# Modular Curve as Moduli Spaces

Kirk Bonney

## Overview

The goal of this project is to develop a clear picture of how modular curves form moduli spaces for elliptic curves with additional torsion data. Along the way, I hope to build a 'compendium' of facts and perspectives for modular curves that can serve as a useful resource for myself and others.

## 1 Motivation

Modular curves are quotients of the upper half space by a subgroup of the full modular group $\mathrm{SL}_2(\mathbb{Z})$. These objects are compact Riemann surfaces, which immediately endows them with topological, differential, and algebraic structure. Relating modular curves to a problem means we can bring a breadth of mathematics long with it. For number theorists, modular curves are useful for (at least) two problems: studying modular forms and studying torsion of elliptic curves. The focus of this paper is the latter.

### 1.1 Modular Forms (Modularity and $\mathcal{M}_k(\Gamma)$)

The Modularity Theorem is well known for its role in Wiles' proof of Fermat's Last Theorem. A form of its statement can be made using modular curves.

**Theorem 1.1.** *Let $E$ be a complex elliptic curve with $j(E) \in \mathbb{Q}$. Then for some positive integer $N$ there exists a surjective, holomorphic function of compact Riemann surfaces from the modular curve $X_0(N)$ to the elliptic curve $E$,*

$$X_0(N) \to E.$$

In other words, the theorem states that all elliptic curves with rational invariant can be parametrized by a modular curve. The typical form of the modularity theorem involves modular forms, however, recalling that spaces of modular forms are defined, just as modular curves are, using a congruence subgroup, it might not be surprising that either object can be used to state the theorem. In fact, modular curves have an important role in understanding modular forms. The space of modular forms of weight $k$ with respect to the congruence subgroup $\Gamma$ is a finite dimensional vector space, denoted $\mathcal{M}_k(\Gamma)$). The dimension of this space can be calculated using the genus of the modular curve corresponding to $\Gamma$.

### 1.2 Elliptic Curves (Mazur's theorem)

When studying Elliptic Curves over the rational numbers, the structural possibilities for the torsion subgroup is rather restricted.

**Theorem 1.2 (Mazur (1977)).** *Let $E$ be an elliptic curve.*

$$E(\mathbb{Q})_{tors} \cong \mathbb{Z}_N \text{ where } N = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12$$

*or*

$$E(\mathbb{Q})_{tors} \cong \mathbb{Z}_N \oplus \mathbb{Z}_N \text{ where } N = 1, 2, 3, 4$$

This theorem is proven using modular curves. As we will discuss later, modular curves can be viewed as moduli spaces of elliptic curves with additional torsion data. Then we can morph questions like "is it possible for $G$ to arise as a torsion subgroup of an elliptic curve?" to "is there a rational point on this modular curve?' So understanding modular curves well means understanding torsion of elliptic curves well. Currently, there is a great deal of research being done on torsion for elliptic curves over other fields such as number fields using modular curves.

## 2 Modular Curves

This section will lay out the fundamental definitions and concepts necessary for working with modular curves. A natural starting point is the modular curve for $\mathrm{SL}_2(\mathbb{Z})$.

### 2.1 The modular curve for $\mathrm{SL}_2(\mathbb{Z})$, briefly

First recall the definition of the full modular group is the set of all integral $2 \times 2$ matrices with determinant 1. This group is finitely generated by

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

There is a transitive action[1] of $\mathrm{SL}_2(\mathbb{Z})$ on the upper half plane $\mathcal{H}$ defined by

$$\gamma \cdot \tau = \frac{a\tau + b}{c\tau + d} \; \forall \gamma \in \mathrm{SL}_2(\mathbb{Z}) \; \forall \tau \in \mathcal{H}.$$

A fun resource to get acquainted with how this action looks can be found at here.

Via this action, we form a quotient of the upper half plane $\mathcal{H}/\mathrm{SL}_2(\mathbb{Z})$. This quotient can be made into a Riemann surface with appropriate charts, and a point at infinity can be added to $\mathcal{H}$ to make the Riemann surface compact.

We can construct a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$. Define

$$\mathcal{D} := \{\tau \in \mathcal{H} : |\tau| \geq 1 \text{ and } |\mathrm{Re}(\tau)| \leq 1/2\}.$$

[1] This action surjects into the set of isometries for the upper-half plane model of hyperbolic space.

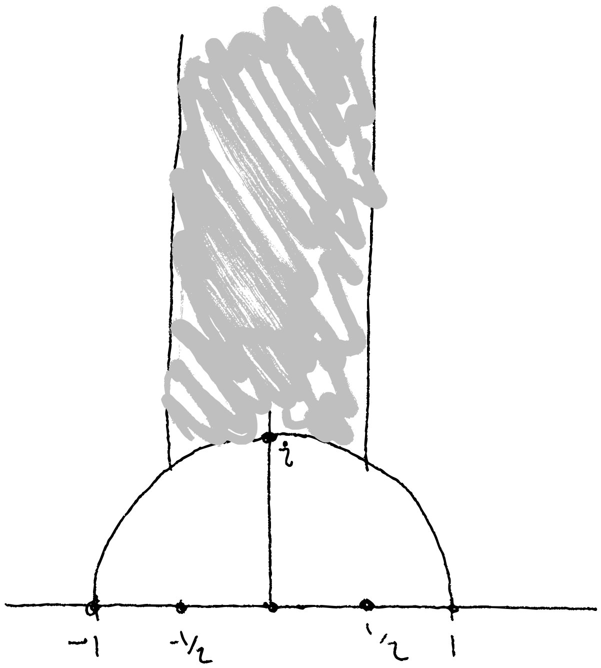

Figure 1: The fundamental domain $\mathcal{D}$.

Then every point in the upper half plane is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to exactly one point in $\mathcal{D}$, with the exception of those points equivalent to a point on the boundary of $\mathcal{D}$. The generator $T$, mentioned above, gives us the action $\tau \mapsto \tau + 1$, so the portion of the boundary on the line $\mathrm{Re}(\tau) = 1/2$ is equivalent to the portion on the line $\mathrm{Re}(\tau) = -1/2$. Further, the generator $S$ gives us the action $\tau \mapsto -\frac{1}{\tau}$, which identifies the two portions of the unit circle that are split by the imaginary axis. Note that $i$ is not glued to anything else via $S$, as it is actually fixed by $S$. This will be relevant later.

### 2.2 Congruence subgroups & their modular curves

By examining subgroups of $\mathrm{SL}_2(\mathbb{Z})$, we can form other modular curves.

**Definition 2.1.** A congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is any subgroup which contains the principal congruence subgroup $\Gamma(N)$, defined by

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

for some $N$. The lowest such $N$ is called its level.

We will be concerned with three particular families of congruence subgroups. In addition to $\Gamma(N)$ we can define

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

and

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

Note the containments $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$. For any $\Gamma$, there is a natural map $\pi : Y(\Gamma) \to Y(\mathrm{SL}_2(\mathbb{Z}))$ defined by

$$\pi(\Gamma\tau) = \mathrm{SL}_2(\mathbb{Z})\tau,$$

which is well-defined as $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$. We can describe much of the structure of $\Gamma$ using this mapping.

For an in-depth introduction to $\mathrm{SL}_2(\mathbb{Z})$ and its congruence subgroups, we invite the reader to read this well-written document from Keith Conrad.

Congruence subgroups inherit the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$, so we may form quotients of the upper half plane as we did before. For any congruence subgroup $\Gamma$, we can write $Y(\Gamma) = \mathcal{H}/\Gamma$. We also use the notations $Y_0(N) = Y(\Gamma_0(N))$, $Y_1(N) = Y(\Gamma_1(N))$, and $Y(N) = Y(\Gamma(N))$ Note that $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$, so for concision we write $Y(1)$ for $Y(\mathrm{SL}_2(\mathbb{Z}))$.

A quotient by a congruence subgroup $Y(\Gamma)$ can be made into a Riemann surface using charts, which we refer to as the modular curve for $\Gamma$. We will see that we can compactify this surface, in this case we write $X(\Gamma) = \mathcal{H}^*/\Gamma$ and generalize the notation from before to $X_0(N) = X(\Gamma_0(N))$ and so on. The creation of charts for a modular curve is *almost* straightforward. With the right lemmas in hand, we can cover most of $X(\Gamma)$ using the natural quotient map $\phi : \mathcal{H} \to \mathcal{H}/\Gamma$. The only difficulty arises at a set of finitely many problematic points which are characterized by having non-trivial stabilizers in $\Gamma$. These points are important to the understanding of modular curves, so we will spend some time discussing them. However, we won't describe how to deal with them while creating charts. For that we refer the reader to Diamond and Shurman's book [2]. In particular, we will see that these points tell us where the natural map $\pi$ ramifies, which will be for determining the genus of a modular curve.

### Compactification

When we compactify $Y(\Gamma)$, the added points turn out to be problem points as they come with non-trivial stabilizers. We'll look at the process of compactification in detail for $\mathrm{SL}_2(\mathbb{Z})$, and see what happens in the more general situation for any $\Gamma$.

If we identify $Y(\mathrm{SL}_2(\mathbb{Z}))$ with its fundamental domain $\mathcal{D}$ (Figure 1), then it is easy to see why it is not compact. As we move up the imaginary axis, we have no accumulation point. This is why it is necessary to add a point at infinity to the upper-half plane. When we do this, we can draw a complete picture of our curve in the following way.

- (I) We start with our fundamental domain $\mathcal{D}$.
- (II) When we add $\infty$, we can image the surface pinching off at the top to represent the compactification.
- (III) The fundamental domain has redundant points along the boundary. In particular, the lines with real part equal to $\pm 1/2$ must be identified since they are equivalent under the action $\tau \mapsto \tau + 1$. There identification is represented by the dashed line.
- (IV) Finally, we have to 'zip up' the open circle, since two halves of the circle are identified under the action $\tau \mapsto \frac{-1}{\tau}$. This is a bit hard to draw, so the reader is invited to use their imagination in place of our illustration.

Indeed, modular curves are something our 3D-bound brains can visualize (Figure 2). Hopefully this illustration makes it clear that our compactified curve is topologically equivalent to the Riemann sphere. However, we have to be careful now that we have added a point to $\mathcal{H}$. We have only defined the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$, so for the quotient to make sense we must extend it to $\infty$. Take an arbitrary $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and let it act on $\infty$

$$\gamma \cdot \infty = \frac{a \cdot \infty + b}{c \cdot \infty + d}.$$

How should we make sense of this? The most straightforward way is to "take the limit" and view the quantity as $a/c$, a rational number. This isn't actually in the upper-half plane though, so along with $\infty$ we must add $\mathbb{Q}$ so that the orbit of $\infty$ under the action is well-defined. So really what we called $\mathcal{H}^*$ is not just the upper half-plane and infinity; it includes a copy of $\mathbb{Q}$ as well. We write $\mathcal{H}^* = \mathcal{H} \cup \mathbb{Q} \cup \{\infty\}$ and call $\mathbb{Q} \cup \{\infty\}$ the cusps. Since the orbit $\mathrm{SL}_2(\mathbb{Z})\infty$ hits all of the added points, we sometimes abuse notation and write it as just $\infty$.

When we look at a subgroup $\Gamma$, the cusps no longer lives in a single orbit.

**Example 2.1.** Consider $\Gamma(2)$. Then the action of this subgroup cannot carry $\infty$ to 2. Suppose it could,

$$\frac{a \cdot \infty + b}{c \cdot \infty + d} = a/c = 2.$$

This implies that $a = 2$ and $c = 1$, since the two are coprime, but the conditions of $\Gamma(2)$ mean that $a \equiv 1 \bmod (2)$ and $c \equiv 0 \bmod (2)$, a contradiction.

The moral is that as $\Gamma$ becomes smaller[2], its action becomes more restricted and we obtain more distinct orbits among $\mathbb{Q} \cup \infty$.

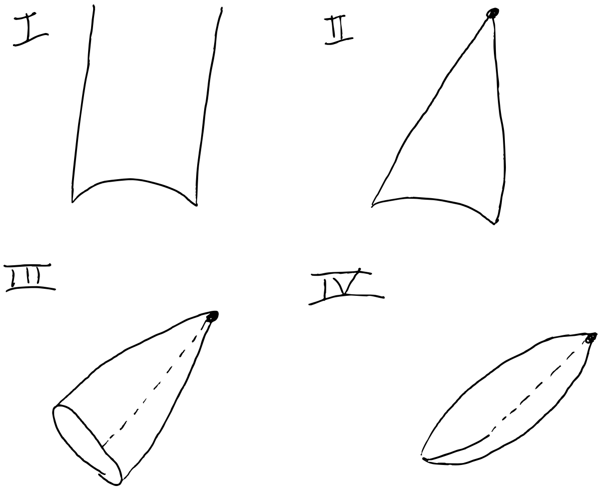

Figure 2: Visualization of the compactification of $Y(\mathrm{SL}_2(\mathbb{Z}))$.

**Definition 2.2.** A point $\Gamma\tau$ of $X(\Gamma)$ is said to be a **cusp** for $X(\Gamma)$ if its image under the natural map $\pi : X(\Gamma) \to X(\mathrm{SL}_2(\mathbb{Z}))$ is $\mathrm{SL}_2(\mathbb{Z})\infty$.

In other words, $\Gamma\tau$ is a cusp of $X(\Gamma)$ if it contains rational numbers and possibly $\infty$.

At the beginning of the section, it was claimed that cusps have non-trivial stabilizer. Before proceeding any further, let's make clear what is meant by this.

**Definition 2.3.** We write the **stabilizer** of $\tau \in \mathcal{H}^*$ in $\Gamma$ to be the set

$$\Gamma_\tau := \{\gamma \in \Gamma \mid \gamma \cdot \tau = \tau\}.$$

In particular, if the contaiment $\{\pm I\}\Gamma_\tau \subset \{\pm I\}$ is proper, we say the stabilizer $\Gamma_\tau$ is non-trivial.

[2] We interpret small as having a large index in $\mathrm{SL}_2(\mathbb{Z})$.

**Example 2.2.** Certainly when $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, we have that $\infty$ is stabilized by all of $\langle T \rangle$ as $\infty + n$ is equal to $\infty$ for any $n \in \mathbb{Z}$. Now suppose that $\Gamma$ is not $\mathrm{SL}_2(\mathbb{Z})$ and has level $N$. Then the matrix $A = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}$ belongs to $\Gamma$[3]. Since the stabilizer $\Gamma_\infty$ is just equal to $\mathrm{SL}_2(\mathbb{Z})_\infty \cap \Gamma$, it follows that it is non-empty as $A \in \mathrm{SL}_2(\mathbb{Z})_\infty$ and $A \in \Gamma$.

Cusps are a necessary addition to modular curves because they turn these objects into compact Riemann surfaces, the theory of which is very useful. Once we give similar exposition for elliptic points, we will be prepared to understand the ramification of the natural map $\pi$, and with the help of the Riemann-Hurwitz theorem, we can calculate the genus of $X(\Gamma)$.

### Elliptic Points

Elliptic points are the non-cuspoidal troublemakers.

**Definition 2.4.** Let $\Gamma$ be a congruence subgroup. Then $\tau$ is an **elliptic point for** $\Gamma$ if $\Gamma_\tau$

That is, they are the points in $\mathcal{H}$ which have non-trivial stabilizers under the action of $\Gamma$. As in the case of cusps, it will be easiest to understand the elliptic points for $\mathrm{SL}_2(\mathbb{Z})$ before approaching the general case. Suppose we have a point $\tau \in \mathcal{H}$ which is fixed by some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ where $\gamma \neq \pm I$. Then

$$\gamma \cdot \tau = \frac{a\tau + b}{c\tau + d} = \tau$$

implies that $\tau$ is the root of an integral quadratic

$$c\tau^2 + (d-a)\tau - b = 0.$$

Applying the quadratic equation and keeping in mind that $\tau$ lives in the upper halfplane, we see that

$$\sqrt{(d-a)^2 + 4cb} \implies |a+d| < 2.$$

Further, if we require that $\tau$ be in our fundamental domain $\mathcal{D}$ we can show that our possible quadratics are $x^2 + 1$ or $x^2 \pm x + 1$. Thus, $\tau$ must either be $i$ or $\mu_3 = e^{2\pi i/3}$. Note, the conjugate root for the first polynomial is not in $\mathcal{H}$, and the conjugate root for the second polynomial is in fact $\mathrm{SL}_2(\mathbb{Z})$ equivalent to $\mu_6$, so it is safe to ignore them. From here one can prove the following.

**Proposition 2.1.** *[2] The elliptic points of the modular curve for $SL_2(\mathbb{Z})$ are $SL_2(\mathbb{Z})i$, with isotropy subgroup*

$$SL_2(\mathbb{Z})_i = \left\langle \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$$

*and $\tau = \mu_3$ with isotropy subgroup*

$$SL_2(\mathbb{Z})_{\mu_3} = \left\langle \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \right\rangle.$$

[3] This fact is why modular forms for congruence subgroups have Fourier expansions.

To move forward into general territory, the following result is key.

**Proposition 2.2.** *[2] Let $\Gamma$ be a congruence subgroup of $SL_2(\mathbb{Z})$. The modular curve $Y(\Gamma)$ has finitely many elliptic points. For each elliptic point $\tau$ of $\Gamma$ the isotropy subgroup $\Gamma_\tau$ is finite cyclic.*

*Proof.* It is a fact that congruence subgroups have finite index inside of $\mathrm{SL}_2(\mathbb{Z})$, so for any $\Gamma$ we may write as a disjoint union

$$\mathrm{SL}_2(\mathbb{Z}) = \bigcup_{j=1}^{d} \Gamma\gamma_j.$$

Clearly, if a point of $\Gamma$ has non-trivial stabilizer in $\Gamma$, then the same is true for the image of that point under the natural map $\pi : X(\Gamma) \to X(\mathrm{SL}_2(\mathbb{Z}))$. When we move backwards through $\pi$, a point in $X(\mathrm{SL}_2(\mathbb{Z}))$ splits into $d$ points of $X(\Gamma)$. With the description of $\mathrm{SL}_2(\mathbb{Z})$ above, we can describe these explicitly. A point $\mathrm{SL}_2(\mathbb{Z})\tau$ has the following pre-image under $\pi$

$$\pi^{-1}(\mathrm{SL}_2(\mathbb{Z})\tau) = \{\Gamma\gamma_j(\tau) \mid 1 \le j \le d\}.$$

It follows that all elliptic points for $X(\Gamma)$ must be in the pre-image of either $i$ or $\mu_3$. Additionally, the stabilizer $\Gamma_\tau$ of an elliptic point is a conjugate to a subgroup of $\mathrm{SL}_2(\mathbb{Z})_i$ or $\mathrm{SL}_2(\mathbb{Z})_{\mu_3}$, both of which are cyclic. □

It is important to note that the converse is not true; every point in the pre-image of an elliptic point for $X(\mathrm{SL}_2(\mathbb{Z}))$ is not necessarily a elliptic point for $X(\Gamma)$. Fortunately still, we know what points to check and that there are finitely many of them. The statement holds analogously for cusps. Using the strategy laid out in the proof, let's work an example for finding both the cusps and elliptic points for $\Gamma(2)$.

**Example 2.3.** Our first step will be finding coset representatives for $\Gamma(2)$ in $\mathrm{SL}_2(\mathbb{Z})$. Using SAGE, we can find that the index of $\Gamma(2)$ in $\mathrm{SL}_2(\mathbb{Z})$ is 6. From here, we find by hands-on calculation that

$$\gamma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \gamma_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \gamma_3 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \gamma_4 = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \gamma_5 = \begin{bmatrix} -1 & 0 \\ -1 & -1 \end{bmatrix}, \gamma_6 = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}$$

is a complete collection of representatives for the cosets. Next we use these to find the inverse images of our three distinguished points: $i$ and $\mu_3$ for elliptic points, and $\infty$ for the cusps. We have that

$$\pi^{-1}(\mathrm{SL}_2(\mathbb{Z})i) = \{\Gamma\gamma_j(i) \mid 1 \le j \le 6\} = \{\Gamma(2)i, \Gamma(2)1 + i, \Gamma(2) - 1/2 + 1/2i\}.$$

Note that the stabilizer of $i$ intersects trivially with $\Gamma(2)$, therefor none of the above points are elliptic. The same is true for the stabilizer of $\mu_3$. In this case we have that none of

$$\pi^{-1}(\mathrm{SL}_2(\mathbb{Z})\mu_3) = \{\Gamma\gamma_j(\mu_3) \mid 1 \le j \le 6\} = \{\Gamma(2)\mu_3, \Gamma(2)\mu_3 + 1, \Gamma(2)\frac{1}{\sqrt{3}}e^{\frac{5\pi i}{6}}\}.$$

are elliptic. Finally, the cusps are given by

$$\pi^{-1}(\mathrm{SL}_2(\mathbb{Z})\infty) = \{\Gamma\gamma_j(\infty) \mid 1 \le j \le 6\} = \{\Gamma(2)\infty, \Gamma(2)0, \Gamma(2)1\}.$$

In each case we have a collapsing of the pre-image down from the expected six points. This is the mark of ramification and demonstrates why elliptic points and cusps are important to study when understanding a map between modular curves.

### 2.3 Genus formula

Our discussion of cusps and elliptic points have set us up to give an explicit formula for the genus of a modular curve $X(\Gamma)$ as a Riemann surface. This formula is built from the Riemann-Hurwitz formula, which describes the relationship of the genus of two Riemann surfaces given by a map between the two. Before we jump into these formulas, we will establish some concepts concerning maps between Riemann surfaces.

**Definition 2.5.** Let $X$ and $Y$ be Riemann surfaces, and let $f : X \to Y$ be a complex analytic mapping between the two. Near each point $z_0$ of the domain we can choose a chart centered at $z_0$ so that, locally, $f(z) = z^n$ for some $n$. We denote this quantity by $\text{mult}_{z_0}(f) = n$. When $n > 1$, we say that the point is **ramified** and that the ramification degree of $\pi$ at $z_0$ is $e_{z_0} = n$.

Note that it is not immediate that the charts mentioned can be found, but they do indeed exist. It is useful to view $z^n$ as the model for ramification, so we use the flexibility of the chart structure to mold our local perspective into this model. Doing so captures special behavoir of the function at a ramified point $z_0$. Here, the function looks like an $n$-to-one mapping in a deleted neighborhood of $z_0$, since non-zero values have a pre-image of size $n$ under $z^n$. At $z_0$, however, the function is 1-1.

Before stating the Riemann-Hurwitz formula, we need to define one last quantity.

**Definition 2.6.** [4] It is a theorem that the quantity

$$d_y(f) = \sum_{p \in f^{-1}(y)} \text{mult}_p(f)$$

does not depend on $y$, and thus is a characteristic of the function. We call this quantity the **degree of** $f$.

Now we state the main tool used to calculate the genus.

**Theorem 2.1.** *(The Riemann-Hurwitz Formula)*[4]

*Let $X$ and $Y$ be compact Riemann surfaces, and let $f : X \to Y$ be a nonconstant holomorphic map of degree $d$. Then*

$$2g_X - 2 = d(2g_Y - 2) + \sum_{x \in X} (e_x - 1)$$

*where $e_x$ is the ramification index of $x$.*

Two immediate corollaries are that there is no non-constant holomorphic map to a surface of higher genus and a non-constant map between surfaces of equal genus $g \geq 1$ is unramified. We will apply this to the specific case where $X = X(\Gamma)$ and $Y = X(\text{SL}_2(\mathbb{Z}))$ and $f = \pi$ is the natural map. Since we know that the genus of $X(\text{SL}_2(\mathbb{Z}))$ is zero, all we must do is understand the degree and ramification of the natural map to be able to calculate the genus of $X(\Gamma)$. The follow lemma describes the ramification of $\pi$, and is a key piece of the genus formula.

[4] I have been told there are a couple dozen different proofs of this theorem.

**Lemma 2.1.** *Let $\Gamma$ be a congruence subgroup, and let $\pi : X(\Gamma) \to X(SL_2(\mathbb{Z}))$ be the natural map.*

1. *Let $\tau \in \mathcal{H}$. Define $h = |SL_2(\mathbb{Z})_\tau|/2$. Then the ramification degree for a point $\tau$ is given by*

$$e_\tau = \begin{cases} h & \text{if } \tau \text{ is an elliptic point for } SL_2(\mathbb{Z}) \text{ but not for } \Gamma \\ 1 & o.w. \end{cases}$$

2. *Let $\tau \in \mathbb{Q} \cup \{\infty\}$. Then the ramification for $\tau$ is given by*

$$e_\tau = [SL_2(\mathbb{Z})_\tau : \{\pm I\}\Gamma_\tau],$$

*the index of the stabilizer of $\tau$ in $\Gamma$ with the trivial action accounted for in the stabilizer of $\tau$ in $SL_2(\mathbb{Z})$.*

To avoid giving a full description of the chart structure on modular curves, we omit the proof of this lemma. The punchline of the theorem is that all ramification for $\pi$ happens at the cusps and elliptic curves. Let's apply this in our pursuit of the genus formula.

**Theorem 2.2.** *(Genus Formula for Modular Curves)*

*Let $\Gamma$ be a congruence subgroup of $SL_2(\mathbb{Z})$ and let $d$ be the degree of the natural map $X(\Gamma) \to X(1)$. Let $\epsilon_2$ denote the number of elliptic points of period 2, $\epsilon_3$ the number of elliptic points of period 3, and $\epsilon_\infty$ the number of cusps of $\Gamma$ (i.e., the number of orbits of $\Gamma$ on $\mathbb{Q} \cup \{\infty\}$. Then then genus $G$ of $X(\Gamma)$ is*

$$G = 1 + \frac{d}{12} - \frac{\epsilon_2}{4} - \frac{\epsilon_3}{3} - \frac{\epsilon_\infty}{2}$$

*Proof.* Plugging our scenario into Riemann-Hurwitz gives us

$$2 - 2G = 2d - \sum_{x \in X(\Gamma)} (e_x - 1).$$

Since we know that ramification only happens at the cusps and elliptic points, we can write

$$\sum_{x \in X(\Gamma)} = \sum_{x \in \pi^{-1}(i)} e_x - 1 + \sum_{x \in \pi^{-1}(\mu_3)} e_x - 1 + \sum_{x \in \pi^{-1}(\infty)} e_x - 1.$$

Next we address each sum individually. Consider the points living above $i$. Those which are elliptic will not ramify, so they do not contribute to the sum. Write $\epsilon_2$ for the number of points which are elliptic in $X(\Gamma)$. Those which are not elliptic, do ramify with degree 2 (Prop 2.1 & Lemma 2.1). Recalling that $d$ is the number of points in $\pi^{-1}(i)$ counted with multiplicity, it follows that the number of ramified points is $(d - \epsilon_2)/2$ and that

$$\sum_{x \in \pi^{-1}(i)} e_x - 1 = (d - \epsilon_2)/2.$$

Now consider the points living above $\mu_3$. Those which are elliptic will not ramify, so they do not contribute to the sum. Write $\epsilon_3$ for the number of points which are elliptic in $X(\Gamma)$. Those which are not elliptic, do ramify with degree 3 (Prop 2.1 & Lemma 2.1). As before, we can calculate that the number of ramified points is $(d - \epsilon_3)/3$ and that

$$\sum_{x \in \pi^{-1}(\mu_3)} e_x - 1 = 2(d - \epsilon_2)/3.$$

Finally, we examine the points above $\infty$. Here we don't have a distinction of cusps versus non-cusps; everything in $\pi^{-1}(\infty)$ is a cusp by definition. Write $\epsilon_\infty$ for the number of cusps in $X(\Gamma)$. Then we have

$$\sum_{x \in \pi^{-1}(\infty)} e_x - 1 = d - \epsilon_\infty.$$

Putting this all together, we have

$$2 - 2G = 2d - \frac{d - \epsilon_2}{2} - \frac{2(d - \epsilon_3)}{3} - d + \epsilon_\infty.$$

□

To see this formula in action, we will continue with our example of $\Gamma(2)$.

**Example 2.4.** We know from before that the degree of the natural map $\pi : X(2) \to X(1)$ is 6. We also know that the pre-image of $i$ has 3 elements, none of which were elliptic. The same is true for $\mu_3$. The number of cusps is 3. We apply the formula.

$$G = 1 + \frac{6}{12} - \frac{0}{4} - \frac{0}{3} - \frac{3}{2} = 0.$$

So we have successfully calculated that the genus of $X(2)$ is 0.

Using the `fareysymbol` package from SAGE, we can print out a visual for a fundamental domain of $\Gamma(2)$. We leave it as an exercise to visualize how this pastes together to the Riemann sphere. We'll end the discussion on the genus of modular curves with a more general result.

**Example 2.5.** We can find the genus of $X_1(N)$ where $N$ is a prime greater than 3. In this case, one can prove that the degree of $\pi$ is $\frac{N^2-1}{2}$, there are no elliptic points in $X_1(N)$, and that there are $N - 1$ cusps. Hence the genus is

$$G = \frac{(N-7)(N+1)}{12}.$$

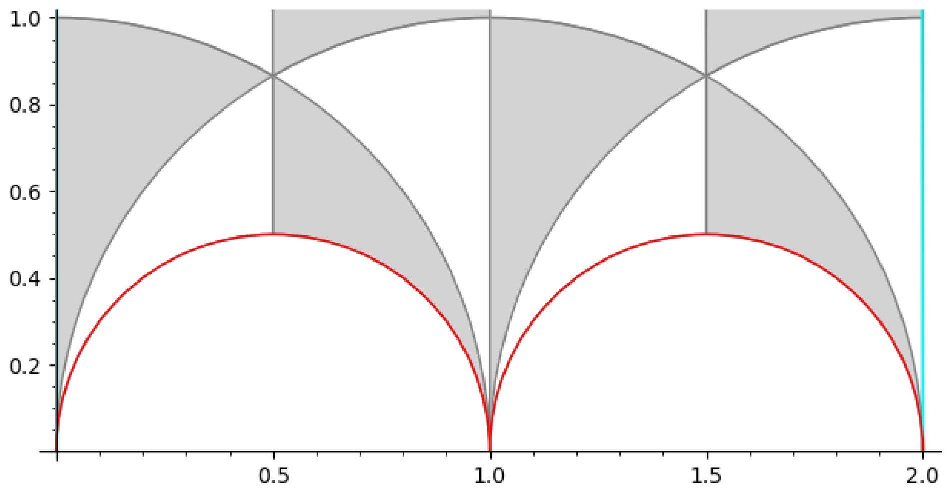

Figure 3: A fundamental domain for $\Gamma(2)$.

## 3 Modular Curves as Moduli Spaces

Now that we have become acquainted with some of the fundamental features of modular curves, we will take a look at how modular curves can be used in the study of elliptic curves. In particular, we will see that modular curves serve as moduli spaces for elliptic curves with attached torsion data. Typical definitions of moduli spaces involve a general perspective situated in the language of category theory. We won't need this for our purposes. Instead it will suffice to define a **moduli space** for a set of geometric objects $\mathcal{P}$ to be a some geometric object $X$ whose points are in bijective correspondence with $\mathcal{P}$.

In our case, $\mathcal{P}$ will be elliptic curves under various kinds of isomorphism involving torsion and $X$ will a modular curve $X(\Gamma)$. To prepare ourselves for this point of view, let's first examine the simple case where $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Recall that an elliptic curve $E$ may be identified with a point $\tau$ in the upper-half plane by viewing it as the torus $\mathbb{C}/\Lambda_\tau$, where $\Lambda_\tau$ is the $\mathbb{Z}$-lattice generated by 1 and $\tau$. There are many possible $\tau$ for a fixed $E$, so the map $\phi$ taking $\tau$ to its corresponding elliptic is surjective, but not injective. To make it injective, the right thing to do is to mod out by the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$ because $\tau$ and $\tau'$ give the same elliptic curve if and only if they are $\mathrm{SL}_2(\mathbb{Z})$ equivalent. Then, writing $\mathcal{S}$ as the set of elliptic curves under typical isomorphism,

$$\phi' : \mathcal{H}/\mathrm{SL}_2(\mathbb{Z}) \to \mathcal{S}$$

is a bijection. Therefor, we can call $Y(\mathrm{SL}_2(\mathbb{Z}))$ a moduli space for $\mathcal{S}$.

To expand this approach to the rest of the congruence subgroups we are familiar with, $\Gamma_0(N), \Gamma_1(N), \Gamma(N)$, we must describe different kinds of equivalence classes of elliptic curves.

**Definition 3.1.** [2] An **enhanced elliptic curve for** $\Gamma_0(N)$ is an ordered pair $(E, C)$ where $E$ is a complex elliptic curve and $C$ is a cyclic subgroup of $E$ with order $N$. Two enhanced elliptic curves for $\Gamma_0(N)$, $(E, C)$ and $(E', C')$ are said to be isomorphic if there exists an isomorphism $f : E \to E'$ such that $f(C) = C'$. The set of equivalence classes is denoted

$$\mathcal{S}_0(N) = \{\text{enhanced elliptic curves for } \Gamma_0(N)\}/\sim .$$

An **enhanced elliptic curve for** $\Gamma_1(N)$ is an ordered pair $(E, Q)$ where $E$ is a complex elliptic curve and $Q$ is a point of $E$ with order $N$. Two enhanced elliptic curves for $\Gamma_1(N)$, $(E, Q)$ and $(E', Q')$ are said to be isomorphic if there exists an isomorphism $f : E \to E'$ such that $f(Q) = Q'$. The set of equivalence classes is denoted

$$\mathcal{S}_1(N) = \{\text{enhanced elliptic curves for } \Gamma_1(N)\}/\sim .$$

An **enhanced elliptic curve for** $\Gamma(N)$ is an ordered pair $(E, (P, Q))$ where $E$ is a complex elliptic curve and $(P, Q)$ is a pair of points of $E$ which generate the torsion subgroup $E[N]$. Two enhanced elliptic curves for $\Gamma(N)$, $(E, (P, Q))$ and $(E', (P', Q'))$ are said to be isomorphic if there exists an isomorphism $f : E \to E'$ such that $f(P) = P'$ and $f(Q) = Q'$. The set of equivalence classes is denoted

$$\mathcal{S}(N) = \{\text{enhanced elliptic curves for } \Gamma(N)\}/\sim .$$

The more restrictions we put on $\Gamma$, the more information its corresponding set of elliptic curves will contain. Note that $\mathcal{S}(N)$ surjects to $\mathcal{S}_1(N)$ by forgetting the second generator, $\mathcal{S}_1(N)$ surjects to $\mathcal{S}_0(N)$ by taking the cyclic subgroup generated by the specific point, and that $\mathcal{S}_0(N)$ surjects to $\mathcal{S}$ by forgetting the torsion data entirely. This reflects the natural maps we have between the congruence subgroups.

The key statement is that the modular curve $Y(\Gamma)$ is a moduli space for the set of enhanced elliptic curves for $\Gamma$. We state this formally while also giving explicit descriptions of the sets of enhanced elliptic curves. In what follows, we write $E_\tau$ to mean $\mathbb{C}/\Lambda_\tau$.

**Theorem 3.1.** *[2]*

*(a) We can describe the enhanced elliptic curves for* $\Gamma_0(N)$ *by*

$$\mathcal{S}_0(N) = \{(E_\tau, \langle 1/N + \Lambda_\tau \rangle) \mid \tau \in \mathcal{H}\}.$$

*Two points* $(E_\tau, \langle 1/N + \Lambda_\tau \rangle)$ *and* $(E_{\tau'}, \langle 1/N + \Lambda_{\tau'} \rangle)$ *are equal if and only if* $\Gamma_0(N)\tau = \Gamma_0(N)\tau'$. *So there exists a bijection*

$$\psi_0 : \mathcal{S}_0(N) \to Y_0(N), \quad (E_\tau, \langle 1/N + \Lambda_\tau \rangle) \mapsto \Gamma_0(N)\tau$$

*(b) We can describe the enhanced elliptic curves for* $\Gamma_1(N)$ *by*

$$\mathcal{S}_1(N) = \{(E_\tau, 1/N + \Lambda_\tau) \mid \tau \in \mathcal{H}\}.$$

*Two points* $(E_\tau, 1/N + \Lambda_\tau)$ *and* $(E_{\tau'}, 1/N + \Lambda_{\tau'})$ *are equal if and only if* $\Gamma_1\tau = \Gamma_1(N)\tau'$. *So there exists a bijection*

$$\psi_1 : \mathcal{S}_1(N) \to Y_1(N), \quad (E_\tau, 1/N + \Lambda_\tau) \mapsto \Gamma_1(N)\tau$$

*(c) We can describe the enhanced elliptic curves for* $\Gamma(N)$ *by*

$$\mathcal{S}(N) = \{(E_\tau, (\tau/N + \Lambda_\tau, 1/N + \Lambda_\tau)) \mid \tau \in \mathcal{H}\}.$$

*Two points* $(E_\tau, (\tau/N + \Lambda_\tau, 1/N + \Lambda_\tau))$ *and* $(E_{\tau'}, (\tau'/N + \Lambda_{\tau'}, 1/N + \Lambda_{\tau'}))$ *are equal if and only if* $\Gamma(N)\tau = \Gamma(N)\tau'$. *So there exists a bijection*

$$\psi : \mathcal{S}(N) \to Y(N), \quad (E_\tau, (\tau/N + \Lambda_\tau, 1/N + \Lambda_\tau)) \mapsto \Gamma(N)\tau$$

*Proof.* We will prove part $(a)$ of the theorem. Essentially, what we must show is that the additional torsion information carried within $S_0(N)$ is exactly the information preserved by the action of $\Gamma_0(N)$, and that for any point in $S_0(N)$ we may choose a representative of the form $(E_\tau, \langle 1/N + \Lambda_\tau \rangle)$.

First we'll find the distinguished representative for an arbitrary point $(E, C) \in S_0(N)$. We know that $E$ is isomorphic to $E_{\tau'}$ for some $\tau' \in \mathcal{H}$, so we may assume WLOG that $E = E_{\tau'}$. Let $k$ be a generator for $C$. Then we know $k$ will take the form

$$K = \frac{c}{N}\tau' + \frac{d}{N} + \Lambda_{\tau'}$$

where $\gcd(c, d, N) = 1$, since $k$ must have order $N$. Then we can find an $a$ and $b$ so that

$$ad - bc - lN = 1$$

for some $l$ so that the matrix $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. We may modify the entries as we please modulo $N$ without changing $K$ as defined. So we can assume that $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, since $\mathrm{SL}_2(\mathbb{Z})$ naturally maps onto $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Defining $\tau = \gamma \cdot \tau'$ and $m = c\tau' + d$ [5] one can check that $m\Lambda_\tau = \Lambda_{\tau'}$, which implies $E_\tau \cong E_{\tau'}$. Further, $m\left(1/N + \Lambda_\tau\right) = \frac{c}{N}\tau' + \frac{d}{N} + \Lambda_{\tau'} = Q$. Thus, in $S_0(N)$, $(E, \langle K \rangle) \cong (E_\tau, \langle 1/N + \Lambda_\tau \rangle)$.

It is worth noting that the above argument made no special reference to $\Gamma_0(N)$, so we can use the same line of reasoning to find the distinguished representatives in the other cases. In fact, what was proved is the choice of representative for part $(b)$, which was immediately applied to prove it also for $(a)$. Now we proceed to show that the bijection holds.

Suppose $\tau, \tau' \in \mathcal{H}$ satisfy $\Gamma_0(N)\tau = \Gamma_0(N)\tau'$. We can write $\gamma \cdot \tau' = \tau$ for some $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma_0(N)$. As before define $m = c\tau' + d$. Then $m\Lambda_\tau = \Lambda_{\tau'}$ and

$$m\left(1/N + \Lambda_\tau\right) = c\tau'/N + d/N + \Lambda_{\tau'}.$$

Here is where $\Gamma_0(N)$ becomes relevant. We have assumed that $c \equiv 0 \bmod N$, so then the above expression is equal to $d/N + \Lambda_{\tau'}$. A consequence of $\gamma \in \Gamma_0(N)$ is that $\gcd(d, N) = 1$, so it follows that $d/N + \Lambda_{\tau'}$ has order $N$, thus it generates the same cyclic group as $1/N + \Lambda\tau'$. So then the homothety given by $m$ gives an isomorphism for the enhanced elliptic curves $(E_\tau, \langle 1/N + \Lambda_\tau \rangle)$ and $(E_{\tau'}, \langle 1/N + \Lambda_{\tau'} \rangle)$.

Lastly, suppose that $(E_\tau, \langle 1/N + \Lambda_\tau \rangle)$ and $(E_{\tau'}, \langle 1/N + \Lambda_{\tau'} \rangle)$. Then we have some homothety $m$ such that $m\Lambda_\tau = \Lambda_{\tau'}$ and $m\langle 1/N + \Lambda_\tau \rangle = \langle 1/N + \Lambda_{\tau'} \rangle$. Further, since $E_\tau$ and $E_{\tau'}$ must be isomorphic, we have some $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma \cdot \tau' = \tau$. Then the correspondence between homothety and $\mathrm{SL}_2(\mathbb{Z})$-equivalence tells us that $m = c\tau' + d$. As before, we know that

$$m\left(1/N + \Lambda_\tau\right) = c\tau'/N + d/N + \Lambda_{\tau'}.$$

However we know that $m(1/N + \Lambda_\tau) = k/N + \Lambda_{\tau'}$ for some $k$ with $\gcd(k, N) = 1$, since $m(1/N + \Lambda_\tau)$ generates $\langle 1/N + \Lambda_{\tau'} \rangle$. Thus it follows that $c \equiv 0 \bmod N$ and $k = d$, proving that $\gamma \in \Gamma_0(N)$.

□

[5] I like to think of this $m$ as the $\mathrm{SL}_2(\mathbb{Z})$ equivalence of $\tau$ and $\tau'$ translated to the world of lattices, since $m$ gives a homothety between $\Lambda_\tau$ and $\Lambda_{\tau'}$.

# 4 Modular Curves as Tools

Consider Mazur's theorem on the structure of torsion subgroups of elliptic curves over $\mathbb{Q}$.

**Theorem 4.1 (Mazur (1977)).** *Let $E$ be an elliptic curve.*

$$E(\mathbb{Q})_{tors} \cong \mathbb{Z}_N \text{ where } N = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12$$

*or*

$$E(\mathbb{Q})_{tors} \cong \mathbb{Z}_N \oplus \mathbb{Z}_N \text{ where } N = 1, 2, 3, 4$$

Using the machinery we have built up, we can immediately rephrase this question in terms of modular curves. Recall that the curve $Y_0(N)$ is a moduli space for elliptic curves with a specified cyclic subgroup of size $N$. It turns out that modular curves are also algebraic, so we can ask about rational points on $Y_0(N)$. Such points $(E, C)$ are those where $E$ is defined over $\mathbb{Q}$ and $C \subset E(\mathbb{Q})$. Now a portion of Mazur's theorem can be rephrased to the non-existence of a rational point on $Y_0(N)$ where $N$ is a prime greater than 7. The following theorem provides a key step towards proving this statement.

**Theorem 4.2.** *[6] Suppose $N > 7$ and there exists an Abelain variety $A/\mathbb{Q}$ and a map of varieties $f : X_0(N) \to A$ defined over $\mathbb{Q}$ such that the following hold*

- *$A$ has good reduction away from $N$.*
- *$f(0) \neq f(\infty)$.*
- *$A(\mathbb{Q})$ has rank 0.*

*Then there is no elliptic curve defined over $\mathbb{Q}$ which has a point of order $N$.*

So the challenge is finding such an $A$. One ends up realizing $A$ as a quotient of the Jacobian of $X_0(N)$. This path requires one to reach deep into the modular forms side of the theory of modular curves. We won't delve far in this direction, but we'll give some definitions to give the flavor of the kind of tools used in this approach.

**Definition 4.1.** The **Jacobian** of a compact Riemann surface $X$ is the quotient group

$$\mathrm{Jac}(X) = \Omega^1_{\mathrm{hol}}(X)^\wedge / \mathrm{H}_1(X, \mathbb{Z})$$

where $\Omega^1_{\mathrm{hol}}(X)^\wedge$ is the dual of the set of holomorphic one forms on $X$ and $\mathrm{H}_1(X, \mathbb{Z})$ is the first homology group of $X$.

The Jacobian of $X$ can also be realized using the theory of divisors. One can understand $\mathrm{Jac}(X)$ as a pointed, universal Abelian variety which $X$ maps through. In the proof of Mazur's theorem, the Jacobian is used to construct the $A$ mentioned in Theorem 4.2. To do so, one also utilizes the Hecke Algebra for $\Gamma_0(N)$.

**Definition 4.2.** [3] Let $n$ be a positive integer and let $M$ be the set of integral matrices with determinant $n$. Let $f$ be a modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight $k$. Define the Hecke operator $T_n$ to be

$$(T_n f)(z) := n^{k-1} \sum_{\mu \in M/\mathrm{SL}_2(\mathbb{Z})} f|[\mu]_k$$

where

$$(f|[\gamma]_k)(z) = (cz - d)^k f\left(\frac{az+b}{cz+d}\right)$$

for $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $z \in \mathcal{H}$.

Clearly, there is a lot more going on here than we have covered. What we can say for now is that the Hecke operators are linear operators on modular forms. We have defined them for $\mathrm{SL}_2(\mathbb{Z})$, but they can be defined for congruence subgroups as well. In each of these cases, one can show that these linear operators form an algebra, given a fixed $\Gamma$. Within this algebra, one seeks out a particular ideal and uses this to form the desired quotient $A$ of $\mathrm{Jac}(X_0(N))$.

## 5 Conclusion

We have explored quotients of the upper-half plane by the action of congruence subgroups $\mathcal{H}/\Gamma$. We call such a space a modular curve $Y(\Gamma)$, which can be endowed with the structure of Riemann surface. This brings about the study of cusps (to compactify) and elliptic points (to build the complex atlas). These distinguished points also play an important role in determining the genus of our now compact Riemann surface $X(\Gamma)$.

After establishing these basic properties of the curve, we finished the discussion on modular curves by surveying their role in Mazur's theorem. There we saw that modular curves are effective in solving number-theoretic problems over $\mathbb{Q}$; it turns out that we can generalize them over cyclotomic and number fields to extract results for elliptic curves over these fields. We focused our attention on modular curves in relation to elliptic curves, but they have widespread application, showing up in the study of modular forms as well as having a large role in the monstrous moonshine conjectures.

## References

[1] Keith Conrad. The modular group. https://kconrad.math.uconn.edu/blurbs/grouptheory/SL(2,Z).pdf. Accessed: 4/01/2021.

[2] Fred Diamond and Jerry Shurman. *A First Course in Modular Forms.* Springer New York, New York, NY, 2016.

[3] L. J. P. (Lloyd James Peter) Kilford. *Modular forms : a classical and computational introduction.* YBP Print DDA. Imperial College Press, London, 2nd edition. edition, 2015.

[4] Rick Miranda. *Algebraic curves and Riemann surfaces*. Graduate studies in mathematics ; v. 5. American Mathematical Society, Providence, R.I, 1995.

[5] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics, 106. Springer New York, New York, NY, 1986.

[6] Andrew Snowden. Course on mazur's theorem. http://www-personal.umich.edu/~asnowden/teaching/2013/679/index.html. Accessed: 4/01/2021.

[7] Lori Watson. An introduction to modular groups. https://www.math.arizona.edu/~swc/. Accessed: 4/01/2021.