Chapter 2

Modular curves

2.1 The action of $SL_2(\mathbb{R})$

We let $GL_2(\mathbb{C})$ act on the Riemann sphere $\mathbb{C} \cup \{\infty\}$ by fractional linear transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau \mapsto \frac{a\tau + b}{c\tau + d}$$

Note that we can identify $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1_{\mathbb{C}}$. With respect to this, the above action of $GL_2(\mathbb{C})$ coincides with the usual action on the projective line by $[v] \mapsto [Av]$.

Lemma 2.1.1. *$SL_2(\mathbb{R})$ acts transitively on the upper half plane $\mathbb{H} = \{\tau \in \mathbb{C} \mid \operatorname{Im} \tau > 0\}$ by fractional linear transformations. The stabilizer of i is $SO(2)$. Therefore, we can identify $\mathbb{H} = SL_2(\mathbb{R})/SO(2)$.*

Proof. If $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R})$ and $\tau \in \mathbb{C}$, then

$$\operatorname{Im} \frac{a\tau + b}{c\tau + d} = \frac{\operatorname{Im} \tau}{|c\tau + d|^2} \tag{2.1}$$

This shows that $SL_2(\mathbb{R})$ preserves $\mathbb{H}$. We have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot i = \frac{(ca + db) + i}{c^2 + d^2}$$

It is now an easy exercise to see that given $\tau \in \mathbb{H}$, we can find a solution to

$$A \cdot i = \tau$$

with $A \in SL_2(\mathbb{R})$, and that if $\tau = i$, we must have $A \in SO(2)$.

□

We can view $\mathbb{H}$ as the upper hemisphere of the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$. The action of $SL_2(\mathbb{R})$ extends to the boundary $\partial\mathbb{H} = \mathbb{P}^1_{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. In order to better visualize the action, it useful to note that $\mathbb{H}$ has a Riemannian metric, called the hyperbolic or Poincaré metric, where the geodesics are lines or circles meeting $\partial\mathbb{H}$ at right angles. The action of $SL_2(\mathbb{R})$ preserves this metric, so it takes a geodesic to another geodesic.

2.2 The modular group $SL_2(\mathbb{Z})$

Let $L_\tau = \mathbb{Z}+\mathbb{Z}\tau$ with $\tau \in \mathbb{H}$ as before. We can see that elliptic curves $E_\tau = \mathbb{C}/L_\tau$ and $E_{\tau'}$ are isomorphic if and only if $L_\tau = L_{\tau'}$.

Lemma 2.2.1.

(a) If $(u,v)^T, (u',v')^T \in B^+$, then $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}u' + \mathbb{Z}v'$ if and only if $(u,v)^T$ and $(u',v')^T$ lie in the same orbit of $SL_2(\mathbb{Z})$.

(b) $L_\tau = L_{\tau'}$ if and only τ, τ' lie in the same orbit under $SL_2(\mathbb{Z})$.

Proof. If $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}u' + \mathbb{Z}v'$, there would be change of basis matrix A taking $(u,v)^T$ to $(u',v')^T$. A is necessarily integral with positive determinant, and this already ensures that $A \in SL_2(\mathbb{Z})$. The converse is easy. (b) follows from (a). □

From this lemma, we can conclude that:

Theorem 2.2.2. *The set of isomorphism classes of elliptic curves (over $\mathbb{C}$) is parameterized by $SL_2(\mathbb{Z})\backslash\mathbb{H}$.*

At the moment, $A_1 = SL_2(\mathbb{Z})\backslash\mathbb{H}$ is just a set. In order to give more structure, we need to analyze the action more carefully. First observe that $-I$ acts trivially on $\mathbb{H}$, so the action factors through $\Gamma = PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$. Consider the closed region $F \subset \mathbb{C}$ bounded by the unit circle and the lines $\operatorname{Im} z = \pm 1/2$ depicted below.

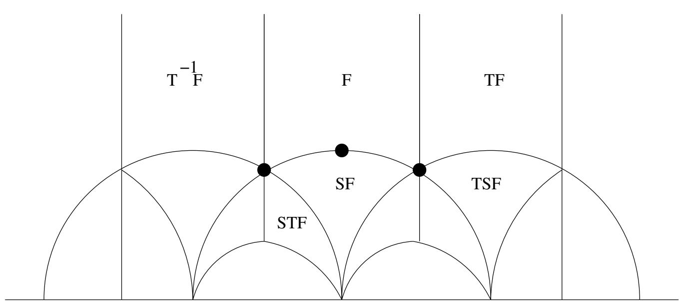

Figure 2.1: Fundamental domain

Let $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. These act by $z \mapsto -1/z$ and $z \mapsto z+1$ respectively. S is a reflection about i which interchanges the regions $|z| \geq 1$ and $|z| \leq 1$. They generate a subgroup $G \subseteq \Gamma$.

Theorem 2.2.3.

(a) The union of translates gF, $g \in G$, covers $\mathbb{H}$.

(b) An interior point of F does not lie in any other translate of F under G.

(c) The isotropy group of $z \in F$ is trivial unless it is one of the points $\{i, e^{\pi i/3}, e^{2\pi i/3}\}$ marked in the diagram. The isotropy group is $\langle S\rangle, \langle ST\rangle, \langle TS\rangle$ respectively.

Proof. The intuition behind this can be understood from the picture. Repeatedly applying S and $T^{\pm 1}$ to F gives a *tiling* of $\mathbb{H}$ by hyperbolic triangles. Choose $\tau \in \mathbb{H}$, we want to find $A' \in SL_2(\mathbb{Z})$ and $\tau' \in F$ such that $A' \cdot \tau' = \tau$. Using (2.1), we can see that $\{\operatorname{Im} A \cdot \tau \mid A \in SL_2(\mathbb{Z})\}$ has a maximum M. Choose an A which realizes this maximum. Choose an integer n so that $\tau' = T^n A\tau$ has real part in $[-1/2, 1/2]$. Observe that $\operatorname{Im} \tau' = M$. If $|\tau'| < 1$ then $-1/\tau'$ would have imaginary bigger than M which is impossible. It follows that $\tau' \in F$, and τ lies in its orbit. This proves (a). For the remaining parts, see Serre [Se, pp 79].

□

The set F is called a *fundamental domain* for the action of G. We can draw a number of useful conclusions.

Corollary 2.2.4. *$G = PSL_2(\mathbb{Z})$, i.e. S and T generate $PSL_2(\mathbb{Z})$.*

Proof. Let $z \in F$ be an interior point, and $h \in \Gamma$. Then $hz = gz$ for some $g \in G$. Since $z \in h^{-1}gF$, we must have $h^{-1}g = I$. □

Corollary 2.2.5. *The nontrivial elements of finite order in Γ are conjugate to S or $(ST)^{\pm 1}$.*

Proof. A nontrivial element of finite must lie in the isotropy group of some point in $\mathbb{H}$. The points in the plane with nontrivial isotropy groups must be a translate of i or $e^{2\pi i/3}$. Their isotropy groups must be conjugate to the isotropy groups of one these two points. □

Corollary 2.2.6. *The action of $PSL_2(\mathbb{Z})$ is properly discontinuous, which means that for every point $p \in \mathbb{H}$, there is a neighbourhood U such that $gU \cap U = \emptyset$ for all but finitely many g.*

We can give A_1 the quotient topology where $U \subseteq A_1$ is open if and only its pullback to $\mathbb{H}$, under the projection $\pi : \mathbb{H} \to A_1$ is open.

Proposition 2.2.7. *The topology on A_1 is Hausdorff. In fact, it is homeomorphic to $\mathbb{C}$*

Proof. The first statement follows immediately from the last corollary. Using the above results, one can see that A_1 is obtained by gluing the two bounding lines of F and folding the circular boundary in half. This is easily seen to be homeomorphic to the sphere minus the north pole. □

A_1 has a natural compactification $\bar{A}_1$ given by adding single point at infinity to make it a sphere. We will follow the convention of the automorphic form literature and call it a *cusp*. It is important to keep in mind that this clashes with the usual terminology in algebraic geometry, that a cusp is a singularity of the form $y^2 = x^3$. We will refer the last thing as cuspidal singularity in order to avoid confusion. We can construct this a quotient as follows. Let $\mathbb{H}^* = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\} \subset \mathbb{P}^1$. The action of Γ on $\mathbb{P}^1$ stabilizes $\mathbb{H}^*$. On $\mathbb{H}$ it coincides with the standard action, and on $\partial\mathbb{H}^* = \mathbb{Q} \cup \{\infty\}$ it consists of a single orbit. Thus $\Gamma\backslash\mathbb{H}^* = \bar{A}_1$ as a set. In order to get the correct topology on the quotient, one needs a somewhat exotic topology of $\mathbb{H}^*$. On $\mathbb{H}$ it's the usual one, but on $\partial\mathbb{H}^*$ a fundamental system of neighbourhoods of (a translate of) ∞ are (translates of) strips $\operatorname{Im} z > n$, $n \in \mathbb{N}$.

2.3 Modular forms

Since A_1 has a topology, we can talk about continuous functions on it. We can see that $f : A_1 \to \mathbb{C}$ is continuous if and only if it's pullback $\pi^* f := f \circ \pi$ is continuous. Let us also declare that a function on an open subset of $\mathcal{A}_1$ is holomorphic or meromorphic if its pullback to $\mathbb{H}$ has the same property. This means that such functions correspond to Γ-invariant functions on $\mathbb{H}$. Before constructing nontrivial examples, we want to relax the condition. We say that f is automorphic, with automorphy factor $\phi_\gamma(z)$, if it satisfies the functional equation

$$f(\gamma z) = \phi_\gamma(z) f(z)$$

This is very similar to what we did with theta functions. If we have two such functions with the same factor, their ratio would be invariant. Note that for this to work, we need to impose a consistency condition

$$\phi_{\gamma\xi}(z) f(z) = f(\gamma\xi z)$$

$$= \phi_\gamma(\xi z) f(\xi z) = \phi_\gamma(\xi z)\phi_\xi(z) f(z)$$

Cancelling f, leads to a so called cocycle condition on the automorphy factor

$$\phi_{\gamma\xi}(z) = \phi_\gamma(\xi z)\phi_\xi(z)$$

As the terminology suggests, ϕ_γ does give an element of a certain cohomology group. Rather than pursuing this direction, let us look for natural automorphic forms/factors in nature. Given a meromorphic differential form $\omega = f(z)dz$ on $\mathbb{H}$, let us see how it transforms under $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. We can see that

$$\omega \mapsto f(\gamma \cdot z) d\left(\frac{az+b}{cz+d}\right) = (cz+d)^{-2} f(\gamma \cdot z) dz$$

We say that $f(z)$ is a weakly modular form of weight 2, with respect to Γ, if $f(z)dz$ is invariant. We say that f is *weakly modular of weight* $2k$ if it

$$f(z) = (cz+d)^{-2k} f\left(\frac{az+b}{cz+d}\right) \tag{2.2}$$

This means that the tensor $f(z)dz^{\otimes k}$ is invariant. More generally, it makes sense to consider weakly modular forms of arbitrary integer weight ℓ, satisfying

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

However, when ℓ is odd, taking $\gamma = -I$, shows that $f = -f$, so it's zero! Natural nonzero examples do exist for other groups however, as we shall see shortly.

To drop the "weakly", we impose holomorphy conditions on $\mathbb{H}$ but also at infinity. To understand what the last part means, we first note that by using S and T, (2.2) is equivalent to

$$\begin{aligned} f(z+1) &= f(z) \\ f(-1/z) &= z^k f(z) \end{aligned} \tag{2.3}$$

The first condition means that we have a Fourier expansion

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

where $q = e^{2\pi i z}$. Note that as $z \to i\infty$, $q \to 0$. So we want to think of q as the local parameter at infinity. Then the Fourier series becomes the Laurent series in q. f is a *modular form of weight* $2k$ if it is holomorphic in $\mathbb{H}$, (2.2) holds, and the Fourier coefficients $a_n = 0$ for $n < 0$. It is called a *cusp form* of weight $2k$ if in addition $a_0 = 0$.

Theorem 2.3.1. *The Eisenstein series*

$$G_{2k}(z) = \sum_{\mathbb{Z}^2 - 0} \frac{1}{(mz+n)^{2k}}$$

is a modular form of weight $2k$*, when* $k \geq 2$*.*

$$\Delta(z) = (60G_4(z))^3 - 27(140G_6(z))^2$$

is a cusp form of weight 12*.*

Proof. The sum can be seen to converge uniformly on compact sets, so it must converge to a holomorphic function on $\mathbb{H}$. One has

$$G_{2k}(\frac{az+b}{cz+d}) = (cz+d)^{2k} \sum \frac{1}{(ma+ndc)z + (mb+nd))^{2k}}$$

The vectors $(ma + ndc, mb + nd)$ can be seen to run over $\mathbb{Z}^2 - 0$. So the right side can be rewritten as

$$(cz + d)^{2k} G_{2k}(z)$$

as required.

We have to check holomorphy at infinity. By uniform convergence, we can evaluate the limit as $z \to \infty$ term by term. When $m \neq 0$, have $(mz+n)^{-2k} \to 0$ as $z \to \infty$. Therefore

$$\lim_{z\to\infty} G_{2k}(z) = 2\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = 2\zeta(2k)$$

where ζ is the Riemann zeta function. Euler gave explicit formulas for the values

$$\zeta(4) = \frac{\pi^4}{90}$$

$$\zeta(6) = \frac{\pi^6}{945}$$

This allows us to evaluate $\lim_{z\to\infty} \Delta(z)$ and check that it's zero. $\square$

Corollary 2.3.2.

$$j(z) = 1728\frac{(60G_4(z))^3}{\Delta}$$

is weakly modular of weight 0.

Finally, let us consider Jacobi's theta function. This is a function of two variables $\theta(z, \tau)$. We already studied the behaviour in the first, now we consider the second where we set $z = 0$.

$$\theta(0, \tau) = \sum_{n\in\mathbb{Z}} \exp(\pi i n^2 \tau)$$

From this formula, we see that

$$\theta(0, \tau + 2) = \theta(0, \tau)$$

There is also a somewhat subtler functional equation.

Theorem 2.3.3. *We have*

$$\theta(0, -1/\tau) = \sqrt{-i\tau}\,\theta(0, \tau)$$

where the complex square root needs to be handled with the usual care.

Sketch. We need the Poisson summation formula [DM], which tells us that

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

where f is a rapidly decreasing smooth (aka Schwartz) function, and

$$\hat{f}(v) = \int_{-\infty}^{\infty} f(u)e^{-2\pi iuv}du$$

is its Fourier transform. The Fourier transform of the Gaussian $e^{-\pi u^2\tau}$ is $\tau^{-1/2}e^{-\pi v^2/\tau}$. Therefore the Poisson summation formula shows that

$$T(1/y) = \sqrt{y}T(y)$$

where $T(y) = \theta(0, iy)$. The theorem follows by analytic continuation. □

Corollary 2.3.4.

$$\theta(0, -1/\tau)^2 = -i\tau\theta(0, \tau)^2$$

The last equation plus the previous periodicity suggests that $\theta(0, 1/\tau)^2$ is a modular form of some kind. In fact, it is a modular form of weight one for a subgroup $\Gamma(4)$ to be defined below. See [MT, p 39].

2.4 Modular curves

With the topology of $X(1) = \bar{\mathcal{A}}_1$ constructed earlier, which is homeomorphic to $\mathbb{P}^1$, we can construct a sheaf of functions $\mathcal{O}_{X(1)}$ as follows. Let $\Gamma(1) = SL_2(\mathbb{Z})$. Given a $\Gamma(1)$-invariant open set $\tilde{U} \subset \mathbb{H}$, let us say that a holomorphic function f on it is modular of weight $2k$ if (2.2) holds and the negative Fourier coefficients vanish when $\infty \in \overline{\tilde{U}}$. Given an open set $U \subset X(1)$, let $f \in \mathcal{O}_{X(1)}(U)$ be a modular form on the preimage $\pi^{-1}U \cap \mathbb{H}$ of weight 0. We can view f as a function on U, where the value at $z \in U - \{\infty\}$ is the value at any of the preimages, and the value at ∞ is the zeroth Fourier coefficient.

Proposition 2.4.1. *The ringed space $(X(1), \mathcal{O}_{X(1)})$ is a Riemann surface.*

Sketch. The key point is to show that any point $x \in X(1)$ has a neighbourhood D with a homeomorphism z, called a local coordinate or parameter, to a disk in $\mathbb{C}$, such that holomorphic functions on both disks coincide. There are three cases: $x = \infty$, x is an image of one of the fixed points $i, e^{2\pi i/3}$, or x is any other point. The first case was essentially done in the last section, q is the local coordinate at ∞. The third case is straight forward. The map $\pi : \mathbb{H} \to X(1)$ is unramified over x A local coordinate z at a point $y \in \mathbb{H}$ lying over x will give a local coordinate at x. The map π is ramified at i and $e^{2\pi i/3}$ with ramification index $e = 2$ and 3 respectively. z^e will give a local coordinate at the image. □

It is worth noting that the images of i and $e^{2\pi i/3}$ are nonsingular, and therefore no different from any other point from this point of view. However, these points clearly are special. One way to keep track of this, is the to use the language of orbifolds or stacks. To simplify our story, we won't do this here.

Given an integer $N > 0$, the *principal congruence subgroup of level N* of $\Gamma(1) = SL_2(\mathbb{Z})$ is

$$\Gamma(N) = \ker \Gamma(1) \to SL_2(\mathbb{Z}/N) = \{M \in \Gamma(1) \mid M \equiv I \mod N\}$$

A *congruence* group is a subgroup of $\Gamma(1)$ containing some $\Gamma(N)$. It therefore has finite index in $\Gamma(1)$. Some other important examples are

$$\Gamma_1(N) = \{M \in \Gamma(1) \mid M \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod N\}$$

$$\Gamma_0(N) = \{M \in \Gamma(1) \mid M \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod N\}$$

We have inclusions

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma(1)$$

We can compute the indices.

Proposition 2.4.2.

(a)

$$[\Gamma(1) : \Gamma(N)] = N^3 \prod \left(1 - \frac{1}{p^2}\right)$$

where p runs over primes dividing N.

(b)

$$[\Gamma(1) : \Gamma_1(N)] = N^2 \prod \left(1 - \frac{1}{p^2}\right)$$

(c)

$$[\Gamma(1) : \Gamma_0(N)] = N \prod \left(1 + \frac{1}{p}\right)$$

Proof. We have $[\Gamma(1) : \Gamma(N)] = |SL_2(\mathbb{Z}/N)|$, $[\Gamma_1(N) : \Gamma(N)| = |Z/N|$ and $[\Gamma_0(N), \Gamma_1(N)] = |(\mathbb{Z}/N)^*|$. These can be checked to yield the above formulas. □

Lemma 2.4.3. *$\Gamma(N)$ is torsion free once $N \geq 3$.*

Given such a group, it will act on $\mathbb{H}^*$, let $Y(\Gamma') = \Gamma' \backslash \mathbb{H}$ and let $X(\Gamma') = \Gamma' \backslash \mathbb{H}^*$. The points of $X(\Gamma') - Y(\Gamma')$ are called cusps. We write $Y(N), Y_1(N)$ etc. when the groups are $\Gamma(N), \Gamma_1(N)$. A meromorphic function f on $\mathbb{H}$ is weakly modular form of weight $2k$, with respect to Γ', if (2.2) holds for matrices in Γ'. The isotropy group of ∞ is a finite index subgroup of $\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ so it is of the form $\langle \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \rangle$ for some n. This implies that a weakly modular form satisfies

$$f(z + n) = f(z)$$

So that it has a Fourier expansion in $q = e^{2\pi iz/n}$. A similar Fourier expansion occurs at all the other cusps. We say that f is a modular (resp. cusp) form if it is holomorphic in $\mathbb{H}$ and the negative (resp. nonpositive) Fourier coefficients at each cusp vanish. When extend this to the case where the domain of f is an invariant open set. Then we can turn $Y(\Gamma') \subset X(\Gamma')$ into Riemann surfaces exactly as above. These are called *modular curves*. We have a holomorphic map

$$X(\Gamma') \to X(\Gamma(1)) = A_1 \cong \mathbb{P}^1$$

induced by inclusion $\Gamma' \subset \Gamma(1)$. This is a branched covering. So we can compute the genus using the Riemann-Hurwitz formula, which says that if $Y \to X$ is a degree d branched covering of compact Riemann surfaces of genus $g(Y)$ and $g(X)$, then

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

where e_y is the ramification index which counts the number of sheets which "come together" at y. We will use this to compute for most of the principal congruence groups. More general formulas can be found in [DS, S].

Theorem 2.4.4. *When $N \geq 3$, the genus of $X(N) = X(\Gamma(N))$ is*

$$g = 1 + \frac{d(N-6)}{12N}$$

where

$$d = \frac{1}{2}[\Gamma(1) : \Gamma(N)] = \frac{N^3}{2} \prod \left(1 - \frac{1}{p^2}\right)$$

The genus of $X(2)$ is zero.

Proof. The covering $\pi : X(N) \to X(1)$ is Galois with group $G = PSL_2(\mathbb{Z})/\operatorname{im}\Gamma(N) = PSL_2(\mathbb{Z}/N)$. The degree of this covering $|G| = d$, when $N \geq 3$, and $d = 6$ when $N = 2$. Let p_2 and p_3 represent the images of i and $e^{2\pi i/3}$ in $X(1)$. Then p_2, p_3, ∞ are the ramification points. Given one of these points p, and $q \in \pi^{-1}(p)$, e_q is the order of the isotropy group $G_q = \{g \in G\ gq = q\}$. This independent of q, because all the isotropy groups are conjugate. It also follows that $|\pi^{-1}(p)| = d/|G_q|$. So we can make a table consisting of p, $|\pi^{-1}(p)|$, $|e_q|$:

$$p_2, d/2, 2$$

$$p_3, d/3, 3$$

$$\infty, d/N, N$$

Putting these into Riemann-Hurwitz and simplifying proves the theorem. □

Using this formula, we can see that the first nonzero value for g occurs at $N = 7$, then $g = 3$. Note that $X(7)$ has an action of $PSL_2(\mathbb{Z}/7)$. The cardinality of this $168 = 84(g-1)$, which is the maximal possible size for an automorphism group by a theorem of Hurwitz. Formulas for the genera of other modular curves can be found in [DS, S].

2.5 Dimension of spaces of modular forms

Given a smooth curve X and a divisor D, let $\Omega^1_X(D) = \Omega^1_X \otimes \mathcal{O}_X(D)$. It can be identified with $\mathcal{O}_X(K+D)$, where K is a canonical divisor. The space of global sections $\Gamma(X, \Omega^1_X(D))$ can be identified with the space of meromorphic 1-forms ω satisfying $\operatorname{div}\omega + D \geq 0$.

Theorem 2.5.1. *Suppose that Γ' is a torsion free congruence group. Let $X = X(\Gamma')$ and let $D = \sum p_i$ be the sum of cusps. The space weight $2k$ modular forms (resp. cusp forms) $M_{2k}(\Gamma')$ (resp. $S_{2k}(\Gamma')$) is isomorphic to $\Gamma(X, \mathcal{O}(kK + kD))$ (resp. $\Gamma(X, \mathcal{O}(kK + (k-1)D))$. In particular, $S_2(\Gamma') \cong \Gamma(X, \Omega^1_X)$*

Proof. Let $f(z) \in M_{2k}(\Gamma')$. Then $f(z)(dz)^{\otimes k}$ is a Γ'-invariant holomorphic section of $(\Omega^1_{\mathbb{H}})^{\otimes k}$, so it descends to a holomorphic section of $(\Omega^1_{Y(\Gamma')})^{\otimes k}$. We have to check what happens near a cusp. We have a local coordinate $q = e^{2\pi i z/n}$. By assumption f can be expanded as $\sum_0^\infty a_m q^m$, with $a_0 = 0$ for a cusp form. We have $dz = (n/2\pi i)dq/q$. So

$$f(z)(dz)^{\otimes k} = \left(\frac{n}{2\pi i}\right)^k (a_0 q^{-k} + a_1 q^{1-k} + \ldots) dq^{\otimes k}$$

So the theorem follows.

□

Corollary 2.5.2. *Suppose that X has genus g with m cusps, then*

$$\dim S_{2k}(\Gamma') = \begin{cases} g & \text{if } k = 1 \\ (2k-1)(g-1) + (k-1)m & \text{if } k > 1 \end{cases}$$

Proof. The first case is an immediate consequence of the theorem. For the second, we use Riemann-Roch.

$$\begin{aligned} h^0(\mathcal{O}(kK + (k-1)D)) &= h^0(\mathcal{O}(kK + (k-1)D)) - h^0(\mathcal{O}((1-k)K - (k-1)D) \\ &= \deg(kK + (k-1)D) + 1 - g \end{aligned}$$

□

A product of a modular form of weight $2k$ and 2ℓ is clearly a modular form of weight $2(k+\ell)$. Therefore $\bigoplus_k S_{2k}(\Gamma')$ is a graded $\mathbb{C}$-algebra.

Corollary 2.5.3. *The algebra of modular forms is finitely generated.*

Proof. This follows from the standard fact that the algebra

$$\bigoplus_k H^0(X, \mathcal{O}(kE))$$

is finitely generated, whenever X is a compact Riemann surface and E is divisor with $\deg E \geq 0$. □

We refer to [DS, S] for more general formulas allowing k to be odd and Γ' to have torsion. Using these formulas, one can show that the algebra of modular forms for $SL_2(\mathbb{Z})$ is generated by the Eisenstein series G_4 and G_6.

2.6 Moduli interpretation

As we saw, $Y(1)$ parameterizes elliptic curves. While it's intuitively clear what this means, the actual statement requires a bit more precision. Let us define an analytic family of (compact) complex manifolds to be a (proper) holomorphic submersion of complex manifolds $f : E \to B$. We recall that a submersion is map such that derivative is surjective on tangent spaces. This implies that fibres $E_b = f^{-1}(b)$ are complex submanifolds. By an analytic family of elliptic curves we mean an analytic family of compact complex manifolds $f : E \to B$ with a holomorphic section $s : B \to E$ such that each fibre E_b is a compact Riemann surface of genus one. We can regard E_b as an elliptic curve with origin $s(b)$.

Theorem 2.6.1. *$Y(1)$ has the following properties:*

(a) The map $E \to j(E)$ gives a bijection between the set of isomorphism classes of elliptic curves over $\mathbb{C}$ and points of $Y(1)$,

(b) Given an analytic family elliptic curves $E \to B$, the map $B \to Y(1)$, called the classifying map, given by $b \mapsto j(E_b)$ is holomorphic.

The statement can be strengthened to completely characterize $Y(1)$, but we need a bit of terminology. Let $Ell^{an}(B)$ be the set of isomorphism classes of analytic families of elliptic curves over B, where isomorphism has the obvious meaning. Given a holomorphic map $B' \to B$, the pullback $E \mapsto E \times_B B'$ gives a map $Ell^{an}(B) \to Ell(B')$ which makes it into a contravariant functor. More generally, let $M(-)$ be contravariant functor from the category of complex manifolds (or schemes or...) to sets; one thinks of elements of $M(B)$ as families of objects over B of interest. We say that M is *representable* by U, or that U is a *fine moduli space* for M, if there is a natural isomorphism of functors

$$M(B) \cong Hom(B, U)$$

Yoneda's lemma, in category theory, tells us that U is completely determined by this property, and moreover it carries a universal family such that any object in $M(B)$ is the pullback of it under some map $B \to U$. Although this is ideal scenario for any moduli problem, it fails for Ell^{an}. This is because there exists nontrivial families in $Ell^{an}(B)$ with constant j-invariant. Here is a general construction.

Example 2.6.2. *Let E be either E_i or $E_{\exp(2\pi i/3)}$ Either curve has a nontrivial automorphism group G, which is cyclic in both cases. Choose a manifold $\tilde{B}$ on which G acts freely, e.g. $\mathbb{C}^*$. The quotient $(E \times \tilde{B})/G \to \tilde{B}/G$ is a nontrivial family with constant j-invariant.*

In spite of this bad news, we do have a natural transformation $Ell^{an}(B) \to Hom(B, Y(1))$, which is universal in an appropriate sense, and which induces a bijection when B is a point. We say that $Y(1)$ is the *coarse moduli space* for Ell^{an}.

The other modular curves have similar interpretations. Let us explain the characterizations of $Y(N) = Y(\Gamma(N))$ in a somewhat informal way. Given $E_\tau = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$, the image of $(\frac{1}{N}, \frac{\tau}{N})$ gives a basis for the N-torsion points in E_τ. We refer to this as a level N-structure. If $A \in \Gamma(N)$, then the induced isomorphism $E_\tau \cong E_{A\cdot\tau}$ takes the above level N-structure of the first curve to the level structure of the second. In order to make this notion independent of our representation of E_τ as a quotient, we note that the lattice is isomorphic to homology $L_\tau \cong H_1(E_\tau, \mathbb{Z})$. Thus a level N-structure is a choice of basis for $H_1(E, \mathbb{Z}/N\mathbb{Z}) = H_1(E, \mathbb{Z}) \otimes \mathbb{Z}/N\mathbb{Z}$, but not just any basis. The group carries an intersection pairing

$$H_1(E, \mathbb{Z}/N) \times H_1(E, \mathbb{Z}/N\mathbb{Z}) \to \mathbb{Z}/N$$

So now we can say that a level N-structure is a basis for $H_1(E, \mathbb{Z}/N\mathbb{Z})$ which is symplectic in the sense that the matrix of the above pairing is

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

Theorem 2.6.3. *$Y(N)$ is the coarse moduli space of elliptic curves with level N-structure. When $N \geq 3$ it is a fine moduli space.*

Recall that the assumption $N \geq 3$ is precisely the condition to guarantee that $\Gamma(N)$ is torsion free. This same condition also allows us to kill the automorphism groups which created the problem in example 2.6.2.

For the other moduli spaces, we have similar interpretations. $Y_1(N) = Y(\Gamma_1(N))$ is the coarse moduli space of pairs (E, P) consisting of an elliptic curve E and a point P of order N. $Y_0(N)$ is the moduli space of pairs (E, C) consisting of an elliptic curve and a cyclic subgroup of the group of N-torsion points. The projections

$$Y(N) \to Y_1(N) \to Y_0(N) \to Y(1)$$

induced by the inclusions

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma$$

have moduli interpretations. Given E a level N-structure is a pair of N-torsion points P, Q satisfying suitable conditions. The map $Y(N) \to Y_1(N)$ corresponds to the forgetful map $(E, P, Q) \mapsto (E, P)$.

2.7 Models over number fields

So far we have considered modular curves as Riemann surfaces, but in fact they are algebraic curves. This is true of any compact Riemann surface minus a finite set of points. However, a more natural way to see this is to consider algebraic versions of the moduli problems consider earlier. As a bonus this will show that these curves are naturally defined over number fields and even over rings of integers. This is very important for applications to number theory. Let us start with $Y(1)$. We consider the corresponding moduli problem in the algebraic setting. Given a scheme B, an elliptic curve over it is a smooth proper map is a smooth proper map $f : E \to B$, with a section, such that the closed fibres of f are genus one curves. Let $Ell(B)$ denote the isomorphism classes of elliptic curves over B. Then $Y(1)_{\mathbb{Z}} = \operatorname{Spec} \mathbb{Z}[j]$ is a coarse moduli scheme for $Ell(-)$, and this gives a model for $Y(1)$ over the integers, i.e. $Y(1)$ is the complex manifold associated to $Y(1)_{\mathbb{Z}} \times_{\operatorname{Spec} \mathbb{Z}} \operatorname{Spec} \mathbb{C}$.

Next let us turn to $Y(N)$. We can formulate the definition of level structure of an elliptic curve over an arbitrary field k. In this case, we need N to be prime to the characteristic. Then a level N-structure is a pair of N-torsion points $P, Q \in E(k)$ such that they generate the group of N-torsion points and such that $e_N(P, Q)$ is a primitive N-root of unity. Here e_N is the Weil pairing whose definition can be found in [Si]. Note that the condition forces k to contain a primitive N-root of unity. More generally, there is a notion of a level structure for an elliptic curve over a base scheme. This is basically a pair of sections which induces a level structure on the closed fibres.

Theorem 2.7.1. *There exists a scheme $Y(N)$ defined over $\mathbb{Z}[1/N, e^{2\pi i/N}]$ which is the coarse moduli space of elliptic curves with level N-structure. It is a fine moduli when $N \geq 3$. The set of complex points is the Riemann surface $\Gamma(N)\backslash\mathbb{H}$ considered before.*

See Deligne-Rapoport [DR] for the construction in general. They also give a more general construction which would include the $Y_i(N)$. $Y_0(N)$ is particularly interesting because it is defined over $\mathbb{Q}$. When N is small, $Y(N)$ can be made very explicit. We have

$$Y(2) = \operatorname{Spec} \mathbb{Z}[t, \frac{1}{t(t-1)}]$$

Although this is not fine, there is an "almost" universal family called the Legendre family

$$y^2 z = x(x - z)(x - tz)$$

in $\mathbb{P}^2_{\mathbb{Z}}$. Over a field, this curve has 4 branch points over $0, 1, t, \infty$. Take the first to be the origin, and the next two to be the level 2-structure.

When $N = 3$, let $R = \mathbb{Z}[1/3, e^{2\pi i/3}]$, then

$$Y(N) = \operatorname{Spec} R[t, \frac{1}{t^3 - 1}]$$

The universal family is given by the elliptic curve

$$x^3 + y^3 + z^3 = 3txyz$$

in $\mathbb{P}^2_R$ with section $[1, -1, 0]$. The level 3-structure is given by the sections $[-1, 0, 1]$ and $[-1, e^{2\pi i/3}, 0]$.