Modular Curves and Modular Forms

Yonatan Harpaz

November 12, 2008

Contents

1 Introduction

A starting motivation for studying the upper half plain

$$\mathbb{H} = \{z \in \mathbb{Z} | \Im(z) > 0\}$$

comes from the famous classification of elliptic curves over $\mathbb{C}$. We can associate to a point $\tau \in \mathbb{H}$ the elliptic curves corresponding to the lattice in $\mathbb{C}$ generated by τ and 1. Two points τ_1, τ_2 give the same elliptic curve if and only if there exist an element $\rho \in \Gamma \stackrel{def}{=} \mathrm{PSL}_2(\mathbb{Z})$ such that $\rho(\tau_1) = \tau_2$ where Γ acts on $\mathbb{H}$ via Mobius transformations.

The topological space $\mathbb{H}/\Gamma$ is a (non-compact) topological manifold, and it also inherits a **complex structure** from $\mathbb{H}$. After compactification it even admits an **algebraic structure** making it isomorphic to $\mathbb{P}^1$. This corresponds to the fact that elliptic curves over a general field K are classified by their j-invariant, which one can think of as a point in $\mathbb{A}^1(K)$. Adding a point in infinity ("compactification") we can say that elliptic curves over K are classified by $\mathbb{P}^1(K)$.

The j-invariant then gives us a map from the compactification of $\mathbb{H}/\Gamma$ to $\mathbb{P}^1(\mathbb{C})$ which gives $\mathbb{H}/\Gamma$ an algebraic structure defined over any subfield of $\mathbb{C}$, i.e. defined over $\mathbb{Q}$. This remarkable situation turns $\mathbb{H}/\Gamma$ (and some other quotient spaces of $\mathbb{H}$) from topological/analytical objects into arithemtic objects! This connection leads to a beautiful and surprising theory which we wish to present in these notes.

2 The Upper Half Plain and its Quotients

In the modern approach we defines a **complex manifold** as a pair $(M, \mathcal{F})$ of a topological manifold M and a sheaf $\mathcal{F}$ of complex functions on M (i.e. $\mathcal{F}(U)$ is an algebra of functions $U \longrightarrow \mathbb{C}$ with the usual restriction maps) which is locally isomorphic to the sheaf of holomorphic functions on the unit disk

$$\mathbb{D} = \{z | |z| \leq 1\} \subseteq \mathbb{C}$$

We then call $\mathcal{F}$ a **complex structure** on M.

If X is a **smooth** algebraic variety over $\mathbb{C}$ then the set $X(\mathbb{C})$ can be endowed with the topology inherited from $\mathbb{C}$ and admits a natural complex structure by replacing local polynomial functions by holomorphic functions. We then say that the complex manifold $X(\mathbb{C})$ has an **algebraic structure**.

Note that in general, non-isomorphic algebraic varieties may give isomorphic complex manifolds, and some complex manifold may fail to have an algebraic structure. The situation is however much nicer for **riemann surfaces**, i.e. one dimensional complex manifolds:

Theorem 2.1. *Let M be a **compact** riemann surface. Then there exists a unique projective algebraic curve X over $\mathbb{C}$ such that $X(\mathbb{C}) \cong M$ as complex manifolds.*

Proof. **(Sketch)** First show that the field $\mathbb{C}(M)$ of global meromorphic functions on M is a finitely generated extension of $\mathbb{C}$ with transcendence degree 1. $\mathbb{C}(M)$ then corresponds to a unique smooth projective algebraic curve X over $\mathbb{C}$ whose field of rational functions is isomorphic to $\mathbb{C}(M)$. By using the isomorphism on the level of function fields one can obtain an isomorphism on the level of points $M \cong X(\mathbb{C})$. □

The above theorem is **not true** for non-compact riemann surfaces. For example, the upper half plain $\mathbb{H}$ does not admit any algebraic structure. This is one of the reasons we shall usually try to compactify the riemann surfaces we stumble upon before trying to find algebraic structures for them.

The spaces we shall be interested in are obtained from $\mathbb{H}$ as quotients by an action of some finite index subgroup $\Gamma' \subseteq \Gamma$. By analyzing what happens in points with non-trivial stabilizers one can show that the resulting space is a topological manifold. Now let $p : \mathbb{H} \longrightarrow Y(\Gamma')$ be the quotient map. In order to define a complex structure we declare a function

$$f : U \longrightarrow \mathbb{C}$$

(where $U \subseteq Y(\Gamma')$ open) as holomorphic if

$$f \circ p : p^{-1}(U) \longrightarrow \mathbb{C}$$

is holomorphic on $\mathbb{H}$. Note that this is definition gives us a bijection between the set of holomorphic functions on U and the set of Γ'-invariant functions on $p^{-1}(U)$.

Now, in order to find an **algebraic structure** on $Y(\Gamma')$ we shall first find a compactification $Y(\Gamma') \subseteq X(\Gamma')$ such that $X(\Gamma')$ is a compact riemann surface. How do we find such an $X(\Gamma')$?

First assume that $\Gamma' = \Gamma$. Define $X(\Gamma) = Y(\Gamma) \cup \{\infty\}$ and declare $\infty \in U$ as a neighborhood of ∞ if the pullback $p^{-1}(U \backslash \{\infty\})$ contains a set of the form

$$V_C = \{z \in \mathbb{H} | \Im(z) > C\}$$

for some $0 < C \in \mathbb{R}$. We then define $q(z) = e^{2\pi i z}$ as our local coordinate at ∞, i.e. we declare a function f on a neighborhood $\infty \in U$ as holomorphic if the pullback $f \circ p$ is a holomorphic function of q, i.e. if it is of the form $g(q(z))$ for some holomorphic function g. Similarly we shall say that a meromorphic function f on a puctured neighborhood U of ∞ has a **pole** at ∞ if $f \circ p$ can be expressed as $g(z) = f(q(z))$ where f has a pole at 0.

In particular the global **meromorphic** functions on $X(\Gamma)$ are in one-to-one correspondence with meromorphic functions on $\mathbb{H}$ which are invariant under Γ and have an expansion of the form

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

This expansion is called the **Fourier expansion** of f. It can be shown that after adding this single point, the space $X(\Gamma)$ is already a compact riemann surface.

By analyzing this space topologically we can see that it is homeomorphic to the sphere, and thus it has to be $\mathbb{P}^1(\mathbb{C})$, i.e. its field of meromorphic functions should be generated over $\mathbb{C}$ by a single transcedental element. This element can be taken to be the j-invariant function. We shall return to this soon.

For the case of a general finite index $\Gamma' \subseteq \Gamma$ we first find the least h such that the map $T(z) = z + h$ is in Γ', and add a point at infinity in the same way only using

$$q\left(\frac{z}{h}\right) = e^{\frac{2\pi i z}{h}}$$

as a local coordinate. The rest of the "points at infinity", called **cusps**, can be obtained by applying elements of Γ to ∞ (i.e. applying them to the defining neighborhood base and to the local coordiante)

The global meromorphic functions on $X(\Gamma')$ can now be identified with what is called **modular functions**, which we define as:

Definition 2.2. A **modular function** for a finite index subgroup $\Gamma' \subseteq \Gamma$ is a meromorphic function on $\mathbb{H}$ which is invariant under Γ' and is meromorphic at the cusps.

There are many ways to identify $X(\Gamma)$ with $\mathbb{P}^1(\mathbb{C})$, as $\mathbb{P}^1(\mathbb{C})$ has a large group of automorphisms (think of this is different choices of transcedental generators for the field of meromorphic fucntions). However, we like to think of $X(\Gamma)$ as a space which parametrizes elliptic curves over $\mathbb{C}$. Thus we have a **preferred** choice for such an identification - we can choose the function j which associates to each $\tau \in \mathbb{H}$ the j-invariant of the corresponding elliptic curve.

j is clearly invariant under Γ and hence defines a holomorphic function on $Y(\Gamma)$. This function can be shown to to have a simple pole at the cusp. Its Fourier expansion begins as

$$j = q^{-1} + 744 + 196884q + ...$$

Thus j defines a meromorphic function on $X(\Gamma)$ (note that the coefficients of j are integer numbers and not general complex numbers. It is not at all obvious at first glance why this should be. We shall return to this phenomenon later).

The choice of j as our identifier between $X(\Gamma)$ and $\mathbb{P}^1(\mathbb{C})$ allows us to copy any additional structure $\mathbb{P}^1$ might have into $X(\Gamma)$. In particular, we obtain an algebraic structure on $X(\Gamma)$ which is **defined over** $\mathbb{Q}$, and not only over $\mathbb{C}$.

We saw above that we can represent meromorphic functions on $X(\Gamma')$ by Γ'-invariant meromorphic functions on $\mathbb{H}$. This is a useful mechinary and we wish to generalize it a bit:

Definition 2.3. A **modular form** of weight $2k$ for a finite index subgroup $\Gamma' \subseteq \Gamma$ is a **holomorphic** function f on $\mathbb{H}$ which satisfies

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

and is **holomorphic** and the cusps. If f is 0 at the cusps then it is called a **cusp** form.

We denote by $M_{2k}(\Gamma)$ the vector space of modular forms of weight $2k$ for Γ' and by $S_{2k}(\Gamma')$ the space of cusp forms of weight $2k$ for Γ'.

Exmaples:

1. A modular function for Γ' is a modular form for Γ' of weight 0.
2. For $k \geq 2$ we have the **Eisenstein series**:

$$G_k(\tau) = \sum_{a,b\in\mathbb{Z},(a,b)\neq(0,0)} \frac{1}{(a\tau+b)^{2k}}$$

which are modular forms (but not cusp forms) for Γ of weight $2k$. These forms appear naturally in the theory of elliptic curves as they give coefficients for a weierstrass equation for $E(j(\tau))$:

$$y^2 = x^3 - 15G_2(\tau) - 35G_3(\tau)$$

In this lecture we will be particularly interested in cusp forms of weight 1, because of the following theorem:

Theorem 2.4. *There is a natural isomorphism between the vector space of cusp forms of weight* 2 *for* Γ' *and the vector space of holomorphic* 1*-forms on* $X(\Gamma')$.

Proof. We shall describe a map $\Omega^1(X(\Gamma')) \longrightarrow M_2(\Gamma')$ and show that it is injective. It will be left as an exercise to show that its image is exactly $S_2(\Gamma')$. Let $\omega \in \Omega^1(X(\Gamma'))$ be a 1-form. Consider the pull back $\pi^*\omega$ of ω to $\mathbb{H}$. Since $\mathbb{H}$ is a domain in $\mathbb{C}$ we can write $\pi^*\omega$ using a the global coordiante τ:

$$\pi^*\omega = f(\tau)d\tau$$

where f is a holomorphic function. Our map will be then $\omega \mapsto f$. It is clear that if $\omega \neq 0$ then $f \neq 0$, so this is an injective map of vector spaces. We need to show that f is a modular form.

$\pi^*\omega$ is invariant under Γ' which are the deck transformations of π, so for $\rho(\tau) = \frac{a\tau+c}{c\tau+d} \in \Gamma'$ we get

$$f(\tau)d\tau = \rho^*(f(\tau)d\tau) = f\left(\frac{a\tau+c}{c\tau+d}\right)\frac{\partial\frac{a\tau+c}{c\tau+d}}{\partial\tau}d\tau = f\left(\frac{a\tau+c}{c\tau+d}\right)\frac{d\tau}{(c\tau+d)^2}$$

Thus f satisfies

$$f\left(\frac{a\tau+c}{c\tau+d}\right) = (c\tau+d)^2 f(\tau)$$

It can be shown that f extends holomorphically to the cusps and actually has 0 in the cusps. This gives us the desired isomorphism

$$\Omega^1(X(\Gamma')) \cong S_2(\Gamma')$$

□

3 Modular Curves

3.1 Introduction

The main riemann surfaces we shall be interested in are obtained as quotients of $\mathbb{H}$ by one of the subgroups $\Gamma(N) \subseteq \Gamma$ defined by

$$\Gamma_0(N) = \left\{ \frac{az+b}{cz+d} \middle| a,b,c,d \in \mathbb{Z}, c = 0(\mod N) \right\}$$

The resulting riemann surface is denoted by $Y_0(N)$ and its compactifiaction by $X_0(N)$. The algebraic curve which corresponds to $X_0(N)$ is called the **modular curve** of conductor N. The reason for the name conductor comes from elliptic curves, as we shall see later. In particular $X_0(1) = X(\Gamma) \cong \mathbb{P}^1$ is the basic example we described before.

We shall now want to understand the field $\mathbb{C}(X_0(N))$ of meromorphic functions on $X_0(N)$. First of all note that the quotient map $\mathbb{H} \longrightarrow X_0(1)$ factors through the quotient $\mathbb{H} \longrightarrow X_0(N)$. This gives us a natural quotient map $\pi : X_0(N) \longrightarrow X_0(1)$. By pulling back meromorphic functions on $X_0(1)$ via π we get a subfield $\pi^*\mathbb{C}(X_0(1))$ of $\mathbb{C}(X_0(N))$. By abuse of notation we shall also refer to j as a function on $X_0(N)$ via this pullback.

This is not the whole field of meromorphic functions. The whole field is in fact a finite extension of $\pi^*\mathbb{C}(X_0(1))$. In order to describe this we need the following observation. For a natural number m consider the (obviously holomorphic) function $\varphi_m : \mathbb{H} \longrightarrow \mathbb{H}$ defined by

$$\varphi_m(\tau) = m\tau$$

Then we have the following

Lemma 3.1. *For each k, φ_m decends to a well-defined (holomorphic) map*

$$\widetilde{\varphi_m} : X_0(mk) \longrightarrow X_0(k)$$

Proof. Let $\rho \in \Gamma(mk)$ act on $\mathbb{H}$ by

$$\rho(\tau) = \frac{a\tau + b}{c\tau + d}$$

such that $c = 0(\mod mk)$. Then there exists an integer c' such that $c = mkc'$. Now:

$$\varphi_m(\rho(\tau)) = \frac{ma\tau + mb}{c\tau + d} = \frac{a(m\tau) + mb}{kc'(m\tau) + d} = \rho'(\varphi_m(\tau))$$

where

$$\rho'(z) = \frac{az + mb}{kc'z + d}$$

which is given by an element in $\Gamma(N)$ since $ad - mbkc' = ad - bc = 1$.

Thus we have a well defined map $Y_0(mk) \longrightarrow Y_0(k)$. Note that since the holomorphic structure on $Y_0(N)$ is inherited from $\mathbb{H}$ it is clear that this map will be holomorphic. It is left to show that it extends holomorphically to the cusps but we shall omit this part. □

By choosing $m = N, n = 1$ we obtain a map $\widetilde{\varphi_N} : X_0(N) \longrightarrow X_0(1)$. This map is different from the map π we had before. Thus by pulling back j through $\widetilde{\varphi_N}$ we get a new function which turns out to generate $\mathbb{C}(X_0(N))$ over $\pi^*\mathbb{C}(X_0(1))$. This function is naturally denoted by $j(N\tau)$.

To conclude, the field of meromorphic functions on $X_0(N)$ is generated over $\mathbb{C}$ by the functions $j(\tau)$ and $j(N\tau)$, which satisfy a polynomial relation

$$F_N(j(\tau), j(N\tau)) = 0$$

It can be shown that this polynomial can be chosen to have coefficients in $\mathbb{Q}$, which gives us an algebraic structure defined over $\mathbb{Q}$. At this points, though, nothing surprises us anymore.

3.2 Interpetation as Moduli Spaces

We know that for $N = 1$, $X_0(N) = X(\Gamma)$ can be interpreted as the compactification of the moduli space of elliptic curves over $\mathbb{C}$. Can we geenralize this interpetation to $X_0(N)$ for a general N? the answer is yes. The key fact here is to note that if we know τ up to an element of $\Gamma_0(N)$ then we know $N\tau$ up to an element of Γ. Thus the elliptic curve $E(j(N\tau))$ is well defined an since the lattice $L(1, N\tau)$ spanned by 1 and $N\tau$ is a sublattice of index N in the lattice $L(1, \tau)$ we get an isogeny of degree N:

$$E(j(N\tau)) \longrightarrow E(j(\tau))$$

The space $X_0(N)$ is in fact a compactification of the moduli space of degree N isogenies of elliptic curves.

Let us now interpret the natural maps $\pi : X(Nm) \longrightarrow X(N)$ and $\widetilde{\varphi_m} : X(Nm) \longrightarrow X(N)$ in the moduli setting. Recall that the point in $X_0(Nm)$ which corresponds to $\tau \in \mathbb{H}$ represents the natural degree Nm isogeny

$$E(j(Nm\tau)) \xrightarrow{f} E(j(\tau))$$

such an isogeny can be factored uniquely as $f = f_N \circ f_m$ with $deg(f_N) = N$ and $deg(f_m) = m$. In fact these are the natural isogenies

$$E(j(Nm\tau)) \xrightarrow{f_m} E(j(N\tau)) \xrightarrow{f_N} E(j(\tau))$$

The point in $X_0(N)$ corresponding to τ represents the isogeny f_N, so that π is the "right extract" map sending f to f_N.

But f can also be uniquely factorized as $f = f'_m \circ f'_N$ with $deg(f'_N) = N, deg(f'_m) = m$. These are the natural isogenies

$$E(j(Nm\tau)) \xrightarrow{f'_N} E(j(m\tau)) \xrightarrow{f'_m} E(j(\tau))$$

The point in $X_0(N)$ corresponding to $m\tau$ represents the isogeny f'_N, so that $\widetilde{\varphi_m}$ is the "left extract" map sending f to f_N.

4 The Hecke Operators

4.1 Pulling Back and Forth

Recall that a surjective map $f : X \longrightarrow Y$ between two smooth projective curves over a field K induces maps on their Jacobians in both directions. On the level of divisors we define it on points and extend linearly

$$f_*(P_i) = f(P_i)$$

$$f^*(Q_i) = \sum_{f(P)=Q} P$$

To show that both these maps preserve principle divisors, recall that f induces an injection $i : K(Y)^* \hookrightarrow K(X)^*$ and on the other direction we have a norm map: $N : K^*(X) \longrightarrow K(Y)$. It is easy to show the commutivity relations $div \circ N = f_* \circ div$ and $div \circ i = f^* \circ div$ (where div is the function associating to a function its divisor). Thus we have well defined maps on the Jacobians. It is also clear that these maps are morphisms of algebraic varieties.

There is an alternative way to describe these maps, which goes through the space of 1-forms. Recall that the Jacobian of a smooth projective curve X over $\mathbb{C}$ has a dual representation as the complex torus

$$\Omega^1(X)/H^1(X, \mathbb{Z})$$

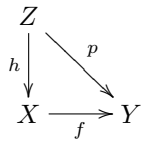
where $\Omega^1(X)$ is the g-dimensional $\mathbb{C}$-vector space of global 1-forms and $H^1(X, \mathbb{Z})$ is the singular integral cohomology group which is isomorphic to $\mathbb{Z}^{2g}$ and is naturally embedded as a lattice in $\Omega^1(X)$.

Now consider a surjective map $f : X \longrightarrow Y$ of smooth projective curves. Then we have a natural pull-back map $f^* : \Omega^1(Y) \longrightarrow \Omega^1(X)$ which always exists. But in this case we can also push forward 1-forms $f_* : \Omega(X)^1 \longrightarrow \Omega^1(Y)$.

This push-forward can be defined as follows. First suppose that f is a galois covering (which might be ramified - we don't care). Then the pull-back f^* identifes the space $\Omega^1(Y)$ with the space of G-invariant 1-forms on X. Thus we can define f_* by defining $f^* \circ f_*$ as

$$f^* f_*(\omega) = \sum_{g \in G} g^* \omega$$

If f is not a galois covering then we can always find a curve Z and maps



Such that p is a galois covering with galois group G and h is a galois covering with galois group $H \subseteq G$ such that $[G : H] = \deg(f)$. To make this construction less surprising recall that on the category of smooth projective curves, the

function field functor induces an equivalence of categories (with an appropriate subcategory of fields over K). The curve Z corresponds to taking the galois closure of $K(Y)$ over $K(Y)$. The group G is the galois group of $K(Z)$ over $K(Y)$ and H the galois group of $K(Z)$ over $K(X)$.

Now if we have a 1-form $\omega \in \Omega^1(X)$ and we want to push it over to Y then we start by pulling it back to Z and symmetrizing by taking

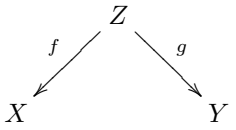
$$\sum_{g \in S} g^* h^* \omega$$

where $S \subseteq G$ is a complete set of H-coset representatives in G. Then we obtain a G-invariant form, so it comes from a unique form $\omega' \in \Omega^1(Y)$. This is the push-forward of ω. Note the similarity with defining the norm map from $K(X)$ to $K(Y)$.

It is an exercise for the reader to show that these two definitions induce the same maps on the Jacobians.

4.2 Correspondences

Now let X, Y be two complete smooth algebraic curves. A **correspondence** from X to Y is a third curve Z and two surjective maps

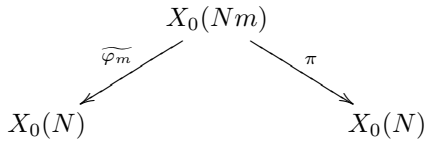


As we saw above this diagram induces a map of Jacobians

$$g_* \circ f^* : J(X) \longrightarrow J(Y)$$

Note that the Jacobians of curves are self-dual abelian varieties in a natural way. This gives a duality map from $\mathrm{Hom}(A, B)$ to $\mathrm{Hom}(B, A)$ (where A, B are Jacobians of curves). This duality already exists on the level of correspondences, as we can define a dual correspondense by swiching the role of f and g. A morphsim which is the dual of itself is called **self-adjoint**.

Now let us return to our modular curves $X_0(N)$. For a natural number m, We know two natural (surjective) maps $X_0(Nm) \longrightarrow X_0(N)$. One is the natural projection π and the second is the map $\widetilde{\varphi_m}$ defined above which is induced by multiplication by m in $\mathbb{H}$. This gives a **correspondence**:



from $X_0(N)$ to itself. This correspondence is called the **Hecke Correspondence** of m. Let us denote by $J_0(N)$ the Picard variety (jacobian) of $X_0(N)$.

Then this correspondence induces an endomorphism $\mathbb{T}_m : J_0(N) \longrightarrow J_0(N)$ called the **Hecke Operator** of m.

In order to understand the Hecke operators in the moduli setting recall that the maps π and φ_m can be interpreted as "right extract" and "left extract" respectively. Use this description in order to solve the next exercise:

Exercise 1. *Prove that the Hecke operators commute with each other and satisfy the following relations*

1. *If* $gcd(m,k) = 1$ *then* $\mathbb{T}_m \circ \mathbb{T}_k = \mathbb{T}_{mk}$.
2. *if* $p \not| N$ *then* $\mathbb{T}_{p^k} \circ \mathbb{T}_p = \mathbb{T}_{p^{k+1}} + p\mathbb{T}_{p^{k-1}}$
3. *if* $p|N$ *then* $\mathbb{T}_{p^k} = (\mathbb{T}_p)^k$.

4.3 The Eichler Shimura Theorem

We are now ready to prove the important **Eichler-Shimura** theorem:

Theorem 4.1. *For* p *not dividing* N, *let* $\widetilde{J_0}(N)$ *be the mod-*p *reduction of* $J_0(N)$ *and* $\widetilde{\mathbb{T}}_p$ *the reduction of the Hecke operator. Let* $\Phi_p : \widetilde{J_0}(N) \longrightarrow \widetilde{J_0}(N)$ *be the Frobenious endomorphism and* Φ_p^* *its dual. Then*

$$\widetilde{\mathbb{T}}_p = \Phi_p + \Phi_p^*$$

Proof. Let E, F be two elliptic curves defined over $\mathbb{Q}$ with smooth mod-p reductions $\widetilde{E}, \widetilde{F}$ which are not supersingular. Let $f : E \longrightarrow F$ be an isogeny of degree N. Then the isomorphism class $[f : E \longrightarrow F]$ defines a point in $X_0(N)$ and the isomorphism class $[\widetilde{E} \xrightarrow{\tilde{f}} \widetilde{F}]$ defines its reduction in $\widetilde{X}_0(N)$.

Let $S_0, S_1, ..., S_p \subseteq E$ be the subgroups of size p in E and $T_i = f(S_i)$ their images in F. Then we have commutative diagrams of isogenies

$$\begin{array}{ccc} E & \xrightarrow{f} & F \\ {\scriptstyle g_i}\downarrow & & \downarrow{\scriptstyle h_i} \\ E_i & \xrightarrow[f_i]{} & F_i \end{array}$$

where $E_i = E/S_i$ and $F_i = F/T_i$. Then by definition

$$\mathbb{T}_p([E \xrightarrow{f} F]) = \sum_{i=0}^{p} [E_i \xrightarrow{f_i} F_i]$$

Now consider the multiplication by p map which has degree p^2. It induces a commutative diagram

$$\begin{array}{ccc} \widetilde{E} & \xrightarrow{f} & \widetilde{F} \\ {\scriptstyle [p]}\downarrow & & \downarrow{\scriptstyle [p]} \\ \widetilde{E} & \xrightarrow[f]{} & \widetilde{F} \end{array}$$

For each $i = 0, ..., p$, we can factor this diagram as

$$\begin{array}{ccc} \widetilde{E} & \xrightarrow{f} & \widetilde{F} \\ {\scriptstyle \widetilde{g}_i}\downarrow & & \downarrow{\scriptstyle \widetilde{h}_i} \\ \widetilde{E_i} & \xrightarrow[f_i]{} & \widetilde{F_i} \\ {\scriptstyle G_i}\downarrow & & \downarrow{\scriptstyle H_i} \\ \widetilde{E} & \xrightarrow{f} & \widetilde{F} \end{array}$$

where the G_i's and H_i's have degree p.

Since $\widetilde{E}$ is not supersingular, the reduction map $E \longrightarrow \widetilde{E}$ has a kernel of size p. Assume without loss of generality that it is S_0. Then the maps g_0, h_0 and G_i, H_i for $i \neq 0$ have no kernel and hence are purely inseperable. But there is only one purly inseperable isogeny of degree p (up to maybe isomorphism, which we don't care about) - its the frobenious map!

To avoid confusion, let us denote this Frobenious by Ψ_p, to distinguish it from Φ_p which is the Frobenious map on $\widetilde{J}_0(N)$.

Now, up to maybe changing the isogenies $\widetilde{E_i} \xrightarrow{\widetilde{f}_i} \widetilde{F}_i$ by an isomorphism, we get a commutative diagram

$$\begin{array}{ccc} \widetilde{E} & \xrightarrow{\widetilde{f}} & \widetilde{F} \\ {\scriptstyle \Psi_p}\downarrow & & \downarrow{\scriptstyle \Psi_p} \\ \widetilde{E_0} & \xrightarrow[\widetilde{f}_0]{} & \widetilde{F_0} \end{array}$$

and also commutative diagrams

$$\begin{array}{ccc} \widetilde{E_i} & \xrightarrow{\widetilde{f}_i} & \widetilde{F}_i \\ {\scriptstyle \Psi_p}\downarrow & & \downarrow{\scriptstyle \Psi_p} \\ \widetilde{E} & \xrightarrow[\widetilde{f}]{} & \widetilde{F} \end{array}$$

for each $i \neq 0$.

This means that on $\widetilde{X_0}(N)$ we have the relation

$$\Phi_p([\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}]) = [\widetilde{E_0} \xrightarrow{\widetilde{f}_0} \widetilde{F_0}]$$

and for each $i \neq 0$ the relation

$$\Phi_p([\widetilde{E_i} \xrightarrow{\widetilde{f}_i} \widetilde{F}_i]) = [\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}]$$

Since the Frobenious map has no kernel, we see that the points $[\widetilde{E_i} \xrightarrow{\widetilde{f_i}} \widetilde{F_i}]$ are **all equal** on $\widetilde{X_0}(N)$. Further more since the dual of Frobenious Φ_p^* satisfies $\Phi_p^* \circ \Phi_p = [p]$ (where $[p]$ is multiplication by $[p]$ on $\widetilde{J_0}(N)$) we get that

$$\sum_{i=1}^{p} [\widetilde{E_i} \xrightarrow{\widetilde{f_i}} \widetilde{F_i}] = [p][\widetilde{E_1} \xrightarrow{\widetilde{f_1}} \widetilde{F_1}] = \Phi_p^*(\Phi_p([\widetilde{E_1} \xrightarrow{\widetilde{f_1}} \widetilde{F_1}])) = \Phi_p^*([\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}])$$

To conclude, we have found that

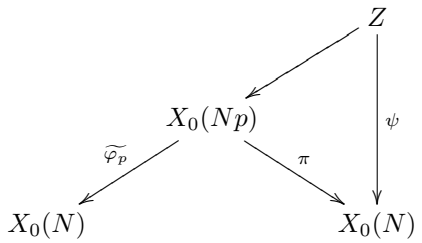
$$\widetilde{\mathbb{T}_p}([\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}]) = \sum_{i=0}^{p} [\widetilde{E_i} \xrightarrow{\widetilde{f_i}} \widetilde{F_i}] = \Phi_p([\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}]) + \Phi_p^*([\widetilde{E} \xrightarrow{\widetilde{f}} \widetilde{F}])$$

In order to complete the proof we need to argue that in some sense, almost all rational points on $X_0(N)$ can be represented by an isogeny of curves having non-supersingular smooth reduction at p, and thus $\widetilde{\mathbb{T}_p}$ has to be equal to $\Phi_p + \Phi_p^*$. □

4.4 The Action on Cusp Forms

As we saw in the previous section, a correspondence can also act on the space of 1-forms, and this action is compatible with the action on the Jacobians. We also know that the space of holomorphic 1-forms on $X_0(N)$ can be identified with the space of cusp forms of weight 2 for $\Gamma_0(N)$. We shall now calculate the action of the correspondence $\mathbb{T}_m$ on a cusp form f under the assumption that m is coprime to N.

Recall the definition in the previous section of pushing forward a 1-form using a "galois-closure" curve Z:



What is Z in our case? Define

$$\Gamma' \subseteq \Gamma_0(Nm) \subseteq \Gamma_0(N)$$

to be the largest subgroup which is normal in $\Gamma_0(N)$ and $Z = X(\Gamma')$. Then ψ is a galois covering with galois group $\Gamma_0(N)/\Gamma'$. Let $S \subseteq \Gamma_0(N)$ be a complete set of representatives of left $\Gamma_0(Nm)$-cosets.

Now take a 1-form on $X_0(N)$ and represent it by a 1-form $\omega(z) = f(z)dz$ on $\mathbb{H}$ which is invariant under $\Gamma_0(N)$ (i.e., such that f is a cusp form of weight 2). Pulling it back via φ_m we get the form

$$\varphi_m^*\omega(z) = f(\varphi_m(z))\frac{\partial \varphi_m}{\partial z}dz = mf(mz)dz$$

We now symmetrize it by the left-coset representatives $\sigma \in S$:

$$\sum_{\sigma \in S} \sigma^* \varphi_m^* \omega = \sum_{\sigma \in S} (\varphi_m \sigma)^* \omega$$

Lemma 4.2.

$$\sum_{\sigma \in S} (\varphi_m \sigma)^* \omega = \sum_{\{a,b,d | ad=m, 0 \leq b \leq d\}} \frac{m}{d^2} f\left(\frac{az+b}{d}\right) dz$$

Proof. Cosider the set $O = \varphi_m \Gamma_0(N)$. The element $\alpha = \varphi_m \sigma \in O$ is in the **left** coset $\Gamma_0(N)\varphi_m$ if and only if $\sigma \in \varphi_m^{-1}\Gamma_0(N)\varphi_m$. But

$$\varphi_m^{-1}\Gamma_0(N)\varphi_m = \left\{ \frac{az + m^{-1}b}{mcz+d} \middle| a,b,c,d \in \mathbb{Z}, c = 0(\mod N) \right\}$$

Since m is coprime to N we see that

$$\Gamma_0(N) \bigcap \varphi_m^{-1}\Gamma_0(N)\varphi_m = \Gamma(Nm)$$

Thus the left $\Gamma_0(N)$-cosets of O are in one-to-one correspondence with the left $\Gamma_0(Nm)$-cosets of $\Gamma_0(N)$. This means that the set

$$\{\varphi_m \sigma | \sigma \in S\}$$

is a complete set of representatives of left $\Gamma_0(N)$-coset in O. But it is an easy exercise to show that the following set of mobious transformations:

$$\left\{ \frac{az+b}{d} \middle| ad = m, 0 \leq b \leq d \right\}$$

is a complete set of representatives of left $\Gamma_0(N)$-cosets in O. Since ω is $\Gamma_0(N)$ invariant we get that

$$\sum_{\sigma \in S} (\varphi_m \sigma)^* \omega = \sum_{a,b,d | ad=m, 0 \leq b < d} \frac{m}{d^2} f\left(\frac{az+b}{d}\right) dz$$

□

Now if f is a cusp form of weight 2 given by Fourier expansion

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

Then we get the explicite expression for the action of $\mathbb{T}_m$ on f:

$$\mathbb{T}_m(f)(z) = \sum_{ad=m, 0 \leq b < d} \frac{m}{d^2} f\left(\frac{az+b}{d}\right) =$$

$$\sum_{ad=m,0\leq b\leq d} \frac{m}{d^2} \sum_{n=1}^{\infty} c_n e^{\frac{2\pi i n(az+b)}{d}} =$$

$$\sum_{ad=m,a\geq 1} a \sum_{n=1}^{\infty} c_n e^{\frac{2\pi i n a z}{d}} \frac{1}{d} \sum_{b=0}^{d-1} e^{\frac{2\pi i n b}{d}} =$$

But the last expression is 0 if n is not divisible by d and 1 if it is. Thus by defining $n' = \frac{n}{d}$ we get the final answer

$$\mathbb{T}_m(f) = \sum_{1\leq a|m} \sum_{n'=1}^{\infty} a c_{dn'} q^{an'}$$

In particular for p prime not dividing N we get the formula

$$\mathbb{T}_p(f) = \sum_{n'=1}^{\infty} p c_{n'} q^{pn'} + \sum_{n'=1}^{\infty} c_{pn'} q^{n'}$$

Exercise 2. *Show that for $p|N$ we have*

$$\mathbb{T}_p(f) = \sum_{n'=1}^{\infty} c_{pn'} q^{n'}$$

5 Eigenforms and newforms

Our basic aim is to study the modular curves. A basic strategy for investigating a curve X is by studying the maps from X to simpler curves. The simplest curve is $\mathbb{P}^1$, and studying maps $X \longrightarrow \mathbb{P}^1$ is just like studying the field of rational functions on X, which we touched upon above. The next class of curves is the genus 1 curves, or elliptic curves.

The way to study maps $X \longrightarrow E$ with E an elliptic curve is to study the Jacobian $J(X)$. The reason is that we have a unique factorization theorem for abelian varieties: each abelian varity A is isogenous to a unique product

$$A \simeq \prod A_i^{e_i}$$

where each A_i is a **simple** abelian variety, i.e. contains no subabelian varities.

Now a noncontant map $f : X \longrightarrow E$ induces a nonzero map $f^* : E \longrightarrow J(X)$ and a nonzero projection $f_* : J(X) \longrightarrow E$ with kernel $\widehat{E}$. This gives us an isogeny

$$J(X) \simeq E \times \widehat{E}$$

which means that E is one of the simple factors of $J(X)$. In particular there is only a finite set of elliptic curves which admit a nonconstant map $X \longrightarrow E$ and in order to understand them we need to anaylize the Jacobian $J(X)$.

How do we find simple factors of $J(X)$ which are 1-dimensional? Suppose we were working with curves over $\mathbb{F}_p$. Then we would have the Frobenious endomorphism $\Phi_p : J(X) \longrightarrow J(X)$. Since Φ preserves the simple factors we can look for one dimensional simple factors like looking for eigenvectors. More precisely, if one considers the Tate module

$$M_l(J(X)) = \varprojlim J(X)/lJ(X)$$

then each simple 1-dimensional factor E of $J(X)$ would corespond to an eigenvector in $M_l(J(X))$ of the operator $\Phi_p + \Phi_p^*$ with eigenvalue a_p.

Lucky for us, we don't have to reduce mod p in order to do this, because we have an endomorphism whose mod p reduction is exactly $\Phi_p + \Phi_p^*$ - the Hecke operators $\mathbb{T}_p$!

Since we are working over $\mathbb{Q}$ it wil be more convenient and just as effective to replace the Tate module of $J_0(N)$ with the vector space of holomorphic 1-forms which we identified with the space $S^2(N)$ of cusp forms of weight 2. Each simple factor of $J_0(N)$ will correspond to a cusp form which is an eigenvector for all the $\mathbb{T}_m$'s with m coprime to N.

With a little but of effort one can also show that if E is a simple factor of $J_0(N)$ and not a simple factor of any $J_0(m)$ for $m|N$ then the corresponding cusp form will be a eigenvector for all the $\mathbb{T}_m$'s.

Let us make the last statement more precise. If $m|N$ then we have a natural projection map $\pi : X_0(N) \longrightarrow X_0(m)$ which induces a map $S^2(m) \longrightarrow S^2(N)$. We define the subspace $S^2_{old}(N) \subseteq S^2(N)$ to be the space which is generated by the images of all these maps. We define $S^2_{new}(N)$ to be its orthogonal complement. The cusp forms in $S^2_{new}(N)$ are called **newforms**.

The set of operators $\{\mathbb{T}_m\}$ for m coprimes to N is a commuting family of self adjoint operators which preserve $S^2_{new}(N)$. Thus we can find a basis for $S^2_{new}(N)$ consisting of simultanous eigenvectors of all the $\mathbb{T}_m$'s for m coprime to N. This is called a spectral decomposition. Now we have a theorem

Theorem 5.1. *Each sequence of eigenvalues which appears in the spectral decomposition appears with multiplicity* 1*, i.e. appears in a* 1*-dimensional subspace.*

Since all the $\mathbb{T}_m$'s commute (even those which are not coprime to N) we see that these eigenvectors must actualy be simultanous eigenvectors for **all** the $\{\mathbb{T}_m\}$'s. These correspond exactly to simple factors of $J_0(N)$ which don't appear in any $J_0(m)$ for $m|N$.

Now suppose that $f = \sum_{n=1}^{\infty} c_n q^n \in S^2_{new}(N)$ is cusp form of weight 2 which is an eigenvector for all the $\mathbb{T}_n$'s. Then by the formula above we see that

$$\lambda_m c_n = \sum_{0 \leq a | n,m} a \cdot c_{\frac{nm}{a^2}}$$

and in particular $\lambda_m c_1 = c_m$. This means that c_1 must be non-zero, otherwize all the c_m's would be zero, i.e. f would be zero. Since c_1 is non-zero we can

normalize f so that $c_1 = 1$. Such an eigenform is called **normalized**. Now for a normalized eigenform we get

$$\lambda_m = c_m$$

i.e. the eigenvalues of f are the coefficients themselves!

Now recall that $f(z)dz$ is a $\Gamma_0(N)$-invariant 1-form on $\mathbb{H}$. Consider the mobius map

$$\rho(\tau) = \frac{-1}{N\tau}$$

We can calculate and check that $\rho^{-1}\Gamma_0(N)\rho = \Gamma_0(N)$. This implies that ρ acts on $S^2(N)$ and it preserves $S^2_{new}(N)$. It can be shown that its action commutes with action of all the Hecke operators $\mathbb{T}_m$ and so from the theorem above we see that each eigenform is a also an eigenvector of ρ. Since $\rho^2 = 1$ this eigenvalue is ± 1. This is called the parity of the eigenform.

6 Modular L-functions

Let $f = \sum_{n=1}^{\infty} c_n q^n \in S_{new}(N)$ be an eigenform. Then we know that the eigenvalue of $\mathbb{T}_n$ is c_n. Since the $\mathbb{T}_n$'s satisfy the relations given in exercise 1 we get the same relations on the c_n's, namely:

1. If $gcd(m,k) = 1$ then $c_m c_k = c_{mk}$.
2. If $p \not| N$ then $c_{p^k} c_p = c_{p^{k+1}} + p c_{p^{k-1}}$
3. If $p|N$ then $c_{p^k} = c_p^k$.

Define the L-function associated with f to be the complex function

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

Then from the properties above we can transform this infinite sum into an **Euler product**:

$$L(f,s) = \prod_{p \not| N} \frac{1}{1 - a_p p^{-s} + p^{1-2s}} \prod_{p|N} \frac{1}{1 - a_p p^{-s}}$$

From the Eichler Shimura theorem we know that a_p should also be an eigenvalue of $\Phi_p + \Phi_p^*$. Since Φ_p commutes with Φ_p^* (i.e. Φ_p is a normal operator) and since $\Phi_p \circ \Phi_p^* = p$ on any abelian variety we get the Hasse-Weil bound $|a_p| \leq 2\sqrt{p}$. This means that the Euler product converges for $Re(s) > \frac{3}{2}$.

Theorem 6.1. *Let f be an eigenform which is an ϵ-eigenvector of ρ for $\epsilon = \pm 1$. Then the L-function $L(E,s)$ admits a holomorphic continuation to the whole plain and the function*

$$\Lambda(f,s) = N^{1-s}\Gamma(s)L(f,s)$$

satisfies the functional equation

$$\Lambda(f, s) = \epsilon\Lambda(f, 2 - s)$$

Proof. The key point to showing this is that there is an analytic connection between $L(f, s)$ and f via the mellin transform, which is a particular case of a Fourier transform for the multiplicative group $\mathbb{R}_{>0}$ with the Haar measure $\frac{dt}{t}$. Specifically one has

$$\int_0^\infty f(it)t^s \frac{dt}{t} = \Gamma(t)L(f, s)$$

□