - xz^2$ on three affine patches

which agrees with the usual q-expansion map after tensoring with $\mathbf{C}$. (For us $X_1(N)$ is the curve that parameterizes pairs $(E, \mu_N \hookrightarrow E)$, so that there is a q-expansion map with values in $\mathbf{Z}[[q]]$.)

Let φ_A be the composition of (13.2.1) with (13.2.2).

Definition 13.2.3 (Manin Index). The *Manin index* c_A of A is the index of $\varphi_A(\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}))$ in its saturation. I.e., it is the order of the quotient group

$$\left(\frac{\mathbf{Z}[[q]]}{\varphi_A(\mathrm{H}^0(\mathcal{A}, \Omega^1_{\mathcal{A}/\mathbf{Z}}))}\right)_{\mathrm{tor}}.$$

Open Problem 13.2.4. Find an algorithm to compute c_A.

Manin conjectured that $c_A = 1$ when $\dim A = 1$, and I think $c_A = 1$ in general.

Conjecture 13.2.5 (Agashe, Stein). $c_A = 1$.

This conjecture is false if A is not required to be attached to a newform, even if $A_f \subset J_1(N)^{\mathrm{new}}$. For example, Adam Joyce, a student of Kevin Buzzard, found an $A \subset J_1(431)$ (and also $A' \subset J_0(431)$) whose Manin constant is 2. Here A is isogenous over $\mathbf{Q}$ to a product of two elliptic curves. Also, the Manin index for $J_0(33)$ (viewed as a quotient of $J_0(33)$) is divisible by 3, because there is a cusp form in $S_2(\Gamma_0(33))$ that has integer Fourier expansion at ∞, but not at one of the other cusps.

Theorem 13.2.6. *If $f \in S_2(\Gamma_0(N))$ then the Manin index c of $A_f^\vee$ can only divisible by 2 or primes whose square divides N. Moreover, if $4 \nmid N$, then $\mathrm{ord}_2(c) \leq \dim(A_f)$.*

The proof involves applying nontrivial theorems of Raynaud about exactness of sequences of differentials, then using a trick with the Atkin-Lehner involution, which was introduced by Mazur in [Maz78], and finally one applies the "q-expansion principle" in characteristic p to deduce the result (see [ASb]). Also, Edixhoven claims he can prove that if A_f is an elliptic curve then c_A is only divisible by 2, 3, 5, or 7. His argument use his semistable models for $X_0(p^2)$, but my understanding is that the details are not all written up.

13.2.3 The Real Volume Ω_A

Definition 13.2.7 (Real Volume). The *real volume* Ω_A of $A(\mathbf{R})$ is the volume of $A(\mathbf{R})$ with respect to a measure obtained by wedging together a basis for $\mathrm{H}^0(\mathcal{A}, \Omega^1)$.

If A is an elliptic curve with *minimal* Weierstrass equation

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

then one can show that

$$\omega = \frac{dx}{2y + a_1x + a_3} \tag{13.2.3}$$

is a basis for $\mathrm{H}^0(\mathcal{A}, \Omega^1)$. Thus

$$\Omega_A = \int_{A(\mathbf{R})} \frac{dx}{2y + a_1x + a_3}.$$

There is a fast algorithm for computing Ω_A, for A an elliptic curve, which relies on the quickly-convergent Gauss arithmetic-geometric mean (see [Cre97, §3.7]). For example, if A is the curve defined by $y^2 = x^3 - x$ (this is a minimal model), then

$$\Omega_A \sim 2 \times 2.62205755429211981046483958989.$$

For a general abelian variety A, it is an open problem to compute Ω_A. However, we can compute Ω_A/c_A, where c_A is the Manin index of A, by explicitly computing A as a complex torus using the period mapping Φ, which we define in the next section.

13.2.4 The Period Mapping

Let

$$\Phi : \mathrm{H}_1(X_1(N), \mathbf{Z}) \to \mathrm{Hom}_{\mathbf{C}}(\mathbf{C}f_1 + \cdots + \mathbf{C}f_d, \mathbf{C})$$

be the *period mapping* on integral homology induced by integrating homology classes on $X_0(N)$ against the $\mathbf{C}$-vector space spanned by the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-conjugates f_i of f. Extend Φ to $\mathrm{H}_1(X_1(N), \mathbf{Q})$ by $\mathbf{Q}$-linearity. We normalize Φ so that $\Phi(\{0, \infty\})(f) = L(f, 1)$. More explicitly, for $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, we have

$$\Phi(\{\alpha, \beta\})(f) = -2\pi i \int_\alpha^\beta f(z)dz.$$

The motivation for this normalization is that

$$L(f, 1) = -2\pi i \int_0^{i\infty} f(z)dz, \tag{13.2.4}$$

which we see immediately from the Mellin transform definition of $L(f, s)$:

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

13.2.5 Manin-Drinfeld Theorem

Recall the Manin-Drinfeld theorem, which we proved long ago, asserts that $\{0,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Q})$. We proved this by explicitly computing $(p+1-T_p)(\{0,\infty\})$, for $p \nmid N$, noting that the result is in $\mathrm{H}_1(X_0(N), \mathbf{Z})$, and inverting $p+1-T_p$. Thus there is an integer n such that $n\{0,\infty\} \in \mathrm{H}_1(X_0(N), \mathbf{Z})$.

Suppose that $A = A_f^\vee$ is an elliptic curve quotient of $J_0(N)$. Rewriting (13.2.4) in terms of Φ, we have $\Phi(\{0,\infty\}) = L(f,1)$. Let ω be a minimal differential on A, as in (13.2.3), so $\omega = -c_A \cdot 2\pi i f(z)dz$, where c_A is the Manin index of A, and the equality is after pulling ω back to $\mathrm{H}^0(X_0(N), \Omega) \cong S_2(\Gamma_0(N))$. Note that when we defined c_A, there was no factor of $2\pi i$, since we compared ω with $f(q)\frac{dq}{q}$, and $q = e^{2\pi i z}$, so $dq/q = 2\pi i dz$.

13.2.6 The Period Lattice

The *period lattice* of A with respect to a nonzero differential g on A is

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

and we have $A(\mathbf{C}) \cong \mathbf{C}/\mathcal{L}_g$. This is the Abel-Jacobi theorem, and the significance of g is that we are choosing a basis for the one-dimensional $\mathbf{C}$-vector space $\mathrm{Hom}(\mathrm{H}^0(A,\Omega), \mathbf{C})$, in order to embed the image of $\mathrm{H}_1(A,\mathbf{Z})$ in $\mathbf{C}$.

The integral $\int_{A(\mathbf{R})} g$ is "visible" in terms of the complex torus representation of $A(\mathbf{C}) = \mathbf{C}/\mathcal{L}_g$. More precisely, if $\mathcal{L}_g$ is not rectangular, then $A(\mathbf{R})$ may be identified with the part of the real line in a fundamental domain for $\mathcal{L}_g$, and $\int_{A(\mathbf{R})} g$ is the length of this segment of the real line. If $\mathcal{L}_g$ is rectangular, then it is that line along with another line above it that is midway to the top of the fundamental domain.

The real volume, which appears in Conjecture 13.2.1, is

$$\Omega_A = \int_{A(\mathbf{R})} \omega = -c_A \cdot 2\pi i \int_{A(\mathbf{R})} f.$$

Thus Ω_A is the least positive real number in $\mathcal{L}_\omega = -c_A \cdot 2\pi i \mathcal{L}_f$, when the period lattice is not rectangular, and twice the least positive real number when it is.

13.2.7 The Special Value $L(A,1)$

Proposition 13.2.8. *We have $L(f,1) \in \mathbf{R}$.*

Proof. With the right setup, this would follow immediately from the fact that $z \mapsto -\overline{z}$ fixes the homology class $\{0,\infty\}$. However, we don't have such a setup, so we give a direct proof.

Just as in the proof of the functional equation for $\Lambda(f,s)$, use that f is an eigenvector for the Atkin-Lehner operator W_N and (13.2.4) to write $L(f,1)$ as the sum of two integrals from $i/\sqrt{N}$ to $i\infty$. Then use the calculation

$$\begin{aligned}\overline{2\pi i \int_{i/\sqrt{N}}^{i\infty} \sum_{n=1}^{\infty} a_n e^{2\pi i n z} dz} &= -2\pi i \sum_{n=1}^{\infty} a_n \overline{\int_{i/\sqrt{N}}^{i\infty} e^{2\pi i n z} dz} \\ &= -2\pi i \sum_{n=1}^{\infty} a_n \overline{\frac{1}{2\pi i n} e^{-2\pi n/\sqrt{N}}} \\ &= 2\pi i \sum_{n=1}^{\infty} a_n \frac{1}{2\pi i n} e^{2\pi n/\sqrt{N}}\end{aligned}$$

to see that $\overline{L(f,1)} = L(f,1)$. □

Remark 13.2.9. The BSD conjecture implies that $L(f,1) \geq 0$, but this is unknown (it follows from GRH for $L(f,s)$).

13.2.8 Rationality of $L(A,1)/\Omega_A$

Proposition 13.2.10. *Suppose $A = A_f$ is an elliptic curve. Then $L(A,1)/\Omega_A \in \mathbf{Q}$. More precisely, if n is the smallest multiple of $\{0,\infty\}$ that lies in $\mathrm{H}_1(X_0(N),\mathbf{Z})$ and c_A is the Manin constant of A, then $2n \cdot c_A \cdot L(A,1)/\Omega_A \in \mathbf{Z}$.*

Proof. By the Manin-Drinfeld theorem $n\{0,\infty\} \in \mathrm{H}_1(X_0(N),\mathbf{Z})$, so

$$n \cdot L(f,1) = -n \cdot 2\pi i \cdot \int_0^{i\infty} f(z)dz \in -2\pi i \cdot \mathcal{L}_f = \frac{1}{c_A}\mathcal{L}_\omega.$$

Combining this with Proposition 13.2.8, we see that

$$n \cdot c_A \cdot L(f,1) \in \mathcal{L}_\omega^+,$$

where $\mathcal{L}_\omega^+$ is the submodule fixed by complex conjugation (i.e., $\mathcal{L}_\omega^+ = \mathcal{L} \cap \mathbf{R}$). When the period lattice is not rectangular, Ω_A generates $\mathcal{L}_\omega^+$, and when it is rectangular, $\frac{1}{2}\Omega_A$ generates. Thus $n \cdot c_A \cdot L(f,1)$ is an integer multiple of $\frac{1}{2}\Omega_A$, which proves the proposition. □

Proposition 13.2.10 can be more precise and generalized to abelian varieties $A = A_f^\vee$ attached to newforms. One can also replace n by the order of the image of $(0) - (\infty)$ in $A(\mathbf{Q})$.

Theorem 13.2.11 (Agashe, Stein). *Suppose $f \in S_2(\Gamma_1(N))$ is a newform and let $A = A_f^\vee$ be the abelian variety attached to f. Then we have the following equality of rational numbers:*

$$\frac{|L(A,1)|}{\Omega_A} = \frac{1}{c_\infty \cdot c_A} \cdot [\Phi(H_1(X_1(N),\mathbf{Z}))^+ : \Phi(\mathbf{T}\{0,\infty\})].$$

Note that $L(A,1) \in \mathbf{R}$, so $|L(A,1)| = \pm L(A,1)$, and one expects, of course, that $L(A,1) \geq 0$.

For V and W lattices in an $\mathbf{R}$-vector space M, the *lattice index* $[V : W]$ is by definition the absolute value of the determinant of a change of basis taking a basis for V to a basis for W, or 0 if W has rank smaller than the dimension of M.

Proof. Let $\tilde{\Omega}_A$ be the measure of $A(\mathbf{R})$ with respect to a basis for $S_2(\Gamma_1(N), \mathbf{Z})[I_f]$, where I_f is the annihilator in $\mathbf{T}$ of f. Note that $\tilde{\Omega}_A \cdot c_A = \Omega_A$, where c_A is the Manin index. Unwinding the definitions, we find that

$$\tilde{\Omega}_A = c_\infty \cdot [\mathrm{Hom}(S_2(\Gamma_1(N), \mathbf{Z})[I_f], \mathbf{Z}) : \Phi(H_1(X_0(N), \mathbf{Z}))^+].$$

For any ring R the pairing

$$\mathbf{T}_R \times S_2(\Gamma_1(N), R) \to R$$

given by $\langle T_n, f\rangle = a_1(T_n f)$ is perfect, so $(\mathbf{T}/I_f) \otimes R \cong \mathrm{Hom}(S_2(\Gamma_1(N), R)[I_f], R)$. Using this pairing, we may view Φ as a map

$$\Phi : H_1(X_1(N), \mathbf{Q}) \to (\mathbf{T}/I_f) \otimes \mathbf{C},$$

so that

$$\tilde{\Omega}_A = c_\infty \cdot [\mathbf{T}/I_f : \Phi(H_1(X_0(N), \mathbf{Z}))^+].$$

Note that $(\mathbf{T}/I_f) \otimes \mathbf{C}$ is isomorphic as a ring to a product of copies of $\mathbf{C}$, with one copy corresponding to each Galois conjugate f_i of f. Let $\pi_i \in (\mathbf{T}/I_f) \otimes \mathbf{C}$ be the projector onto the subspace of $(\mathbf{T}/I_f) \otimes \mathbf{C}$ corresponding to f_i. Then

$$\Phi(\{0, \infty\}) \cdot \pi_i = L(f_i, 1) \cdot \pi_i.$$

Since the π_i form a basis for the complex vector space $(\mathbf{T}/I_f) \otimes \mathbf{C}$, if we view $\Phi(\{0, \infty\})$ as the operator "left-multiplication by $\Phi(\{0, \infty\})$", then

$$\det(\Phi(\{0, \infty\})) = \prod_i L(f_i, 1) = L(A, 1),$$

Letting $H = H_1(X_0(N), \mathbf{Z})$, we have

$$\begin{aligned}
[\Phi(H)^+ : \Phi(\mathbf{T}\{0, \infty\})] &= [\Phi(H)^+ : (\mathbf{T}/I_f) \cdot \Phi(\{0, \infty\})] \\
&= [\Phi(H)^+ : \mathbf{T}/I_f] \cdot [\mathbf{T}/I_f : \mathbf{T}/I_f \cdot \Phi(\{0, \infty\})] \\
&= \frac{c_\infty}{\tilde{\Omega}_A} \cdot |\det(\Phi(\{0, \infty\}))| \\
&= \frac{c_\infty c_A}{\Omega_A} \cdot |L(A, 1)|,
\end{aligned}$$

which proves the theorem. □

Remark 13.2.12. Theorem 13.2.11 is false, in general, when A is a quotient of $J_1(N)$ not attached to a single $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-orbit of newforms. It could be modified to handle this more general case, but the generalization seems not to has been written down.

13.3 General Refined Conjecture

Conjecture 13.3.1 (Birch and Swinnerton-Dyer). *Let* $r = \mathrm{ord}_{s=1} L(A,s)$*. Then* r *is the rank of* $A(\mathbf{Q})$*, the group* $Ш(A)$ *is finite, and*

$$\frac{L^{(r)}(A,1)}{r!} = \frac{\#Ш(A) \cdot \Omega_A \cdot \mathrm{Reg}_A \cdot \prod_{p|N} c_p}{\#A(\mathbf{Q})_{\mathrm{tor}} \cdot \#A^{\vee}(\mathbf{Q})_{\mathrm{tor}}}.$$

13.4 The Conjecture for Non-Modular Abelian Varieties

Conjecture 13.3.1 can be extended to general abelian varieties over global fields. Here we discuss only the case of a general abelian variety A over $\mathbf{Q}$. We follow the discussion in [Lan91, 95-94] (Lang, Number Theory III), which describes Gross's formulation of the conjecture for abelian varieties over number fields, and to which we refer the reader for more details.

For each prime number ℓ, the ℓ-adic *Tate module* associated to A is

$$\mathrm{Ta}_\ell(A) = \varprojlim_n A(\overline{\mathbf{Q}})[\ell^n].$$

Since $A(\overline{\mathbf{Q}})[\ell^n] \cong (\mathbf{Z}/\ell^n\mathbf{Z})^{2\dim(A)}$, we see that $\mathrm{Ta}_\ell(A)$ is free of rank $2\dim(A)$ as a $\mathbf{Z}_\ell$-module. Also, since the group structure on A is defined over $\mathbf{Q}$, $\mathrm{Ta}_\ell(A)$ comes equipped with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$:

$$\rho_{A,\ell} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(\mathrm{Ta}_\ell(A)) \approx \mathrm{GL}_{2d}(\mathbf{Z}_\ell).$$

Suppose p is a prime and let $\ell \neq p$ be another prime. Fix any embedding $\overline{\mathbf{Q}} \hookrightarrow \overline{\mathbf{Q}}_p$, and notice that restriction defines a homorphism $r : \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let $G_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be the image of r. The inertia group $I_p \subset G_p$ is the kernel of the natural surjective reduction map, and we have an exact sequence

$$0 \to I_p \to \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \to 0.$$

The Galois group $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ is isomorphic to $\widehat{\mathbf{Z}}$ with canonical generator $x \mapsto x^p$. Lifting this generator, we obtain an element $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, which is well-defined up to an element of I_p. Viewed as an element of $G_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, the element Frob_p is well-defined up I_p and our choice of embedding $\overline{\mathbf{Q}} \hookrightarrow \overline{\mathbf{Q}}_p$. One can show that this implies that $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is well-defined up to I_p and conjugation by an element of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

For a G_p-module M, let

$$M^{I_p} = \{x \in M : \sigma(x) = x \text{ all } \sigma \in I_p\}.$$

Because I_p acts trivially on M^{I_p}, the action of the element $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on M^{I_p} is well-defined up to conjugation (I_p acts trivially, so the "up to I_p" obstruction vanishes). Thus the characteristic polynomial of Frob_p on M^{I_p} is well-defined, which is why $L_p(A,s)$ is well-defined. The *local L-factor* of $L(A,s)$ at p is

$$L_p(A,s) = \frac{1}{\det\left(I - p^{-s}\,\mathrm{Frob}_p^{-1} \mid \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Ta}_\ell(A), \mathbf{Z}_\ell)^{I_p}\right)}.$$

Definition 13.4.1. $L(A,s) = \prod_{\text{all } p} L_p(A,s)$

For all but finitely many primes $\mathrm{Ta}_\ell(A)^{I_p} = \mathrm{Ta}_\ell(A)$. For example, if $A = A_f$ is attached to a newform $f = \sum a_n q^n$ of level N and $p \nmid \ell \cdot N$, then $\mathrm{Ta}_\ell(A)^{I_p} = \mathrm{Ta}_\ell(A)$. In this case, the Eichler-Shimura relation implies that $L_p(A,s)$ equals $\prod L_p(f_i,s)$, where the $f_i = \sum a_{n,i} q^n$ are the Galois conjugates of f and $L_p(f_i,s) = (1 - a_{p,i} \cdot p^{-s} + p^{1-2s})^{-1}$. The point is that Eichler-Shimura can be used to show that the characteristic polynomial of Frob_p is $\prod_{i=1}^{\dim(A)}(X^2 - a_{p,i}X + p^{1-2s})$.

Theorem 13.4.2. $L(A_f,s) = \prod_{i=1}^{d} L(f_i,s)$.

13.5 Visibility of Shafarevich-Tate Groups

Let K be a number field. Suppose

$$0 \to A \to B \to C \to 0$$

is an exact sequence of abelian varieties over K. (Thus each of A, B, and C is a complete group variety over K, whose group is automatically abelian.) Then there is a corresponding long exact sequence of cohomology for the group $\mathrm{Gal}(\overline{\mathbf{Q}}/K)$:

$$0 \to A(K) \to B(K) \to C(K) \xrightarrow{\delta} \mathrm{H}^1(K, A) \to \mathrm{H}^1(K, B) \to \mathrm{H}^1(K, C) \to \cdots$$

The study of the Mordell-Weil group $C(K) = \mathrm{H}^0(K, C)$ is popular in arithmetic geometry. For example, the Birch and Swinnerton-Dyer conjecture (BSD conjecture), which is one of the million dollar Clay Math Problems, asserts that the dimension of $C(K) \otimes \mathbf{Q}$ equals the ordering vanishing of $L(C, s)$ at $s = 1$.

The group $\mathrm{H}^1(K, A)$ is also of interest in connection with the BSD conjecture, because it contains the Shafarevich-Tate group

$$Ш(A) = Ш(A/K) = \mathrm{Ker}\left(\mathrm{H}^1(K, A) \to \bigoplus_v \mathrm{H}^1(K_v, A)\right) \subset \mathrm{H}^1(K, A),$$

where the sum is over all places v of K (e.g., when $K = \mathbf{Q}$, the fields K_v are $\mathbf{Q}_p$ for all prime numbers p and $\mathbf{Q}_\infty = \mathbf{R}$).

The group $A(K)$ is fundamentally different than $\mathrm{H}^1(K, C)$. The Mordell-Weil group $A(K)$ is finitely generated, whereas the first Galois cohomology $\mathrm{H}^1(K, C)$ is far from being finitely generated—in fact, every element has finite order and there are infinitely many elements of any given order.

This talk is about "dimension shifting", i.e., relating information about $\mathrm{H}^0(K, C)$ to information about $\mathrm{H}^1(K, A)$.

13.5.1 Definitions

Elements of $\mathrm{H}^0(K, C)$ are simply points, i.e., elements of $C(K)$, so they are relatively easy to "visualize". In contrast, elements of $\mathrm{H}^1(K, A)$ are Galois cohomology classes, i.e., equivalence classes of set-theoretic (continuous) maps $f : \mathrm{Gal}(\overline{\mathbf{Q}}/K) \to A(\overline{\mathbf{Q}})$ such that $f(\sigma\tau) = f(\sigma) + \sigma f(\tau)$. Two maps are equivalent if their difference is a map of the form $\sigma \mapsto \sigma(P) - P$ for some fixed $P \in A(\overline{\mathbf{Q}})$. From this point of view H^1 is more mysterious than H^0.

There is an alternative way to view elements of $\mathrm{H}^1(K, A)$. The WC group of A is the group of isomorphism classes of principal homogeneous spaces for A, where a principal homogeneous space is a variety X and a map $A \times X \to X$ that satisfies the same axioms as those for a simply transitive group action. Thus X is a twist as variety of A, but $X(K) = \emptyset$, unless $X \approx A$. Also, the nontrivial elements of $Ш(A)$ correspond to the classes in WC that have a K_v-rational point for all places v, but no K-rational point.

Mazur introduced the following definition in order to help unify diverse constructions of principal homogeneous spaces:

Definition 13.5.1 (Visible). The *visible subgroup* of $\mathrm{H}^1(K,A)$ in B is

$$\begin{aligned}\mathrm{Vis}_B\,\mathrm{H}^1(K,A) &= \mathrm{Ker}(\mathrm{H}^1(K,A)\to \mathrm{H}^1(K,B))\\ &= \mathrm{Coker}(B(K)\to C(K)).\end{aligned}$$

Remark 13.5.2. Note that $\mathrm{Vis}_B\,\mathrm{H}^1(K,A)$ *does* depend on the embedding of A into B. For example, suppose $B = B_1\times A$. Then there could be nonzero visible elements if A is embedding into the first factor, but there will be no nonzero visible elements if A is embedded into the second factor. Here we are using that $\mathrm{H}^1(K,B_1\times A) = \mathrm{H}^1(K,B_1)\oplus \mathrm{H}^1(K,A)$.

The connection with the WC group of A is as follows. Suppose

$$0\to A\xrightarrow{f} B\xrightarrow{g} C\to 0$$

is an exact sequence of abelian varieties and that $c\in \mathrm{H}^1(K,A)$ is visible in B. Thus there exists $x\in C(K)$ such that $\delta(x) = c$, where $\delta : C(K)\to \mathrm{H}^1(K,A)$ is the connecting homomorphism. Then $X = \pi^{-1}(x)\subset B$ is a translate of A in B, so the group law on B gives X the structure of principal homogeneous space for A, and one can show that the class of X in the WC group of A corresponds to c.

Lemma 13.5.3. *The group* $\mathrm{Vis}_B\,\mathrm{H}^1(K,A)$ *is finite.*

Proof. Since $\mathrm{Vis}_B\,\mathrm{H}^1(K,A)$ is a homomorphic image of the finitely generated group $C(K)$, it is also finitely generated. On the other hand, it is a subgroup of $\mathrm{H}^1(K,A)$, so it is a torsion group. The lemma follows since a finitely generated torsion abelian group is finite. □

13.5.2 Every Element of $\mathrm{H}^1(K,A)$ *is Visible Somewhere*

Proposition 13.5.4. *Let* $c\in \mathrm{H}^1(K,A)$. *Then there exists an abelian variety* $B = B_c$ *and an embedding* $A\hookrightarrow B$ *such that* c *is visible in* B.

Proof. By definition of Galois cohomology, there is a finite extension L of K such that $\mathrm{res}_L(c) = 0$. Thus c maps to 0 in $\mathrm{H}^1(L,A_L)$. By a slight generalization of the Shapiro Lemma from group cohomology (which can be proved by dimension shifting; see, e.g., Atiyah-Wall in Cassels-Frohlich), there is a canonical isomorphism

$$\mathrm{H}^1(L,A_L)\cong \mathrm{H}^1(K,\mathrm{Res}_{L/K}(A_L)) = \mathrm{H}^1(K,B),$$

where $B = \mathrm{Res}_{L/K}(A_L)$ is the Weil restriction of scalars of A_L back down to K. The restriction of scalars B is an abelian variety of dimension $[L:K]\cdot\dim A$ that is characterized by the existence of functorial isomorphisms

$$\mathrm{Mor}_K(S,B)\cong \mathrm{Mor}_L(S_L,A_L),$$

for any K-scheme S, i.e., $B(S) = A_L(S_L)$. In particular, setting $S = A$ we find that the identity map $A_L\to A_L$ corresponds to an injection $A\hookrightarrow B$. Moreover, $c\mapsto \mathrm{res}_L(c) = 0\in \mathrm{H}^1(K,B)$. □

Remark 13.5.5. The abelian variety B in Proposition 13.5.4 is a twist of a power of A.

13.5.3 Visibility in the Context of Modularity

Usually we focus on visibility of elements in $Ш(A)$. There are a number of other results about visibility in various special cases, and large tables of examples in the context of elliptic curves and modular abelian varieties. There are also interesting modularity questions and conjectures in this context.

Motivated by the desire to understand the Birch and Swinnerton-Dyer conjecture more explicitly, I developed (with significant input from Agashe, Cremona, Mazur, and Merel) computational techniques for unconditionally constructing Shafarevich-Tate groups of modular abelian varieties $A \subset J_0(N)$ (or $J_1(N)$). For example, if $A \subset J_0(389)$ is the 20-dimensional simple factor, then

$$\mathbf{Z}/5\mathbf{Z} \times \mathbf{Z}/5\mathbf{Z} \subset Ш(A),$$

as predicted by the Birch and Swinnerton-Dyer conjecture. See [CM00] for examples when $\dim A = 1$. We will spend the rest of this section discussing the examples of [ASc, AS02] in more detail.

Tables 13.5.1–13.5.4 illustrate the main computational results of [ASc]. These tables were made by gathering data about certain arithmetic invariants of the 19608 abelian varieties A_f of level ≤ 2333. Of these, exactly 10360 have satisfy $L(A_f, 1) \neq 0$, and for these with $L(A_f, 1) \neq 0$, we compute a divisor and multiple of the conjectural order of $Ш(A_f)$. We find that there are at least 168 such that the Birch and Swinnerton-Dyer Conjecture implies that $Ш(A_f)$ is divisible by an odd prime, and we prove for 37 of these that the odd part of the conjectural order of $Ш(A_f)$ really divides $\#Ш(A_f)$ by constructing nontrivial elements of $Ш(A_f)$ using visibility.

The meaning of the tables is as follows. The first column lists a level N and an isogeny class, which uniquely specifies an abelian variety $A = A_f \subset J_0(N)$. The nth isogeny class is given by the nth letter of the alphabet. We will not discuss the ordering further, except to note that usually, the dimension of A, which is given in the second column, is enough to determine A. When $L(A, 1) \neq 0$, Conjecture 13.2.1 predicts that

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

We view the quotient $L(A, 1)/\Omega_A$, which is a rational number, as a single quantity. We can compute multiples and divisors of every quantity appearing in the right hand side of this equation, and this yields columns three and four, which are a divisor S_ℓ and a multiple S_u of the conjectural order of $Ш(A)$ (when $S_u = S_\ell$, we put an equals sign in the S_u column). Column five, which is labeled odd $\deg(\varphi_A)$, contains the odd part of the degree of the polarization

$$\varphi_A : (A \hookrightarrow J_0(N) \cong J_0(N)^\vee \to A^\vee). \qquad (13.5.1)$$

The second set of columns, columns six and seven, contain an abelian variety $B = B_g \subset J_0(N)$ such that $\#(A \cap B)$ is divisible by an odd prime divisor of S_ℓ and $L(B, 1) = 0$. When $\dim(B) = 1$, we have verified that B is an elliptic curve of rank 2. The eighth column $A \cap B$ contains the group structure of $A \cap B$, where e.g., $[2^2 302^2]$ is shorthand notation for $(\mathbf{Z}/2\mathbf{Z})^2 \oplus (\mathbf{Z}/302\mathbf{Z})^2$. The final column, labeled Vis, contains a divisor of the order of $\mathrm{Vis}_{A+B}(Ш(A))$.

The following proposition explains the significance of the odd $\deg(\varphi_A)$ column.

Proposition 13.5.6. *If $p \nmid \deg(\varphi_A)$, then $p \nmid \mathrm{Vis}_{J_0(N)}(\mathrm{H}^1(\mathbf{Q}, A))$.*

Proof. There exists a complementary morphism $\hat{\varphi}_A$, such that $\varphi_A \circ \hat{\varphi}_A = \hat{\varphi}_A \circ \varphi_A = [n]$, where n is the degree of φ_A. If $c \in \mathrm{H}^1(\mathbf{Q}, A)$ maps to 0 in $\mathrm{H}^1(\mathbf{Q}, J_0(N))$, then it also maps to 0 under the following composition

$$\mathrm{H}^1(\mathbf{Q}, A) \to \mathrm{H}^1(\mathbf{Q}, J_0(N)) \to \mathrm{H}^1(\mathbf{Q}, A^\vee) \xrightarrow{\hat{\varphi}_A} \mathrm{H}^1(\mathbf{Q}, A).$$

Since this composition is $[n]$, it follows that $c \in \mathrm{H}^1(\mathbf{Q}, A)[n]$, which proves the proposition. □

Remark 13.5.7. Since the degree of φ_A does not change if we extend scalars to a number field K, the subgroup of $\mathrm{H}^1(K, A)$ visible in $J_0(N)_K$, still has order divisible only by primes that divide $\deg(\varphi_A)$.

The following theorem explains the significance of the B column, and how it was used to deduce the Vis column.

Theorem 13.5.8. *Suppose A and B are abelian subvarieties of an abelian variety C over $\mathbf{Q}$ and that $A(\overline{\mathbf{Q}}) \cap B(\overline{\mathbf{Q}})$ is finite. Assume also that $A(\mathbf{Q})$ is finite. Let N be an integer divisible by the residue characteristics of primes of bad reduction for C (e.g., N could be the conductor of C). Suppose p is a prime such that*

$$p \nmid 2 \cdot N \cdot \#((A+B)/B)(\mathbf{Q})_{\mathrm{tor}} \cdot \#B(\mathbf{Q})_{\mathrm{tor}} \cdot \prod_{\ell} c_{A,\ell} \cdot c_{B,\ell},$$

where $c_{A,\ell} = \#\Phi_{A,\ell}(\mathbf{F}_\ell)$ is the Tamagawa number of A at ℓ (and similarly for B). Suppose furthermore that $B(\overline{\mathbf{Q}})[p] \subset A(\overline{\mathbf{Q}})$ as subgroups of $C(\overline{\mathbf{Q}})$. Then there is a natural injection

$$B(\mathbf{Q})/pB(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_C(\text{Ш}(A)).$$

A complete proof of a generalization of this theorem can be found in [AS02].

Sketch of Proof. Without loss of generality, we may assume $C = A + B$. Our hypotheses yield a diagram

$$\begin{tikzcd}
0 \arrow[r] & B[p] \arrow[r] \arrow[d] & B \arrow[r, "p"] \arrow[d] & B \arrow[r] \arrow[d, dotted] & 0 \\
0 \arrow[r] & A \arrow[r] & C \arrow[r] & B' \arrow[r] & 0,
\end{tikzcd}$$

where $B' = C/A$. Taking $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-cohomology, we obtain the following diagram:

$$\begin{tikzcd}
0 \arrow[r] & B(\mathbf{Q}) \arrow[r, "p"] \arrow[d] & B(\mathbf{Q}) \arrow[r] \arrow[d] & B(\mathbf{Q})/pB(\mathbf{Q}) \arrow[r] \arrow[d, dotted] & 0 \\
0 \arrow[r] & C(\mathbf{Q})/A(\mathbf{Q}) \arrow[r] & B'(\mathbf{Q}) \arrow[r] & \mathrm{Vis}_C(\mathrm{H}^1(\mathbf{Q}, A)) \arrow[r] & 0.
\end{tikzcd}$$

The snake lemma and our hypothesis that $p \nmid \#(C/B)(\mathbf{Q})_{\mathrm{tor}}$ imply that the rightmost vertical map is an injection

$$i : B(\mathbf{Q})/pB(\mathbf{Q}) \hookrightarrow \mathrm{Vis}_C(\mathrm{H}^1(\mathbf{Q}, A)), \tag{13.5.2}$$

since $C(A)/(A(\mathbf{Q})+B(\mathbf{Q}))$ is a sub-quotient of $(C'/B)(\mathbf{Q})$.

We show that the image of (13.5.2) lies in $Ш(A)$ using a local analysis at each prime, which we now sketch. At the archimedian prime, no work is needed since $p \neq 2$. At non-archimedian primes ℓ, one uses facts about Néron models (when $\ell = p$) and our hypothesis that p does not divide the Tamagawa numbers of B (when $\ell \neq p$) to show that if $x \in B(\mathbf{Q})/pB(\mathbf{Q})$, then the corresponding cohomology class $\mathrm{res}_\ell(i(x)) \in \mathrm{H}^1(\mathbf{Q}_\ell, A)$ splits over the maximal unramified extension. However,

$$\mathrm{H}^1(\mathbf{Q}_\ell^{\mathrm{ur}}/\mathbf{Q}_\ell, A) \cong \mathrm{H}^1(\overline{\mathbf{F}}_\ell/\mathbf{F}_\ell, \Phi_{A,\ell}(\overline{\mathbf{F}}_\ell)),$$

and the right hand cohomology group has order $c_{A,\ell}$, which is coprime to p. Thus $\mathrm{res}_\ell(i(x)) = 0$, which completes the sketch of the proof. □

13.5.4 Future Directions

The data in Tables 13.5.1-13.5.4 could be investigated further.

It should be possible to replace the hypothesis that $B[p] \subset A$, with the weaker hypothesis that $B[\mathfrak{m}] \subset A$, where $\mathfrak{m}$ is a maximal ideal of the Hecke algebra $\mathbf{T}$. For example, this improvement would help one to show that 5^2 divides the order of the Shafarevich-Tate group of **1041E**. Note that for this example, we only know that $L(B,1) = 0$, not that $B(\mathbf{Q})$ has positive rank (as predicted by Conjecture 13.1.5), which is another obstruction.

One can consider visibility at a higher level. For example, there are elements of order 3 in the Shafarevich-Tate group of **551H** that are not visible in $J_0(551)$, but these elements are visible in $J_0(2 \cdot 551)$, according to the computations in [Ste03] (Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using MAGMA).

Conjecture 13.5.9 (Stein). *Suppose $c \in Ш(A_f)$, where $A_f \subset J_0(N)$. Then there exists M such that c is visible in $J_0(NM)$. In other words, every element of $Ш(A_f)$ is "modular".*

TABLE 13.5.1. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	Vis
389E$*$	20	5^2	$=$	5	**389A**	1	$[20^2]$	5^2
433D$*$	16	7^2	$=$	$7\cdot 111$	**433A**	1	$[14^2]$	7^2
446F$*$	8	11^2	$=$	$11\cdot 359353$	**446B**	1	$[11^2]$	11^2
551H	18	3^2	$=$	169	NONE			
563E$*$	31	13^2	$=$	13	**563A**	1	$[26^2]$	13^2
571D$*$	2	3^2	$=$	$3^2\cdot 127$	**571B**	1	$[3^2]$	3^2
655D$*$	13	3^4	$=$	$3^2\cdot 9799079$	**655A**	1	$[36^2]$	3^4
681B	1	3^2	$=$	$3\cdot 125$	**681C**	1	$[3^2]$	–
707G$*$	15	13^2	$=$	$13\cdot 800077$	**707A**	1	$[13^2]$	13^2
709C$*$	30	11^2	$=$	11	**709A**	1	$[22^2]$	11^2
718F$*$	7	7^2	$=$	$7\cdot 5371523$	**718B**	1	$[7^2]$	7^2
767F	23	3^2	$=$	1	NONE			
794G	12	11^2	$=$	$11\cdot 34986189$	**794A**	1	$[11^2]$	–
817E	15	7^2	$=$	$7\cdot 79$	**817A**	1	$[7^2]$	–
959D	24	3^2	$=$	583673	NONE			
997H$*$	42	3^4	$=$	3^2	**997B**	1	$[12^2]$	3^2
					997C	1	$[24^2]$	3^2
1001F	3	3^2	$=$	$3^2\cdot 1269$	**1001C**	1	$[3^2]$	–
					91A	1	$[3^2]$	–
1001L	7	7^2	$=$	$7\cdot 2029789$	**1001C**	1	$[7^2]$	–
1041E	4	5^2	$=$	$5^2\cdot 13589$	**1041B**	2	$[5^2]$	–
1041J	13	5^4	$=$	$5^3\cdot 21120929983$	**1041B**	2	$[5^4]$	–
1058D	1	5^2	$=$	$5\cdot 483$	**1058C**	1	$[5^2]$	–
1061D	46	151^2	$=$	$151\cdot 10919$	**1061B**	2	$[2^2 302^2]$	–
1070M	7	$3\cdot 5^2$	$3^2\cdot 5^2$	$3\cdot 5\cdot 1720261$	**1070A**	1	$[15^2]$	–
1077J	15	3^4	$=$	$3^2\cdot 1227767047943$	**1077A**	1	$[9^2]$	–
1091C	62	7^2	$=$	1	NONE			
1094F$*$	13	11^2	$=$	$11^2\cdot 172446773$	**1094A**	1	$[11^2]$	11^2
1102K	4	3^2	$=$	$3^2\cdot 31009$	**1102A**	1	$[3^2]$	–
1126F$*$	11	11^2	$=$	$11\cdot 13990352759$	**1126A**	1	$[11^2]$	11^2
1137C	14	3^4	$=$	$3^2\cdot 64082807$	**1137A**	1	$[9^2]$	–
1141I	22	7^2	$=$	$7\cdot 528921$	**1141A**	1	$[14^2]$	–
1147H	23	5^2	$=$	$5\cdot 729$	**1147A**	1	$[10^2]$	–
1171D$*$	53	11^2	$=$	$11\cdot 81$	**1171A**	1	$[44^2]$	11^2
1246B	1	5^2	$=$	$5\cdot 81$	**1246C**	1	$[5^2]$	–
1247D	32	3^2	$=$	$3^2\cdot 2399$	**43A**	1	$[36^2]$	–
1283C	62	5^2	$=$	$5\cdot 2419$	NONE			
1337E	33	3^2	$=$	71	NONE			
1339G	30	3^2	$=$	5776049	NONE			
1355E	28	3	3^2	$3^2\cdot 2224523985405$	NONE			
1363F	25	31^2	$=$	$31\cdot 34889$	**1363B**	2	$[2^2 62^2]$	–
1429B	64	5^2	$=$	1	NONE			
1443G	5	7^2	$=$	$7^2\cdot 18525$	**1443C**	1	$[7^1 14^1]$	–
1446N	7	3^2	$=$	$3\cdot 17459029$	**1446A**	1	$[12^2]$	–

TABLE 13.5.2. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	Vis
1466H$*$	23	13^2	$=$	$13\cdot 25631993723$	**1466B**	1	$[26^2]$	13^2
1477C$*$	24	13^2	$=$	$13\cdot 57037637$	**1477A**	1	$[13^2]$	13^2
1481C	71	13^2	$=$	70825	NONE			
1483D$*$	67	$3^2\cdot 5^2$	$=$	$3\cdot 5$	**1483A**	1	$[60^2]$	$3^2\cdot 5^2$
1513F	31	3	3^4	$3\cdot 759709$	NONE			
1529D	36	5^2	$=$	535641763	NONE			
1531D	73	3	3^2	3	**1531A**	1	$[48^2]$	–
1534J	6	3	3^2	$3^2\cdot 635931$	**1534B**	1	$[6^2]$	–
1551G	13	3^2	$=$	$3\cdot 110659885$	**141A**	1	$[15^2]$	–
1559B	90	11^2	$=$	1	NONE			
1567D	69	$7^2\cdot 41^2$	$=$	$7\cdot 41$	**1567B**	3	$[4^4 1148^2]$	–
1570J$*$	6	11^2	$=$	$11\cdot 228651397$	**1570B**	1	$[11^2]$	11^2
1577E	36	3	3^2	$3^2\cdot 15$	**83A**	1	$[6^2]$	–
1589D	35	3^2	$=$	6005292627343	NONE			
1591F$*$	35	31^2	$=$	$31\cdot 2401$	**1591A**	1	$[31^2]$	31^2
1594J	17	3^2	$=$	$3\cdot 259338050025131$	**1594A**	1	$[12^2]$	–
1613D$*$	75	5^2	$=$	$5\cdot 19$	**1613A**	1	$[20^2]$	5^2
1615J	13	3^4	$=$	$3^2\cdot 13317421$	**1615A**	1	$[9^1 18^1]$	–
1621C$*$	70	17^2	$=$	17	**1621A**	1	$[34^2]$	17^2
1627C$*$	73	3^4	$=$	3^2	**1627A**	1	$[36^2]$	3^4
1631C	37	5^2	$=$	6354841131	NONE			
1633D	27	$3^6\cdot 7^2$	$=$	$3^5\cdot 7\cdot 31375$	**1633A**	3	$[6^4 42^2]$	–
1634K	12	3^2	$=$	$3\cdot 3311565989$	**817A**	1	$[3^2]$	–
1639G$*$	34	17^2	$=$	$17\cdot 82355$	**1639B**	1	$[34^2]$	17^2
1641J$*$	24	23^2	$=$	$23\cdot 1491344147471$	**1641B**	1	$[23^2]$	23^2
1642D$*$	14	7^2	$=$	$7\cdot 123398360851$	**1642A**	1	$[7^2]$	7^2
1662K	7	11^2	$=$	$11\cdot 16610917393$	**1662A**	1	$[11^2]$	–
1664K	1	5^2	$=$	$5\cdot 7$	**1664N**	1	$[5^2]$	–
1679C	45	11^2	$=$	6489	NONE			
1689E	28	3^2	$=$	$3\cdot 172707180029157365$	**563A**	1	$[3^2]$	–
1693C	72	1301^2	$=$	1301	**1693A**	3	$[2^4 2602^2]$	–
1717H$*$	34	13^2	$=$	$13\cdot 345$	**1717B**	1	$[26^2]$	13^2
1727E	39	3^2	$=$	118242943	NONE			
1739F	43	659^2	$=$	$659\cdot 151291281$	**1739C**	2	$[2^2 1318^2]$	–
1745K	33	5^2	$=$	$5\cdot 1971380677489$	**1745D**	1	$[20^2]$	–
1751C	45	5^2	$=$	$5\cdot 707$	**103A**	2	$[505^2]$	–
1781D	44	3^2	$=$	61541	NONE			
1793G$*$	36	23^2	$=$	$23\cdot 8846589$	**1793B**	1	$[23^2]$	23^2
1799D	44	5^2	$=$	201449	NONE			
1811D	98	31^2	$=$	1	NONE			
1829E	44	13^2	$=$	3595	NONE			
1843F	40	3^2	$=$	8389	NONE			
1847B	98	3^6	$=$	1	NONE			
1871C	98	19^2	$=$	14699	NONE			

TABLE 13.5.3. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	Vis
1877B	86	7^2	$=$	1	NONE			
1887J	12	5^2	$=$	$5\cdot 10825598693$	**1887A**	1	$[20^2]$	–
1891H	40	7^4	$=$	$7^2\cdot 44082137$	**1891C**	2	$[4^2 196^2]$	–
1907D$*$	90	7^2	$=$	$7\cdot 165$	**1907A**	1	$[56^2]$	7^2
1909D$*$	38	3^4	$=$	$3^2\cdot 9317$	**1909A**	1	$[18^2]$	3^4
1913B$*$	1	3^2	$=$	$3\cdot 103$	**1913A**	1	$[3^2]$	3^2
1913E	84	$5^4\cdot 61^2$	$=$	$5^2\cdot 61\cdot 103$	**1913A**	1	$[10^2]$	–
					1913C	2	$[2^2 610^2]$	–
1919D	52	23^2	$=$	675	NONE			
1927E	45	3^2	3^4	52667	NONE			
1933C	83	$3^2\cdot 7$	$3^2\cdot 7^2$	$3\cdot 7$	**1933A**	1	$[42^2]$	3^2
1943E	46	13^2	$=$	62931125	NONE			
1945E$*$	34	3^2	$=$	$3\cdot 571255479184807$	**389A**	1	$[3^2]$	3^2
1957E$*$	37	$7^2\cdot 11^2$	$=$	$7\cdot 11\cdot 3481$	**1957A**	1	$[22^2]$	11^2
					1957B	1	$[14^2]$	7^2
1979C	104	19^2	$=$	55	NONE			
1991C	49	7^2	$=$	1634403663	NONE			
1994D	26	3	3^2	$3^2\cdot 46197281414642501$	**997B**	1	$[3^2]$	–
1997C	93	17^2	$=$	1	NONE			
2001L	11	3^2	$=$	$3^2\cdot 44513447$	NONE			
2006E	1	3^2	$=$	$3\cdot 805$	**2006D**	1	$[3^2]$	–
2014L	12	3^2	$=$	$3^2\cdot 126381129003$	**106A**	1	$[9^2]$	–
2021E	50	5^6	$=$	$5^2\cdot 729$	**2021A**	1	$[100^2]$	5^4
2027C$*$	94	29^2	$=$	29	**2027A**	1	$[58^2]$	29^2
2029C	90	$5^2\cdot 269^2$	$=$	$5\cdot 269$	**2029A**	2	$[2^2 2690^2]$	–
2031H$*$	36	11^2	$=$	$11\cdot 1014875952355$	**2031C**	1	$[44^2]$	11^2
2035K	16	11^2	$=$	$11\cdot 218702421$	**2035C**	1	$[11^1 22^1]$	–
2038F	25	5	5^2	$5^2\cdot 92198576587$	**2038A**	1	$[20^2]$	–
					1019B	1	$[5^2]$	–
2039F	99	$3^4\cdot 5^2$	$=$	13741381043009	NONE			
2041C	43	3^4	$=$	61889617	NONE			
2045I	39	3^4	$=$	$3^3\cdot 3123399893$	**2045C**	1	$[18^2]$	–
					409A	13	$[9370199679^2]$	–
2049D	31	3^2	$=$	29174705448000469937	NONE			
2051D	45	7^2	$=$	$7\cdot 674652424406369$	**2051A**	1	$[56^2]$	–
2059E	45	$5\cdot 7^2$	$5^2\cdot 7^2$	$5^2\cdot 7\cdot 167359757$	**2059A**	1	$[70^2]$	–
2063C	106	13^2	$=$	8479	NONE			
2071F	48	13^2	$=$	36348745	NONE			
2099B	106	3^2	$=$	1	NONE			
2101F	46	5^2	$=$	$5\cdot 11521429$	**191A**	2	$[155^2]$	–
2103E	37	$3^2\cdot 11^2$	$=$	$3^2\cdot 11\cdot 874412923071571792611$	**2103B**	1	$[33^2]$	11^2
2111B	112	211^2	$=$	1	NONE			
2113B	91	7^2	$=$	1	NONE			
2117E$*$	45	19^2	$=$	$19\cdot 1078389$	**2117A**	1	$[38^2]$	19^2

TABLE 13.5.4. Visibility of Nontrivial Odd Parts of Shafarevich-Tate Groups

A	dim	S_l	S_u	odd deg(φ_A)	B	dim	$A \cap B$	Vis
2119C	48	7^2	=	89746579	NONE			
2127D	34	3^2	=	$3\cdot 18740561792121901$	**709A**	1	$[3^2]$	–
2129B	102	3^2	=	1	NONE			
2130Y	4	7^2	=	$7\cdot 83927$	**2130B**	1	$[14^2]$	–
2131B	101	17^2	=	1	NONE			
2134J	11	3^2	=	1710248025389	NONE			
2146J	10	7^2	=	$7\cdot 1672443$	**2146A**	1	$[7^2]$	–
2159E	57	13^2	=	31154538351	NONE			
2159D	56	3^4	=	233801	NONE			
2161C	98	23^2	=	1	NONE			
2162H	14	3	3^2	$3\cdot 6578391763$	NONE			
2171E	54	13^2	=	271	NONE			
2173H	44	199^2	=	$199\cdot 3581$	**2173D**	2	$[398^2]$	–
2173F	43	19^2	$3^2\cdot 19^2$	$3^2\cdot 19\cdot 229341$	**2173A**	1	$[38^2]$	19^2
2174F	31	5^2	=	$5\cdot 21555702093188316107$	NONE			
2181E	27	7^2	=	$7\cdot 7217996450474835$	**2181A**	1	$[28^2]$	–
2193K	17	3^2	=	$3\cdot 15096035814223$	**129A**	1	$[21^2]$	–
2199C	36	7^2	=	$7^2\cdot 13033437060276603$	NONE			
2213C	101	3^4	=	19	NONE			
2215F	46	13^2	=	$13\cdot 1182141633$	**2215A**	1	$[52^2]$	–
2224R	11	79^2	=	79	**2224G**	2	$[79^2]$	–
2227E	51	11^2	=	259	NONE			
2231D	60	47^2	=	91109	NONE			
2239B	110	11^4	=	1	NONE			
2251E$*$	99	37^2	=	37	**2251A**	1	$[74^2]$	37^2
2253C$*$	27	13^2	=	$13\cdot 14987929400988647$	**2253A**	1	$[26^2]$	13^2
2255J	23	7^2	=	15666366543129	NONE			
2257H	46	$3^6\cdot 29^2$	=	$3^3\cdot 29\cdot 175$	**2257A**	1	$[9^2]$	–
					2257D	2	$[2^2 174^2]$	–
2264J	22	73^2	=	73	**2264B**	2	$[146^2]$	–
2265U	14	7^2	=	$7^2\cdot 73023816368925$	**2265B**	1	$[7^2]$	–
2271I$*$	43	23^2	=	$23\cdot 392918345997771783$	**2271C**	1	$[46^2]$	23^2
2273C	105	7^2	=	7^2	NONE			
2279D	61	13^2	=	96991	NONE			
2279C	58	5^2	=	1777847	NONE			
2285E	45	151^2	=	$151\cdot 138908751161$	**2285A**	2	$[2^2 302^2]$	–
2287B	109	71^2	=	1	NONE			
2291C	52	3^2	=	427943	NONE			
2293C	96	479^2	=	479	**2293A**	2	$[2^2 958^2]$	–
2294F	15	3^2	=	$3\cdot 6289390462793$	**1147A**	1	$[3^2]$	–
2311B	110	5^2	=	1	NONE			
2315I	51	3^2	=	$3\cdot 4475437589723$	**463A**	16	$[13426312769169^2]$	–
2333C	101	83341^2	=	83341	**2333A**	4	$[2^6 166682^2]$	–

References

[ASa] A. Agashe and W. A. Stein, Appendix to Joan-C. Lario and René Schoof: *Some computations with Hecke rings and deformation rings*, To appear in Exp. Math.

[ASb] ———, *The manin constant, congruence primes, and the modular degree*, In progress.

[ASc] ———, *Visible Evidence for the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties of Analytic Rank* 0, To appear in Math. of Computation.

[AS02] ———, *Visibility of Shafarevich-Tate groups of abelian varieties*, J. Number Theory **97** (2002), no. 1, 171–185. MR 2003h:11070

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic). MR 2002d:11058

[Bir65] B. J. Birch, *Conjectures concerning elliptic curves*, Proceedings of Symposia in Pure Mathematics, VIII, Amer. Math. Soc., Providence, R.I., 1965, pp. 106–112. MR 30 #4759

[Bir71] B. J. Birch, *Elliptic curves over* $\mathbf{Q}$*: A progress report*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Amer. Math. Soc., Providence, R.I., 1971, pp. 396–400.

[BK90] S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, The Grothendieck Festschrift, Vol. I, Birkhäuser Boston, Boston, MA, 1990, pp. 333–400.

[BLR90] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990. MR 91i:14034

[BpR91] N. Boston, H.W. Lenstra, Jr., and K.A. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sci. Paris Sér. I Math. **312** (1991), no. 4, 323–328.

[Buz96] Kevin Buzzard, *On the eigenvalues of the Hecke operator T_2*, J. Number Theory **57** (1996), no. 1, 130–132. MR 96m:11033

[CF99] J.B. Conrey and D.W. Farmer, *Hecke operators and the nonvanishing of L-functions*, Topics in number theory (University Park, PA, 1997), Math. Appl., vol. 467, Kluwer Acad. Publ., Dordrecht, 1999, pp. 143–150. MR 2000f:11055

[CM98] R. Coleman and B. Mazur, *The Eigencurve*, Galois representations in arithmetic algebraic geometry (Durham, 1996), Cambridge Univ. Press, Cambridge, 1998, pp. 1–113. MR 1 696 469

[CM00] J.E. Cremona and B. Mazur, *Visualizing elements in the Shafarevich-Tate group*, Experiment. Math. **9** (2000), no. 1, 13–28. MR 1 758 797

[Con01] B. Conrad, *The shimura construction in weight 2 (appendix to Ribet-Stein, Lectures on Serre's Conjecture)*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 205–232. MR 2002h:11047

[Con03] K. Conrad, *Partial Euler products on the critical line*, Preprint (2003).

[Cp86] J.W.S. Cassels and A. Fröhlich (eds.), *Algebraic number theory*, London, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1986, Reprint of the 1967 original.

[CR62] C.W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1962, Pure and Applied Mathematics, Vol. XI.

[Cre97] J.E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[DDT94] H. Darmon, F. Diamond, and R. Taylor, *Fermat's last theorem*, Current developments in mathematics, 1995 (Cambridge, MA), Internat. Press, Cambridge, MA, 1994, pp. 1–154.

[DI95] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[DR73] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972) (Berlin), Springer, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349.

[dSL97] B. de Smit and Jr. Lenstra, H.W., *Explicit construction of universal deformation rings*, Modular forms and Fermat's last theorem (Boston, MA, 1995), Springer, New York, 1997, pp. 313–326.

[Edi92] B. Edixhoven, *Néron models and tame ramification*, Compositio Math. **81** (1992), no. 3, 291–306. MR 93a:14041

[Eis95] D. Eisenbud, *Commutative algebra with a view toward algebraic geometry*, Springer-Verlag, New York, 1995. MR 97a:13001

[Elk98] N. D. Elkies, *Elliptic and modular curves over finite fields and related computational issues*, Computational perspectives on number theory (Chicago, IL, 1995), Amer. Math. Soc., Providence, RI, 1998, pp. 21–76.

[Ell02] J. Ellenberg, $\mathbf{q}$-*curves and Galois Representations*, `http://www.math.princeton.edu/ ellenber/papers.html#MCAV` (2002).

[ES00] J. Ellenberg and C. Skinner, *On the Modularity of* $\mathbf{Q}$-*curves*, `http://www.math.princeton.edu/ ellenber/papers.html#QCURVE` (2000).

[FJ02] D. W. Farmer and K. James, *The irreducibility of some level 1 Hecke polynomials*, Math. Comp. **71** (2002), no. 239, 1263–1270 (electronic). MR 2003e:11046

[FM99] G. Frey and M. Müller, *Arithmetic of modular curves and applications*, Algorithmic algebra and number theory (Heidelberg, 1997), Springer, Berlin, 1999, pp. 11–48.

[FpS^{+}01] E. V. Flynn, F. Leprévost, E. F. Schaefer, W. A. Stein, M. Stoll, and J. L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. Comp. **70** (2001), no. 236, 1675–1697 (electronic). MR 1 836 926

[Gol82] D. Goldfeld, *Sur les produits partiels eulériens attachés aux courbes elliptiques*, C. R. Acad. Sci. Paris Sér. I Math. **294** (1982), no. 14, 471–474. MR 84d:14031

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[Har77] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[Igu56] J. Igusa, *Fibre systems of Jacobian varieties*, Amer. J. Math. **78** (1956), 171–199.

[KM85] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Princeton University Press, Princeton, N.J., 1985.

[Kna92] A. W. Knapp, *Elliptic curves*, Princeton University Press, Princeton, NJ, 1992.

[Kol88a] V. A. Kolyvagin, *Finiteness of* $E(\mathbf{q})$ *and* $SH(E,\mathbf{q})$ *for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671. MR 89m:11056

[Kol88b] ______, *The Mordell-Weil and Shafarevich-Tate groups for Weil elliptic curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 6, 1154–1180, 1327. MR 90f:11035

[Lan91] S. Lang, *Number theory. III*, Springer-Verlag, Berlin, 1991, Diophantine geometry. MR 93a:11048

[Lan93] ______, *Algebra*, third ed., Addison-Wesley Publishing Co., Reading, Mass., 1993.

[Lan94] ______, *Algebraic number theory*, second ed., Springer-Verlag, New York, 1994.

[Lan95] ______, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[Li75] W-C. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.

[Man72] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66. MR 47 #3396

[Maz72] B. Mazur, *Rational points of abelian varieties with values in towers of number fields*, Invent. Math. **18** (1972), 183–266.

[Maz77] ______, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978).

[Maz78] ______, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[Maz89] ______, *Deforming Galois representations*, Galois groups over **Q** (Berkeley, CA, 1987), Springer, New York, 1989, pp. 385–437.

[Mer94] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations, Springer, 1994, pp. 59–94.

[MFK94] D. Mumford, J. Fogarty, and F. Kirwan, *Geometric invariant theory*, third ed., Springer-Verlag, Berlin, 1994.

[Mil86] J. S. Milne, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[Miy89] T. Miyake, *Modular forms*, Springer-Verlag, Berlin, 1989, Translated from the Japanese by Yoshitaka Maeda.

[MO02] W. J. McGraw and K. Ono, *Modular form Congruences and Selmer groups*, preprint (2002).

[MR91] B. Mazur and K. A. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque (1991), no. 196-197, 6, 215–255 (1992), Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).

[MS01] L. Merel and W. A. Stein, *The field generated by the points of small prime order on an elliptic curve*, Internat. Math. Res. Notices (2001), no. 20, 1075–1082. MR 1 857 596

[MSD74] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61. MR 50 #7152

[Mum70] D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.

[Ogg71] A. P. Ogg, *Rational points of finite order on elliptic curves*, Invent. Math. **12** (1971), 105–111. MR 45 #178

[Rib75] K. A. Ribet, *Endomorphisms of semi-stable abelian varieties over number fields*, Ann. Math. (2) **101** (1975), 555–562. MR 51 #8120

[Rib92] ______, *Abelian varieties over* **Q** *and modular forms*, Algebra and topology 1992 (Taejŏn), Korea Adv. Inst. Sci. Tech., Taejŏn, 1992, pp. 53–79. MR 94g:11042

[Rib99] ______, *Torsion points on $J_0(N)$ and Galois representations*, Arithmetic theory of elliptic curves (Cetraro, 1997), Springer, Berlin, 1999, pp. 145–166. MR 2001b:11054

[Ros86] M. Rosen, *Abelian varieties over* **C**, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 79–101.

[RS01] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, pp. 143–232. MR 2002h:11047

[Sch06] I. Schur, *Arithmetische Untersuchungen über endliche Gruppen linearer Substitutionen*, Sitz. Pr. Akad. Wiss. (1906), 164–184, Gesam. Abhl., **I**, 177–197, Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1973.

[Sch65] O. F. G. Schilling (ed.), *Arithmetical algebraic geometry. (Proceedings of a Conference held at Purdue University, December 5–7, 1963)*, Harper & Row Publishers, New York, 1965.

[SD67] P. Swinnerton-Dyer, *The conjectures of Birch and Swinnerton-Dyer, and of Tate*, Proc. Conf. Local Fields (Driebergen, 1966), Springer, Berlin, 1967, pp. 132–157. MR 37 #6287

[SD74] H. P. F. Swinnerton-Dyer, *Analytic theory of abelian varieties*, Cambridge University Press, London, 1974, London Mathematical Society Lecture Note Series, No. 14. MR 51 #3180

[Ser73] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[Ser77] ______, *Linear representations of finite groups*, Springer-Verlag, New York, 1977, Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.

[Ser79] ———, *Local fields*, Springer-Verlag, New York, 1979, Translated from the French by Marvin Jay Greenberg.

[Ser87] ———, *Sur les représentations modulaires de degré* 2 *de* Gal($\overline{\mathbf{Q}}/\mathbf{Q}$), Duke Math. J. **54** (1987), no. 1, 179–230.

[Ser88] ———, *Algebraic groups and class fields*, Springer-Verlag, New York, 1988, Translated from the French.

[Shi59] G. Shimura, *Sur les intégrales attachées aux formes automorphes*, J. Math. Soc. Japan **11** (1959), 291–311.

[Shi73] ———, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), no. 3, 523–544.

[Shi94] ———, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.

[Sil92] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original.

[Sil94] ———, *Advanced topics in the arithmetic of elliptic curves*, Springer-Verlag, New York, 1994.

[ST68] J-P. Serre and J. T. Tate, *Good reduction of abelian varieties*, Ann. of Math. (2) **88** (1968), 492–517.

[Ste03] W. A. Stein, *Studying the Birch and Swinnerton-Dyer Conjecture for Modular Abelian Varieties Using MAGMA*, To appear in J. Cannon, ed., *Computational Experiments in Algebra and Geometry*, Springer-Verlag (2003).

[Tat66] J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1965/66, pp. Exp. No. 306, 415–440.

[TW95] R. Taylor and A. J. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[Wal85] J.-L. Waldspurger, *Quelques propriétés arithmétiques de certaines formes automorphes sur* GL(2), Compositio Math. **54** (1985), no. 2, 121–171. MR 87g:11061a

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Wil00] ———, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[Zag85a] D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. **28** (1985), no. 3, 372–384. MR 86m:11041

[Zag85b] ———, *Modular points, modular curves, modular surfaces and modular forms*, Workshop Bonn 1984 (Bonn, 1984), Springer, Berlin, 1985, pp. 225–248.