Contents

Chapter 1

Preliminaries

1.1 Definition of a Modular Form

Part of the power of modular forms comes from the fact that they can be seen as functions on several different spaces. A modular form can be seen as a function on lattices in the complex plane, on the upper half plane, or on a Riemann surface known as the modular curve. We will first give a definition of a modular function as a function on the space of lattices which is not quite correct but is the most intuitive and will then refine it.

1.1.1 Modular Forms as Functions on Lattices

A lattice in $\mathbf{C}$, by which we will always mean a rank two lattice embedded in a discrete manner, can be specified by giving a basis (ω_1, ω_2) for it. We will then define a modular form as a holomorphic function of two variables; we will assume that ω_1, ω_2 is an oriented basis, meaning that ω_1/ω_2 has positive imaginary part. A modular form is then a holomorphic funtion $f(\omega_1, \omega_2)$, defined for $\mathrm{Im}(\omega_1/\omega_2) > 0$, which depends only on the lattice spanned by ω_1 and ω_2 and which has the folowing homogenity property: there is an integer k such that $f(\alpha\omega_1, \alpha\omega_2) = \alpha^{-k} f(\omega_1, \omega_2)$ for all complex α. k is called the *weight* iof the modular form.

There two things wrong with this definition: a generality we have not included and a condition we have not imposed. To see the first, note that the requirement that $f(\omega_1, \omega_2)$ depend only on the lattice spanned by ω_1 and ω_2 means explicitly that $f(\omega_1, \omega_2) = f(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbf{Z})$. The generaliztion is to replace $SL_2(\mathbf{Z})$ with a subgroup of $SL_2(\mathbf{Z})$ of the sort known as a congruence subgroup. For convenience, we introduce the notation

$$([\gamma]|_k f)(\omega_1, \omega_2) = f(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$$

where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. $[\gamma]|_k$ is a right action of $GL_2(\mathbf{R})$ on the space of modular forms.

For any positive integer N, define the group

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}.$$

(By this we mean $a = 1$, $b = 0$, $c = 0$ and $d = 1$.) We will frequently write modular conditions on matrix entries in this shorthand.) A congruence subgroup is a subgroup of $SL_2(\mathbf{Z})$ that contains some $\Gamma(N)$. We now give the expanded Definition: given an integer N, a group Γ with $\Gamma(N) \subseteq \Gamma \subseteq SL_2(\mathbf{Z})$ and an integer k, a modular form is a holomorphic function f of two variables such that

$$f = [\gamma]|_k f \text{ for } \gamma \in \Gamma$$

and

$$f(\alpha\omega_1, \alpha\omega_2) = \alpha^{-k} f(\omega_1, \omega_2).$$

f is called a modular form of weight k for Γ, the least N for which this Definition holds is called the *level* of f.

Just as the symmetry under $SL_2(\mathbf{Z})$ corresponds to requiring that the modular form depend only on the matrix spanned by ω_1 and ω_2, the symmetry under $\Gamma(N)$ has a geometric interpretation as well. Let Λ be the lattice spanned by ω_1 and ω_2. Color the points of Λ with N^2 colors based on which coset of $\Lambda/N\Lambda$ they lie in. Being symmetric under $\Gamma(N)$ means that a function is dependen only on the colored lattice arising from ω_1 and ω_2.

There is one subgroup lying in between $\Gamma(N)$ and $SL_2(\mathbf{Z})$ that will be of importance in the future. Set

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} c = 0 \mod N \right\}$$

Being a modular form for $\Gamma_0(N)$ also has coloring interpretations. Color a point of $\Lambda(N)$ with one color if it lies in the sublattice spanned by $(\omega_1, N\omega_2)$ and another color otherwise; f is a modular form for $\Gamma_0(N)$ if it depends only on the coloring arising from ω_1 and ω_2.

1.1.2 Modular Forms as Functions on the Upper Half Plane

We are still missing one condition before we give the completely correct definition of a modular form. In order to easily give this, we now show how to look at a modular form as a function on the upper half plane and then on a Riemann surface. Note that, by our homogenity condition, the values of a modular form of given weight k are determined by its values on points of the form $(z, 1)$. We must now interpret $[\gamma]|_k$ in this context. Writing f both for the function of one and two variables,

$$([\gamma]|_k f)(z) = ([\gamma]|_k f)(z,1) = f(az+b, cz+d) = (cz+d)^{-k} f((az+b)/(cz+d), 1) = (cz+d)^{-k} f((az+b)/(cz+d)).$$

Our requirement that ω_1/ω_2 have positive imaginary part means we can consider functions solely on the upper half plane.

1.1.3 The Modular Curve

Now, in the case of forms of weight 0, this definition can be restated by saying that a modular form is a holomorphic function on the Riemann surface produced by taking the quotient of the upper half plane by the action on the upper half plane by Γ by $z \to (az+b)/(cz+d)$. It can be easily shown that this Riemman surface can be uniquely compactified by adding a finite number of points, known as the cusps. Add to the upper half plane the rational points of the real access and a point at infinity, the neighborhoods of which are of the form $\{z \text{ s.t. } \mathrm{Im}(z) > R\}$. The the quotient map from the upper half plane to the curve extends to a map from the added points to the cusps. We call the resulting curve the modular curve correspoding to Γ and denote it X. (At times it will be denoted $X(N)$, $X_0(N)$, etc. to show what Γ we are considering.) This makes plausible the condition we have previously been missing in the case of weight 0: modular forms should extend to meromorphic functions on the modular curve.

1.1.4 The Meromorphic Condition

The condition in general is the following. A modular form must have the property that, when viewed as a function in the upper half plane, it must be bounded on any path whose imaginary part approaches infinity. By the general theory of Riemann surfaces, this will mean that in a neighborhood of infinity, it will extend to a holomorphic function on the modular curve. The corresponding condition at the other cusps is messier to state. Perhaps the simplest statement is that, if $\gamma \in GL_2(\mathbf{R})$ maps the cusp in question to infinity, that $[\gamma]f$ should be bounded at infinity. Another statement is that f shuold be bounded on arcs orthogonal to the real axis. A form is called a cusp form if it's value at each of the cusps is 0. If f is a form of weight two, then $f(z)dz$ gives rise to a differential form on the modular curve; oddly, it is the cusp forms, and not simply the modular forms, that correspond to holomorphic differential fomrs in this way.

For most purposes we can ignore the meromorphic condition, as it is preserved by every operation we normally perform and most formal identities hold with out it. We can thus think of modular forms in three ways. We can think of a modular form simply as a function on lattices with additional structure and the appropriate homogenity. Alternatively, one can view them as objects on the upper half plane with less intuitive symmetries. Finally, one can view a modular form as an object on the modular curve. A weight 0 form is simply a fucntion on the modular curve. Although it is less obvious, a weight 2 worm is a differential one form. This is because $d((az+b)/(cz+d)) = (cz+d)^{-2}dz$ so, if f is a weight 2 form, then $f(z)dz$ is invariant under the pull back action by members of Γ, so it descends to a one fomr on the modular curve. Modular forms of other weights are sections of less familiar line bundles. (The forms of weight $2k$ correspond to sections of the k-fold tesor product of the cotangent bundle with itself; the odd weight forms are less familiar.)

Especially important are modular forms which, viewed as sections of bundles on the modular curve, are holomorphic at the cusps. These are called cusp forms. The cusp forms of weight k are denoted S_k.

1.1.5 The Definitions

We now gather together the equivalent definitions above.

Definition. *A subgroup $\Gamma \in \mathrm{SL}_2(\mathbf{Z})$ is a congruence subgroup if there is some positive integer N such that, if*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \text{ and } \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \mod N$$

then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma \iff \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

The most important congruence subgroup for our purposes is

Definition.

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \ : \ c = 0 \mod N \right\}.$$

Definition. *Let $\mathcal{U}$ be*

$$\{(\omega_1, \omega_2) \in \mathbf{C}^2 \ : \ \mathrm{Im}(\omega_1/\omega_2) > 0\}.$$

Let $\mathcal{H}$ be the upper half plane in $\mathbf{C}$. (Note: as far as I know, there is no standard ntoation for $\mathcal{U}$.)

We now define the action of $\gamma \in \mathrm{Mat}(\mathbf{Z})$ on various spaces.

Definition. *For $(\omega_1, \omega_2) \in \mathbf{C}^2$, and $\gamma \in \mathrm{Mat}\mathbf{Z}$*

$$\gamma(\omega_1, \omega_2) = (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2).$$

For $z \in \mathbf{C}$,

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

This second action allows us to discuss the modular curve associated to a congruence subgroup.

Definition. *The modular curve X associated to a congruence subgroup Γ is the unique smooth Riemmann surface obtainable by compactifying the quotient of the $\mathcal{H}$ under the second action of γ above for all $\gamma \in \Gamma$. The points added in the compactification are called the cusps of the curve and can be associated to orbits of the above action on $\mathbf{Q} \cup \{\infty\}$. The modular curve associated to $\Gamma_0(N)$ is denoted $X_0(N)$.*

We now can define a modular form. Let k be an integer.

Definition. *A modular form f of weight k for Γ is equivalently,*

1. *A holomorphic function on $\mathcal{U}$ such that $f(\lambda\omega_1, \lambda\omega_2) = \lambda^{-k} f(\omega_1, \omega_2)$, $f(\gamma(\omega_1, \omega_2)) = f(\omega_1, \omega_2)$ and, for any a and $b \in \mathbf{Z}$, not both 0, $f(a + it, b) = O(t^{-n})$ as $t \to 0^+$ for some n.*
2. *A holomorphic function on $\mathcal{H}$ such that $(cz+d)^{-k} f(\frac{az+b}{cz+d}) = f(z)$ for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, $f(q + it) = O(t^{-n})$ as $t \to 0^+$ and $f(it) = O(t^n)$ as $t \to \infty$ for some n and all $q \in QQ$.*

3. *A section of the line bundle arising from the above transition maps (if k is even, this is the $k/2$ tensor power of the canonical bundle) holomorphic away from the cusps and meromorphic at the cusps.*

The space of modular forms is denoted M_k if Γ is clear; the space of modular forms for $\Gamma_0(N)$ is denoted $M_k(N)$.

Definition. *A cusp form f of weight k for Γ is a modular form of weight k for Γ obeying one of the following equivalent conditions.*

1. *For all a and $b \in \mathbf{Z}$ not both 0, $\lim_{t\to 0^+} f(a+it, b) = 0$.*
2. *For all $q \in \mathbf{Q}$ we have $\lim_{t\to 0^+} f(q+it) = 0$ and also $\lim^{t\to\infty} f(it) = 0$.*
3. *The section f is everywhere holomorphic.*

The space of cusp forms is denoted S_k if Γ is clear; the cusp forms for $\Gamma_0(N)$ is denoted $S_k(N)$.

1.2 Basic Tools

1.2.1 q-Series

We now introduce a representation of a modular forms known as a q-series. Suppose that Γ contains the matrix $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. Then f is periodic with period 1 and can thus be written as

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

We will regularly make the substitution $q = e^{2\pi i z}$ when writing this series. By the meromorphic condition, this sum contains only finitely many negative powers of q. This manner of expressing a modular form will become more important later.

1.2.2 The Petersson Inner Product

There is a Hermitian inner product on S_k known as the Petersson inner product. For $k = 2$, cusp forms can be viewed as differential one forms ω and η on the modular curve and the inner product is simply $<\omega, \eta>= \int_X \omega \wedge \overline{\eta}$. For general k, the Petersson inner product is slightly more complicated. We begin by noting that both Γ and $\mathbf{C}^*$ act on $\mathcal{U}$; the action of Γ has already been defined and the action of $\mathbf{C}^*$ is by $\lambda : (x, y) \to (\lambda x, \lambda y)$.

Definition. *Define a one form Ω on $\mathcal{U}$ by*

$$\Omega = \omega_1 d\omega_2 - \omega_2 d\omega_1.$$

Define a real valued function det *by*

$$\det(\omega_1, \omega_2) = \mathrm{Im}(\omega_1 \overline{\omega_2}).$$

The name det *is motivated by the fact that* $\det(\omega_1, \omega_2)$ *is the volume of the fundamental parrelleogram of* $\mathrm{span}(\omega_1, \omega_2)$. *View our modular forms f and g as functions on $\mathcal{U}$. Then all of Ω,* det *and f are Γ invariant. We have*

$$\begin{aligned} f(\lambda(\omega_1,\omega_2)) &= \lambda^{-k} f(\omega_1,\omega_2) && (1.1)\\ g(\lambda(\omega_1,\omega_2)) &= \lambda^{-k} g(\omega_1,\omega_2) && (1.2)\\ \Omega(\lambda(\omega_1,\omega_2)) &= \lambda^{2}\Omega(\omega_1,\omega_2) && (1.3)\\ \det(\lambda(\omega_1,\omega_2)) &= |\lambda|^2 \det(\omega_1,\omega_2) && (1.4)\\ &= \lambda\overline{\lambda}\Omega(\omega_1,\omega_2). && (1.5)\end{aligned}$$

Thus, the following differential $(1,1)$ *form*

$$f\overline{g}\ \det{}^{k-2}\Omega\overline{\Omega}$$

is $\mathbf{C}^*$ *invariant as well as* Γ *invariant and thus gives rise to a* $(1,1)$ *form on* X. *We define*

$$< f, g >= \int_X f\overline{g} \overset{k-2}{\det} \Omega\overline{\Omega}.$$

If we are viewing our modular forms f and g as functions on $\mathcal{H}$, this formula reduces to

$$\int_\Delta f\overline{g}y^{k-2}dzd\overline{z}$$

where Δ is a fundamental domain for the action of Γ and $y = \mathrm{Im}(z)$. The only part of this which might not be obvious is that $\det(z, 1) = y$; this follows either from the formula $\det(z, 1) = \mathrm{Im}(z\dot{1})$ or geometrically by noting that the lattice parallelogram has base 1 and height y.

Clearly, $< f, g >= \overline{< g, f >}$ and $< f, f >\geq 0$ with equality iff $f = 0$.

If f and g are just modular forms and not cusp forms, the integral above will not converge, this is why they were specified to be cusp forms. (Interestingly, if either f or g is a cusp form, this is enough to gaurantee convergence. The orthogonal complement of the cusp forms in the moduar forms is known as the Eisenstein series. These will play now significant role in the following.)

Chapter 2

Hecke Operators

We now move to one of the main subjects of this paper, the Hecke operators. These are linear operators which act on formal sums of lattices and thus induce a linear action on modular forms. There is a Hecke operater associated with each positive integer, denoted $T(n)$.

We will first define Hecke operators for $SL_2(\mathbf{Z})$ on formal sums of lattices. We will then show how this leads them to act on modular forms, including giving explicit formulas in terms of q-series. Of especial importance will be modular forms that are eigenvectors for every Hecke operater; these are caled eigenforms. The Hecke operators will be deliberately normalized so that, if $f = \sum a_n q^n$ with $a_1 = 1$ is an eigenform then $T(n)f = a_n f$. We will then derive certain relations between the Hecke operators and between the Hecke operators and the Petersson inner product. Using these, we will be able to derive relations between the coefficients of the q-expansion of an eigenform and show the eigenforms are a basis for the modular forms.

Finally, we will translate this theory to an arbitrary congruent subgroup. There will be some difficulties, but the sequence of events will be similar. Definition acting on formal sums of colored lattices, resulting action on modular forrms, formulas in terms of q-series, implications for eigenforms, relations between Hecke operators, relations between Hecke operators Petersson inner product and linear algebra consequences. It will not be true in general that a basis of eigenvectors will exist, but it will be close to true.

2.1 Hecke Operators for $\mathrm{SL}_2(\mathbf{Z})$

In this section, all Hecke operators will be for the group $\mathrm{SL}_2(\mathbf{Z})$.

2.1.1 Definitions and Formulae

Definition. *The nth Hecke operator of weight k, $T(n)$, to act on a lattice Λ by*

$$T(n) : \Lambda \to n^{k-1} \sum_{\mathcal{M} \subset \Lambda,\ [\Lambda:\mathcal{M}]=n} \mathcal{M}.$$

(The n^{k-1} is a normalization factor whose importance will emerge later). $T(n)$ is defined on formal sums of lattices by linearity. It is not immediately clear that this sum is finite, but it will follow from other formulas we will derive from this expression that it is. This formula is very useful theoretically, but not computationally; our next task will be to devise a computational form of this formula.

For this, we need a good way to describe all of the index n sublattices of Λ. Let $\mathcal{M}$ be such a sublattice, we know that $\mathcal{M} = [\gamma]\Lambda$ where γ is a member of $GL_2(\mathbf{Z})$ with determinant n. Let ω_1, ω_2 be a basis for Λ. Then $\mathrm{span}(\omega_2) \cap \Lambda = \mathrm{span}(d\omega_2)$ for some $d|n$, and we can take $d\omega_2$ to be one member of a basis for $\mathcal{M}$. Let $a\omega_1 + b\omega_2$ be another member of $\mathcal{M}$ which forms a basis for $\mathcal{M}$ together with $d\omega_2$. As $[\Lambda : \mathcal{M}] = n$, we have $ad = n$. Moreover, we can subtract multiples of $d\omega_2$ from $a\omega_1 + b\omega_2$ to make $0 \le b < d$. Thus, every $\mathcal{M}$ is of the form $[\gamma]\Lambda$ where $\gamma = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, $ad = n$, $0 \le b < d$ and it is easy to check that none of these are identical. So we get the new formula

$$T(n)\mathrm{span}(\omega_1, \omega_2) = n^{k-1} \sum_{ad=n, 0 \le b < d} \mathrm{span}(a\omega_1 + b\omega_2, d\omega_2).$$

To see how this acts on modular forms, viewed as functions on the upper half plane, we just plug in, we get

$$(T(n)f)(z) = n^{k-1} \sum_{ad=n, 0\le b<d} \frac{1}{(0z+d)^k} f(\frac{az+b}{0z+d}) = n^{k-1} \sum_{ad=n, 0\le b<d} \frac{1}{d^k} f(\frac{az+b}{d}).$$

Since the definition of the operater on lattices depended only on the lattice and not its basis and clearly commutes with scaling, $T(n)f$ will clearly be a modular form of weight k if f was.

Translating the previous formulae into a formula for q-series is also easy, although it gets complicated. Let $q = e^{2\pi i z}$ and write $f = \sum_j a_j q^j$. Then

$$T(n)f = n^{k-1} \sum_{ad=n, 0\le b<d} \frac{1}{d^k} \sum_j a_j e^{2\pi i \frac{az+b}{d} j} = n^{k-1} \sum_{ad=n, 0\le b<d} \frac{1}{d^k} \sum_j a_j q^{\frac{a}{d} j} e^{2\pi i \frac{b}{d} j}.$$

(Which (a/d)-th root of 1 is chosen is not important as long as the same one is used every time, because the sum on b has the effect of causing the expression $q^{a/d} e^{2\pi i b/d}$ to take each possible value once.)

Significant simplification can be obtained by regrouping to sum first on b to give

$$n^{k-1} \sum_{ad=n} \sum_j a_j q^{aj/d} \sum_{0\le b<d} \frac{1}{d^k} e^{2\pi i b j/d}.$$

The inner sum, $\sum_{0\le b<d} e^{2\pi i b j/d}$ is 0 unless $d|j$, in which case it is d. Writing $j = dl$ and using the above observation, we get

$$n^{k-1} \sum_{ad=n} \frac{1}{d^k} \sum_l a_{dl} q^{al} d = \sum_{ad=n} a^{k-1} \sum_l a_{dl} q^{al}.$$

Summarizing the above, we get

Proposition 1. *The coefficient of q^j in $T(n) \sum a_i q^i$ is*

$$\sum_{al=j, ad=n} a^{k-1} a_{dl}.$$

(The sum is over all ordered triples (a, l, d) such that $a = j$ and $ad = n$.)

This formula is not especially worth remembering, but some aspects of it are. The first is that the coefficients of the new q-series are integer combinations of the coefficients of the old one. The second is that, if the previous q-series contains no terms with negative exponent, neither does the new and if the old series has constant term 0 so does the new. Thus, $T(n)$ preserves modular forms and cusp forms.

Finally, the special case $j = 1$ is very important. In this case, the formula simplifies to just a_n. (This was the point of the n^{k-1} normalization.) This has several important consequences. Suppose that f is a cuspform and is an eigenform for every $T(n)$. Then the complete sequence of $T(n)$ determine f up to scaling. If (with the same assumptions on f) a_1 is 0, then every a_n is 0. For $f \neq 0$ and meeting the above conditions, we can assume therefore than $a_1 = 1$. We thus have

Proposition 2. *Let $f \neq 0$ be an eigenvector of every $T(n)$. Then we can normalize f so that $a_1 = 1$. We then have*

$$T(n)f = a_n f.$$

Such an f will be called a normalized eigenform. For such an f, we have

2.1.2 Relations Between the Hecke Operators

We will now derive two important relations between the $T(n)$. To experss them, we will need the operater $R(n)$, which acts on formal sums of lattices by sending $\Lambda \to n^{2k-2}(n \cdot \Lambda)$. (The simplest way to remember these peculiar factors up front is that $T(n)$ has an image consisiting of lattices of index n, so there is an n^{k-1} in front; $R(n)$ has an image of latttices of index n^2 so it has a $n^{2(k-1)}$.) By definition, this induces an action on modular forms by $R(n)f = n^{2k-2} \cdot n^{-k} f = n^{k-1} f$. The two relations are

Proposition 3.

$$T(m)\circ T(n)=T(mn)\ (m,n)=1$$

and

$$T(p^{\alpha+1})=T(p)\circ T(p^{\alpha})-pR(p)\circ T(p^{\alpha-1}).$$

We will delay proving these because it will be better to prove them at the same time as proving the slightly more complicated relations obeyed by Hecke operators for general congruence subgroups. Instead, we note several consequences.

By recursively using the two formulas, we see that every Hecke operater can be expressed as a polynomial in $R(p)$ and $T(p)$. The first relation shows that $T(p)$ and $T(q)$ commute where p and q are distinct primes and it is onbiovus that the $R(p)$ commute with each other and with all $T(p)$. Therefore, the Hecke operators all commute with each other. Also, if f is an normalized eigenform for all of the $T(n)$, plugging in the formulae $T(n)f=a_nf$ and $R(n)f=n^{k-2}f$ give

$$a_{mn}=a_ma_n\ (m,n)=1$$

and

$$a_{p^{\alpha+1}}=a_pa_{p^{\alpha}}-p^{k-1}a_{p^{\alpha-1}}.$$

These can be regrouped into the simpler identity

$$\sum_n\frac{a_n}{n^s}=\prod_p\frac{1}{1-a_pp^{-s}+p^{k-1-2s}}.$$

(Note: historically, this was how the Hecke operators were discovered. Ramanujan observed the above identities in the case of a certain normalized eigenform of weight 12 and Hecke invented his operators to prove Ramanujan's observations.)

2.1.3 Self-Adjointness of the Hecke Operators

We now state the important relation between the Hecke operators and the Petersson inner product:

$$<T(n)f,g>=<f,T(n)g>.$$

The term for this behavior is that $T(n)$ is self-adjoint. This is important because it is known that a commuting ring of self-adjoint operaters on a finite dimensional vector space always decomposes into a sum of simultaneous eigenspaces for all of the operaters. Applying this result to the space of cusp forms, we see that the eigenforms we have proved results about above do exist. Moreover, since we showed that the set of eigenvalues of a normalized eigenform determines it, we know that each of these spaces is one-dimensional. We postpne proving this as well until we can deal with Hecke operators for a general $\Gamma_0(N)$.

2.2 General Hecke Operaters

2.2.1 Definitions and Formulas

We now define the Hecke operaters on modular forms for $\Gamma_0(N)$. We can interpret the members of $\Gamma_0(N)$ as functions on lattices with a cyclic sublattice of index N being colored in a different color from the other points. Let (Λ,C) represent such a lattice and its coloring. Then we can define the $T(n)$ as a formal sum on such pairs (Λ,C) by

$$T(n)(\Lambda,C)=\sum_{\mathcal{M}\subset\Lambda,[\Lambda:\mathcal{M}]=n,D=C|_{\mathcal{M}}}(\mathcal{M},D).$$

We work out the formula explicitly. In this case, a point $a\omega_1+b+\omega_2$ is colored one color, say blue, if $a=0\mod N$ and another, say red, otherwise. A sublattice $\mathcal{M}$ of index n will have the appropriate coloring induced iff there are points in $\mathcal{M}$ with ω_1 coordinate relatively prime to N. (Otherwise, the blue points of $\mathcal{M}$ will have index $N/\mathrm{GCD}(N,n)<N$ in $\mathcal{M}$.) We have seen before that every index n sublattice of $\mathcal{M}$ has

a unique basis of the form $a\omega_1 + b\omega_2$, $d\omega_2$, $ad = n$, $0 \le b < d$. It is clear that the opssible ω_1 coefficients in such a lattice are the multiples of a, so this lattice is aceptable iff $(a, N) = 1$ in which case this basis is a compatible one with the coloring. Thus we get

$$T(n)(\omega_1, \omega_2) = n^{k-1} \sum_{ad=n, 0\le b<d, (a,N)=1} (a\omega_1 + b\omega_2, d\omega_2).$$

As before, this gives an action on the upper half plane and on q-series. As 1 is relatively prime to everything, we still get all the $q^{1/n}$ terms and thus again get that the coefficient of q in $T(n)f$ is the coefficient of q^n in f.

In general, congruence subgroups for which the $\begin{pmatrix} 1 & b \\ 0 & n \end{pmatrix}$ are not acceptable sublattices do not have interesting behavior with respect to the Hecke operaters as this property no longer holds and we thus can not use the eigenforms of the Hecke operaters to extract coefficients of the q-series. Requiring that these matrices be acceptable basis changes amounts to requiring we consider colorings in which lines parallel to ω_2 recieve the same color, and thus congruence subgroups containing some $\Gamma_1(N)$. For any Γ containing a $\Gamma_1(N)$, we have that the coefficient of q in $T(n)f$ is that of q^n in f by computations similar to the above.

2.2.2 Proof of the Relations

We will now prove the relations

Proposition 4.

$$T(mn) = T(m) \circ T(n)$$
$$T(p^{\alpha+1}) = T(p) \circ T(p^\alpha) - pR(p) \circ T(p^{\alpha-1}),\ (p, N) = 1$$
$$T(p^{\alpha+1}) = T(p) \circ T(p^\alpha)$$

where $R(n)(\Lambda, C)$ now means $n^{2k-2}(n\Lambda, C|_{n\Lambda})$ if $C|_{n\Lambda}$. Note that, for a normalized eigenform $\sum a_n q^n$, these give

$$\sum \frac{a_n}{n^s} = \prod_p \frac{1}{1 - a_p p^s + \delta_p p^{k-1-2s}}$$

where $\delta_p = 0$ if $p|N$, 1 otherwise.

Proof. Apply each side of the formula to (Λ, C). In both cases, the extra factors of n^{k-1} and n^{2k-2} cancel out. The first formula thus becomes

$$\sum_{\Lambda''\subset\Lambda, [\Lambda:\Lambda'']=mn, C''=C|_{\Lambda''}} (\Lambda'', C'') = \sum_{\Lambda''\subset\Lambda', [\Lambda':\Lambda'']=m, C''=C'|_{\Lambda''}} \sum_{\Lambda'\subset\Lambda, [\Lambda:\Lambda']=n, C'=C|_{\Lambda'}} (\Lambda'', C'')$$

In other words, we are being asked to show that each (Λ'', C'') on the left arises from a unique chain $\Lambda \supset \Lambda' \supset \Lambda''$ on the right. If we disregard the colorings, this is easy. Λ/Λ'' is an abelian group of order mn. It follows from the classification fo abelian groups that Λ/Λ'' contains a unique subgroup of order m (using that $(m, n) = 1$). This corresponds to a unque lattice Λ' with $\Lambda \supset \Lambda' \supset \Lambda''$ and $[\Lambda' : \Lambda''] = n$. We now must see that Λ' is appropriately colored. But this is simple. A sublattice of Λ will have an appropriate subcoloring induced iff it contains points with ω_1 coordinate relatively prime to N. If Λ'' has this property, then so must Λ' which contains it.

For the second property, regroup this equation as $T(p^{\alpha+1}) + pT(p^{\alpha-1}R(p)) = T(p)T(p^\alpha)$. Note that, as $(p, N) = 1$, any sublattice of Λ with index a power of p will automatically be appropriately colored, so we just need to check that the same lattices appear with the same multiplicities on each side. For any lattice $\Lambda'' \subset \Lambda$ of index $p^{\alpha+1}$, the number of times Λ'' appears on the right is 1 if Λ'' is not a sublattice of $p\Lambda$ and $p + 1$ if it is.

On the right, the number of times Λ'' appears is the number of lattices Λ' such that $\Lambda \subset \Lambda' \subset \Lambda''$ and $[\Lambda' : \Lambda''] = p$. Using the equivalence between lattices in between Λ and Λ'' and subgroups of Λ'', the right hand side is the number of subgroups of order p in Λ/Λ''. Now, Λ/Λ'' is an abelian group of order $p^{\alpha+1}$. Moreover, it can be generated by two elements, as it is a quotient of $\Lambda = \mathbf{Z}^2$. So $\Lambda/\Lambda'' = (\mathbf{Z}/p^{\alpha+1})$ or

$(\mathbf{Z}/p^\beta) \oplus (\mathbf{Z}/p^\gamma)$ for $\beta, \gamma > 0$, $\beta + \gamma = \alpha + 1$. In the first case, there is one subgroup of order p. In the second, there are $p+1$. Moreover, in the first case, $p\Lambda$ is not a sublattice of Λ'' and in the second it is, so the conclusion is proven.

To show the result where $p|N$, we use the same proof as the above, except that we now have to consider whether the coloring will be appropriate. Any sublattice of $p\Lambda$ can not be correctly colored, as $p|N$. On the other hand, if Λ'' is correctly colored, so is any Λ' containing it. Thus we get that each side is simply the correctly colored subleattices of Λ of index $p^{\alpha+1}$. $\square$

2.2.3 Proof of Self-Adjointness

We now need to discuss the relation between the Hecke operaters and the Petersson inner product. Specifically, we shall prove that

Proposition 5. *If* $\mathrm{GCD}(N,n) = 1$ *then* $< f, T(n)g > = < T(n)f, g >$.

Proof. We need to set up some geometry. Let $\tilde{X}$ the quotient of Γ acting on $\mathcal{U}$. Let X in turn be the modular curve. We saw before that we could write the Petersson inner product as

$$\int_X f(\Lambda)\overline{g}(\Lambda)\Omega(\Lambda)\overline{\Omega}(\Lambda)\det(\Lambda)^{k-2}$$

where Λ is short for (ω_1, ω_2) and $\Omega = \omega_2 d\omega_1 - \omega_1 d\omega_2$. Note that this can also be seen as the integral in $\tilde{X}$ over any section of the projection map to X. The advaantage of this second form is that we have a simlpe way of thinking of $T(n)$ as acting on formal sums of points in $\tilde{X}$.

Let $\tilde{\Pi}_n \subset \tilde{X} \times \tilde{X}$ be

$$\{(\Lambda_1, \Lambda_2) | \Lambda_2 \text{ appears in } T(n)\Lambda_2\}$$

and Π_n be the image of $\tilde{\Pi}_n$ under the projection to $X \times X$. Then we can write

$$< f, T(n)g > = \int_{\text{any section of } X \text{ in} \tilde{X}} f(\Lambda)\Omega(\Lambda)\det(\Lambda)^{k/2-1}T(n)\overline{g}(\Lambda)\overline{\Omega}(\Lambda_2)\det(\Lambda)^{k/2-1}.$$

Now, any for any $\mathcal{M}$ appearing in $T(n)\Lambda$, $[\Lambda : \mathcal{M}] = n$ and therefore $\det(\mathcal{M}) = n\det(\Lambda)$ and $\Omega(\mathcal{M}) = n\Omega(\Lambda)$ so we can rewrtie the above as

$$\int_{\text{any section of } X \text{ in} \tilde{X}} f(\Lambda)\Omega(\Lambda)\det(\Lambda)^{k/2-1}n^{k-1} \sum_{\mathcal{M} \in T(n)\Lambda} \overline{g}(\mathcal{M})\frac{1}{n}\overline{\Omega}(\mathcal{M})\frac{1}{n^{k/2-1}}\det(\mathcal{M})^{k/2-1} =$$

$$n^{k/2-1} \int_{\text{any section of } \Pi_n \text{ in } \tilde{\Pi}_n} f(\Lambda)\Omega(\Lambda)\det(\Lambda)^{k/2-1}\overline{g}(\mathcal{M})\frac{1}{n}\overline{\Omega}(\mathcal{M})\frac{1}{n^{k/2-1}}\det(\mathcal{M})^{k/2-1}.$$

This integrand is symmetric under scaling and vainshes on the fibers of $\tilde{\Pi}_n \to \Pi_n$ so it can be rewritten as

$$n^{k/2-1} \int_{\Pi_n} f(\Lambda)\Omega(\Lambda)\det(\Lambda)^{k/2-1}\overline{g}(\mathcal{M})\frac{1}{n}\overline{\Omega}(\mathcal{M})\frac{1}{n^{k/2-1}}\det(\mathcal{M})^{k/2-1}.$$

Now, the integrand is clearly conjugate-symmetric in Λ and $\mathcal{M}$, so if we prove that Π_n is symetric under the interchange of Λ and $\mathcal{M}$, we will have $< f, T(n)g > = < T(n)f, g >$ as promised. We will prove this by showing that $\mathcal{M}$ appears in $T(n)\Lambda$ iff $n\Lambda$ apears in $T(n)\mathcal{M}$ when $(N,n) = 1$.

As $(N,n) = 1$, the colorings on $\mathcal{M}$ and $n\Lambda$ will automatically be correct so we just need to check that we get the same sets of lattices. If $\mathcal{M} \subset \Lambda$, $[\Lambda : \mathcal{M}] = n$, then $\Lambda/\mathcal{M}$ is a group of order n, so every element in it has order dividing n which means $n\Lambda \subset \mathcal{M}$. THe indices work out because $[\mathcal{M} : n\Lambda] = [\Lambda : n\Lambda]/[\mathcal{M} : \Lambda] = n^2/n = n$. Similarly, if $n\Lambda \subset \mathcal{M}$, $[\mathcal{M} : n\Lambda] = n$, then by the above logic $n\mathcal{M} \subset n\Lambda$ so $\mathcal{M} \subset \Lambda$ and the indices work out as before. $\square$

2.2.4 Linear Algebra

The $T(n)$ are all commuting matrices acting on S. Moreover, for n relatively primes to N, they are self-adjoint with respect to $<,>$. In this section, we see what this permits us to deduce about the action of the $T(n)$.

Proposition 6. *We can uniquely decompose S as $\oplus S_i$ such that for each S_i there exists a different seqeunce a_2^i, a_3^i, ... such that $T(n) - a_n^i$ is nilpotent on S_i. Within such a subspace, there is a unique one-dimensional subspace on which every $T(n) - a_n = 0$. For $\mathrm{GCD}(n, N) = 1$, we have $T(n) = a_n$ on all of the subspace.*

Proof. We can find a decomposition of S into a direct sum of subspaces on each of which $T(2) - a_2$ is nilpotent for some a_2 dependant on the subspace. (This is a consequence of the Jordan Canonical Form from standard linear algebra.) Each $T(n)$ will preserve these subspaces, as $T(2)$ commutes with $T(n)$ so $T(n)(T(2) - a_s)^k v = 0 \iff (T(2) - a_2)^k T(n)v$. Thus, on each of these subspaces, we may further subdivide it into subspaces on which there exist numbers a_3 such that $T(3) - a_3$ is nilpotent. Continuing in this way, we find the claimed decomposition.

We now restrict our attention to one term of the decomposition. Consider the kernel of $T(2) - a_2$, we know that it is nnonempty and note that it is preserved by all the $T(n)$. We know that $T(3) - a_3$ is nilpotent and thus remains so restricted to the kernel of $T(2) - a_2$, so $T(3) - a_3$ has a nontrivial kernel restricted to this space. Continuing this way we find a subspace on which all $T(n) - a_n$ are zero.

This subspace is one dimensional by CITE. Finally, the claim when $\mathrm{GCD}(N, n) = 1$ follows because self-adjoint matrices are diaganolizable. □

2.2.5 Old and New Forms

The previous section left it unclear as to whether we would ever actually fail to have a decomposition of S into eigenforms. It did, however, provide a good idea of what a form on which the $T(n) - a_n$ were nilpotent but not zero would have to look like. Let $\sum b_i q^i$ be a form and a_n a sequence of numbers such that $(T(n) - a_n)^k \sum b_i q^i = 0$ for some k. We then know that $\sum b_i q^i$ is also a modular form in S. Moreover, when $\mathrm{GCD}(N, n) = 1$, we know $a_n = b_n$. So $\sum (a_i - b_i) q^i$ is a form only containing coefficients for terms of exponent having a factor dividing N. There is a known way to make such forms: for any $l|N$ and $f \in S(N/l)$, consider $f(q^l)$. It is easy to check that this lies in $S(N)$. We may also make forms for which $T(n) - a_n$ is zero for $\mathrm{GCD}(N, n) = 1$ another way: let $f \in S(N/l)$, simply consider f as a member of $S(N)$. When $\mathrm{GCD}(n, l) = 1$, $T(n)$ is computed by the same formula over $\Gamma_0(N)$ and $\Gamma_0(N/l)$ so $T(n)f$ will equal a_n both times. We thus have shown

Proposition 7. *For some seqeunce a_n let $V \subseteq S$ be the subspace on which $T(n) - a_n$ is nilpotent for every n. Let $v \in V$ be $\sum a_i q^i$. Then there is a subspace W of V such that $V = \mathrm{span}(v, W)$ and, for any $\sum b_i q^i \in W$, we have $\mathrm{GCD}(N, n) = 1 \Leftarrow b_n = 0$.*

Forms created via either of the two constructions above, or linear combinations of such forms, are called *old forms*. There is a more geometric interpretation of this contruction: for $(\Lambda, \mathcal{M})$ a pair of lattices of the sort parameterized by $\Gamma_0(N)$, that is, such that $\Lambda/\mathcal{M} = \mathbf{Z}/N$ and for any $d|N$, define $\Lambda_d \mathcal{M} = \mathrm{span}(d\Lambda, \mathcal{M}) = \Lambda \cap (1/d)\mathcal{M}$. Another definition is that this is the preimage under $\Lambda \to \Lambda/\mathcal{M}$ of the unique order d subgroup of $\mathbf{Z}/N$. Then we get two maps, $\pi_1 : \Gamma_0(d) \to Gamma_0(N)$ and $\pi_2 : \Gamma_0(N/d) \to Gamma_0(N)$ by

$$(\pi_1 f)(\Lambda, \mathcal{M}) = f(\Lambda, \Lambda_d \mathcal{M})$$

$$(\pi_2 f)(\Lambda, \mathcal{M}) = f(\Lambda_(N/d)\mathcal{M}, \mathcal{M})$$

(Note all of this notation is nonstandard, I haven't been able to find a standard notation for most of the relevant maps here.) The old forms are the subspace of $S(N)$ spanned by $\pi_1 S(d)$ and $\pi_2 S(d)$ for all $d|N$, $d < N$. We define the space of *new forms* to consist of those fomrs orthogonal under $<,>$ to the old forns. To recap

Definition. *An old form for $\Gamma_0(N)$ is a form that is a linear combination of forms of the forms $f(z)$ and $f(z^d))$ with f a modular form for $\Gamma_0(N/d)$, $d|N$, $d > 1$. A new form is a modular form g for which $< f, g >= 1$ for all olf forms f.*

Proposition 8. *The old and new forms are stable under the $T(n)$ for* $\mathrm{GCD}(N,n)=1$.

Proof. To say that the old forms are stable is just to note that $T(n)$ commute with the π's. This is true because the hypothesis that $\mathrm{GCD}(N,n)=1$ implies that $T(n)$ on $S(d)$ is given by the same formula for all $d \mid N$. These $T(n)$ are self-adjoint, so they also preserve the orthogonal complement of the old forms. $\square$

Our goal for the next part of this section is to prove the following theorem CITE

Theorem 1. *Using the definitions of the above propostion, W has a basis of forms of the form $f(z^p)$ where $f \in S(N/p)$ for p some prime dividing N. (Different basis elements may have different values of p.)*

Proof. Our proof will be the easy combination of several lemmas. p always denotes a prime dividing N.

Lemma 1. *Let $\Gamma(N/p,p)$ be the group of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $c=0 \mod N/l$ and $b=0 \mod l$. Then $\Gamma_0(N/p)$ is generated by $\Gamma(N/p,p)$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.*

Proof. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N/l)$. First, suppose p does not divide a. Then we can find k such that p divides $ak+b$ and thus $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^k = \begin{pmatrix} a & ak+b \\ c & ck+d \end{pmatrix}$ lies in $\Gamma(N/p,p)$. Now suppose l divides a. Then p can not divide c, as $ad-bc=1$. Thus, $p \nmid a+c$ and we can apply the previous analysis to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix}$. $\square$

Lemma 2. *Let $g(q^p)=\sum b_i q^{ip} \in S(N)$. Then $g(q)=\sum b_i q^i \in S(N/p)$.*

Proof. We switch back to the upper halfplane notation, writing $g(\tau)=\sum b_i e^{2\pi i \tau}$. By the previous lemma, we must check that g is preserved by $\Gamma(N/p,p)$ and by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Clearly, g is preserved by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}: \tau \to \tau+1$. Now, suppose $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(N/p,p)$. Then $\begin{pmatrix} a & b/l \\ cl & d \end{pmatrix} \in \Gamma_0(N)$. By assumption, $g(p\tau) \in S(N)$ which means

$$g(p\tau) = \left(\frac{1}{cp\tau+d}\right)^k g(p\frac{a\tau+b/p}{cp\tau+d}).$$

Substituting τ/p for τ in the above

$$\begin{aligned} g(\tau) &= \left(\frac{1}{c\tau+d}\right)^k g(p\frac{a\tau/p+b/p}{c\tau+d}) \\ &= \left(\frac{1}{c\tau+d}\right)^k g(\frac{a\tau+b}{c\tau+d}). \end{aligned}$$

This precisely says g is preserved by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. $\square$

Lemma 3. *We can define a map U from $S(N)$ to $S(N/p)$ as follows: let π be the map $X(N) \to X(N/p)$. Define $(Ug)(x) = (1/\det \pi)\sum_{\pi(y)=x} g(y)$ for $x \in X(N/p)$ and $g \in S(N)$. If $g(q)=\sum b_i q^{pi}$, then $(Ug)(q) = \sum_i q^i$.*

Proof. The first assertion is obvious. As for the second, we know by the above lemma that $\sum b_i q^i$ is in $S(N/p)$. The map π in the q-coordinate is simply $q \to q^p$ so $\sum b_i q^{pi} = \pi^* \sum b_i q^i$ and the result is obvious. $\square$

(Incidentally, another way to define U is to consider the injection $S(N/p) \to S(N)$. This gives a map $S(N)^* \to S(N/p)^*$. We can then idenitfy $S(N)$ and $S(N/p)$ with their duals by $<,>$.)

Lemma 4. *Let $g=\sum b_i q^i \in S(N)$. Then $\sum b_{pi} q^i \in S(N)$ and $\sum b_{pi} q^{pi} \in S(pN)$.*

Proof. As $p|N$, the Hecke operater $T(p)$ is simply $T(p) : \sum b_i q^i \to \sum b_{pi} \to q^i$. We thus know $\sum b_{pi} q^i$ is a modular form for $\Gamma_0(N)$. We already observed that if $f(q)$ is a modular form for $\Gamma_0(N)$ then $f(q^p)$ is for $\Gamma_0(pN)$. □

Lemma 5. *Let $g = \sum b_i q^i \in S(N)$ and let d be divisible only my those primes divinding N. Then $\sum b_{di} q^i \in S(N)$ and $\sum b_{di} q^{di} \in S(dN)$.*

Proof. This is an easy induction on d from the previous lemma. □

We now prove the theorem proving the following stronger claim by induction on r.

Claim. *Let $p_1, \ldots\ p_r$ be distinct primes dividing N. Suppose $f(q) = \sum a_i q^i \in S(N)$ such that $a_i = 0$ for all i not divisible by any p_k. Then we can find $f_i \in S(N/p_i)$ such that $f(q) = \sum f_i(q^{p_i})$.*

If $r = 0$, the claim is vacuous. Set $p_r = p$ and $M = p_1 \cdots p_{r-1}$ and write $f(q) = g(q) + h(q)$ where

$$\begin{aligned} g(q) &= \sum_{\mathrm{GCD}(i,M)=1} a_i q^i \\ h(q) &= \sum_{\mathrm{GCD}(i,M)\neq 1} a_i q^i \end{aligned}$$

By inclusion-exclusion, we can write $g(q)$ and $h(q)$ as a sum of $\sum a_{di} q^{di}$ for various d dividing M. Thus $g,\ h \in \bigcup_{d|M} S(dN) = S(MN)$. (The level isn't important, we just need to know they are in some $S(K)$ so that we may apply U to them.) All the coefficients of g are divisible by p, so by Lemma CITE we have $Ug \in S(NM/p)$ and $(Ug)(q^p) = g$. Consider $f(q) - (Uf)(q^p) = g(q) + h(q) - (Ug)(q^p) - (Uh)(q^p) = h(q) - (Uh)(q^p)$.

Now, we can use the inclusion exclusion trick to write $h(q) = \sum_{i=1}^{r-1} h_i(q^{p_i})$ for some $h_i \in S(NM)$. We can check that we then have $h(q) - (Uh)(q) = \sum_{i=1}^{r-1} h_i'(q^{p_i})$ where $h_i'(q) = h_i(q) - (Uh_i)(q^p) \in S(NM)$.Using the relation $h(q) - (Uh)(q^p) = f(q) - (Uf)(q^p)$ we see $h(q) - (Uh)(q^p) \in S(N)$. On the other hand, using $h(q) - (Uh)(q) = \sum_{i=1}^{r-1} h_i'(q^{p_i})$ we see that $h(q) - (Uh)(q^p)$ obeys the hypotheses of the claim with r replaced by $r - 1$. So we can write $h(q) - (Uh)(q) = \sum_{i=1}^{r-1} f_i(q^{p_i})$. Seeing $f_r(q) = (Uf)(q)$, we have

$$\begin{aligned} \sum_{i=1}^{r} f_i(q^{p_i}) &= \sum_{i=1}^{r-1} f_i(q^{p_i}) + (Uf)(q^{p_r}) \\ &= h(q) - (Uh)(q^{p_r}) + f(q^{p_r}) \\ &= f(q) - (Uf)(q^{p_r}) + f(q^{p_r}) = f(q) \end{aligned}$$

□

We can now show

Theorem 2. *The new forms have a basis of mutually orthogonal eigenforms.*

Proof. We denote the spaces of old and new forms S^o and S^n. We know that S is the direct sum of orthogonal spaces on which there exist a seqeunce a_n such that the $T(n) - a_n$ are nilpotent. It will be enough to show that the new forms are diagonalizable on there intersection with every such space. Let V be the subspace on which $T(n) - a_n$ are nilpotent for some choice of a_n. Call n *easy* if $\mathrm{GCD}(N, n) = 1$. (This is my terminology, I don't think there is a standard shorthand.)

By the above theorem, $V \cap S^n$ is of dimension at most 1. We claim that in fact, there are only two possibilies: $\dim V = 1$ (in which case its generator is clearly an eigenvector) or $V \cap S^n = \{0\}$. If these are both violated, there is a new form f whose span is $V \cap S^n$ and an old form g lying in $V \cap S^o$, both of which are eigenvectors for the $T(n)$ with eigenvalue a_n when n is easy. g is a linear combination $\sum g_M(q^d)$ of forms induced from lower levels M. By inducing from further down if necesary, we may arrange that all of these forms are new forms and by induction on the level may arrange that they are all eigenforms. Then, deleting terms that combine to zero if necesary, every $g_M(q^d)$ will be an new eigenvector with eigenvalues a_n for the easy n. Let $h(q^d)$ be one such example. Then $h(q)$ also has these eigenvalues and, as h is new on $S(M)$ for some M, h has nonzero leading term.

We can thus find a linear combination of f and h where the coefficient of q^n is 0 for every easy n. Such a form must be old. But then f is a linear combination of this form and the old form h, contradicting that h is new. $\square$

Corrollary. *S^n is preserved by all of the $T(n)$ and the $T(n)$ are selfadjoint when restricted to it.*

Chapter 3

The Abelian Varieties Associated to Modular Forms

3.1 The Action on the Jacobian

Recall that for any Riemann surface X with a specified point x_0 on it, we define an abelian variety J called the Jacobian of X with a map $i : X \to J$ such that for any set of points $y_1, \ldots, y_k \in X$ the divisor $\sum(y_i - x_0)$ is principle iff $\sum i(y_i) = 0$. There is another way to describe the structure of J: let Ω be the space of (holomorphic) differential forms on X and Ω^* its dual. For any path γ on X from x_0 to y we can define a path in Ω^* by integration. That is, $\int_\gamma \cdot$ defines a linear form on Ω and this linear form varies as we grow γ from x_0 to y. Locally, this lets us embed a neighborhood of X into Ω^*. If we consider the images of all closed loops, the collection of which is $H_1(X, Z)$, we get a lattice on Ω^*. We can then write J as the quotient of Ω^* by that lattice, and we will have a genuine map from X to J as the quotienting will remove dependence on pathes. Those not familiar with this picture should consult CITE.

When X is $X_0(N)$ and Ω is thus $S(N)$, the Hecke operaters act on every space in the preceeding paragraph. The Hecke operaters send points on X to formal sums of points on X and thus act on divisors and particularly on degree 0 divisors. The Hecke operaters act on S and on S^*. The Hecke operators act on $H_1(X, Z)$. (If γ is a closed path on X, applying $T(n)$ pointwise to γ will give a finite sum of closed paths.) Hecke operaters act on functions on X and thus act on principal divisors. Seeing J either as degree 0 divisors modulo principal divisors or as $\Omega^*/H_1(X, Z)$, $T(n)$ act on J. And all of these actions are compatible.

We showed in the previous section that the action of $T(n)$ on S decompose S into invariant subspaces. Of particular interest are the subspaces on which all the $T(n)$ act by mutliplication by some a_n. We now look at how this decomposition relates to the embedding of $H_1(X, z)$ in Ω. We will denote $H_1(X, Z)$ by H for short. It always possibel that H has subspaces on which $T(n)$ acts by multiplication by a_n: the A_n need not be and often are not integers.

Simply from the fact that the $T(n)$ acts both on H and S^* compatibly, we can deduce a great deal about the nature of the a_n.

Proposition 9. *The eigenvalues of the $T(n)$ are algebraic integers. For any n the set of its eigenvalues is preserved under Galois conjugacy. Chose some particular seqeunce of a_n which are the eigenvalues of $T(n)$ applied to some $v \in S$, then the a_n generate a number field of finite degree over $\mathbf{Q}$. Let K be the normal closure of this field. We can find a unique subspace V of S containing v of (complex) dimension equal to $[K : \mathbf{Q}]$ on which we may identify V with $K \otimes_{\mathbf{Q}} \mathbf{C}$ such that the $T(n)$ act by a_n. Within V, the lattice $H \cap V$ has rank $2[K : \mathbf{Q}]$. The same comments apply if we instead work with $H^1(X, Z)$ in $S(N)$ rather than $H_1(X, Z)$ inside $S^*(N)$.*

Proof. WRITE THIS $\square$

The V's inside S^*, because they contain sublattices of H of the appropriate rank, give rise to subvarieties of the Jacobian J. Surprisingly, even though we can choose the V to only intersect at 0, they can still give rise to intersecting varieties in J other than at the identity. On the other hand, when we have a V in S, we get a surjection $S^* \to V^*$ and hence an abelian variety as a surjective image of J. This case is generally more interesting, because we have a map $X \to J$ so we get maps from X to abelian varieties by composing. In particular, if the V in question is one dimensional, we get a map from X to an elliptic curve.

Such an elliptic curve is called modular. We get a modular elliptic curve whenever we have an eigenform with rational coefficients. Classically, this situation was understood by saying that an elliptic curve $E = \{y^2 = x^3 + Ax + B\}$ was modular if one could find modular forms y and x obeying this equation; we now see that such modular forms would be the pullbacks to X from E of y and x as functions on E. As yet, modularity of an ellitic curve appears to be at best a peculiarity of the curve. We will eventually see that the modular form, after a slight transformation, gives the L-series of the elliptic curve, which encodes a great deal of the number theoretic properties of the curve.

3.2 Definedness over Q

REWRITE THIS

3.3 Reduction Modulo p

REWRITE THIS

Chapter 4

Elliptic Curves

4.1 Elliptic Curves over Finite Fields

Definition. *Let E be an elliptic curve defined over F a finite field of q elements. Call a divisor (formal sum of points) on E an F rational divisor if it is stable under Galois action. For such a divisor D, let $|D|$ be $q^{\#D}$ where $\#D$ is the number of points in D. The ζ function of the elliptic curve is*

$$\zeta(E,s) = \sum_{\substack{D \text{ effective} \\ F \text{ rational}}} \frac{1}{D^s}.$$

This definition was chosen to make the ζ function of the curve look as much like the classical ζ function as possible, we will now explore some alternate forms. The machinery of the proofs in this section will return later. Let X denote the collection of orbits of $\overline{F_p}$ points under $\mathrm{Gal}(\overline{F_q}/F_q)$. (For those who are comfortable enough with algebraic geometry over finite fields, X simply is the collection of closed points of E considered as a variety over F.) Just as in number theory, we can factor ζ into a product:

$$\zeta(E,s) = \prod_{x \in X} (1 - \frac{1}{|x|^s})^{-1}.$$

Let M_d be the number of orbits in P of length d. This can be rewritten

$$\zeta(E,s) = \prod_{d} (1 - \frac{1}{q^{ds}})^{-M_d}.$$

Taking logs

$$\begin{aligned} \log \zeta(E,s) &= \sum_d M_d \sum_k \frac{1}{k} p^{-kds} \\ &= \sum_n \frac{q^{-ns}}{n} \sum_{d|n} M_d d. \end{aligned}$$

(We have subsituted $n = kd$.) The inner sum is precisely the number of points of E in F_{q^n}, which we will denote N_n. We now must think about how to compute this number.

The obvious observation is that, letting Φ denote the Frobenius, this is the number of fixed points of Φ^n. An elliptic curve has a group law, which gives us a better way to compute this: N_n is the number of points in the kernel of $\Phi^n - \mathrm{Id}$. Let ω be the invariant differential on E. We know $\Phi^*\omega = 0$ so $(\Phi^n - \mathrm{Id})\omega = -\omega \neq 0$ so $\Phi^n - \mathrm{Id}$ is seperable. Thus, the size of its kernel is simply the size of the preimage of 0 which is the degree of the map. Thus, we have

$$\log \zeta(E,s) = \sum_n \frac{\deg(\Phi^n - \mathrm{Id})}{n} q^{-ns}.$$

We now must briefly digress to introduce a techinical tool.

Definition. *Let $f : E_1 \to E_2$ be a map of elliptic curves. Define a map $\hat{f} : E_2 \to E_1$ by*

$$\hat{f}(x) = \sum_{f(y)=x} y - \sum_{f(z)=0} z.$$

Proposition 10. *Let $f, g : E_1 \to E_2$, $h : E_2 \to E_3$ and n be the degree of f. Then*

1. $\hat{f}(x_1 + x_2) = \hat{f}(x_1) + \hat{f}(x_2)$.

2. $\hat{\hat{f}} = f$.

3. $\widehat{f+g} = \hat{f} + \hat{g}$.

4. $\widehat{Ih \circ f} = \hat{f} \circ \hat{h}$.

5. $f \circ \hat{f} = [n]$ *and* $\hat{f} \circ f = n$.

6. $\deg \hat{f} = n$.

WRITE THIS
We can now prove

Proposition 11. *There exist α, $\beta \in \mathbf{C}$ such that $\alpha\beta = q$ and*

$$N_n = q^n - \alpha^n - \beta^n + 1.$$

Conceptually, we may think of E as defined over $\mathbf{C}$ instead of F and think of $\Phi : E \to E$ as lifting to an $\mathbf{R}$-linear map on the universal cover of E. Then $\deg \Phi$ would be te determinant of this map. Let α and β be its eigenvalues, the above relation simply expresses the simple matrix identity that, if Φ is a 2×2 matrix with eigenvalues α and β then $\det(\Phi^n - \mathrm{Id}) = (\alpha^n - 1)(\beta^n - 1) = q^n - \alpha^n - \beta^n + 1$. This intutive argument seems hard to make rigorous.

Proof. Define a by $q - a + 1 = N_1$. So we have

$$\begin{aligned} [q-a+1] &= (\Phi - \mathrm{Id})(\hat{\Phi} - \mathrm{Id}) = [q] - (\Phi + \hat{\Phi}) + [1] \Leftarrow \\ [a] &= (P\hat{h}i + \Phi). \end{aligned}$$

Let α and β be defined such that $\alpha + \beta = a$, $\alpha\beta = q$. We now prove $\hat{\Phi}^n + \Phi^n = [\alpha^n + \beta^n]$ by induction on n. We defined α and β so that this is true for $n = 1$ and it is obvious for $n = 0$. We have

$$\begin{aligned} \hat{\Phi}^{n+1} + \Phi^{n+1} &= (\hat{\Phi} + \Phi)(\hat{\Phi}^n + \Phi^n) - \hat{\Phi}\Phi\Phi^{n-1} - \Phi\hat{\Phi}\hat{\Phi}^{n-1} \\ &= [a][\alpha^n + \beta^n] - [q][\alpha^{n-1} + \beta^{n-1}] \\ &= [(\alpha + \beta)(\alpha^n + \beta^n) - (\alpha\beta)(\alpha^{n-1} + \beta^{n-1})] \\ &= [\alpha^{n+1} + \beta^{n+1}] \end{aligned}$$

□

We thus have

$$\begin{aligned} \log \zeta(E, s) &= \sum_n \frac{q^n - \alpha^n - \beta^n + 1}{n} q^{-ns} \\ &= -\log(1 - q^{1-s}) + \log(1 - \alpha q^{-s}) + \log(1 - \beta q^{-s}) - \log(1 - q^{-s}) \end{aligned}$$

or

$$\zeta(E, s) = \frac{(1 - \alpha q^{-s})(1 - \beta q^{-s})}{(1 - q^{1-s})(1 - q^{-s})} = \frac{1 - aq^{-s} + q^{1-2s}}{(1 - q^{1-s})(1 - q^{-s})}.$$

4.1.1 The L-Series of an Elliptic Curve

Let E be an elliptic curve defined over $\mathbf{Q}$. For all but a finite number of primes p, we can reduce E modulo p to get an elliptic curve over F_p. Let $\zeta_p(E,s) = (1 - a_p p^{-s} + p^{-2s})/(1-p^{1-s})(1-p^{-s})$ be its zeta function. At a finite number of primes, E will reduce to one of

1. a cuspidal cubic
2. a nodal cubic, where the slopes of the tanget lines to the node are defined over F_p
3. a nodal cubic, where the slopes of the tanget lines to the node are not defined over F_p

In these cases we define $\zeta_p(E,s)$ to be

1. $1/(1-p^{-s})(1-p^{1-s})$
2. $1/(1-p^{-s})$
3. $(1+p^{-s})/(1-p^{-s})(1-p^{1-s})$

respectively.

(These definitions are consistent with the definition in terms of sums of degrees of effective divisors in the previous section. We check them using the equivalent form $\log\zeta(E,s) = \sum N_n/nq^{-ns}$. In the first case, the curve minus the cusp is isomorphic to the affine line so $N_n = p^n + 1$. In the second, the curve minus the node is isomorphic to G_m so $N_n = (p^n - 1) + 1 = p^n$. In the third, the curve minus the node is equivalent to the curve $x^2 - Dy^2 = 1$ where D is a quadratic nonresidue for p. When n is even, the F_{p^n} points of this curve are in bijection with $F_{p^n}^\times$; when n is odd they are in bijection with the set of $t \in F_{p^{2n}}^\times$ such that $N(t) = 1$ where $N : F_{p^{2n}}^\times \to F_{p^n}$ is the norm. Thus N_n is $1 + (p^n - 1)$, n even, and $1 + (p^n + 1)$ for n odd. In each case, one can check that ζ is given by the claimed formula.)

We now define the ζ function of the curve to be

$$\zeta(E,s) = \prod_p \zeta_p(E,s).$$

Note that the factors $(1-p^{-s})$ and $(1-p^{1-s})$ occur regularly. Factoring them out lets us write

$$\begin{aligned}\zeta(E,s) &= \prod_p \frac{1}{(1-p^{-s})(1-p^{1-s})} \prod_p (1-\epsilon_p p^{-s}) \prod_p (1 - a_p p^{-s} + p^{1-2s}) \\ &= \zeta(s)\zeta(s-1)/L(E,s)\end{aligned}$$

Here the second product is over p of bad reduction and ϵ_p is 0, 1 or -1 according to the three types of bad reduction above respectively. $\zeta(s)$ is the ordinary Riemann zeta function and $L(E,s)$ is defined to be one over the product of the last two terms. Thus, $L(E,s)$ encapsulates the hard part of $\zeta(E,s)$ to compute. We review what hs proceeded in a definition.

Definition. *Let E be an elliptic curve defined over $\mathbf{Q}$. For a prime p, if E has bad reduction at p set ϵ_p equal to 0, 1 or -1 according to which of the respective three types of bad reduction above E has. Otherwise, if $N_{p,1}$ is the number of points of E over F_p, set $a_p = p + 1 - N_{p,1}$. Then*

$$L(E,s) = \prod_p 1/(1-\epsilon_p p^{-s}) \prod_p 1/(1 - a_p p^{-s} + p^{1-2s})$$

where the first product is over primes of bad reduction and the second over primes of good reduction.

Chapter 5

The Eichler-Shimura Relation

REWRITE THIS

Here is the seeming miracle. We will use $F = T(p)$, the Hecke correspondence defined earlier. We have already shown that $T(p)$, turned into a function, acts on E by muliplication by a_p. While we only need that the sum of the points in $T(p)$ and $\phi + \phi^*$ have the same sum in the curve reduced modulo p, we will actually prove that the sets themselves coincide modulo p.

Recall that we have E defined at the end of a sequence of maps $X_0(N) \to JacC_0(N) \to E$. On each of these, $T(p)$ acts and was in fact defined on the latter two by maknig its action consistent with the action on $X_0(N)$. As all of these are defined over $\mathbf{Q}$ and thus reduce to curves over $\mathbf{Z}/p$, ϕ and thus ϕ^* also act on them. We will thus establish $T = \phi + \phi^*$ on $X_0(N)$.

One might intuitively argue as follows. $Y_0(N)$ (that is, $X_0(n)$ without the cusps) is a quotient of the upper half plane. Moreover, the map from the upper half plane to $Y_0(N)$ factors through its quotient by translations by integers, so we get $Y_0(N)$ as a quotient of the punctured disk. Write τ for a coordinate on the upperhalf plane and $q = e^{2\pi i\tau}$ for a coordinate on the disk as before. Then T can be lifted to a self map of the upper half plane by $\tau \to \{p\tau, \tau/p, \tau/p + 1/p + \dots, \tau/p + (p-1)/p\}$ and to a map from the disk to itself by $q \to \{q^p, q^{1/p}, e^{2\pi i/p}q^{1/p}, \dots e^{2\pi i(p-1)/p}q^{1/p}\}$. This latter map looks a lot like $\phi + \phi^*$; it takes q to the pth power of q and the preimage of q under the pth power map. Unfortunately, the map from the

$$f_i(q) = q^{-n_i} + \sum_{0 \leq k} a_{ik} q^k$$

and can take the a_{ik} to be integers. So any other function on $X_0(N)$ whose only pole is at ∞ is a polynomial in the f_i and if that function's q series has integer coefficients that so does the polynomail in the f_i.

The importaance of this is the following. For any point on $Y_0(N)$, its value is determined by the values of the f_i at it, so we can embed $Y_0(N)$ into projective space by $x \to (f_1(x), \dots, f_n(x))$. So $Y_0(N)$ is cut out algebraically by the f_i and the polynomials relations among them. These polynomials reduced modulo p will cut out the reduction of $Y_0(N)$ modulo p, where the q-series no longer make sense. So, to establish that two correspondences are equal modulo p, it is enough to establish that their effect on q-series is the same when the coefficients are taken modulo p.

So we do get to use our intutition that T lifts to $q \to \{q^p, q^{1/p}e^{2\pi i/p}, \dots q^{1/p}e2\pi i(p-1)/p\}$ after all. We simply note that ϕ takes f_i to f_i^p. Hwover, $f(q)^p = f(q^p) modulop$. Similarly, $(\phi^* f)(q) = \sum (q^{1/p}e^{2\pi ij/p})$ mod p. Neither of these right hand sides by themselves is a function defined on $X_0(N)$, but their sum is and is simply $(Tf)(q)$. Thus we have established the relation.

Chapter 6

Computational Methods

6.1 $X_0(11)$: Prelude to Modular Symbols

In this section, we will compute the geometry of $X_0(11)$ and many of the coefficients of the weight two form for $\Gamma_0(11)$. THe methods we use will forshadow the full method of modular symbols which we will discuss next, and will also show the impracticality of performing any sizeable computations by *ad hoc* means. In this section, X will denote $X_0(11)$ and Γ will denote $\Gamma_0(11)$.

We first compute coset representatives of $\Gamma_0(11)\backslash\mathrm{SL}_2(\mathbf{Z})$. Recalling the interpretation of Γ as the symmetries of lattices with a specified index 11 sublattice, we see that we should have $[\mathrm{SL}_2(\mathbf{Z}) : \Gamma] = 12$, because a planar lattice has 12 sublattices of index 11. The 12 sublattices of $\mathrm{span}((1,0),(0,1))$ are $\mathrm{span}((11,0),(0,1))$ and $\mathrm{span}((1,0),(a,11))$, for $a = 0,\ 1,\ \ldots\ 10$. Thus, to give coset representatives, we just need to give transformations from $\mathrm{SL}_2(\mathbf{Z})$ that take one of the sublattices to the other 11.. We will choose as our base matrix $\mathrm{span}((1,0),(0,11))$. The matrices are

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & a \\ 01 & \end{pmatrix}.$$

Thus, if A is any set on which $\mathrm{SL}_2(\mathbf{Z})$ acts transitively on the left, we can find coset representatives for A under the action of Γ by taking a member of A and applying the 12 matrices above to it, then checking to see if any of them are equivalent under the Γ action.

6.1.1 The Triangulation

We now form a triangulation of the upper half plane, which will give to rise a triangulation on X. Draw vertical lines from 0 and 1 to ∞ and a semicircle joining 0 to 1. The egion bounded by these curves will be called the fundamental triagle. Now consider all of the images of this triangle under the action of $\mathrm{SL}_2(\mathbf{Z})$. This is a triangulation of the upper half plane. Each triangle is stabilized by three maps; for example, the fundamental triangle is stabilized by

$$z \to z,\ z \to \frac{z-1}{z} \text{ and } z \to \frac{z}{1-z}.$$

The two nontrivial maps above do not lie in Γ. Thus, there are four orbits of triangles under the action of Γ. Using the coset representatives already found, we can easily verify them to be the triangles with vertices at

$$(0, 1\ infty)\ (1, 0, 1/2),\ (1/3, 1/2, 0) \text{ and } (1/2, 2/3, 1)$$

One can check that the edges on the boundary of this region are identified as follows

$$(0,\infty) \iff (1,\infty),\ (0,1/3) \iff (2/3,1/2) \text{ and } (1/3,1/2) \iff (1,2/3).$$

The cusps at 0, 1/3, 1/2, 2/3 and 1 are all idenitfied, but the cusp at ∞ is not identified with them. Either by visualizing these identifications, or by knowing X is a compact orientable surface and computing the Euler characteristic, we get that X is a torus.

6.1.2 The Cohomology

Recall that we are identifying the (dimension 1) cohomology of X with its homology by the Petersson inner product. As X is a torus, we know that $H^1(X)$ is two dimensional, and decomposes as a direct sum of the holomorphic and antiholomorphic forms. There is an action on X, which we will refer to as complex conjugation, which can be viewed as (1) applying complex conjugation to the coordinates in some mebedding of X, (2) reflecting the upper half plane over the imaginary axis or (3) applying complex conjugation to the colored lattices X parameterizes. The holomorphic forms will be fixed by this action; the antiholomorphic forms will be negated. We thus have a one dimensional vector space of holomorphic forms, which will be mapped to itself by every Hecke operater. A properly normalized member of this vector space will automatically be an eigenform; we now attempt to compute its coefficients.

The integer homology of X is generated by the paths $c_1 = (0, 1/3) = (2/3, 1/2)$ and $c_2 = (1/3, 1/2) = (1, 2/3)$. Complexes conjugation switches c_1 and c_2, so $c_1 + c_2$ is fixed by complex conjugation and must thus by an eigen vector for every Hecke operater. Project onto a one dimensional vector space by sending c_1 and c_2 to the same vector which we will call c, the Hecke operaters will necesarily commute with this projection so we can work in this one dimensional space for our convenience.

6.1.3 $T(5)$: A Sample Hecke Operator

We compute as an example, the result of applying T_5 to $(0, 1/3) = c$. We know

$$T_5 = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 0 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 3 \\ 0 & 5 \end{pmatrix} + \begin{pmatrix} 1 & 4 \\ 0 & 5 \end{pmatrix}.$$

Applying these to $(0, 1/3)$ gives

$$(0, 5/3) + (0, 1/15) + (1/5, 4/15) + (2/5, 7/15) + (3/5, 10/15) + (4/5, 13/15)$$

or

$$\begin{aligned}(0,1) + (1,2) + (2,5/3) + (0,1/15) + (1/5,1/4) + (1/4,4/15) \\ + (2/5,1/2) + (1/2,7/15) + (3/5,2/3) + (4/5,5/6) + (5/6,6/7) + (6/7,13/15).\end{aligned}$$

In this second rewriting, all of the cusps are immediately joined by an edge. Discarding $(0, 1)$ and $(1, 2)$, applying $z \to z - 1$ to shift everything in between 0 and 1, and then applying complex conjugation to put everything between 0 and $1/2$, we get

$$\begin{aligned} & (0,1/3) + (0,1/15) + (1/5,1/4) + (1/4,4/15) \\ + \;& (2/5,1/2) + (1/2,7/15) + (2/5,1/3) + (1/5,1/6) + (1/6,1/7) + (1/7,2/15) \\ = \;& (0,1/3) + (0,1/15) + (1/7,2/15) \\ + \;& (1/5,1/6) + (1/6,1/7) + (1/5,1/4) + (1/4,4/15) + (2/5,1/2) + (2/5,1/3) + (1/2,7/15) \end{aligned}$$

We now apply transformations in Γ to write these as

$$\begin{aligned} & (0,1/3) + \left(z \to \frac{2z+1}{33z+17}\right)(-1/2,-2/3) + \left(z \to \frac{3z-1}{22z-7}\right)(0,1) \\ + \;& \left(z \to \frac{2z+1}{11z+6}\right)((-1,0) + (0,-1/3) + (-1,-2/3)) \\ + \;& \left(z \to \frac{3z+1}{11z+4}\right)(0,1) \\ + \;& \left(z \to \frac{4z+1}{11z+3}\right)((-1/2,-1/3) + (-1/2,0)) + \left(z \to \frac{5z-1}{11z-2}\right)(0,-1/2) \end{aligned}$$

Recalling that we my shift things by $z \to z + 1$, we get

$$c - c + c + c - c - 2c + 2c = c$$

So $a_5 = 1$.

This basically shows all that can happen, except for one thing: it is possible for a Hecke operater to take the cusp at 0 to the sucp at ∞. However, we know that the final result will be a cycle, so we we will be able to combine arcs to deal only with the cusp at 0 as here.

6.2 Modular Symbols

The previous section computed the action of the Hecke operater for $p = 5$, at a form of level $N = 11$ and was fairly intricate. We will want to be able to deal with levels in the tens of thousands and primes in the hundreds, so it is clear that an organizing scheme and some mechanization is essential. Modular symbols are a notation meant to keep track of the geometric objects that turn up in computations like the above.

6.2.1 The Edges

The most awkward part of the above computation was computing the equivalence classes of lines unde the action of Γ. We thus need to determine what the different equivalence classes are and find a notation that will handle them concisely. $\{\alpha, \beta\}$ will denote the line from α to β, α, $\beta \in \mathbf{Q}$. We need to know when $\{a/b, c/d\}$ is equivalent to $\{a'/b', c'/d'\}$ under $\Gamma_0(N)$. As in the above computation, we will only need to solve the problem for lines in the triangulation, for which $ad - bc = 1$.

Suppose we are given c and d relatively prime and let $a_0 d - b_0 c = 1$. Then all other solutions to $ad - bc = 1$ are of the form $a = a_0 + kc$, $b = b_0 + kd$ for some k. As $z \to z + k$ is in $\Gamma_0(N)$, all of these give equivalent lines $\{a_0/c + k, b_0/d + k\}$. We denote this line $(c : d)$. We now need to know in this notation when two such symbols $(c : d)$ and $(c' : d')$ represent equivalent symbols. Let $\{a/b, c/d\}$ and $\{a'/b', c'/d'\}$ be such that $ad - bc = a'd' - b'c' = 1$. We want to know if there is a matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_0(N)$ such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}.$$

Now, if $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ has this property then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d' & -b' \\ -c' & a' \end{pmatrix} = \begin{pmatrix} ad' - bc' & -ab' + ba' \\ -cd' + dc' & -cb' + da' \end{pmatrix}$$

where we have used $a'd' - b'c' = 1$ to invert $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$. So, we want to know if this matrix lies in $\Gamma_0(N)$. Equivalently, we want to know if

$$cd' - c'd = 0 \mod N.$$

Thus we have our criterion: two lines $(c : d)$ and $(c' : d')$ are equivalent iff $cd' = c'd \mod N$.

6.2.2 The Relations

We now want to know how to write the relations between the lines in our new notation. There will be two kinds of relations. The trivial one will be simply recognizing lines that are going in opposite directions, that is

$$(c : d) + (-d : c) = 0.$$

(One might think that the relation should be $(c : d) + (d : c) = 0$. This is incorrect. The reason is that, if $ad - bc = 1$, then $bc - ad = -1$. Thus, if we must represent the reverse of the line $\{a/c, b/d\}$ by $\{(-b)/(-d), a/c\}$.) The slightly less trivial one will be recognizing when lines make up the three sides of a triangle. We know that 0, 1, ∞ is a triangle. Suppose we have an edge $\{a/b, c/d\}$ and we want to know what the third vertex of the triangle is. The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ takes $\infty \to a/b$ and $0 \to c/d$. This matrix takes $1 \to (a + b)/(c + d)$ so the third vertex of the triangle is $(a + b)/(c + d)$. We have

$$\{a/c, b/d\} + \{(-b)/(-d), (a + b)/(c + d)\} + \{(a + b)/(c + d), (-a)/(-c)\}$$

where we have written each $\{w/y, x/z\}$ such that $wx - yz = 1$. Thus

$$(c : d) + (-d : c + d) + (c + d : -c) = 0.$$

With the results above, we can find a basis for the homology of $X_0(N)$. We begin with a single edge $(0 : 1)$ and the two cusps it contains. We then use the above relation to generate a triangle with side $(0 : 1)$. We now have three edges and three cusps, with one relation between them. We take one of these edges and generate the other triangle containing it, getting two new edges a new cusp and another relation. We continue in this way, building a table of edges and relations. As we go, we check the edges we generate to se if they are equivalent to edges already found, using the above test. When we identify two edges, we also idenitfy the corresponding cusps. When every edge has been found in two triangles, our cusps and triangles will give a triangulation of $X_0(N)$, and it is then easy to compute the homology.

6.2.3 Assorted Other Formulae

6.2.4 The Graphical Method

6.2.5 The Hecke Operators and Modular Symbols

As yet we have found a hard solution to an easy problem: if we had just wanted to calculate the cohomology of $X_0(N)$ we could have just noted that it will be $\mathbf{Z}^{2g}$ for some g and used an explicit formulas for g. The benefit of the above results is that it will be easy top compute the effect of the Hecke operators on the edges of our triangulation, and thus it will be easy to compute the action of the Hecke operators on homology. We now explain this in detail. Until we say otherwise, we are dealing with operators for $p \nmid N$.

There is an obvious difficulty; while the Hecke operaters take the cusps to cusps, they do *not* always take the edges to edges. For example, in the example in section REFERENCE, $\begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$ takes the edge $\{0, 1/3\}$ to the path $\{0, 5/3\}$ which is not an edge of the triangulation. We dealt with this by writing it as $\{0, 1\} + \{1, 2\} + \{2, 5/3\}$, each of which is an edge of the triangulation. The same phenomenon will happen in general: Hecke operaters will take single edges to sums of edges.

Assuming then that we have a good way to replace a path $\{a/c, b/d\}$ by a sum of edges in the triangulation, our algorithm is simple. To compute $T_p(c : d)$, find $\{a/c, b/d\}$ an edge of the triangulation, apply the matrices of T_p to $\{a/c, b/d\}$, express the results as sums of edges of the triangulation and convert back to the (:) notation. Note that it is important to use the same representative $\{a/c, b/d\}$ for each matrix in T_p to act on, as it is only the total sum of the products, not each of the individual results that is $\Gamma_0(N)$ invariant.

We thus address the issue: given a path between a pair of cusps $\{a/c, b/d\}$ how to we go about finding edges of the triangulation from one cusp to the other. As was noted in the preceeding paragraph, we will need to find an exact path between a/c and b/d, not between some equivalent pair of cusps. We show how to find a path from ∞ to a/c, applying this process twice will solve the problem.

Let the convergents of the continued fraction to a/c be $a_1/c_1,\ a_2/c_2,\ \ldots,\ a/c$ where $c_1 = 1$. (So, if necesarry, $0/1$ is the first convergent.) Then it is well known REFERENCE that $a_i c_{i+1} - a_{i+1} c_i = 1$. Thus we get a path

$$\{1/0, a_1/c_1\} + \{a_1/c_1, a_2/c_2\} + \cdots + \{a_{k-1}/c_{k-1}, a/c\} = \{\infty, a/c\}.$$

Computing the convergents can be performed rapidly, as in REFERENCE.

In our sample calculations with $\Gamma_0(11)$ we used that the differential forms split into holomorphic and antiholomorphic forms and performed our computations after projecting into the space of holomorphic forms. To do this, we used the action of complex conjugation: $\{a/c, b/d\} \to \{(-a)/c, b/(-d)\}$. Thus, we see that we want to identify $(c : d) = (c : -d)$. Near the end of our computations, this will introduce some difficulties, but until then we can make this idenitfication and cut the dimensions of all of our spaces in half.

6.3 Diagonalization

As working in large dimensional spaces is inconvenient, we want to find the eigenspaces for the Hecke operators at rapidly as possible. Our procedure is to compute T_p for some small values of p in order to find the eigenspace decomposition of the space of new forms. We will then apply the higher T_p to these eigen vectors to find many more Eigenvalues. Let V be the vector space of modular symbols.

We can use the partial decompositions we've found along the way to speed the process as follows: for some small, T_p, use standard linear algebra packages to decompose V as $\bigoplus V_i$ where the V_i are the jordan blocks assocsciate to various eigenvalues λ_i and find bases for the V_i. (That is, subspaces on which $T_p - \lambda_i$ is nilpotent.) Every T_q will fix the V_i, so if we apply them to the basis for the V_i found before we will decompose the T_q into block diagonal form. We thus get smaller matrices to diagonalize and can use these to split V as a fruther direct sum. In this way, we can find all the eigenforms.

We do not have to worry about the parts of V that are not in the span of the eigenforms, they correspond to old forms that we do not care about. The converse is not true: it is possible for an old form to also be a simultaneous eigenform of all the T_p. This situation may be a problem for us (for example, it will lead to an elliptic curve with lower conductor than expected, and it will produce redundancy in compiling tables.) Let p be a prime that divides N, it is easy to test whether an eigenform $\sum a_i q^i$ is induced from a form $\sum b_i q^i$ of level N/p as follows. If it is, we either have $a_i = b_i$ or $a_{pi} = b_i$, $a_i = 0$ if $p \nmid i$ and $\sum b_i q^i$ wil be an eigenform at level N/q. Assuming then that we haev already found all of the eigenforms at level N/q, for each such form look at the subspace of V for which all computed a_i agree with the porceeding relations. If we ever find such a space to be one dimensional, it must be the induced form and we can discard it. If such a space is more than one dimensional, we continue to apply the splitting routine. Since we know that to distinct modular forms can not have the same a_p, this is guaranteed to eventually work.

Once we have a basis of eigenforms, we could compute additional a_p by applying the T_p to the eigenvector v we have already found. A faster method sometimes is to choose a symbol x for which the action of the Hecke operators is very easy to compute, *e.g.* $(0:1) = \{\infty, 0\}$ which has a nonzero component along v. Then the v component of $T_p x$ will be a_p times the v component of x.

6.4 Finding the Periods

6.4.1 Undoing the projection to holomorphic forms

FIX NOTATION

Recall that for we reduced the dimensions of our spaces by projecting the homology H of $X_0(N)$, a group of the form $\mathbf{Z}^{2g}$, onto a group H^+ of the form $\mathbf{Z}^g$ by setting $(c:d) = (-c:d)$. At this point in the process, we will want to use the diagnolization we have found to undo that process. Suppose then that we have found a one dimensional subspace V of H^+ which is preserved by the Hecke operaters, or more generally a d dimensional subspace on which the Hecke operaters act like an order of the ring of integers in a degree d number field. (We will eventually abandon the higher dimensional case as too difficult, but it does not harm to consider it here.) We know that the preimage of V in H will be a $d+g$ dimensional space. There will be some $2d$ dimensional subspace of this space which concerns us. It will split into the direct sum of two d dimensional spaces on which the Hecke operators will acts in the same manner; one on which complex conjugation acts by 1 and one on which it acts by -1. We would like to recover this space.

On the vector space level, this is easy. First note that if $(c_1 : d_1)$, $(c_2 : d_2)$, ... , $(c_g : d_{2g})$ are a basis of H then $(c_i : d_i) + (-c_i : d_i)$ span H^1 and $(c_i : d_i) - (-c_i : d_i)$ span H^{-1}, so finding basis of these spaces is easy., Now, the projection $H^1 \to H^+$ is bijective and has an easy inverse via $(c:d) \to ((c:d) + (-c:d))/2$. We can thus immediately compute V^1. On the other hand $H^{-1} \to H^+$ is 0, so the knowledge of V does not directly help us find V^{-1}. However, we know what eigenvalues the T_p must have on V^{-1}, so we can find V^{-1} by applying T_p to H^{-1} and finding the vectors with the appropriate eigen values. Note that this is *not* simply repeating the computation of the previous section in H^{-1}, we now know what eigenvalues we are looking for. Looking for an eigenvector of a matrix A with known eigenvalue λ simply amounts to finding $\ker(A - \lambda)$ which is much easier than diagonalizing A. (If $d > 1$ and A is known to act by an algebraic integer λ obeying $\sum a_i \lambda^i$, we can avoid floating point arithmetic by instead looking at $\ker \sum a_i A^i$.)

The above methods allow us to find V^1 and V^{-1}, and hence V as vector spaces. However, they do not immediately tell us what a $\mathbf{Z}$ basis for V will be. It is of no difficulty to find a $\mathbf{Z}$ basis for V^1 and V^{-1}; we get one for V^1 directly from pulling back the $\mathbf{Z}$ basis for V^+ we have already found and it is easy to make sure we find $\mathbf{Z}$ basises when evaluating kernels of operators. The difficulty is that we may not have $V = V^1 \oplus v^{-1}$ as lattices. The following example may help: Let $\sigma : \mathbf{R}^2 \to \mathbf{R}^2$ via $(x,y) \to (x,-y)$. Then

$H^1 = \text{span}((1,0))$ is the space on which σ acts by 1 and $H^{-1} = \text{span}((0,1))$ is the space on which it acts by -1. The two lattices $\text{span}((0,1),(1,0))$ and $\text{span}((1,1),(1,-1))$ both have the smae projection to H^1 and H^{-1}, but they are not the same lattice.

As the diagram above suggests, however, there is ambiguity but not a great deal of it. Let $\gamma_1^1, \ldots \gamma_d^1$, $\gamma_1^{-1}, \ldots, \gamma_d^{-1}$ be bases for V^1 and V^{-1} respectively. Then it is clear that $V \supseteq \text{span}(\gamma_i^1, \gamma_i^{-1})$. On the other hand, let us suppose that $\delta = \sum a_i^1\gamma_i^1 + a_i^{-1}\gamma_i^{-1} \in V$. Then $\delta + \sigma\delta = \sum 2a_i^1\gamma_i^1 \in V^1$ so $a_i^1 \in (1/2)\mathbf{Z}$ and similarly, $a_i^{-1} \in (1/2)\mathbf{Z}$. So we get

$$(1/2)(V^1 \oplus V^{-1}) \subseteq V \subseteq V^1 \oplus V^{-1}.$$

In the case $d = 1$, there are only two possibilities. Write γ^1 and γ^{-1} for bais elements of V^1 and V^{-1} respectively, we either have $V = \text{span}(\gamma^1, \gamma^{-1})$ or $V = \text{span}((\gamma^1 + \gamma^{-1})/2, (\gamma^1 - \gamma^{-1})/2)$ and we can check which is the case simply by seeing whether or not $((\gamma^1 + \gamma^{-1})/2$ lies in the integral homology. A similar brute force approach will work in the case of general d. In practice, we do not usually deal with large d, so this is a reasonable method.

6.4.2 An Integral for the Periods

At the end of the last section, we have a modular form $f = \sum a_i q^i$ (or, more accurately, we have a large number of a_i for such a modular form) and we have a basis γ^+, γ^- for the sublattice of H lying in the one complex dimensional vector space spanned by f. We would now like to understand what this lattice looks like drawn in the oridnary complex plane. (The case of higher dimensional abelian varieties is still analogous: we have d modular forms and a $2d$ dimensional lattice, we want to draw this lattice correctly in $\mathbf{C}^d$.) To do this, we must think through what al of our identifications mean: whe we identified homology with cohomology and when we then embedded homology in the space of holomoprhic differential forms by the Abel-Jacobi map.

Our modular form f is of weight 2 and hence corresponds to a differential form on $X_0(N)$. Pulled back to the upper half plane, this form is $f(z)dz$. Let γ be some path in H which we now want to understand as a member of $\mathbf{C}$. The Abel-Jacobi map associate to γ a functional on the space of modular forms $g \to \int_\gamma g$. If we let f_i be an orthonormal basis for the space of modular forms, then the we idnentified M_2 with it's dual via $v \to \sum_i v(f_i) f_i$.

The eigenforms and old forms form an orthogonal basis for M. It is not orthonormal because we normalized that $a_1 = 1$, and this is incompatible with $< f, f >= 1$. Ignoring this issue for now, we see that saying γ lies in the complex sapn of f is saying that $\int_\gamma g = 0$ for g any old form or any newform other than f. Moreover, the map $\gamma \to \int_\gamma f$ will be a complex linear map taking $H \cap \text{span}(f)$ to $\mathbf{C}$ and will hence replicate our lattice up to scaling. We note that we have now ignored scaling data twice: once when we brushed over the normalization of f and again just now. We will see later that there is a correct scaling to choose and that, in terms of that choice, these issues cancel.

For now, we discuss how to calculate the image of $H \cap \text{span}(f)$ to $\mathbf{C} = \text{span}(\gamma^+, \gamma^-)$ under $\gamma \to \int_\gamma f$. We can pull γ back to a path in the upper half plane, we now have the problem of evaluating the integral $\int_\gamma f(z)dz$. We discuss slightly the issue of choosing the preimage of γ carefully: suppose $\gamma = \{a/c, b/d\}$. Clearly, we may chose any $(a/c, b/d)$ under $\Gamma_0(N)$ equivalence. However, we actually have more freedom than this. Thinking geometricaly, we can slide γ around $X_0(N)$ without altering $\int_\gamma f$. What this means in terms of the upper half plane is that, if $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_0(N)$ is a matrix taking the cusp a/c to b/d then we may compute $\int_z^{\begin{pmatrix} A & B \\ C & D \end{pmatrix} z} f(z)dz$ for any x. ($\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ lies in $\Gamma_0(N)$ because γ is a closed path in $X_0(N)$.) Later we will discuss how to optimize z, for now we just compute $\int_{x_1+iy_1}^{x_2+iy_2} f(z)dz$ in general.

6.4.3 Computing the Integral

So we want to compute

$$\int_{x_0+iy_0}^{x_1+iy_1} \sum_i a_i e^{2\pi i z} dz.$$

As f is an entire function on the upperhalf plane, the sum converges uniformly as long as our path is bounded away from the real axis. It will turn out that this is true for the value of $x_0 + iy_0$ that we will choose in optimizing our computation. Thus, we may interchange integration and summation to get

$$\sum_i a_i(e^{-2\pi y_1 + 2\pi x_1 i} - e^{-2\pi y_0 + 2\pi x_0 i}).$$

It may appear that we need to compute two such integrals: one for γ^+ and one for γ^-. In fact, we will only need to compute one for the following reason : complex conjugation takes $\gamma^+ + \gamma^- \iff \gamma^+ - \gamma^-$. We also know that that complex conjugation acts on the lattice we are interested in by complex conjugation so if we compute the real and imaginary components of the image of $\gamma^+ + \gamma^-$ we can determine that of $\gamma^+ - \gamma^-$. (Note that this implies that the lattice is either rectangular or of index 2 in a rectangular lattice. This is not a surprise: we are looking for an elliptic curve that is defined over $\mathbf{Q}$, the lattice of an elliptic curve that is defined over $\mathbf{R}$ must be of this form.)

We now come to the optimization of the integral. Recall that $x_1 + iy_1$ and $x_0 + iy_0$ are required to be related by

$$\frac{A(x_0 + iy_0) + B}{C(x_0 + iy_0) + D} = x_1 + iy_1$$

where $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is required to lie in $\Gamma_0(N)$ and obey

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \binom{a}{c} = \binom{b}{d}.$$

We want to make our sum converge rapidly. There are two reasons for this. The minor reason is that this means we will have fewer exponentials to evaluate and add. The more significant reason is that we will have to compute fewer a_i in order to get enough accuracy. The magnitude of the terms is controlled by $\min(y_0, y_1)$, we want it to be as large as possible.

Rewriting the relation between $x_0 + iy_0$ and $x_1 + iy_1$, we get

$$\frac{(Ax_0 + iAy_0 + B)(Cx_0 - Ciy_0 + D)}{(Cx_0 + D)^2 + (Cy_0)^2} = \frac{AC(x_0^2 + y_0^2) + BD + (AD - BC)iy_0}{(Cx_0 + D)^2 + (Cy_0)^2}$$

So

$$y_1 = (AD - BC)\frac{y_0}{C^2|x_0 + iy_0 + C/D|^2} = \frac{y_0}{C^2|x_0 + iy_0 + C/D|^2}.$$

As x_0 and C/D are real, we get $|x_0 + iy_0 + C/D| \geq y_0$ so

$$y_1 \leq \frac{1}{C^2 y_0} \Leftarrow \min(y_0, y_1) \leq 1/C.$$

Thus, for fixed $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, $\min(y_0, y_1)$ will be as large as possible when $y_0 = y_1 = 1/C$. We see that this can only happen when $y_0 = 1/C$ and $x_0 + C/D = 0$. So we need to take C as small as possible, subject to $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_0(N)$ and

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \binom{a}{c} = \binom{b}{d}.$$

Let $\begin{pmatrix} A_0 & B_0 \\ C_0 & D_0 \end{pmatrix} \binom{a}{c} = \binom{b}{d}$ for $matA_0B_0C_0D_0 \in \Gamma_0(N)$. Then any other $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ with this property must be of the form $\begin{pmatrix} A_0 + ka & B_0 - kc \\ C_0 + la & D_0 - lc \end{pmatrix}$. We want our matrix to have determinant 1, so we want

$$(A_0 + ka)(D_0 - lc) - (B_0 - ka)(C_0 + lc) = 1$$

or

$$(A_0 D_0 - B_0 C_0) - l(A_0 a + B_0 c) + k(C_0 a + D_0 c) = 1.$$

Using $\det matA_0B_0C_0D_0 = 1$ and $matA_0B_0C_0D_0 \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}$, we can simplify this to $kd = lb$. As b and d are relatively prime, we get that $k = bm$ and $l = dm$ for some m. Plugging in, we can write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A_0 + mba & B_0 - mbc \\ C_0 + mda & D_0 - mdc \end{pmatrix}.$$

We want to make $C = C_0 + mda$ as small as possible via the correct choice of m. We are also required to ensure that $C = 0 \mod N$. So we take m to be the closest integer to $-C_0/ad$ that lies in the appropriate residue class modulo $\mathrm{GCD}(N, ad)$ as possible and plug into the proceeding.

6.4.4 Guessing the Lattice

Suppose we are presented with the form $\sum a_i q^i$ and know (or conjecture) that it is an eigenform but have not come to it through the proceeding process so we do not know the associated paths γ^+ and γ^- and perhaps do not even have a basis for the homology worked out at all. There is still a method we can use to generate plausible values for the lattice of E. This situation can occur in practice if the a_i are some other arithmetic sequence which may be the coefficients of a modular form, for example, if the a_p are the traces of some representation not arising from an elliptic curve. Alternatively, we might be able to save time by working only with $\mathbf{Q}$ bases and not $\mathbf{Z}$ bases in the proceeding sections and then guessing the $\mathbf{Z}$ basis.

This technique relies on the following

Manin's Theorem 1. *Let M be the group generated by all the modular symbols $(c : d)$, subject to the relations CITE and CITE. Let H be the homology of $X_0(N)$. (So H is the subgroup of M where we require the pathes to close.) Then $[M : H] < \infty$.*

Proof. Let $\mathrm{rank}X$ denote the rank of a lattice X. We already know $\mathrm{rank}M \geq \mathrm{rank}H$ and we want to prove equality. Let p be a prime which is $1 \mod N$. We claim that $(1 - T_p + p)M \subseteq H$ and that $1 - T_p + p$ is injective. This implies that $\mathrm{rank}M = \mathrm{rank}(1 - T_p + p)M \leq \mathrm{rank}H$ and hence concludes the proof.

We first show that $(1 - T_p + p)M \subseteq H$. The paths $\{a/b, \infty\}$ generate M so it is enough to show that $(1 + p - T_p)\{a/b, \infty\} \in H$. We have

$$\begin{aligned} (1 + p - T_p)\{a/b, \infty\} &= (p+1)\{a/b, \infty\} - \{pa/b, \infty\} - \sum_k (a + kb)/(pb), \infty \\ &= \{a/b, pa/b\} + \sum_k \{a/b, (a + kb)/b\} \\ &\quad + \sum_k \{(a + kb)/b, (a + kb)/(pb)\} \end{aligned}$$

We claim that each of the term $\{x_1/y_1, x_2/y_2\}$ above is a closed path. For the middle sum, we use that $(z \to z + k) \in \Gamma_0(N)$. The others will follow immediately from

Lemma 6. *Let $x_1 = y_1 \mod N$, $x_2 = y_2 \mod N$ and $\mathrm{GCD}(X_1, y_1) = \mathrm{GCD}(x_2, y_2) = 1$. Then the cusps x_1/y_1 and x_2/y_2 are $\Gamma_0(N)$ equivalent.*

Proof. Let $x_2 = x_1 + Nx$ and $y_2 = y_1 + Ny$. As $\mathrm{GCD}(x_1, y_1) = 1$, there is a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\begin{pmatrix} 1 + Na & Nb \\ Nc & 1 + Nd \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

□

We now show that $1 - T_p + p$ is injective. As p is relatively prime to N, T_p is diagonalizable. By the Riemann hypothesis for function fields, the eigenvalues λ of T_p obey $|\lambda| \leq \sqrt(p) < 1 + p$. □

The importance of this theorem for our problem is that, if take an arbitrary path $\gamma = \{a/b, c/d\}$ and compute $\int_\gamma f$ and $\overline{\int_\gamma}$, assuming that $\int_\gamma f$ is not purely real or imaginary, the lattice the generate will intersect the one we want in a finite index sublattice and, thus, the lattice we want has a $\mathbf{Z}$ multiple contained in this lattice.

Our method is thus to guess a γ at random and hope that $\int_\gamma f$ is neither purely real nor imaginary. We then try sublattices of the given lattice to see if they have the desired form. For this method to work, we must have a way of knowing when an answer is correct. There are two techniques we can use. The first, which has very high odds of detecting any errors, is to compute the conductor of the elliptic curve you get by the algorithm of Weil CITE. The second is to compute the number of points on E modulo various primes and compare them with the a_i. (There are many algorithms for this, CITE, although often the primes in question are small enough that this can be done by hand.) It is worth knowing how many a_p must be checked to actually prove that two forms of the same level are equal. Recall that a modular form of weight 2 is equivalent to a differential form on $X_0(N)$. Such a form has $2g-2$ zeroes where g is the genus of $X_0(N)$. Thus two modular forms agree if they agree for the first $2g-1$ terms. (Of course, if whatever procedure gave us $\sum a_i q^i$ als gave us series for the other eigenforms at level N, we can look at them and explicitly see how many terms we'll need to check.)

6.5 Getting the Elliptic Curve

6.5.1 Finding an Equation for the Curve

6.5.2 The Scaling Issue

It is not immediately clear what scaling a lattice does to the associated elliptic curve. The two curves are certainly isomorphic as complex manifolds. Moreover, any two elliptic curves that are isomorphic as complex manifolds have the same lattice up to scaling, because the lattce can by recovered from the elliptic curve by taking some holomofrphic differential one-form ω on it and integrating ω over every closed curve. (Up to scaling, there is only one holomorphic one-form on an elliptic curve.)

On the other hand, we know that there are elliptic curves that have the same complex structure but are not isomorphic over $\mathbf{Q}$. A simple example is the pair of curves $y^2 = x^3 - x$ and $y^2 = x^3 - Dx$, D not a square. These curves are isomorphic via the complex change of variables $(x, y) \iff (D^{1/2}x, D^{3/4}y)$. However, they are not isomorphic over $\mathbf{Q}$, because the first curve has all of it's 2-torsion points rational ($(0,0)$ and $(\pm 1, 0)$) while for the second the 2-torsion points $(0, \pm\sqrt(D))$ are not rational. These curves will have the same lattice up to scaling, but we somehow want to recognize that over $\mathbf{Q}$ they are still distinct curves.

The way out is to recognize that the choice of one form ω above is not an arbitrary one. If we write our curve in Weirstrauss form as $y^2 = x^3 + Ax + B$, then there is a specific one-form given by

$$\omega = \frac{dx}{2y} = \frac{dy}{3x^2 + A}.$$

The Weirstrauss form of a curve is not unique, of course. All other equivalent curves are of the form $y^2 = x^3 + Au^4x + Bu^6$ where the corresponding change of coordinates is $(x, y) \iff (u^2x, u^3y)$ for some *rational* u. These changes of cordinates replace ω by $u\omega$. Thus, ω is well defined up to rational scaling.

The questions about scaling that arose in our construction can now be stated precisely. Let f by an eigenform with rational coefficients, normailzed by setting $a_1 = 1$. Let Λ be the lattice formed by computing $\int_\gamma f$. Let E be the elliptic curve computed by the formulas in the previous section. Is Λ the lattice of E in the sense of the previous section?

We will denote by $\mathcal{M}$ the lattice of E. Note that Λ and $\mathcal{M}$ are obtained by integrating a form on $X_0(N)$ and E respectively against all closed paths. The correspondence between these differential forms is that the form ω on f is up to scaling the pullback of ω along the modular parameterization $X_0(N) \to E$; we want to know this scale factor. Normalizing f by $a_1 = 1$ means that, in terms of the local coordinate $1/q$, $f = d(1/q) + o(1)$ at ∞. In terms of the local coordinate $d(y/x)$, $\omega = d(y/x) + o(1)$ at ∞. Therefore, what we need to know is, in terms of these local coordinates, the derivative of $X_0(N) \to Gamma_0(N)$ at ∞. As this map is rational, the derivative is at least rational, which is enough to tell us that we have the correct curve.

Thus, we theoretically have enough to know that our algorithm finds the correct curve. However, our computations of the lattice and hence of A and B are floating point computations. If in practice all we know are that A and B are rational numbers giving the correct curve, it is impossible to recognize a rational number from a finite accuracy computation. In fact, for reasons that are too difficult to cover in this thesis CITE, the algorithm we have given in fact gives integers for A and B. It is conjectured, but not known, that these integers are in fact the minimal possible integers. In any case, as will be discussed in the nex section, once we have a conjectured curve, it is possible to confirm it is correct without too much difficulty, so the reader uncomfortable with the assertion that A and B will always be integers will at least be able to check that the nearest integers to our decimal computations are right in any given case.