CUSP WIDTH IN MODULAR CURVES

AVERY GIRSKY

Abstract. An elliptic curve over the complex field is isomorphic to the quotient of the complex plane by a lattice Λ. The quotient of the upper half plane by an action of subgroups of $\mathrm{SL}_2(\mathbb{Z})$ define moduli spaces of the isomorphism classes of elliptic curves and their N-torsion points. These quotient spaces, known as modular curves, are compactified by gluing in a finite number of points, known as cusps. In this paper we will explore the isomorphism classes of elliptic curves under varying group actions, study interesting properties of cusps and their width, and examine how cusps behave under maps between modular curves.

1. Introduction

Elliptic Curves have many important applications in mathematics, including consequences in group theory, factorization, primality testing, and cryptography. As a result, we are motivated to study these curves in general and their moduli spaces in particular.

Given that $\mathbb{F}$ is a field, the most elementary representation is the set of solutions in a field $\mathbb{F}$ of the cubic equation

$$y^2 = x^3 + Ax + B,$$

where $A, B \in \mathbb{F}$. Specifically considering elliptic curves where $\mathbb{F} = \mathbb{C}$, there are two additional and equivalent representations.

- An elliptic curve over $\mathbb{C}$ is a genus 1 Riemann surface together with a distinguished point.
- An elliptic curve over $\mathbb{C}$ is the quotient of the complex plane by a lattice.

In the latter case, one may imagine a lattice of points in the complex plane, which defines repeating parallelograms formed where each pair of opposite sides will be identified with each other. In our study of elliptic curves, we will focus primarily on the complex torus representation.

For two elliptic curves $E_\tau = \mathbb{C}/\Lambda_\tau$ and $E_{\tau'} = \mathbb{C}/\Lambda_{\tau'}$ viewed as complex tori, if they are isomorphic curves with $\tau' = \frac{a\tau+b}{c\tau+d}$, then $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Therefore we are able to define the group action of $SL_2(\mathbb{Z})$ on the complex upper half plane $\mathcal{H}$ (where $E_\tau \mapsto \tau \in \mathcal{H}$). This action reduces the upper half plane to a space known as the moduli space of elliptic curves where each orbit of curves is represented. We will note the same process can be performed using congruence subgroups of $SL_2(\mathbb{Z})$, for example

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod N \right\}.$$

Considering the space Y_Γ formed from the quotient of a congruence subgroup Γ on $\mathcal{H}$, we will come to define cusps of the subgroup Γ. We then define the compactified modular curve X_Γ to be

Date: May 12, 2023.

the space Y_Γ union the cusps. On top of defining cusps, we will come to study them in depth. We will especially focus on the topological impact of cusps, computing the width of a cusp, and counting the number of cusps in X_Γ.

Eventually we will place an emphasis on the cusps of $\Gamma_1(N)$ and especially how the cusps map from $X_1(N) \to X_1(M)$ for $M|N$. We will develop theorems that count the number of cusps in $X_1(N)$ that map to a specific cusp of width w in $X_1(M)$.

2. Elliptic Curves

The most familiar definition of elliptic curves is the first representation provided in the introduction. In this representation, we would define an elliptic curve as the curve generated by solutions to equations of the form $y^2 = x^3 + Ax + B$ together with a point at infinity, where $A, B \in \mathbb{F}$ for some field $\mathbb{F}$. Our other two equivalent representations for elliptic curves over the complex field are a genus 1 Riemann surface with a distinguished point and the quotient of $\mathbb{C}$ by a lattice. In developing the theory for the latter representation, we now look to define a lattice in the complex plane.

Definition 2.1. A *lattice* in the complex plane is a set $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ where $\{\omega_1, \omega_2\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$.

It is worth noting that by definition, 0 is a point on any lattice in $\mathbb{C}$. From an arbitrary lattice Λ with given $\{\omega_1, \omega_2\}$, we can construct an isomorphic lattice $\Lambda_\tau = \mathbb{Z} \oplus \mathbb{Z}\tau$, where $\tau = \omega_2/\omega_1$. Without loss of generality, we may assume the imaginary part of τ is positive because ω_1, ω_2 can be switched. Further, $\mathrm{Im}(\tau) \neq 0$ as otherwise ω_1, ω_2 would not form a basis. Now, we present the definition of a complex torus.

Definition 2.2. A *complex torus* is the quotient of the complex plane by a lattice Λ. We will denote a complex torus as $E = \mathbb{C}/\Lambda$.

As was hinted at above, an important result shown in Diamond and Shurman is that an isomorphic map can be constructed between complex tori with distinguished points at 0 and elliptic curves over $\mathbb{C}$ [1]. We now will equivalently refer to $E = \mathbb{C}/\Lambda$ as a complex elliptic curve.

From any complex curve $E = \mathbb{C}/\Lambda$, we can construct an isomorphic curve $E_\tau := \mathbb{C}/\Lambda_\tau$. Recall that $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ and $\Lambda_\tau = \mathbb{Z} \oplus \mathbb{Z}(\omega_1/\omega_2)$. We understand that the elliptic curves E and E_τ are isomorphic because the map from $E \to E_\tau$ that takes $z \mapsto z/\omega_1$ is obviously holomorphic and invertible, hence an isomorphism.

This result leads to the question: when are two complex elliptic curves E_τ and $E_{\tau'}$ isomorphic in general?

Suppose that $\phi : E_\tau \to E_{\tau'}$ is an isomorphism. Then ϕ must map some lattice point $a\tau + b$ with $a, b \in \mathbb{Z}$ from E_τ to the point $\tau' \in \Lambda_{\tau'}$. Meanwhile, ϕ maps another lattice point $c\tau + d$ with $c, d \in \mathbb{Z}$ to $1 \in \Lambda_{\tau'}$. In order to learn more about this isomorphism, we will show the specific structure that ϕ must take.

Theorem 2.3. If $\phi : E_\tau \to E_{\tau'}$ is a complex isomorphism between two elliptic curves, then $\phi(z) = mz$, where $m \in \mathbb{C}$.

Proof. We first define the lift of ϕ, denoted $\tilde{\phi}$, as a map from $\mathbb{C} \to \mathbb{C}$ that preserves the lattice points by mapping points in Λ_τ to points in $\Lambda_{\tau'}$.

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\tilde{\phi}} & \mathbb{C} \\ \downarrow & & \downarrow \\ E_\tau & \xrightarrow{\phi} & E_{\tau'} \end{array}$$

We will now fix an arbitrary lattice point $\lambda \in \Lambda_\tau$. Next, we define the complex function

$$g_\lambda(z) = \tilde{\phi}(z + \lambda) - \tilde{\phi}(z).$$

From the continuity of $\tilde{\phi}$, we know g_λ is continuous. We also know that $g_\lambda(z) \in \Lambda_{\tau'}$ because $z + \lambda$ and z are in the same coset in Λ_τ and $\tilde{\phi}$ takes a lattice point in Λ_τ to a lattice point in $\Lambda_{\tau'}$. From this fact joined with the continuity of g_λ and the lattice being a discrete set, we get the fact that g_λ is a constant function. Now by differentiating g_λ, we see that

$$g'_\lambda(z) = \tilde{\phi}'(z + \lambda) - \tilde{\phi}'(z) = 0$$

because g_λ is constant. Then we realize

$$\tilde{\phi}'(z + \lambda) = \tilde{\phi}'(z).$$

The important consequence of this statement is that when working with the function $\tilde{\phi}'$, we can consider only the points $z \in P$, where P is the parallelogram spanned by the basis $1, \tau$ for the lattice Λ_τ. If we have a point $z \in \mathbb{C} - P$, we may translate the point into P by adding a lattice point.

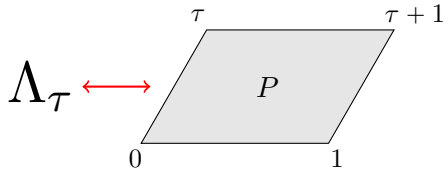


FIGURE 1. The parallelogram P (fundamental domain) that accounts for all cosets of $\mathbb{C}/\Lambda_\tau$ under $\tilde{\phi}'$

Now we have a continuous function $\tilde{\phi}'$ on a bounded space P, so we can conclude by the Extreme Value Theorem that $\tilde{\phi}'$ attains a maximum M. Hence for any $z \in \mathbb{C}$,

$$|\tilde{\phi}'(z)| \leqslant \max_{z \in P} |\tilde{\phi}'(z)| = M.$$

Now by Liouville's Theorem, our holomorphic and bounded function $\tilde{\phi}'$ is constant, so

$$\tilde{\phi}'(z) = m.$$

Lastly, by integrating $\tilde{\phi}'$, we get that

$$\tilde{\phi}(z) = mz + C.$$

However because the origin is fixed in a map between elliptic curves, the additional constant C disappears and we are left with our result $\tilde{\phi}(z) = mz$. □

After proving that $\phi(z) = mz$, we return to our notes about where ϕ takes lattice points to then realize that $m(a\tau + b) = \tau'$ and $m(c\tau + d) = 1$ for some integers a, b, c, and d. This implies that $m = (c\tau + d)^{-1}$ and $\tau' = \frac{a\tau+b}{c\tau+d}$. We conclude that E_τ and $E_{\tau'}$ are isomorphic if and only if $\tau' = \frac{a\tau+b}{c\tau+d}$.

To simplify this construction we will study the matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

Theorem 2.4. If E_τ and $E_{\tau'}$ are isomorphic elliptic curves with $\tau' = \frac{a\tau+b}{c\tau+d}$, then $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, the special linear group of degree 2 over $\mathbb{Z}$.

Proof. First, recall that $\mathrm{Im}(\tau)$, $\mathrm{Im}(\tau') > 0$. We now realize that γ is invertible, because if it were not, τ' would be rational, implying $\mathrm{Im}(\tau') = 0$ and we arrive at a contradiction. Next we seek to find the determinant of γ.

Accomplishing this requires us to study the map between $\Lambda_\tau \to \Lambda_{\tau'}$ as a map between $\mathbb{Z}^2 \to \mathbb{Z}^2$. Hence we note the correspondence between a lattice point $k\tau + l \in \Lambda_\tau$ and the point $(k, l) \in \mathbb{Z}^2$. As we specified, $a\tau + b = \tau'$ and $c\tau + d = 1$, so $(a, b) \mapsto (1, 0) \in \Lambda_{\tau'}$ and $(c, d) \mapsto (0, 1) \in \Lambda_{\tau'}$. Now note that an arbitrary $(k, l) \in \Lambda_\tau$ multiplied by γ will map to $(k, l) \in \Lambda_{\tau'}$. Here we invert this map so $(k, l) \to (k, l)\gamma^{-1}$. Next recognize that

$$\gamma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det(\gamma)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We reach the conclusion that $\det(\gamma) = \pm 1$ because otherwise the image of

$$(k, l) \mapsto (k, l) \frac{1}{\det(\gamma)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

would no longer necessarily be in $\mathbb{Z}^2$. Further, the identity

$$\tau' = \frac{a\tau + b}{c\tau + d} \cdot \frac{c\bar{\tau} + d}{c\bar{\tau} + d} = \frac{ac\tau\bar{\tau} + ad\tau + bc\bar{\tau} + bd}{|c\tau + d|^2}$$

leads us to

$$\mathrm{Im}(\tau') = \frac{(ad - bc)\mathrm{Im}(\tau)}{|c\tau + d|^2}.$$

We now discover that $\mathrm{Im}(\tau') > 0$ if and only if $\det(\gamma) = ad - bc > 0$. We have already asserted $\mathrm{Im}(\tau') > 0$, so we conclude that $\det(\gamma) > 0$. Combined with our previous result, $\det(\gamma) = 1$ and thus $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. □

3. Moduli of Elliptic Curves

The goal of this section is to construct $\mathcal{M}_{1,1}$, the moduli space of elliptic curves. This space will consist of points that have a one-to-one correspondence with isomorphism classes of elliptic curves. We will start by proposing the complex upper half plane as a candidate.

Definition 3.1. The *upper half plane* is $\mathcal{H} = \{\tau \in \mathbb{C} : \mathrm{Im}(\tau) > 0\}$.

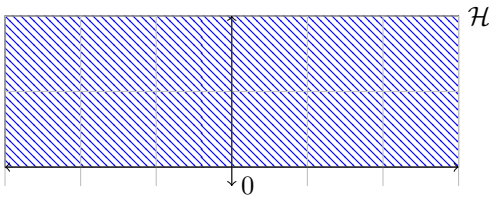


FIGURE 2. The upper half plane $\mathcal{H}$

It is then easy to see that an elliptic curve $E_\tau = \mathbb{C}/\Lambda_\tau$ corresponds to the point τ in $\mathcal{H}$. However, an elliptic curve can correspond to more than one point $\tau \in \mathcal{H}$, specifically any two curves

$$E_\tau = \mathbb{C}/\Lambda_\tau \text{ and } E_{\gamma\tau} = \mathbb{C}/\Lambda_{\gamma\tau}, \gamma \in \mathrm{SL}_2(\mathbb{Z}).$$

We know this to be true because the action of matrices in $\mathrm{SL}_2(\mathbb{Z})$ induces an isomorphism of elliptic curves.

While $\mathcal{M}_{1,1} \neq \mathcal{H}$, we can define a group action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$, where the action will send an elliptic curve $E_\tau, \tau \in \mathcal{H}$ to its isomorphism class of curves.

The obvious and correct candidate for $\mathcal{M}_{1,1}$ is the quotient group $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

Definition 3.2. The *moduli space of elliptic curves* is $\mathcal{M}_{1,1} = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

Our focus now shifts towards a graphical depiction of $\mathcal{M}_{1,1}$.

Note that the group $SL_2(\mathbb{Z})_2(\mathbb{Z})$, otherwise known as the *modular group*, is generated by the two matrices [1, p. 2]

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

If we consider the elliptic curve E_τ corresponding to $\tau \in \mathcal{H}$, we can examine the action of the matrices S and T on τ. First, S acts on τ by

$$S\tau = \frac{1\tau + 1}{0\tau + 1} = \tau + 1.$$

When considering the moduli space $\mathcal{M}_{1,1}$, this result shows us that τ is in the same isomorphism class as $\tau+1$, and subsequently every $\tau+k$ for all $k \in \mathbb{Z}$. Thus, when reducing the upper half plane to this moduli space we notice that the elliptic curves begin repeating themselves (up to isomorphism) every vertical strip of width 1.

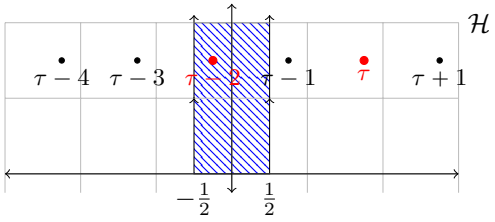


FIGURE 3. The upper half plane reduced to a vertical strip of length 1

Looking at the other generator of $SL_2(\mathbb{Z})_2(\mathbb{Z})$, T acts on τ as follows

$$T\tau = \frac{0\tau + (-1)}{1\tau + 0} = \frac{-1}{\tau}.$$

Now we find that every τ is in an isomorphism class with $\frac{-1}{\tau}$, allowing us to place further restrictions on the width 1 strip of the upper half plane that resulted from the action of S. Specifically, we realize that each point below the unit circle (radius 1) are in the same orbit as a point above. We will also note that the action of S identifies the left boundary with the right boundary, while the action of T identifies the bottom left and bottom right components. We now have all the information to visualize the moduli space $\mathcal{M}_{1,1}$.

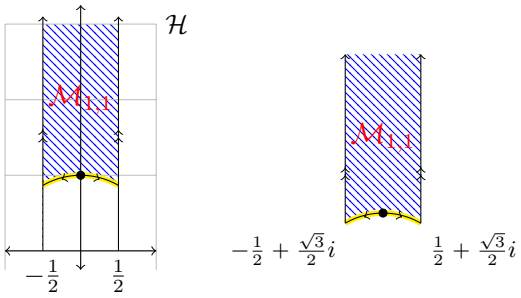


FIGURE 4. A visualization of the moduli space of elliptic curves $\mathcal{M}_{1,1}$

4. Modular Curves

Recall from above that we created the moduli space of elliptic curves from the quotient of the upper half plane by the modular group $SL_2(\mathbb{Z})$. In this section we will look at similarly designed spaces derived from the quotient of $\mathcal{H}$ by subgroups of $SL_2(\mathbb{Z})$ that will be called modular curves. Considering a subgroup of $SL_2(\mathbb{Z})$ indicates that there will be more restrictions placed on which elliptic curves are isomorphic to each other. Consequently there will be fewer curves in an isomorphism class and then the quotient will be a larger space than $\mathcal{M}_{1,1}$.

Definition 4.1. A point P in an elliptic curve $E = \mathbb{C}/\Lambda$ is an *N-torsion point* if $NP \in \Lambda$ (that is, $NP = 0$ in the elliptic curve).

We will now define the three subgroups of $SL_2(\mathbb{Z})$ that we will be working with. Following these definitions, we will make claims regarding how a matrix γ from each of these subgroups act on the N-torsion points between the elliptic curves E_τ and $E_{\gamma\tau}$.

The first of these subgroups is $\Gamma(N)$.

Definition 4.2 ([1, p. 13])**.** Let N be a positive integer. Then the *principal congruence subgroup of level N is*

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}.$$

Note that the matrix congruence is interpreted for each entry, so $a \equiv 1 \mod N$ and so on. We also then observe that $\Gamma(1) = SL_2(\mathbb{Z})$.

Definition 4.3. [1, p. 13] A subgroup Γ of $SL_2(\mathbb{Z})$ is considered a *congruence subgroup* if there exists an $N \in \mathbb{Z}$ such that $\Gamma(N) \subset \Gamma$. In this case, Γ is called a congruence subgroup of level N.

It can then be verified that the two following definitions for $\Gamma_0(N)$ and $\Gamma_1(N)$ are congruence subgroups of $SL_2(\mathbb{Z})$. Note that the $*$ represents any integer value mod N.

$$\Gamma_0(N) \coloneqq \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \bmod N \right\}.$$

$$\Gamma_1(N) \coloneqq \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod N \right\}.$$

We can easily convince ourselves that if N is a positive integer,

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset SL_2(\mathbb{Z})$$

This is clear by the decreasing amount of restrictions placed on the matrices within the groups as we progress from left to right.

As was mentioned at the beginning of this section we will look to take the quotient of the upper half plane by these subgroups, returning what will be defined as a modular curve.

Definition 4.4. If Γ is a congruence subgroup, then $Y_\Gamma = SL_2(\mathbb{Z})\backslash\mathcal{H}$ is a *modular curve*.

The following three definitions will be the modular curves that correspond to the previously defined subgroups of $SL_2(\mathbb{Z})$.

$$Y(N) = \Gamma(N)\backslash\mathcal{H}, \qquad Y_0(N) = \Gamma_0(N)\backslash\mathcal{H}, \qquad Y_1(N) = \Gamma_1(N)\backslash\mathcal{H}.$$

From the previous statement that $\Gamma(1) = SL_2(\mathbb{Z})$, we can now see that $\mathrm{Y}(1) = \mathcal{M}_{1,1}$. This notation will be interchangeable going forward.

Now, if we look at an orbit of elliptic curves in $Y(N)$, we see that these curves are isomorphic by action from $\Gamma(N)$. This indicates that a curve E_τ is isomorphic to $E_{\gamma\tau}$ when $\gamma \in \Gamma(N)$. More information on the geometric implications of this action will come soon.

For the moment, we realize that the orbit of curves isomorphic to E_τ, which corresponds to $\tau \in Y(N)$ has more restrictions placed on it than the orbit $[\tau] \in Y_1(N)$. Then $[\tau] \in Y_1(N)$ has more restrictions than $[\tau] \in Y_0(N)$ which has more restrictions than $[\tau] \in Y(1)$. In this discussion, the additional restrictions on an orbit in a modular curve indicates that the orbit will be smaller. Hence we can consider a function that is defined from

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

This function will take

$$\Gamma(N)\tau \mapsto \Gamma_1(N)\tau \mapsto \Gamma_0(N)\tau \mapsto SL_2(\mathbb{Z})\tau,$$

where each subsequent map is adding more curves to the orbit of τ because the amount of restrictions on the modular curve are decreasing.

Additionally, if we considered a specific modular curve, say $Y_1(N)$, we may define the function

$$Y_1(N) \to \mathrm{Y}_1(M),$$

where $M \mid N$, taking $\Gamma_1(N)\tau \mapsto \Gamma_1(M)\tau$. Here we see that $\tau \in Y_1(N)$ has the restriction of congruence mod N. Any matrix satisfying these congruences (mod N) will be guaranteed to satisfy them for (mod M) where M divides N. So the point τ in the modular curve $\mathrm{Y}_1(\mathrm{M})$ has more

elements in its orbit than its counterpart in $\Gamma_1(N)$. We will study this function in depth at a later point. Also note that this logic applies to $\Gamma(N)$ and $\Gamma_0(N)$ as well, not just $\Gamma_1(N)$.

Let's now describe what the orbits of τ look like for a τ in each of our defined modular curves. We already know that the action of all of $SL_2(\mathbb{Z})$ on τ reduces the space to $\mathcal{M}_{1,1}$ as described before.

Lemma 4.5. The reductive map from $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/N\mathbb{Z})$ is surjective.

Proof. See [3, Lemma 5.2.6]. □

Theorem 4.6.

- $Y(N)$ is the moduli space of elliptic curves together with a specified ordered basis of the N-torsion points.
- $Y_1(N)$ is the moduli space of elliptic curves together with a specified torsion point of order N.
- $Y_0(N)$ is the moduli space of elliptic curves together with a choice of an order N cyclic subgroup of N-torsion points.

A matrix $\gamma \in \Gamma(N)$ will take a curve τ to $\gamma\tau$ where these two curves will of course be isomorphic in the moduli space of elliptic curves because $\gamma \in SL_2(\mathbb{Z})$. However, the additional congruence restrictions will force the action of γ to preserve an ordered basis of N-torsion points.

For a $\gamma \in \Gamma_1(N)$ we have that τ and $\gamma\tau$ are isomorphic, but now add the condition that the specified N-torsion point in E_τ is preserved by the action of γ.

If we consider $\gamma \in \Gamma_0(N)$, we again have that τ and $\gamma\tau$ are in the same orbit, but with the least amount of restrictions from $\Gamma_0(N)$. We see that τ and $\gamma\tau$ may map a single N-torsion point from τ to a different N-torsion point in $\gamma\tau$. These N-torsion points are part of a cyclic subgroup of order N, the group $S = \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$ is an example.

Proof. As we will focus more intently on $\Gamma_1(N)$ later, we will prove this claim for only $\Gamma_1(N)$. We will start with the descriptive definition: the moduli space of elliptic curves with a specified N-torsion point of order N. This is equivalent to the set of isomorphism classes of curves with this N-torsion point of order N. We will now show that each of these isomorphism classes contains an elliptic curve with the specified N-torsion point at $\frac{1}{N}$. Let the specified N-torsion point be $P = \frac{k\tau+l}{N}$. Since the point has order N, we know that $\gcd(k, l, N) = 1$. Now we construct a matrix $\gamma \in SL_2(\mathbb{Z})$ that takes the point P on the curve E_τ to the point $\frac{1}{N}$ on the curve $E_{\gamma\tau}$. Note that $\gamma = \left(\begin{smallmatrix} * & * \\ k & l \end{smallmatrix}\right)$ will take P to $\frac{1}{N}$ because

$$\gamma\left(\frac{k\tau + l}{N}\right) = (k\tau + l)^{-1}\left(\frac{k\tau + l}{N}\right) = \frac{1}{N}.$$

We need to fill the $*$ positions with terms that force γ to have determinant 1. From the fact that $\gcd(k, l, N) = 1$, Bezout's lemma guarantees there exist $r, s, t \in \mathbb{Z}$ such that $sl - rk - tN = 1$. Then

$$\det\left(\begin{smallmatrix} s & r \\ k & l \end{smallmatrix}\right) = 1 + tN \equiv 1 \pmod{N}.$$

Now $\gamma = \left(\begin{smallmatrix} s & r \\ k & l \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}/N)$. By lemma 4.5, we know that there is a matrix in $SL_2(\mathbb{Z})$ that maps to γ when reducing the entries (mod N). We can take any of the matrices in $SL_2(\mathbb{Z})$ that satisfies this condition, and that becomes the matrix that shows E_τ and $E_{\gamma\tau}$ are isomorphic.

Knowing that there is a curve with specified torsion point $\frac{1}{N}$ in each isomorphism class, we can consider the isomorphism classes to just be of elliptic curves with the point $\frac{1}{N}$. The next question to answer is what needs to be true of a new matrix γ so that a curve E_τ with point $\frac{1}{N}$ maps to the curve $E_{\gamma\tau}$ with $\frac{1}{N}$ still specified. We start we the knowledge that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$,

$$\frac{c\tau + d}{N} \mapsto \frac{1}{N}$$

and we want

$$\frac{1}{N} \mapsto \frac{1}{N} \pmod{\Lambda_{\gamma\tau}}.$$

This means that

$$\frac{c\tau + d}{N} \equiv \frac{1}{N} \pmod{\Lambda_\tau} \implies \frac{c\tau + d}{N} - \frac{1}{N} \in \Lambda_\tau.$$

Then

$$\frac{c}{N}\tau + \frac{d-1}{N} \in \Lambda_\tau \implies \frac{c}{N}, \frac{d-1}{N} \in \mathbb{Z}.$$

Thus, we reach the conditions that

$$c \equiv 0 \pmod{N} \qquad d \equiv 1 \pmod{N}.$$

Since γ must have determinant 1, we realize $a \equiv 1 \pmod{N}$ as well. We illustrated exactly the conditions that places $\gamma \in \Gamma_1(N)$. Thus the isomorphism classes of curves with specified point $\frac{1}{N}$ corresponds to the orbits of $\Gamma_1(N)\backslash\mathcal{H} = Y_1(N)$.

□

5. Cusps

In this section we will introduce the concept of a cusp of a modular curve. Let's return to our depiction of $\mathcal{M}_{1,1}$ from Figure 4. If we consider the image topologically, we can glue together the two halves of the base of the ribbon, as well as the two sides. This process will give us a sleeve structure that is completely closed outside of the opening at the "top". Here we can consider the point at infinity, denoted $i\infty$. If we add $i\infty$ to the moduli space of elliptic curves, we now have compactified the space into a sphere. The point $i\infty$ is an example of a cusp and this topological process is the motivation behind studying these points.

Definition 5.1. The *extended upper half* plane is $\mathcal{H}^* := \mathcal{H} \cup \mathbb{Q} \cup \{i\infty\}$.

Now we will look to extend the action of $SL_2(\mathbb{Z})$ to $\mathcal{H}^*$. If we take a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ then we have our standard map of $\gamma\tau = \frac{a\tau+b}{c\tau+d}$. However we now have to define how γ acts on points in the extended upper half plane that are not in the standard upper half plane. We define $\gamma(i\infty) = \frac{a}{c}$ and $\gamma(-\frac{d}{c}) = i\infty$.

As we defined before, $Y(1) = SL_2(\mathbb{Z})\backslash\mathcal{H}$. Now we will define $X(1) = SL_2(\mathbb{Z})\backslash\mathcal{H}^*$. The X notation in replacement of the Y will be common across all the modular curves we described when considering an action on $\mathcal{H}^*$ as opposed to $\mathcal{H}$. We will formalize this shortly after defining the term cusp.

Definition 5.2. [1, p. 58] Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Note that

$$X_\Gamma = \Gamma\backslash\mathcal{H}^* = Y_\Gamma \cup \Gamma\backslash(\mathbb{Q} \cup \{i\infty\}).$$

The *cusps* of X_Γ are the orbits of $\Gamma\backslash(\mathbb{Q}\cup\{i\infty\})$.

We now revisit the space $X(1)$ and ask the question, how many cusps are in $X(1)$, and what are they? We are looking for the number of orbits that come from points in $\mathbb{Q}\cup\{i\infty\}$ from action under matrices in $SL_2(\mathbb{Z})$. Take $s=-\frac{d}{c}\in\mathbb{Q}$ and then create $\gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in SL_2(\mathbb{Z})$, then $\gamma s=i\infty$. We see that every rational number, as s is arbitrary, is in the same orbit as the point at infinity. Thus, there is only one cusp in $X(1)$ which is the orbit of the point at infinity $[i\infty]$.

Now we will define compactified modular curves for the subgroups $\Gamma(N)$, $\Gamma_1(N)$, and $\Gamma_0(N)$ of $SL_2(\mathbb{Z})$.

$$X(N)\coloneqq\Gamma(N)\backslash\mathcal{H}^*=Y(N)\cup\Gamma(N)\backslash(\mathbb{Q}\cup\{i\infty\}) \tag{5.1}$$

$$X_1(N)\coloneqq\Gamma_1(N)\backslash\mathcal{H}^*=Y_1(N)\cup\Gamma_1(N)\backslash(\mathbb{Q}\cup\{i\infty\}) \tag{5.2}$$

$$X_0(N)\coloneqq\Gamma_0(N)\backslash\mathcal{H}^*=Y_0(N)\cup\Gamma_0(N)\backslash(\mathbb{Q}\cup\{i\infty\}) \tag{5.3}$$

The next step in this discussion of cusps is determining the number of cusps of an modular curve X_Γ. We will restrict our cusp counting efforts to $X_1(N)$ before providing a brief commentary on the other two modular curves.

Theorem 5.3 ([1, Prop 3.8.3])**.** Let $s=\frac{a}{c}$ and $s'=\frac{a'}{c'}$ be elements of $\mathbb{Q}\cup\{i\infty\}$ with $\gcd(a,c)=\gcd(a',c')=1$. Then

$$\Gamma_1(N)s=\Gamma_1(N)s' \iff \begin{bmatrix}a'\\ c'\end{bmatrix}\equiv\pm\begin{bmatrix}a+cj\\ c\end{bmatrix}\pmod N$$

for some $j\in\mathbb{Z}$.

This theorem is stating that two elements of $\mathbb{Q}\cup\{i\infty\}$ are representatives of the same cusp of $\Gamma_1(N)$ exactly when there is some $j\in\mathbb{Z}$ that the congruence condition defined in the theorem is satisfied. From this equation we are able to systematically determine if two rational points are in the same cusp, whether by hand or program. We also are able to find a rational representative for each cusp, where a representative is simply one element of the orbit. Let's now look at an example of finding cusps.

Example 5.4. Find the number of cusps and representatives of each cusp in $\mathrm{X}_1(9)$.

Solution. We will start with one cusp representative being the point at infinity. A matrix $\gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in\Gamma_1(9)$ acts on $i\infty$ by sending it to $\frac{a}{c}$. Then we know since $\gamma\in\Gamma_1(9)$ that $a\equiv 1\pmod 9$ and $c\equiv 0\pmod 9$. So the cusp $\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]$ is in the same orbit as the point at infinity. From Theorem 5.3, we add that the cusp $\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]$ can be negated modulo 9 to the cusp $\left[\begin{smallmatrix}8\\ 0\end{smallmatrix}\right]$. Since $c\equiv 0\pmod 9$, there is no integer j that can be multiplied with c to change the numerator of the cusp modulo 9.

Now we look at the cusp $\left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$ and can negate it to get the cusp $\left[\begin{smallmatrix}7\\0\end{smallmatrix}\right]$. Again, we will be unable to change the numerator and then we can choose a representative, say $\frac{2}{9}$.

Going in order, the next cusp would be $\left[\begin{smallmatrix}3\\0\end{smallmatrix}\right]$, but we can note that since we are working modulo 9, this vector is not fully reduced modulo 9 and thus we do not consider it.

So our third cusp will be $\left[\begin{smallmatrix}4\\0\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}5\\0\end{smallmatrix}\right]$ and we choose the representative $\frac{4}{9}$. We have now exhausted all possibilities of $c \equiv 0 \pmod 9$.

Let's now consider the cusps where $a \equiv 0 \pmod 9$. First we have $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ where we can add any integer j multiplied with $c \equiv 1 \pmod 9$ and end up with any integer in the numerator. Thus $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\1\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\8\end{smallmatrix}\right]$ and we choose representative 1. We can note that all the integers will belong to this cusp.

Similarly we have the cusp $\left[\begin{smallmatrix}0\\2\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\2\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\7\end{smallmatrix}\right]$ with representative $\frac{1}{2}$.

The last cusp with $a \equiv 0 \pmod 9$ will be the cusp $\left[\begin{smallmatrix}0\\4\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\4\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}*\\5\end{smallmatrix}\right]$, again skipping 3 because it would not be fully reduced. We choose representative $\frac{1}{4}$.

Up to this point, we have found representatives for 6 distinct cusps. There were 3 cusps with $c \equiv 0 \pmod 9$ as well as 3 cusps with $a \equiv 0 \pmod 9$. We do not have enough evidence to believe this pattern yet, but we will note that $3 = \frac{\phi(9)}{2}$, where ϕ is Euler's totient function.

Returning to the cusp counting, we have covered all of our bases for cusps with $c \equiv 1$, 2, 4, and their negated counterparts modulo 9. We also already established that for $c \equiv 3 \pmod 9$, $a \equiv 0, 3, 6 \pmod 9$ is not fully reduced. So let's consider the cusp $\left[\begin{smallmatrix}1\\3\end{smallmatrix}\right]$. By adding integer multiples of 3 to the numerator we see that $\left[\begin{smallmatrix}1\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}4\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}7\\3\end{smallmatrix}\right]$ and that by negating $\left[\begin{smallmatrix}1\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}8\\6\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}5\\6\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}2\\6\end{smallmatrix}\right]$. We choose a representative to be $\frac{1}{3}$.

Lastly, notice the only remaining option for a cusp is $\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}5\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}8\\3\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}1\\6\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}7\\6\end{smallmatrix}\right] \sim \left[\begin{smallmatrix}4\\6\end{smallmatrix}\right]$. We can take the representative $\frac{2}{3}$ for this cusp.

Thus we have found the 8 cusps in $\mathrm{X}_1(9)$ with representatives $1, \frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{1}{2}, \frac{1}{4}, \frac{1}{3},$ and $\frac{2}{3}$.

□

To illustrate a pattern in the counting of cusps for $X_1(N)$, consider $N = 20$. We do not have the space to discuss finding all the cusps for $\mathrm{X}_1(20)$, so we will instead use code to evaluate all cusp possibilities (all fully reduced rationals) and then reduce them into orbits based on the condition in Theorem 5.3. Representatives of the cusps of $\mathrm{X}_1(20)$ are as follows:

• $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\8\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\16\end{smallmatrix}\right]$
• $\left[\begin{smallmatrix}0\\3\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}3\\0\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\4\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\10\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}2\\5\end{smallmatrix}\right]$
• $\left[\begin{smallmatrix}0\\7\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}7\\0\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\5\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\12\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}2\\15\end{smallmatrix}\right]$
• $\left[\begin{smallmatrix}0\\9\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}9\\0\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\6\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}1\\15\end{smallmatrix}\right]$	• $\left[\begin{smallmatrix}3\\10\end{smallmatrix}\right]$

We notice that the first column has all 4 cusps where $a \equiv 0 \pmod{20}$ and the second column has 4 cusps with $c \equiv 0 \pmod{20}$. Again the pattern mentioned before stands where there are 4 cusps of each of these types and $4 = \frac{\phi(20)}{2}$.

In the following theorem that provides a cusp counting formula for $X_1(N)$, we will exclude $\mathrm{X}_1(2)$ and $\mathrm{X}_1(4)$. The cases where $N = 2$ or 4 behave strangely, so going forward we will not be considering these cases in our results.

Theorem 5.5. For a positive integer $N \neq 1, 2, 4$,

$$\#\text{ of cusps in } X_1(N) = \tfrac{1}{2} \sum_{d|N} \phi(d)\phi(\tfrac{N}{d}).$$

In addition to this simple formula for counting cusps, we remark that the divisor d corresponds to the cusps with a particular denominator. In particular, $d = \gcd(c, N)$. This formula matches up with the pattern we saw before. When $d = 1$, we have c as a positive integer that is coprime with N. In this case we know that the numerator can be any value leading to the cusp $\left[\begin{smallmatrix} * \\ c \end{smallmatrix}\right] \sim \left[\begin{smallmatrix} 0 \\ c \end{smallmatrix}\right]$. So the formula for cusps with $a \equiv 0 \pmod{N}$ is $\frac{1}{2}\phi(1)\phi(N) = \frac{\phi(N)}{2}$. Likewise when $d = N$, we have $c = N$ as well. These are the cusps where $c \equiv 0 \pmod{N}$ and will get counted as $\frac{1}{2}\phi(N)\phi(1) = \frac{\phi(N)}{2}$.

Proof. Fix a divisor d of N. Let us count the cusps of $\Gamma_1(N)$ with denominator c such that $\gcd(c, N) = d$. The number of such denominators is $\phi(N/d)$. If a/c is such a cusp with $1 \leqslant c \leqslant N$, it follows from Theorem 5.3 that we may assume $0 \leqslant a < \gcd(c, N) = d$. Thus, the number of possible a such that $\gcd(a, c) = 1$ is exactly $\phi(d)$. Taking into account negation, we conclude the number of such cusps is $\frac{1}{2}\phi(d)\phi(N/d)$. Summing over all divisors d gives the result. □

Remark 5.6. There are similar formulas for counting the cusps of $X_0(N)$ and $X(N)$ [1, §3.8].

$$\#\text{ of cusps in } X_0(N) = \sum_{d|N} \phi(\gcd(d, \tfrac{N}{d})).$$
$$\#\text{ of cusps in } X(N) = \tfrac{1}{2}N^2 \prod_{p|N} (1 - \tfrac{1}{p^2}).$$

We can now easily compute, for example, that $X_1(52)$ has 60 cusps or that $X_1(75)$ has 112 cusps. Next, we seek to understand the behavior of each of these cusps. In the counting process, we sectioned our counting into each $d|N$. It seems natural that there would be some relationship between the method of counting the cusps and the behavior of the cusps when glued into their modular curve. As we alluded to in the beginning of this section, the cusp of a modular curve is essentially a point that is filling what was previously a hole or gap in the structure of the modular curve. It becomes interesting to see how the type of cusp we are working with behaves when pasted in that hole, or more specifically how a neighborhood of the cusp behaves in the pasting process.

6. Cusp Width

To create X_Γ from Y_Γ, one glues in small disks in the neighborhood of each cusp. The exact formula of this gluing is determined by the *width* of the cusp. For example the cusp of $Y(1)$, $[i\infty]$, will have width 1 because the map of the cusp neighborhood will be 1:1. We will describe this gluing in detail later in the section.

First, we will define width by considering which matrices in a congruence subgroup Γ fix the cusp, or sends the cusp to itself. In defining width, we will first look at specifically the cusp at infinity.

Definition 6.1. The *width* of the cusp $i\infty$ with respect to a congruence subgroup Γ is the smallest positive integer w such that $\left(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$, thus fixing the cusp.

Example 6.2. Observe that $[i\infty]$ has width N in $\Gamma(N)$ because $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma(N)$. Also noting that the cusp $[i\infty]$ has width 1 in $\Gamma_1(N)$, it is apparent that the width of a cusp is dependent on the choice of congruence subgroup.

To expand the width definition to a general cusp, consider a point $\frac{a}{c} \in \mathbb{Q}$ with $\gcd(a,c) = 1$. Then we can define a matrix $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ so that $\alpha(i\infty) = \frac{a}{c}$ and $\alpha^{-1}(\frac{a}{c}) = i\infty$. Thus to fix the cusp $[\frac{a}{c}]$ we need to use α to send $\frac{a}{c}$ to $i\infty$, the matrix $\left(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix}\right)$ to fix the cusp at infinity, then α^{-1} to bring the cusp $i\infty$ back to $\frac{a}{c}$.

Definition 6.3. Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $a, c \in \mathbb{Z}$ with $\gcd(a,c) = 1$. Then the *width* of a cusp $[\frac{a}{c}]$ of Y_Γ is the smallest positive integer w such that $\alpha \left(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix}\right) \alpha^{-1} \in \Gamma$.

We will now dive into an example on computing the width of cusps.

Example 6.4. Find the width of all cusps in $X_1(9)$.

Solution. Recall the cusps in $X_1(9)$ have representatives

$$\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 4 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right].$$

When $d = 9$ in the counting formula, we have the first three listed cusps that have representatives of the form $\frac{k}{9}$ where $\gcd(k,9) = 1$. Then we create $\alpha = \left(\begin{smallmatrix} k & b \\ 9 & d \end{smallmatrix}\right)$ where $b, d \in \mathbb{Z}$. Consequently, $\alpha^{-1} = \left(\begin{smallmatrix} d & -b \\ -9 & k \end{smallmatrix}\right)$. So we need to find w such that the product

$$\begin{pmatrix} k & b \\ 9 & d \end{pmatrix} \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d & -b \\ -9 & k \end{pmatrix}$$

is in $\Gamma_1(9)$. This matrix product evaluates to

$$\begin{pmatrix} kd - 9kw - 9b & -bk + k^2w + bk \\ -81w & -9b + 9wk + dk \end{pmatrix}.$$

In order to be in $\Gamma_1(9)$, the top right can be anything and the bottom right already is congruent to 0 (mod 9). So the top left and bottom right need to both be congruent to 1 (mod 9). Note that since this matrix product must be in $SL_2(\mathbb{Z})$, the determinant $kd - 9b = 1$ so the top right is the same as $1 - 9kw$ and bottom left is $1 + 9wk$. We have

$$1 - 9kw \equiv 1 + 9kw \equiv 1 \pmod 9 \implies 9kw \equiv 0 \pmod 9 \implies w = 1.$$

Next, the three cusps with 0 in the numerator correspond to $d = 1$ in the counting formula. These three cusps require that α has a number in the bottom left that is coprime with 9. When evaluating $\alpha \left(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix}\right) \alpha^{-1}$ the bottom left will be a constant that is coprime with 9 multiplied by w. Thus, $w = 9$ so that the entry is congruent to 0 modulo 9. Note that $w = 9$ will also satisfy the congruences in top right and bottom left, so the width is 9.

The last two cusps that correspond to $d = 3$ in the counting formula will lead to an $\alpha = \left(\begin{smallmatrix} k & b \\ 3 & d \end{smallmatrix}\right)$ where $k = 1, 2$. Then

$$\alpha \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} \alpha^{-1} = \begin{pmatrix} 1 - 3kw & k^2w \\ -9w & 1 + 3kw \end{pmatrix}$$

where the bottom left already satisfies its congruence condition. Now we need to find the smallest w such that

$$1 - 3kw \equiv 1 + 3kw \equiv 1 \pmod 9.$$

We conclude $w = 3$. □

Theorem 6.5. If $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ is a cusp of $X_1(N)$ with $d = \gcd(N, c)$, then the width of $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ is $\frac{N}{d}$.

Proof. We will follow a similar process as we did in the above example to prove this width formula for a general cusp in a general $X_1(N)$. Set $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and we then can reduce to the two congruence conditions

$$awc \equiv 0 \pmod{N} \qquad \text{and} \qquad c^2 w \equiv 0 \pmod{N}.$$

From the first of these equations we can say that the smallest w is

$$w = \frac{N}{d \cdot \gcd(a, \frac{N}{d})}.$$

And from the second we find

$$w = \frac{N}{d \cdot \gcd(c, \frac{N}{d})}.$$

Then we know that

$$w = \operatorname{lcm}\left(\frac{N}{d \cdot \gcd(a, \frac{N}{d})}, \frac{N}{d \cdot \gcd(c, \frac{N}{d})}\right).$$

Set $\ell = \frac{N}{d}$. Then we can simplify to

$$w = \operatorname{lcm}\left(\frac{\ell}{\gcd(a, \ell)}, \frac{\ell}{\gcd(c, \ell)}\right).$$

By using the relationship between lcm and gcd and simplifying further, the next step is

$$w = \frac{\ell}{\gcd(\gcd(a, \ell), \gcd(c, \ell))}.$$

Since we know that the $\gcd(a, c) = 1$,

$$w = \ell = \frac{N}{d}.$$

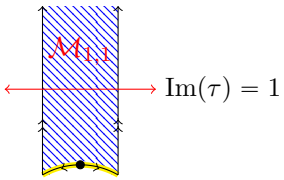
□

Following this concise formula to compute the width of cusps, we return to the cusp counting formula from Theorem 5.5,

$$\#\text{ of cusps in } X_1(N) = \tfrac{1}{2} \sum_{d|N} \phi(d)\phi(\tfrac{N}{d}).$$

Now we can notice that by dividing the summation into $d|N$, we are also effectively computing the number of cusps of width $\frac{N}{d}$ for each divisor d in the summation.

Following our development of the width of cusps, we will revisit the topological cusp pasting process. Recall the fundamental domain of the moduli space $\mathcal{M}_{1,1}$. For reference, we also include the horizontal line $\operatorname{Im}(\tau) = 1$.



Let $\mathcal{D}$ denote the open disk of radius $e^{2\pi}$. Then define a map

$$\{\tau \in \mathcal{H} \mid \operatorname{Im}(\tau) > 1\} \longrightarrow \mathcal{D}$$

given by $\tau \mapsto e^{2\pi i\tau}$. This transforms points above the line $\mathrm{Im}(\tau) = 1$ to an open disk. The line $\mathrm{Im}(\tau) = 1$ will map to the boundary of the disk and each horizontal line at an increasing $\mathrm{Im}(\tau)$ will map to a circle of smaller radius. Note that image of this map does not include the center of $\mathcal{D}$. This missing point corresponds to the cusp at $i\infty$.

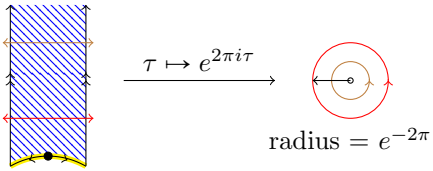


FIGURE 5. The moduli space of elliptic curves $Y(1)$ mapped to a disk $\mathbb{D}$ with a point missing in the center.

Consider the region $\mathcal{R} = \{\tau \in \mathbb{H} \mid \mathrm{Im}(\tau) > 1\}$. If we define

$$\Gamma_\infty = \{S^n = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\},$$

then $\Gamma_\infty \backslash \mathcal{R} \subset Y(1)$ where $\Gamma_\infty \backslash \mathcal{R} \cong \mathbb{D}^\star$. Taking $Y(1) \cup \mathbb{D}^\star = \mathbb{D}$ with the cusp pasted into the disk.

Working with a cusp of width w at infinity, we instead consider

$$\Gamma_\infty = \{S^{nw} = \left(\begin{smallmatrix} 1 & nw \\ 0 & 1 \end{smallmatrix}\right)\}.$$

Cusps other than the one at infinity will be pasted in a similar way to the computation of width. We apply a matrix α^{-1} to take the cusp $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ to $i\infty$. Then we may glue in a disk using the process we just described. One can then transform α will take $i\infty \mapsto \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$.

7. Maps of Modular Curves

After developing our understanding of cusps and their widths, we will now revisit the quotient maps between modular curves. Specifically, we will focus on modular curves of $\Gamma_1(N)$ and study how cusps are mapped between modular curves of level. In particular, we consider the map

$$X_1(N) \to X_1(M),$$

where $M|N$.

Example 7.1. Describe the map from $X_1(9) \to X_1(3)$.

Solution. The eight cusps in $X_1(9)$ have representatives

$$\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 4 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right].$$

Meanwhile, the two cusps in $X_1(3)$ have representatives

$$\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right].$$

If we consider the class of cusps with width 1 in $X_1(9)$ (that is, those of the form $\left[\begin{smallmatrix} * \\ 0 \end{smallmatrix}\right]$) and reduce them modulo 3 by the map, their image is the cusp $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ in $X_1(3)$ (also of width 1). Now the class of cusps of width 9 in $X_1(9)$ (that is, those of the form $\left[\begin{smallmatrix} 0 \\ * \end{smallmatrix}\right]$) reduces (mod 3) to $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ in $X_1(3)$ (width 3). The final two cusps of width 3 in $X_1(9)$ both reduce to $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ in $X_1(3)$, having width 1. □

These maps will generally be inconsistent for any arbitrary M, N. Using the `gamma_map` function described in the appendix, we can see the cusps in $X_1(M)$, their width in $X_1(M)$, and the number of and representative of the cusps that map to it from $X_1(N)$. This function is useful for testing the maps between varying M, N and discovering patterns based on information about M and N. We are specifically searching for patterns that describe the number of cusps in $X_1(N)$ above a cusp of width w in $X_1(M)$.

The first pattern we will discuss is illustrated in the following theorem.

Theorem 7.2. Fix a prime p and integers $k > \ell$. Consider the map $X_1(p^k) \to X_1(p^\ell)$. There are $p^{k-\ell}$ cusps in $X_1(p^k)$ above cusps of maximal width in $X_1(p^\ell)$.

Proof. Consider a cusp $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ in $X_1(p^k)$ of non-maximal width, i.e. the width is less than p^k. Since width $w < p^k$ and $w = p^k/\gcd(c, p^k)$ by Theorem 6.4, we know that $\gcd(c, p^k) > 1$. This implies that the denominator of the cusp c contains a factor of p. Thus, the non-maximal cusp $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ cannot map to a maximal width cusp in the curve $X_1(p^\ell)$ because $\gcd(c, p^\ell) > 1$ as well. Therefore, only the maximal width cusps in $X_1(p^k)$, of which there are $\frac{\phi(p^k)}{2}$, can map to maximal width cusps in $X_1(p^\ell)$, of which there are $\frac{\phi(p^\ell)}{2}$. Finally, we divide these results to find the number of cusps in $X_1(p^k)$ that map to a maximal width cusp in $X_1(p^\ell)$:

$$\frac{\phi(p^k)}{2} \Big/ \frac{\phi(p^\ell)}{2} = \frac{\phi(p^k)}{\phi(p^\ell)} = p^{k-\ell}.$$

□

Before establishing the next pattern, let's examine an example of the map $X_1(p^k) \to X_1(p^\ell)$ with $p = 3, k = 3, \ell = 2$. In other terms, we will look at the map $X_1(27) \to X_1(9)$.

cusp	cusp width	cusps above	num cusps above
[0, 1]	9	[[0, 1], [0, 8], [0, 10]]	3
[0, 2]	9	[[0, 2], [0, 7], [0, 11]]	3
[0, 4]	9	[[0, 4], [0, 5], [0, 13]]	3
[1, 0]	1	[[1, 0], [1, 9], [1, 18], [8, 0], [10, 0]]	5
[1, 3]	3	[[1, 3], [1, 12], [1, 21]]	3
[1, 6]	3	[[1, 6], [1, 15], [1, 24]]	3
[2, 0]	1	[[2, 0], [2, 9], [7, 0], [7, 9], [11, 0]]	5
[4, 0]	1	[[4, 0], [4, 9], [5, 0], [5, 9], [13, 0]]	5

FIGURE 6. Python code output for map $X_1(27) \to X_1(9)$.

Theorem 7.2 holds true in this instance as the maximal width cusps $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right]$ with width 9 each have $p^{k-1} = 3^{3-2} = 3$ cusps above.

The next pattern we show is more general and applies to cusps of any width in $X_1(M)$, as opposed to just those with maximal width.

Theorem 7.3. Let $K, L \neq 2, 4$ be positive integers with $\gcd(K, L) = 1$. Then consider the modulo L map from $X_1(KL) \to X_1(L)$. All cusps of width w in $X_1(L)$ collectively have

$$2(\#\text{ of cusps of } X_1(L) \text{ of width } w)(\#\text{ cusps of } X_1(K))$$

cusps above.

Proof. We will start by looking at the formula for the number of cusps in $X_1(KL)$.

$$\begin{aligned}
\#\text{ of cusps in } X_1(KL) &= \frac{1}{2}\sum_{d|KL}\phi(d)\phi(KL/d) \\
&= \frac{1}{2}\sum_{\substack{d_1|K \\ d_2|L}}\phi(d_1d_2)\phi(KL/d_1d_2) \\
&= \frac{1}{2}\sum_{\substack{d_1|K \\ d_2|L}}\phi(d_1)\phi(d_2)\phi(K/d_1)\phi(L/d_2) \quad \text{by } \phi \text{ being multiplicative} \\
&= 2\left(\frac{1}{2}\sum_{d_1|K}\phi(d_1)\phi(K/d_1)\right)\left(\frac{1}{2}\sum_{d_2|L}\phi(d_2)\phi(L/d_2)\right) \\
&= 2(\#\text{ of cusps in } X_1(K))(\#\text{ of cusps in } X_1(L)).
\end{aligned}$$

Then for a single cusp of width w in $X_1(L)$, there are

$$2(\#\text{ of cusps in } X_1(K))$$

above it in $X_1(KL)$. Hence our formula counts the number of cusps of width w in $X_1(L)$ multiplied by the number above each. □

We again will look at example, namely when $K = 3, L = 8$. This will be the map from $X_1(24) \to X_1(8)$.

cusp	cusp width	cusps above	num cusps above
[0, 1]	8	[[0, 1], [0, 7], [1, 9], [1, 15]]	4
[0, 3]	8	[[0, 5], [0, 11], [1, 3], [1, 21]]	4
[1, 0]	1	[[1, 0], [1, 8], [1, 16], [7, 0]]	4
[1, 2]	4	[[1, 2], [1, 6], [1, 10], [1, 18]]	4
[1, 4]	2	[[1, 4], [1, 12], [1, 20], [5, 12]]	4
[3, 0]	1	[[3, 8], [3, 16], [5, 0], [11, 0]]	4

FIGURE 7. Python code output for map $X_1(24) \to X_1(8)$.

We see that the collective of cusps of width 8 in $X_1(8)$ have 8 cusps above which is in line with the formula that says 2 multiplied by 2 cusps in $X_1(8)$ of width 8 multiplied by 2 cusps in $X_1(3)$. The same holds true for width 1 cusps. Then each of the width 2 and 4 cusps in $X_1(8)$ have 2

multiplied by 1 cusp of width 2, 4 respectively in $X_1(8)$ multiplied by 2 cusps in $X_1(3)$ for 4 cusps above. We also note that we already have the process to count the number of cusps of a specific width in $X_1(KL)$ so it is easy to modify this formula for cusps above only a single cusp of width w in $X_1(L)$.

We make the concluding remark that the previous two theorems are sufficient for computing the number of cusps above any maximal width cusp for any map $X_1(N) \to X_1(M)$ with $M|N$ by considering the prime factorization of M and N. One issue to be careful of when computing *all* maximal width cusps is the case when M, N, or N/M is 2 or 4 (due to the limitations of Theorem 7.3). However, applying these theorems repeatedly in the general case yields a powerful conclusion.

8. Appendix

The following code block contains the python functions used to compute the number of cusps in $\Gamma_1(N)$, given an argument N.

```
from sympy.ntheory.factor_ import totient

def factors(x):
    factors = []
    for i in range(1, x + 1):
        if x % i == 0:
            factors.append(i)
    return factors

def num_cusps(N):
    num_cusps = 0
    for d in factors(N):
        num_cusps += 1/2 * totient(d) * totient(int(N / d))
    return int(num_cusps)
```

This next block contains the function used to return a list of cusp representatives for $\Gamma_1(N)$, given an argument N.

```
import pandas as pd
import math

def cusp_generator(N):
    q_N = [[k % N, l % N] for k in range(1, N+1) for l in
        range(1, N+1) if math.gcd(k, l) == 1]
    q_N = pd.DataFrame(q_N).sort_values([0,1]).values.tolist()
    for [a, c] in q_N:
        for j in range(N):
```

```
            A = a + c * j
            if ([A % N, c % N] in q_N and a != A % N):
                q_N.remove([A % N, c % N])
            if ([-A % N, -c % N] in q_N):
                q_N.remove([-A % N, -c % N])
    return q_N
```

In the code of the `cusp_generator(N)` function we perform a for loop that checks to see if a cusp is equal to a cusp already in our list `q_N`. That is the inspiration for the following `equals` function which takes two cusps (as lists of length 2) and a positive integer N to check if those two cusps are in the same orbit in $X_1(N)$.

```
def equals(cusp1, cusp2, N):
    [a1, c1] = cusp1
    [a2, c2] = cusp2
    a1 = a1 % N
    a2 = a2 % N
    c1 = c1 % N
    c2 = c2 % N
    for j in range(N):
        A = a1 + c1 * j
        if ([A % N, c1 % N] == [a2, c2]):
            return True
        if ([-A % N, -c1 % N] == [a2, c2]):
            return True
    return False
```

We also include a function that computes the width of a cusp in $X_1(N)$.

```
def cusp_width(cusp, N):
    [a, c] = cusp
    d = math.gcd(N, c)
    return N / d
```

The last function we include is one that returns details regarding the map from $X_1(N) \to \mathrm{X}_1(\mathrm{M})$ where $M|N$. The function takes the values of N and M as arguments and returns a data set with each cusp in $\mathrm{X}_1(\mathrm{M})$, it's width, a list of the cusps above in $X_1(N)$ that are in the same orbit in $\mathrm{X}_1(\mathrm{M})$, and the number of cusps above (the length of the list).

```
def gamma_map(N, M):
    assert N >= M
```

```
N_cusps = cusp_generator(N)
M_cusps = cusp_generator(M)
df = pd.DataFrame(pd.Series(M_cusps), columns=["cusp"])
df["cusp width"] = [int(cusp_width(cusp, M)) for cusp in df["cusp"]]
all_cusps_above = []
for M_cusp in M_cusps:
    cusps_above = []
    for N_cusp in N_cusps:
        if equals(M_cusp, N_cusp, M):
            cusps_above.append(N_cusp)
    all_cusps_above.append(cusps_above)
df["cusps above"] = pd.Series(all_cusps_above)
df["num cusps above"] = df["cusps above"].map(len)
assert len(df) == num_cusps(M)
assert sum(df["num cusps above"]) == num_cusps(N)
return df
```

References

[1] F. Diamond & J. Shurman: *A First Course in Modular Forms*, Graduate Texts in Mathematics, Vol. 228, Springer, Berlin, 2005.

[2] N. Koblitz: *Introduction to elliptic curves and modular forms*, Springer-Verlag, Seattle, 1984.

[3] K. Ribet & W. Stein: Lectures on modular forms and Hecke operators (2017). [https://wstein.org/books/ribet-stein/main.pdf]