MODULAR CURVES

RITIK JAIN

1. INTRODUCTION

An *elliptic curve* over a field k is defined as the set of solutions in k^2 to an equation of the form[1]

$$Y^2 = X^3 + AX + B, \qquad A, B \in k, \tag{1}$$

with nonzero discriminant $\Delta = 4A^3 + 27B^2 \neq 0$, together with a point $\mathcal{O}$ at infinity. Over the past century, these curves have become some of the most intensely studied objects in mathematics, playing a central role in the study of Diophantine equations and, more recently, the security of modern cryptographic systems. Central to their usefulness in both fields is the fact that an elliptic curve carries a natural group structure, whose interaction with the geometry of the space gives rise to an intricate arithmetic structure.

While much can be gained from studying a fixed elliptic curve on its own, many deeper insights come from instead stepping back and considering the entire space of elliptic curves at once. In this note, we consider the *modular curves*, which are parameter spaces of (isomorphism classes) of elliptic curves. As their name suggests, modular curves carry a natural geometric structure: indeed, they are even sometimes elliptic curves themselves! By studying a modular curve as a whole, we are often able to extract subtle data about the underlying elliptic curves.

For the remainder of this note, we will restrict to the case $k = \mathbb{C}$, where tools of complex analysis allow for the fullest study of modular curves. In particular, it is over the complex numbers that one may define *modular forms*, which are holomorphic functions on modular curves satisfying certain symmetries. We will give a brief taste of the fascinating study of these functions in Section 5 via an introduction to the j-invariant, which is in some sense *the* modular form.

2. COMPLEX ELLIPTIC CURVES

As a first step towards the construction of modular curves, we will show that complex elliptic curves are precisely the complex tori, that is, quotients of $\mathbb{C}$ by a lattice. Throughout, we closely follow Diamond and Shurman's book [3].

[1]For simplicity, we assume that the characteristic of k is not 2 or 3.

2.1. **Complex Tori.** We begin with a discussion of *complex tori*, which are compact Riemann surfaces of genus one. More informally, they can be described as donuts.

Definition 2.1. Let $\omega_1, \omega_2 \in \mathbb{C}$. A *complex lattice* is the set of the form

$$\Lambda := \{a\omega_1 + b\omega_2 \mid a, b \in \mathbb{Z}\}.$$

Algebraically, a complex lattice is an abelian subgroup of $\mathbb{C}$, and geometrically, it is a discrete set of equally-spaced points on the complex plane.

Since a lattice Λ is an abelian subgroup of $\mathbb{C}$, we can naturally consider the quotient group $\mathbb{C}/\Lambda$. In particular, for a lattice Λ, we call $\mathbb{C}/\Lambda$ a *complex torus*.

Again, from the perspective of algebra, $\mathbb{C}/\Lambda$ is clearly just an abelian additive group. Moreover, since addition is carried out modulo Λ, it suffices to only consider elements inside the *fundamental parallelogram*

$$\mathcal{F} := \{\, t_1\omega_1 + t_2\omega_2 \mid 0 \leq t_1, t_2 < 1 \,\},$$

which can be illustrated as follows:

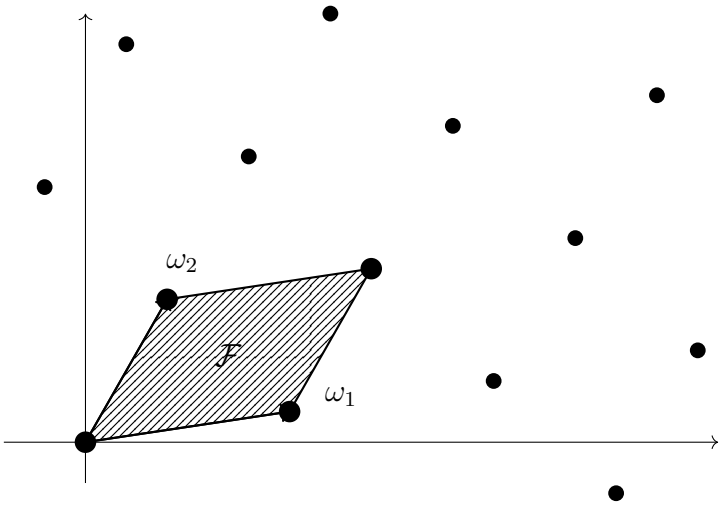

FIGURE 1. The fundamental domain $\mathcal{F}$ of a lattice with generators ω_1, ω_2.

The quotient structure also induces a geometry on $\mathbb{C}/\Lambda$. Indeed, under the induced addition on $\mathbb{C}/\Lambda$, we may identify (or glue) the opposite sides of the fundamental parallelogram to each other, first, gluing the two sides parallel to ω_1 to produce a cylinder, then, gluing the two circular ends of the cylinder together to produce a torus. In particular, this procedure makes $\mathbb{C}/\Lambda$ a *Lie group*, i.e. a manifold with a smooth group structure.

Given any mathematical structure, it is natural to ask for the correct notion of a *morphism*, i.e. a structure-preserving map. In the case of complex tori, these maps are called *isogenies*, and can be defined as follows:

Definition 2.2. A nonzero holomorphic homomorphism between complex tori is called an *isogeny*.

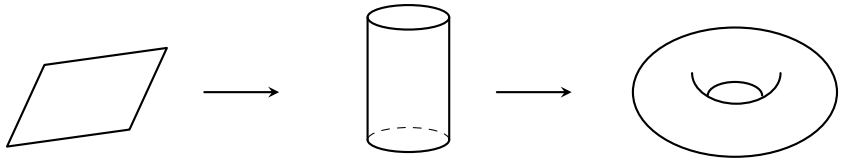

FIGURE 2. The torus as a quotient space.

Note that isogenies preserve both the geometric stucture and algebraic structure of a torus. Moreover, since tori are compact Riemann surfaces, every isogeny is automatically surjective with a finite kernel by complex analysis. If an isogeny is also injective, it is called an *isomorphism*.

As the following theorem illustrates, isogenies and isomorphisms are relatively rigidly determined.

Theorem 2.3. *Let $\Lambda, \Lambda' \subset \mathbb{C}$ be lattices. Then:*

(1) There exists an isogeny $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ if and only if there exists a nonzero $m \in \mathbb{C}$ such that $m\Lambda \subset \Lambda'$.
(2) There exists an isomorphism $\mathbb{C}/\Lambda \xrightarrow{\sim} \mathbb{C}/\Lambda'$ if and only if Λ and Λ' are homothetic*, i.e. there exists $m \in \mathbb{C}$ such that $m\Lambda = \Lambda'$.*

As (2) suggests, much of the behavior of a complex torus $\mathbb{C}/\Lambda$ is governed by its defining lattice Λ. As we will see in Section 3, this perspective in particular allows us to classify complex tori up to isomorphism simply using linear algebra.

2.2. **Complex Elliptic Curves as Tori.** In this section, we sketch the proof of the equivalence between complex tori and elliptic curves. For a more in-depth treatment of this correspondence, see [1, Ch. VI.3].

First, we show that every complex torus E/Λ can be realized as a complex elliptic curve E_Λ. We will need the following technical proposition:

Proposition 2.4. *For a lattice Λ, define the constants*

$$g_2(\Lambda) := 60 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^4}, \qquad g_3(\Lambda) := 140 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^6}.$$

Then, the cubic polynomial $4x^3 - g_2(\Lambda)x - g_3(\Lambda)$ has distinct roots, so the equation

$$y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$$

defines an elliptic curve E_Λ over $\mathbb{C}$.

This immediately allows us to associate an elliptic curve E_Λ to every complex torus $\mathbb{C}/\Lambda$. By analyzing the fields of functions on $\mathbb{C}/\Lambda$ and E_Λ respectively, we can more generally construct an isomorphism $\mathbb{C}/\Lambda \to E_\Lambda$.

The key tool relating the two objects is the *Weierstrass $\wp$-function*, defined for a lattice Λ by

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right).$$

Using a bit of complex analysis, one can show that $\wp_\Lambda(z)$ is meromorphic and Λ-periodic for any lattice Λ, as is its derivative

$$\wp'_\Lambda(z) = -2 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{(z-\omega)^3}.$$

Remarkably, $\wp_\Lambda(z)$ generates all of the meromorphic functions on $\mathbb{C}/\Lambda$, and satisfies the following differential equation:

$$(\wp'(z))^2 = 4(\wp(z))^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda). \tag{2}$$

Now, by Proposition 2.4, the equation $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$ defines an elliptic curve E_Λ over $\mathbb{C}$, thus by (2), the map

$$\phi : \mathbb{C}/\Lambda \setminus \{0\} \longrightarrow E_\Lambda, \qquad z + \Lambda \mapsto (\wp_\Lambda(z), \wp'_\Lambda(z)),$$

is well-defined. Moreover, by setting $\phi(0+\Lambda) = \mathcal{O}$, we can extend this map to the full torus. One can then show that ϕ is an isomorphism

$$\mathbb{C}/\Lambda \longrightarrow E_\Lambda$$

of Lie groups, so every complex torus can be realized as a complex elliptic curve. To see why every complex elliptic curve is a complex torus, we will need to consider the *j-invariant*, which assigns a unique number to each isomorphism class of elliptic curves.

Definition 2.5. Let $E : Y^2 = X^3 + AX + B$ be a complex elliptic curve. The *j-invariant of E* is given by the expression

$$j(E) = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2}.$$

Since $4A^3 + 27B^2 \neq 0$ by (1), j is a well-defined complex number. Moreover, one can show that two complex elliptic curves E, E' are isomorphic if and only if $j(E) = j(E')$, so the j-invariant is indeed an invariant.

Now, to prove that every complex elliptic curve E can be realized as a torus, it suffices to show that the function $j : \mathbb{H} \to \mathbb{C}$ defined as

$$j(\tau) := j(E_{\Lambda_\tau}), \qquad \Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$$

is surjective, which we assume without proof. Surjectivity guarantees the existence of a lattice Λ such that $j(E) = j(E_\Lambda)$, and since j is an isomorphism invariant, this in turn implies $E_\Lambda \cong E$. In this way, we have obtained the following fundamental result:

Theorem 2.6. *For every lattice $\Lambda \subset \mathbb{C}$, the Weierstrass $\wp$-function induces an isomorphism*

$$\mathbb{C}/\Lambda \xrightarrow{\sim} E_\Lambda.$$

On the other hand, for every complex elliptic curve E, there exists a lattice Λ such that $E \cong E_\Lambda$.

This equivalence is of vital importance in the theory of complex elliptic curves, as it allows us to freely pass between complex tori $\mathbb{C}/\Lambda$ and elliptic curves E_Λ, depending on whichever is more convenient. The j-invariant is a fundamental ingredient in the proof of this fact: in Section 5, we will return to it, where it will be realized as a holomorphic function on the moduli space $X(1)$ of elliptic curves.

3. The Moduli Space $X(1)$

In this section, we will construct the modular curve $X(1)$, which gives a geometric parametrization of the space of complex elliptic curves up to isomorphism.

The geometric interpretation of the set of isomorphism classes of complex elliptic curves comes down to a sequence of bijections. First, by Theorem 2.6, every complex elliptic curve can be realized as a complex torus and vice versa. In particular, this correspondence induces a bijection of isomorphism classes

$$\{\text{Complex elliptic curves } E\}/\cong \longleftrightarrow \{\text{Complex tori } \mathbb{C}/\Lambda\}/\cong .$$

On the other hand, by Theorem 2.3 (2), two complex tori $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic if and only if there exists $\alpha \in \mathbb{C}^\times$ such that $\Lambda' = \alpha\Lambda$. Thus, the set of isomorphism classes of complex tori is in bijection with the set of homothety classes of complex lattices:

$$\{\text{Complex tori } \mathbb{C}/\Lambda\}/\cong \longleftrightarrow \{\text{Lattices } \Lambda \subset \mathbb{C} \text{ up to rescaling}\}.$$

It remains to determine when two lattices Λ and Λ' are homothetic, i.e. when they are the same up to rescaling. First, note that for any lattice Λ, we can choose a basis $\{\omega_1, \omega_2\}$ of Λ such that $\tau := \omega_2/\omega_1 \in \mathbb{H}$.

It follows that $\Lambda = \omega_1(\mathbb{Z} + \tau\mathbb{Z})$, so every lattice is homothetic to one of the form $\Lambda_\tau := \mathbb{Z} + \tau\mathbb{Z}$ for $\tau \in \mathbb{H}$. Moreover, by elementary linear algebra[2], two lattices Λ_τ and $\Lambda_{\tau'}$ are homothetic if and only if there exists a matrix $A \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$A \cdot \tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d} = \tau',$$

[2]This is because $\mathrm{SL}_2(\mathbb{Z})$ acts on the space of lattices by change-of-basis.

where $\mathrm{SL}_2(\mathbb{Z}) = \{A \in M_2(\mathbb{Z}) \mid \det(A) = 1\}$.[3] Therefore, there is a third bijection between lattices up to rescaling and *orbits*

$$\mathrm{SL}_2(\mathbb{Z}) \cdot \tau := \{A \cdot \tau : A \in \mathrm{SL}_2(\mathbb{Z})\}.$$

of the following action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$:

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

One can verify that this is a well-defined group action, therefore the set of orbits gives a *partition* of the upper half plane. That is, every point in $\mathbb{H}$ lies in some orbit, and no two distinct orbits have any overlap.

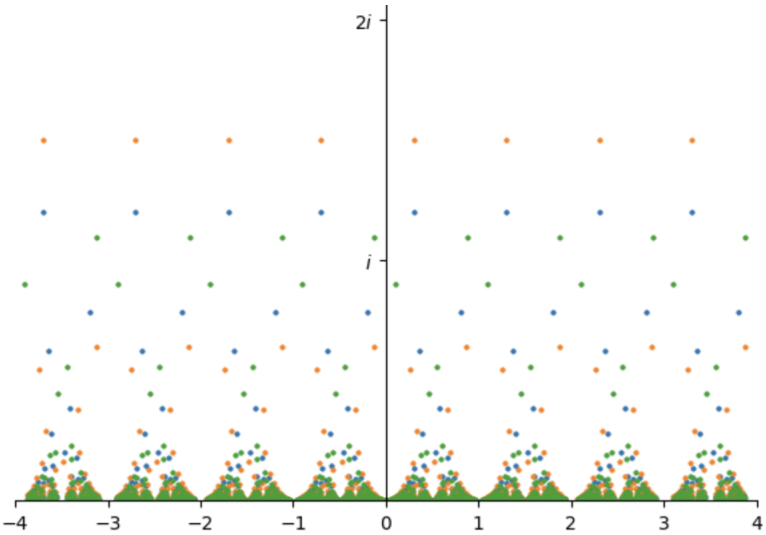

FIGURE 3. Graph of three orbits of the $\mathrm{SL}_2(\mathbb{Z})$-action on $\mathbb{H}$, colored orange, green, and blue respectively.

Thus, letting $Y(1) := \mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$ denote the set of orbits of the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$, we can sum all this up with the bijection

$$\{\text{Complex elliptic curves } E\}/\cong \longleftrightarrow\ Y(1),$$

which gives a parametrization of isomorphism classes of complex elliptic curves.

But what is the geometry of $Y(1)$? First, endowing $Y(1)$ with the quotient topology under the mapping

$$\pi : \mathbb{H} \to Y(1), \qquad \pi(\tau) = \mathrm{SL}_2(\mathbb{Z}) \cdot \tau,$$

we have that $Y(1)$ is the image of $\mathbb{H}$ under the continuous surjection π. Since $\mathbb{H}$ is connected, this in particular implies that $Y(1)$ is a connected topological space. Moreover,

[3]Note that for $A \in \mathrm{SL}_2(\mathbb{Z})$ and $\tau \in \mathbb{H}$, we have $A \cdot \tau = -A \cdot \tau$. Therefore, strictly speaking, the action is really given by $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm I_2\}$ rather than $\mathrm{SL}_2(\mathbb{Z})$.

since $\mathrm{SL}_2(\mathbb{Z})$ is a discrete group, it acts *properly discontinuously* on $\mathbb{H}$, implying that any two orbits can separated by sufficiently small balls. That is, $Y(1)$ is also Hausdorff.

In fact, with some additional work, it can be shown that $Y(1)$ is also second-countable, locally homeomorphic to $\mathbb{C}$, and smooth, so $Y(1)$ is naturally a Riemann surface. Its one-point compactification

$$X(1) := Y(1) \cup \{\infty\}$$

is called the *modular curve*, which, as we will see in Section 5, is biholomorphic to the Riemann sphere $\mathbb{P}^1(\mathbb{C})$. In particular, this implies that $Y(1)$ is biholomorphic to the punctured Riemann sphere $\mathbb{P}^1(\mathbb{C}) \setminus \{\infty\} \cong \mathbb{C}$.

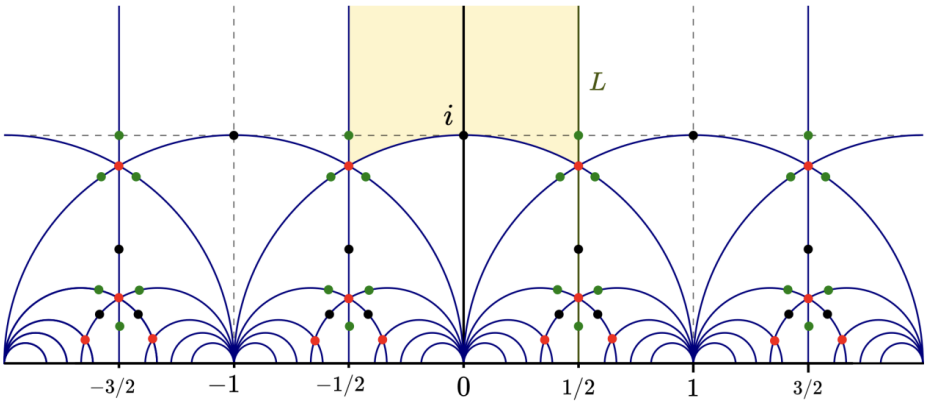

FIGURE 4. A fundamental domain of the $\mathrm{SL}_2(\mathbb{Z})$-action on $\mathbb{H}$. The curve $Y(1)$ can be thought of as the upper half-plane folded along the blue lines.

Therefore, amazingly, we can parametrize the space of elliptic curves up to isomorphism by a manifold! As such, $X(1)$ is a first example of a *moduli space*, which are ubiquitous in algebraic geometry and number theory. As we will see in the following section, we can generalize this construction by replacing $\mathrm{SL}_2(\mathbb{Z})$ by suitable subgroups $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$.

4. CONGRUENCE SUBGROUPS AND MODULAR CURVES

In the previous section, we saw that the quotient

$$Y(1) := \mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$$

parametrizes isomorphism classes of complex elliptic curves. We can refine this construction by replacing $\mathrm{SL}_2(\mathbb{Z})$ with so called *congruence subgroups* Γ, for which the quotients $Y(\Gamma) = \mathbb{H}/\Gamma$ parametrize elliptic curves up to isomorphism and additional torsion data.

Definition 4.1. Let N be a positive integer. The *principal congruence subgroup of level N* is

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

In other words, the principal congruence subgroup $\Gamma(N)$ is the kernel of the natural homomorphism

$$\mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}),$$

so it is a normal subgroup. Since this mapping is also clearly surjective, we have an induced isomorphism $\mathrm{SL}_2(\mathbb{Z})/\Gamma(N) \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. In particular, $\Gamma(1) \cong \mathrm{SL}_2(\mathbb{Z})$, and $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)]$ is finite for all N.

Definition 4.2. A subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is a *congruence subgroup* if $\Gamma(N) \subset \Gamma$ for some $N \in \mathbb{N}$, in which case Γ is said to be a congruence subgroup of level N.

In addition to the principal congruence subgroups, two other important classes of congruence subgroups are as follows:

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

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

where "*" indicates that the entries are unspecified. As one can easily verify, the natural mappings

$$\Gamma_1(N) \longrightarrow \mathbb{Z}/N\mathbb{Z}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto b \pmod{N}$$

and

$$\Gamma_0(N) \longrightarrow (\mathbb{Z}/N\mathbb{Z})^\times, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto d \pmod{N},$$

are surjective homomorphisms with kernels $\Gamma(N)$ and $\Gamma_1(N)$ respectively. Therefore, we have the following series of normal subgroups

$$\Gamma(N) \triangleleft \Gamma_1(N) \triangleleft \Gamma_0(N) \triangleleft \mathrm{SL}_2(\mathbb{Z}).$$

As in the case of $\mathrm{SL}_2(\mathbb{Z})$, each congruence subgroup Γ acts on $\mathbb{H}$ by the action

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

from which one can define the quotient $Y(\Gamma) := \mathbb{H}/\Gamma$.

As in Section 3, these $Y(\Gamma)$ inherit the structure of a Riemann surface, and the compactification $X(\Gamma)$ of such a curve is accordingly called the *modular curve for* Γ. Geometrically, the modular curves are related to one another via *coverings*. In particular, if $\Gamma' \subset \Gamma$ are congruence subgroups, then there is a natural finite covering map

$$Y(\Gamma') \longrightarrow Y(\Gamma), \qquad \Gamma' \cdot \tau \longmapsto \Gamma \cdot \tau,$$

which extends to the respective modular curves $X(\Gamma)$ and $X(\Gamma')$.[4]

The significance of these quotients is that they parametrize elliptic curves together with additional level structure. For instance:

- $Y(N)$ parametrizes elliptic curves together with a choice of ordered basis of the N-torsion subgroup;
- $Y_1(N)$ parametrizes elliptic curves together with a specified point of exact order N;
- $Y_0(N)$ parametrizes elliptic curves together with a cyclic subgroup of order N.

5. The j-invariant

Having constructed the modular curves $X(\Gamma)$ as compact Riemann surfaces, it is natural to study the holomorphic functions defined on them. However, as for any compact Riemann surface, the only globally holomorphic functions are constants; thus we instead pass to the field of *meromorphic functions* defined on them. The space of *modular forms* on a modular curve is a generalization of these functions with a slightly milder symmetry condition.

For simplicity, we restrict our definition to the case $\Gamma = \mathrm{SL}_2(\mathbb{Z})$:

Definition 5.1. Let k be a non-negative integer. A *modular form of weight* k for $\mathrm{SL}_2(\mathbb{Z})$ is a holomorphic function $f : \mathbb{H} \to \mathbb{C}$ satisfying:

(1) *(Modularity)* For all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and all $\tau \in \mathbb{H}$,

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

(2) *(Holomorphicity at the cusp)* The function f is holomorphic as $\tau \to i\infty$.

Note that when $k = 0$, condition (1) reduces to $\mathrm{SL}_2(\mathbb{Z})$-invariance, and condition (2) forces f to extend to the compactification $X(1)$, hence by Liouville's theorem, all modular forms of weight 0 are constant. Moreover, if k is odd, then applying modularity with the matrix $-I_2 \in \mathrm{SL}_2(\mathbb{Z})$, we obtain

$$f(\tau) = (-1)^k f(\tau) = -f(\tau),$$

the only modular form of odd weight is $f \equiv 0$. In this sense, modular forms of even weight $k > 0$ are the interesting objects.

[4]One important reason why the compactifications are emphasized is following amazing fact: every compact Riemann surface is an algebraic curve! That is, all such surfaces are cut out by polynomials.

Moreover, applying modularity with the matrix $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$, we have that every modular form satisfies $f(\tau + 1) = f(\tau)$. Therefore, they all admit a convergent Fourier expansion

$$f(\tau) = \sum_{n=0}^{\infty} a_n q^n, \qquad q := e^{2\pi i \tau},$$

which is called the *q-expansion* of f. If $a_0 = 0$, then f is called a *cusp form*.

An important example of a cusp form is the *modular discriminant*. Recall from Section 2 that for a point $\tau \in \mathbb{H}$, letting $\Lambda_\tau := \mathbb{Z} \oplus \tau\mathbb{Z}$, the expression

$$E_{\Lambda_\tau} : y^2 = 4x^3 - g_2(\Lambda_\tau)x - g_3(\Lambda_\tau)$$

determines an elliptic curve, where $g_2(\Lambda_\tau)$ and $g_3(\Lambda_\tau)$ are defined as in Proposition 2.4. Thus, by definition, the curve has a nonzero discriminant

$$\Delta(\tau) := g_2(\Lambda_\tau)^3 - 27 g_3(\Lambda_\tau)^2 \neq 0.$$

Since g_2 and g_3 are defined by absolutely convergent sums that transform correctly under $\mathrm{SL}_2(\mathbb{Z})$, one can verify that Δ is a modular form of weight 12. That Δ is in fact a *cusp* form can be seen from its remarkable product expansion, due to Jacobi:

$$\Delta(\tau) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24}, \qquad q = e^{2\pi i \tau}.$$

We are now in a position to return to the j-invariant introduced in Section 3 and realize it as a holomorphic function on $X(1)$. Recall from Section 3 that the j-invariant of an elliptic curve E is defined by

$$j(E) = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2},$$

and that two elliptic curves are isomorphic if and only if they have the same j-invariant.

In terms of lattices, letting

$$E_\tau = y^2 = 4x^3 - g_2(\Lambda_\tau)x - g_3(\Lambda_\tau)$$

be the elliptic curve associated to Λ_τ, we have that $A = -g_2/4$ and $B = -g_3/4$. Therefore,

$$j(E_\tau) := j(\tau) = \frac{1728\ g_2(\Lambda_\tau)^3}{g_2(\Lambda_\tau)^3 - 27 g_3(\Lambda_\tau)^2} = \frac{1728\ g_2(\Lambda_\tau)^3}{\Delta(\tau)}.$$

Since $g_2(\Lambda_\tau)^3$ and $\Delta(\tau)$ are defined by lattice sums of the same degree, they transform the same way under $\mathrm{SL}_2(\mathbb{Z})$. Thus in particular, their ratio is $\mathrm{SL}_2(\mathbb{Z})$-invariant. Moreover, since they are both meromorphic, it follows that their ratio is as well:

Definition 5.2. The *j-function* is the meromorphic function on $\mathbb{H}$ defined by

$$j(\tau) := 1728 \cdot \frac{g_2(\Lambda_\tau)^3}{\Delta(\tau)}.$$

Let us verify that j is holomorphic on $\mathbb{H}$ and has a simple pole at the cusp. Since $\Delta(\tau) \neq 0$ on $\mathbb{H}$ and g_2 is holomorphic, j is holomorphic on all of $\mathbb{H}$. At the cusp, the product formula gives $\Delta(\tau) \sim (2\pi)^{12} q$ as $\tau \to i\infty$, while $g_2(\Lambda_\tau) \to 60 \cdot 2\zeta(4) \neq 0$, so

$$j(\tau) \sim \frac{c}{q} \longrightarrow \infty$$

as $\tau \to i\infty$. Therefore, j extends to a meromorphic function on $X(1)$ with a simple pole at ∞.

FIGURE 5. Graph of the j-invariant on the upper half-plane.

Aside from elliptic curves, the j-invariant is also centrally important to the modular curve $X(1)$ for the fact that it generates its function field:

Theorem 5.3. *(i) The j-function induces a biholomorphism*

$$j : X(1) \xrightarrow{\sim} \mathbb{P}^1(\mathbb{C})$$

onto the Riemann sphere. (ii) j generates the field of meromorphic functions on $X(1)$, and takes every value in $\mathbb{C}$ exactly once on $Y(1)$.

Proof sketch. (i) From complex geometry, a nonconstant meromorphic function f on a compact Riemann surface X has a well-defined *degree* $d \geq 1$, meaning it takes each value in $\mathbb{P}^1(\mathbb{C})$ exactly d times counted with multiplicity. The map $f : X \to \mathbb{P}^1(\mathbb{C})$ is then a biholomorphism if and only if $d = 1$. Since j has a single simple pole at $\infty \in X(1)$, its degree is 1, and therefore $j : X(1) \xrightarrow{\sim} \mathbb{P}^1(\mathbb{C})$ is a biholomorphism.

(ii) Since j gives a biholomorphism onto the Riemann sphere, it follows that any meromorphic function on $X(1)$ pulls back via j to a meromorphic function on $\mathbb{P}^1(\mathbb{C})$. Since the meromorphic functions on $\mathbb{P}^1(\mathbb{C})$ are precisely the rational functions, it follows that every meromorphic function on $X(1)$ is of the form $r \circ j(z)$ for some rational function r. Thus, we have an isomorphism of function fields $\mathcal{M}(X(1)) \cong \mathbb{C}(j)$. □

While we have seen two examples of modular forms, its full theory and applications are far more vast and intricate than can be contained in this note. Theoretically, modular forms can be defined in purely algebro-geometric terms as holomorphic sections of the cotangent bundle of $X(\Gamma)$, a perspective which leads, via the Riemann–Roch theorem, to explicit dimension formulas for the spaces of modular forms among other results.

Beyond pure mathematics, modular forms appear naturally in theoretical physics: for instance, the partition function of a free boson on a torus is precisely a modular form, and the modular invariance of string theory amplitudes is nothing more than invariance under the $\mathrm{SL}_2(\mathbb{Z})$-action we have studied throughout.

However, their most natural home remains in number theory, where they were a key ingredient in the proof of the celebrated modularity theorem of Taylor and Wiles, whose most famous corollary, alas, failed to fit in the margins of Fermat's notebook some 400 years ago.

REFERENCES

[1] J. H. Silverman, *The Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics, vol. 106, Springer, New York, 1986. 3

[2] J. H. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics, vol. 151, Springer, New York, 1994.

[3] F. Diamond and J. Shurman, *A First Course in Modular Forms*, Graduate Texts in Mathematics, vol. 228, Springer, New York, 2005. 1