DIPARTIMENTO DI MATEMATICA E FISICA
Corso di Laurea Magistrale in Matematica

An Introduction to Elliptic Curves and Modular Forms

Relatore:
Prof. Francesco Pappalardi

Candidato:
Federico Campanini
n° matricola 428445

Anno Accademico 2014-2015
Ottobre 2015

AMS Classification: 11F06, 11F11, 11F25, 11G05, 33E05.
Keywords: elliptic curves, modular curves, modular groups, modular forms, Hecke operators.

Introduction

The theory of *elliptic curves* and *modular forms* involves some of the most important branches of Mathematics like Complex Analysis, Algebraic Geometry, Representation Theory and Number Theory. Our point of view will be number theoretic. A great amount of conjectures and theorems in Number Theory can be formulated using such simple statements that even a little mathematical knowledge is enough to let them be understood. Despite this exterior simplicity, their proofs are unknown or need some of the twentieth century most proficient mathematical results. The aim of this thesis is to explore the theory of elliptic curves and modular forms and to develop tools that are essential in the resolution of some remarkable problems, like the Ramanujan conjectures and the Fermat's Last Theorem.

Chapter 1 is devoted to the study of elliptic curves, the set of points in the projective complex space $\mathbb{P}^3_{\mathbb{C}}$ which satisfy an algebraic equation (called *Weierstrass equation*) of the form

$$E : Y^2Z = 4X^3 + a_2XZ^2 + a_3Z^3, \qquad a_2, a_3 \in \mathbb{C}.$$

By using the theory of *elliptic functions* we see that we can also regard an elliptic curve as the quotient of the complex plane by a suitable lattice. Our approach emphasizes the connection between elliptic curves and modular forms and this link will be taken up in later chapters.

Chapter 2 introduces the *modular group* $SL_2(\mathbb{Z})$ and its congruence subgroups. We study their action on the upper complex half plane $\mathcal{H} = \{z \in \mathbb{C} \mid Im(z) > 0\}$ and we define the *modular curves* as the spaces of the orbits of the actions of these groups on $\mathcal{H}$. We see that modular curves are in some sense natural domains for modular forms. A significant part of this chapter shows how modular curves can be viewed as Riemann surfaces that can be compactified. Besides being a remarkable result, this fact allows us to study some important aspects of the action of the modular group and of its subgroups on $\mathcal{H}$.

Chapter 3 gives the basic definitions and the main results of modular forms which are complex analytic functions on the upper half-plane that are "essentially invariant" under the action of the modular group. The theory of modular forms therefore belongs to complex analysis but the main importance of the theory has traditionally been in its connections with number theory, as we see in Chapter 5.

Although in this thesis the link between elliptic curves and modular forms is emphasized, the exposure of the first three chapters is designed to make sure that they can be read independently of the others, but for some small measures.

In Chapter 4 we study certain operators on modular forms, the *Hecke operators*. Historically, this operators were used by Mordell in 1917 to prove that the Ramanujan tau function is a multiplicative function (we see a proof in Chapter 5). Mordell used this operators in a paper on the special cusp form of Ramanujan, ahead of the general theory given by Hecke (1937). There are several ways to define the Hecke operators: in this thesis, we define them as *double coset operators*. We see that we can endow the space of *cusp forms* with an inner product (namely, the Petersson inner product) and we define the spaces of *new forms* and *old forms*. In particular we prove that the space of *cusp forms* for the congruence subgroup $\Gamma_1(N)$ has an orthogonal basis of simultaneous *eigenforms* for the family of Hecke operators and that a new form is an eigenform for the Hecke operators and its eigenvalues are essentially its Fourier coefficients.

Chapter 5 explains some applications of these topics and cite some outstanding problems. We prove two of the most important results about the Ramanujan τ function and we see what is the role of elliptic curves and modular forms in the proof of the famous Fermat's Last Theorem.

Contents

CHAPTER 1

Elliptic Curves

In this chapter we will introduce some analytic objects like complex tori and elliptic functions and we will show how they are closely related to some algebraic objects known as *elliptic curves*. Moreover, elliptic curves allows us to introduce in a natural way the notions of modular group and modular forms. We will talk about these topics in the next chapters.

1.1 Elliptic Functions and the Weierstrass Function

We will briefly introduce some basic notions and results about elliptic functions. For more details about these topics see, for example, [Lan87].

Definition 1.1. A **lattice** Λ is a subgroup of $\mathbb{C}$ which is free of dimension 2 over $\mathbb{Z}$ and which generates $\mathbb{C}$ over $\mathbb{R}$. If $\{\omega_1, \omega_2\}$ is a basis of Λ over $\mathbb{Z}$, then we also write $\Lambda = [\omega_1, \omega_2]$ or $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$.

A **complex torus** is a quotient of the complex plane by a lattice,

$$\mathbb{C}/\Lambda = \{z + \Lambda \mid z \in \mathbb{C}\}.$$

Let $\mathcal{H} = \{z \in \mathbb{C} \mid Im(z) > 0\}$ be the *complex upper half plane*. It's easy to see that if $\Lambda = [\omega_1, \omega_2]$ and $\Lambda' = [\omega_1', \omega_2']$ are two lattices with $\omega_1/\omega_2 \in \mathcal{H}$ and $\omega_1'/\omega_2' \in \mathcal{H}$, then $\Lambda = \Lambda'$ if and only if

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

for some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z}) = \{M \in M_2(\mathbb{Z}) \mid det(M) = 1\}$. Unless otherwise specified, if $\Lambda = [\omega_1, \omega_2]$ is a lattice, we assume the normalizing convention that $\omega_1/\omega_2 \in \mathcal{H}$.

Algebraically, a complex torus is an Abelian group under the addition induced by $\mathbb{C}$ and geometrically it is a *compact Riemann surface.* We recall from the complex analysis that any holomorphic map between compact Riemann surfaces is surjective or constant, so any non-constant holomorphic map between complex tori is a surjection.

Definition 1.2. An **elliptic function f with respect to** Λ is a meromorphic function on $\mathbb{C}$ which is Λ-periodic, i.e.

$$f(z+\omega) = f(z) \qquad \forall z \in \mathbb{C}, \quad \forall \omega \in \Lambda.$$

Note that if $\Lambda = [\omega_1, \omega_2]$, a function f is Λ-periodic if and only if $f(z+\omega_1) = f(z) = f(z+\omega_2)$. If an elliptic function f is holomorphic then f must be constant because it is bounded on $\mathbb{C}/\Lambda$, hence it is bounded on $\mathbb{C}$ and therefore it is constant by Liouville theorem.

Let $\Lambda = [\omega_1, \omega_2]$ be a lattice, and $\alpha \in \mathbb{C}$. A *fundamental parallelogram* for Λ is defined as the set of the points $P = \{\alpha + t_1\omega_1 + t_2\omega_2 \mid 0 \leq t_1, t_2 < 1\}$.

Theorem 1.1.1. *Let P be a fundamental parallelogram for Λ and assume that the elliptic function f has no poles on its boundary ∂P. Then the sum of the residues of f in P is zero. In particular an elliptic function has at least two poles on the torus $\mathbb{C}/\Lambda$.*

Proof. Since f is Λ-periodic, we have:

$$2\pi i \sum Res(f) = \int_{\partial P} f(z)dz = 0.$$

□

Theorem 1.1.2. *Let P be a fundamental parallelogram for Λ and assume that the elliptic function f has no zeros or poles on its boundary ∂P. Let $\{a_i\}$ be the singular points of f inside P, and let f have order m_i at a_i. Then $\sum m_i = 0$.*

Proof. If f is an elliptic function then f' and f'/f are elliptic functions, so

$$0 = \int_{\partial P} f'/f(z)dz = 2\pi i \sum Res(f'/f) = 2\pi i \sum m_i.$$

□

Theorem 1.1.3. *Let P be a fundamental parallelogram for Λ and assume that the elliptic function f has no zeros or poles on its boundary ∂P. Let $\{a_i\}$ be the singular points of f inside P, and let f have order m_i at a_i. Then*

$$\sum m_i a_i \equiv 0 \ (mod \ \Lambda).$$

Proof. We consider the integral

$$I = \int_{\partial P} z \frac{f'(z)}{f(z)} dz.$$

On one hand $I = 2\pi i \sum m_i a_i$, since

$$Res_{a_i} \left(z \frac{f'(z)}{f(z)} \right) = m_i a_i.$$

On the other hand, if we compute the integral over ∂P by taking it for two opposite sides at a time we obtain for one pair of such integrals

$$\int_{\alpha}^{\alpha+\omega_1} z \frac{f'(z)}{f(z)} dz - \int_{\alpha+\omega_2}^{\alpha+\omega_1+\omega_2} z \frac{f'(z)}{f(z)} dz.$$

Setting $u = z - \omega_2$ in the second integral we obtain, after a cancellation, the value

$$-\omega_2 \int_{\alpha}^{\alpha+\omega_1} \frac{f'(u)}{f(u)} du = 2\pi i k \omega_2$$

for some integer k. In the same way, for the integral over the other pair of opposite sides we obtain the value $2\pi i h \omega_1$ for some integer h and the result follows. □

Now we introduce the **Weierstrass $\wp$-function**. It is an important example of a non-constant elliptic function that we can use to show how complex tori can also be viewed as a kind of cubic curves, called *elliptic curves*.

Let $\Lambda = [\omega_1, \omega_2]$ be a lattice and put $\Lambda^* = \Lambda \backslash \{0\}$. The Weierstrass $\wp$-function, with respect to Λ, is

$$\wp(z) = \wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right]. \tag{1.1}$$

(In the following we omit the subscript Λ since we may always assume that the lattice is fixed throughout the discussion.) We want to show that this series converges uniformly on compact sets not including the lattice points.

Lemma 1.1.4. *If $k \geq 3$, then $\sum_{\omega \in \Lambda^*} \frac{1}{|\omega|^k}$ converges.*

Proof. Let A_n denote the set $A_n = \{\omega \in \Lambda^* \mid n-1 < |\omega| \leq n\}$. Clearly $A_n \cap A_m = \emptyset$ if $n \neq m$ and $\Lambda^* = \bigcup_n A_n$. Moreover A_n consists at most of $8n$ elements, any of which is in module greater than cn, where $c = \min_{\omega \in A_1} |\omega|$. So, we have

$$\sum_{\omega \in A_n} \frac{1}{|\omega|^k} \leq \frac{8n}{c^k n^k}$$

and hence

$$\sum_{\omega \in \Lambda^*} \frac{1}{|\omega|^k} = \sum_{n \geq 1} \left(\sum_{\omega \in A_n} \frac{1}{|\omega|^k} \right) \leq \sum_{n \geq 1} \frac{8}{c^k n^{k-1}}$$

which converges for $k \geq 3$. □

Proposition 1.1.5. *The series* (1.1) *converges uniformly on compact sets not including the lattice points.*

Proof. Fix $R > 0$. We have $|\omega| \geq R$ for all $\omega \in \Lambda$ except for a finite number of elements. If $|z| < R$, z stays away from the lattice points, we have

$$\left| \frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right| = \left| \frac{2\omega z - z^2}{\omega^2 (z-\omega)^2} \right| = \frac{|z(2 - z/\omega)|}{|\omega|^3 |1 - z/\omega|^2} \leq \frac{cR}{|\omega|^3}$$

for an appropriate constant c. The result follows by the previous lemma. □

The series expression for $\wp$ shows that it is meromorphic, with a double pole at each lattice point, and no other pole. Furthermore, since summing over the lattice points is the same as summing over their negatives, $\wp$ is an even functions, i.e. $\wp(z) = \wp(-z)$. Differentiating term by term we obtain

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

Note that the sum is taken over all $\omega \in \Lambda$. Clearly, $\wp'$ is periodic and odd, i.e. $-\wp'(z) = \wp'(-z)$. From its periodicity, we deduce that there exists a constant C such that $\wp(z+\omega_1) = \wp(z) + C$ and, since $\wp$ is even, if we put $z = -\omega_1/2$, we obtain that $C = 0$. Similarly $\wp(z+\omega_2) = \wp(z)$, hence $\wp$ is itself periodic, and this does not obviously follow from its series expansion. A surprising curiosity of these functions is that the field of elliptic functions with respect to Λ is generated by $\wp$ and $\wp'$. We don't prove this result (and actually we shall not use it).

Now we want to obtain an algebraic relation between $\wp$ and $\wp'$. We do this by computing the power series development at the origin of this two functions. We have

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left[\frac{1}{\omega^2} \left(1 + \frac{z}{\omega} + \left(\frac{z}{\omega}\right)^2 + ... \right)^2 - \frac{1}{\omega^2} \right] =$$

$$= \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \sum_{m=1}^{\infty} (m+1) \left(\frac{z}{\omega}\right)^m \frac{1}{\omega^2} = \frac{1}{z^2} + \sum_{m=1}^{\infty} c_m z^m$$

where

$$c_m = \sum_{\omega \in \Lambda^*} \frac{m+1}{\omega^{m+2}}.$$

Note that $c_m = 0$ if m is odd. Using the notation

$$G_m(\Lambda) = G_m = \sum_{\omega \in \Lambda^*} \frac{1}{\omega^m}$$

we obtain

$$\wp(z) = \frac{1}{z^2} + \sum_{n=1}^{\infty} (2n+1) G_{2n+2} z^{2n}. \tag{1.2}$$

Theorem 1.1.6. *Let $g_2 = 60G_4$ and $g_3 = 140G_6$. Then $\wp$ satisfies the differential relation*

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3. \tag{1.3}$$

Proof. It suffices to expand out the function $\psi(z) = \wp' - 4\wp^3 + g_2\wp + g_3$ at the origin, looking only to the polar terms and the constant term. We want to show that ψ is identically zero. From the expansion (1.2) of $\wp$ we have

$$\wp(z) = \frac{1}{z^2} + 3G_4z^2 + 5G_6z^4 + ...$$

and differentiating term by term, we obtain

$$\wp'(z) = -\frac{2}{z^3} + 6G_4z + 20G_6z^3 + ...$$

It is easy to see that ψ is an elliptic function without poles and with the constant term equal to zero (i.e. with a zero at the origin). It follows that ψ is identically zero, as required. □

The previous theorem shows that the points $(\wp(z), \wp'(z))$ lie on the curve defined by the equation

$$Y^2 = 4X^3 - g_2X - g_3.$$

We want to show that the cubic polynomial on the right-hand side has three distinct roots and so its discriminant $\Delta = g_2^3 - 27g_3^2$ does not vanish. To do this we put

$$e_1 = \wp(\frac{\omega_1}{2}), \qquad e_2 = \wp(\frac{\omega_2}{2}), \qquad e_3 = \wp(\frac{\omega_3}{2})$$

where $\omega_3 = \omega_1 + \omega_2$. Since $\wp'$ is Λ-periodic and odd, we have $\wp'(\omega_i/2) = \wp'(-\omega_i/2) = -\wp'(\omega_i/2)$. It follows that $\wp'(\omega_i/2) = 0$ and hence e_1, e_2, e_3 are the roots of $4X^3 - g_2X - g_3$. Then the function $\psi(z) = \wp(z) - e_i$ has a zero at $\omega_i/2$, which is of order at least 2, because $\wp'(\omega_i/2) = 0$. Since ψ has only one pole of order 2 mod Λ, by Theorem (1.1.2) there are no other points z mod Λ such that $\wp(z) = e_i$ and so $e_i \neq e_j$ for $i \neq j$, as required.

1.2 Elliptic Curves and the Addition Law

In the previous section we have seen that the map

$$z \mapsto (\wp(z), \wp'(z))$$

parametrizes points on the curve $Y^2 = 4X^3 - g_2X - g_3$. If we want to embed the points in the projective space $\mathbb{P}^2_{\mathbb{C}}$, we can write $z + \Lambda \mapsto (\wp(z) : \wp'(z) : 1)$. In this way we have a map from the torus $\mathbb{C}/\Lambda$ to the set $E(\mathbb{C})$ of the complex projective points on the homogenized curve $Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$, where the lattice points are the points going to infinity on the curve. Namely, we have a map

$$\begin{aligned} \Phi\colon \mathbb{C}/\Lambda &\longrightarrow E(\mathbb{C}) \\ 0 + \Lambda &\mapsto (0 : 1 : 0) \\ z + \Lambda &\mapsto (\wp(z) : \wp'(z) : 1) \end{aligned} \tag{1.4}$$

and it is easy to see that it is a bijection. In fact for any $\alpha \in \mathbb{C}$, $\wp(z) - \alpha$ has at most two zeros, and at least one zero, so that already under $\wp$ we cover each complex number α. Now (with abuse of notation), if $\wp(w + \Lambda) = \alpha$, then $\wp(-w + \Lambda) = \alpha$ and the fact that $\wp'(\pm w + \Lambda) = \pm\wp'(w + \Lambda)$ giving us the bijection.

Moreover Φ is an analytic map, meaning that near any point of $\mathbb{C}/\Lambda$ it can be given by a triple of analytic functions. Near non-lattice points the map is given by $z \mapsto (\wp(z) : \wp'(z) : 1)$ while near the lattice points the map is given by $z \mapsto (\frac{\wp(z)}{\wp'(z)} : 1 : \frac{1}{\wp'(z)})$.

Now we introduce the notion of *elliptic curve over* $\mathbb{C}$.

Definition 1.3. An **elliptic curve** over $\mathbb{C}$ is a cubic projective curve, defined over $\mathbb{C}$, given by the equation

$$E : Y^2Z = 4X^3 + a_2XZ^2 + a_3Z^3, \qquad a_2, a_3 \in \mathbb{C}. \tag{1.5}$$

Sometimes we write the equation in its non-homogeneous form, namely $E : Y^2 = 4X^3 + a_2X + a_3$. The equation 1.5 or its non-homogeneous form will be referred to as the **Weierstrass equation**. By definition we also require that E is **non-singular**. It means that the cubic polynomial on the right-hand side has three distinct roots, i.e. its discriminant $\Delta = a_2^3 - 27a_3^2 \neq 0$. The set of the complex projective points on the curve is denoted by $E(\mathbb{C})$.

In the previous section we have proved that for the equation $E : Y^2 = 4X^3 - g_2X - g_3$, the discriminant of the right-hand $\Delta = g_2^3 - 27g_3^2$ does not vanish and hence E defines an elliptic curve. Note that the condition $\Delta \neq 0$ is equivalent to say that the curve is *smooth*, namely we can define the tangent line for each point P on E. We can express this condition by saying that if E is defined by the polynomial equation $F(X,Y) = 0$, where $F(X,Y) = Y^2 - 4X^3 - a_2X - a_3$, then E is smooth at a point (x_0, y_0) if

$$\left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right)\bigg|_{(x_0,y_0)} \neq (0,0).$$

In our case the partial derivatives are $\frac{\partial F}{\partial x} = -12X^2 - a_2$ and $\frac{\partial F}{\partial y} = 2Y$. These are both zero at (x_0, y_0) if and only if $y_0 = 0$ and x_0 is a multiple root of $4X^3 + a_2X + a_3$. Thus a curve defined by $E : Y^2 = 4X^3 + a_2X + a_3$ is non-singular if and only if it is a smooth curve.

We have seen that there is an analytic bijection between $\mathbb{C}/\Lambda$ and the elliptic curve $E : Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$ (recall that $g_2 = g_2(\Lambda)$ and $g_3 = g_3(\Lambda)$ depend on the lattice). Furthermore, $\mathbb{C}/\Lambda$ has a natural group structure, and we now want to see what it looks like when it is transported to $E(\mathbb{C})$.

Let $P_z = \Phi(z+\Lambda)$ denotes a generic point in $E(\mathbb{C})$. We can use the one-to-one correspondence Φ to define a commutative group law on $E(\mathbb{C})$, namely for any $z_1 + \Lambda, z_2 + \Lambda$ in $\mathbb{C}/\Lambda$, we can define

$$P_{z_1} + P_{z_2} = P_{z_1+z_2}.$$

In this way we obtain an addition law that has two remarkable properties: first, there is a simple geometric interpretation for adding points on an elliptic curve and second, we can express the coordinates of $P_{z_1+z_2}$ as rational functions of the coordinates of P_{z_1} and P_{z_2}.

We start to treat the simplest case: $z_2 + \Lambda = 0 + \Lambda$. It is clear that $P_0 + P_{z_1} = P_{z_1} + P_0 = P_{z_1}$, so the identity element is the point at infinity $\infty = P_0 = (0 : 1 : 0)$. Now suppose that neither P_{z_1} nor P_{z_2} is the point at infinity. Using the non-homogeneous coordinates, set

$$P_{z_1} = (x_1, y_1), \qquad P_{z_2} = (x_2, y_2), \qquad P_{z_1} + P_{z_2} = P_{z_3} = (x_3, y_3).$$

We have some special cases. Suppose that $x_1 = x_2$. This means that $y_1 = -y_2$ and so $z_2 = -z_1 \pmod{\Lambda}$, since only in this case we can have the same value for $\wp$. It follows that $P_{z_1} + P_{z_2} = P_0 = \infty$. If $y_1 = -y_2 = 0$, i.e. $P_{z_1} = P_{z_2}$ lies on the x-axis, then we have $P_{z_1} + P_{z_2} = 2P_{z_1} = P_0$. We have just proved that $-(x, y) = (x, -y)$, i.e. the additive inverse of a point $P_z = (x, y)$ is $-P_z = (x, -y)$. Geometrically, we say that two points on the curve which lie on the same vertical line (that is a tangent line to the curve if $P_{z_1} = P_{z_2}$) are additive inverses of one another, namely they have sum ∞. Note that we can think the identity element ∞ as the third point on the vertical line.

Now suppose that the line ℓ through P_{z_1} and P_{z_2} is not a vertical line. We want to find $P_{z_1} + P_{z_2} = P_{z_3}$ and in particular we want to show that $-P_{z_3}$ is the third point of intersection of ℓ with the elliptic curve (note that any line intersects the curve in three points, counting multiplicities). If we write $\ell : y = mx + q$, it is clear that a point $P_z = (\wp(z), \wp'(z))$ lies on ℓ if and only if $\wp'(z) = m\wp(z) + q$. Since the elliptic function

$$F(z) = \wp'(z) - (m\wp(z) + q)$$

has a pole of order three at zero, $F(z)$ has three zeros, counting with their multiplicities, and two of these are at z_1 and z_2. If the third zero lies at z_3, we must have

$$z_1 + z_2 + z_3 \equiv 0 \pmod{\Lambda}$$

by Theorem 1.1.2. It follows that $P_{-(z_1+z_2)} = -P_{z_3}$ is the third point of intersection of ℓ with the curve.

It remains to prove that we can express x_3, y_3 as rational functions of x_1, x_2, y_1 and y_2.

Assume again that if $P_{z_1} \neq P_{z_2}$ then $x_1 \neq x_2$ and that neither of the points is ∞. Then we can write the line through P_{z_1} and P_{z_2} in the form $\ell : y = mx + b$

as above. Its slope is

$$m = \begin{cases} \frac{y_2-y_1}{x_2-x_1} & \text{if } P_{z_1} \neq P_{z_2} \\ \\ \frac{12x_1^2-g_2}{2y_1} & \text{if } P_{z_1} = P_{z_2}. \end{cases}$$

The equation

$$4x^3 - g_2x - g_3 - (mx+b)^2 = 0$$

has three roots, that clearly are $\wp(z_1), \wp(z_2), \wp(z_3)$. So we can factorize the left-hand side as

$$4(x-\wp(z_1))(x-\wp(z_2))(x-\wp(z_3))$$

and comparing the coefficient of x^2 we obtain

$$x_1 + x_2 + x_3 = \wp(z_1) + \wp(z_2) + \wp(z_3) = \frac{m^2}{4}.$$

If $P_{z_1} \neq P_{z_2}$ we have

$$\wp(z_3) = -\wp(z_1) - \wp(z_2) + \frac{1}{4}\left(\frac{\wp'(z_1)-\wp'(z_2)}{\wp(z_1)-\wp(z_2)}\right)^2$$

or equivalently

$$x_3 = -x_1 - x_2 + \frac{1}{4}\left(\frac{y_1-y_2}{x_1-x_2}\right)^2. \tag{1.6}$$

For $P_{z_1} = P_{z_2}$ we take the limit as $z_1 \to z_2$ and get

$$\wp(2z) = -2\wp(z) + \frac{1}{4}\left(\frac{\wp''(z)}{\wp'(z)}\right)^2$$

or in algebraic terms

$$x_3 = -2x_1 + \frac{1}{4}\left(\frac{12x_1^2-g_2}{2y_1}\right)^2. \tag{1.7}$$

For the y-coordinate we have

$$\wp'(z_3) = \begin{cases} -\wp'(z_1) + \frac{1}{4}\left(\frac{\wp'(z_1)-\wp'(z_2)}{\wp(z_1)-\wp(z_2)}\right)^2 & \text{if } P_{z_1} \neq P_{z_2} \\ \\ -\wp'(z_1) + \frac{1}{4}\left(\frac{\wp''(z)}{\wp'(z)}\right)^2 & \text{if } P_{z_1} = P_{z_2} \end{cases}$$

or simply

$$y_3 = -y_1 + m(x_1 - x_3). \tag{1.8}$$

These give us the algebraic addition formulas, as desired. We have proved the following theorem:

Theorem 1.2.1. *Let Λ be a lattice and let $E : Y^2 = 4X^3 - g_2X - g_3$ be an elliptic curve over $\mathbb{C}$, where $g_2 = g_2(\Lambda)$ and $g_3 = g_3(\Lambda)$. Then the one-to-one correspondence* 1.4 *defines a commutative group law on $E(\mathbb{C}) \subseteq \mathbb{P}^2_{\mathbb{C}}$.*

For $z \in \mathbb{C}$ let $P_z = \Phi(z + \Lambda)$ denote the corresponding point. Then

- *$P_0 = (0 : 1 : 0) = \infty$ is the identity element, i.e. $P_0 + P_z = P_z + P_0 = P_z$ for all $P_z \in E(\mathbb{C})$;*
- *Suppose $(x : y : 1) = P_z \neq P_0$. Then the additive inverse of P_z is $-P_z = P_{-z} = (x, -y, 1)$;*
- *A point $P_z = (x : y : 1) \neq P_0$ has order two if and only if $y = 0$;*
- *If $P_{z_1} + P_{z_2} = P_{z_3}$ and ℓ denotes the line joining P_{z_1} and P_{z_2}, then $-P_{z_3}$ is the third point of intersection of ℓ with $E(\mathbb{C})$ (if $P_{z_1} = P_{z_2}$ then ℓ is the tangent line at P_{z_1});*
- *If $P_{z_1} = (x_1 : y_1 : 1)$ and $P_{z_2} = (x_2 : y_2 : 1)$ are not the point P_0 and they are not the inverses of one another, then we can write $P_{z_1} + P_{z_2} = P_{z_3} = (x_3 : y_3 : 1)$, where*

$$x_3 = \begin{cases} x_3 = -x_1 - x_2 + \frac{1}{4}\left(\frac{y_1-y_2}{x_1-x_2}\right)^2 & \textit{if } P_{z_1} \neq P_{z_2} \\ \\ x_3 = -x_1 - x_2 + \frac{1}{4}\left(\frac{12x_1^2-g_2}{2y_1}\right)^2 & \textit{if } P_{z_1} = P_{z_2} \end{cases}$$

and

$$y_3 = \begin{cases} y_3 = -y_1 + \frac{(y_2-y_1)}{x_2-x_1}(x_1 - x_3) & \textit{if } P_{z_1} \neq P_{z_2} \\ \\ y_3 = -y_1 + \frac{(12x_1^2-g_2)}{2y_1}(x_1 - x_3) & \textit{if } P_{z_1} = P_{z_2}. \end{cases}$$

Note that a priori the theorem holds only for the elliptic curves of the form $E : Y^2 = 4X^3 - g_2(\Lambda)X - g_3(\Lambda)$ for some lattice Λ, but in fact any elliptic curve over $\mathbb{C}$ assume that form for a suitable lattice. This result will be proved later in Section 1.4.

Actually, there is a more general context on which we can treat the elliptic curves. The main generalization regards the ground field: instead of working with the field of complex numbers we can work with any field K. Since these topics go beyond our scopes, we will just give some basic notion. A good reference for these arguments is [Was03].

There is a general form for the Weierstrass equation of an elliptic curve, which applies to any field K:

Definition 1.4. Let K be a field and consider the polynomial equation

$$E : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3 \tag{1.9}$$

where $a_1, ..., a_6 \in K$. This is the **Generalized Weierstrass equation**. An **elliptic curve** over K is a cubic projective curve given by an equation of the form 1.9. The set of the points on the curve with coordinates in K is denoted by $E(K)$.

If $char(K) \neq 2$, by the change of variables $Y_1 = Y + a_1X/2 + a_3/2$ we obtain

$$Y_1^2 = X^3 + a_2'X^2 + a_4'X + a_6'$$

for some $a_2', a_4', a_6' \in K$. Moreover, if $char(K) \neq 3$ we can also let $X_1 = X + a_2'/3$ and obtain

$$Y_1^2 = X_1^3 + a_2''X_1 + a_3''$$

for some constants $a_2'', a_3'' \in K$. Finally if we put $Y_2 = Y_1/2$ we obtain an equation of the form

$$Y^2 = 4X^3 + c_2X + c_3, \qquad c_2, c_3 \in K$$

which is the same as the definition 1.3. If $char(K) \neq 2$ we may assume that E is given by an equation of the form $Y^2 = f(X)$ for some cubic polynomial $f(X) = aX^3 + bX^2 + cX + d \in K[X]$ with no multiple roots. We can define a commutative group law on E by using the geometric construction that we have seen above. The formulas 1.6 and 1.7 for the addition law become respectively

$$x_3 = -x_1 - x_2 - \frac{b}{a} + \frac{1}{a}\left(\frac{y_1 - y_2}{x_1 - x_2}\right)^2$$

and

$$x_3 = -2x_1 - \frac{b}{a} + \frac{1}{a}\left(\frac{f'(x_1)}{2y_1}\right)^2.$$

The problem with this general definition, which does not use the group structure over a torus or over some other set, is to check the associativity law. It can be algebraically verified by using the formulas, but there are several cases to consider and this makes the proof rather messy. However there is a different approach, which is discussed in [Was03].

1.3 Isomorphism Classes of Complex Tori

Theorem 1.3.1. *Let Λ and Λ' be two lattices in $\mathbb{C}$ and let*

$$\varphi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$$

be a complex analytic group homomorphism. Then there exists a complex number m such that the following diagram is commutative.

$$\begin{tikzcd}
\mathbb{C} \arrow[r, "m"] \arrow[d] & \mathbb{C} \arrow[d] \\
\mathbb{C}/\Lambda \arrow[r, "\varphi"] & \mathbb{C}/\Lambda'
\end{tikzcd}$$

The top map is multiplication by m, and the vertical maps are the canonical homomorphisms.

In particular, there exists a non-zero analytic group homomorphism between the complex tori $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ if and only if there exists some non-zero $m \in \mathbb{C}$ such that $m\Lambda \subseteq \Lambda'$ and it is an isomorphism if and only if $m\Lambda = \Lambda'$.

Proof. We can expand φ by a power series at the origin,

$$\varphi(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 ...$$

Since φ is a group homomorphism, we have the congruence

$$\varphi(z + z') \equiv \varphi(z) + \varphi(z') \pmod{\Lambda}$$

that can be viewed as an equality for z near 0, because a complex number near 0 represents uniquely its class mod Λ. Hence, locally near 0, we must have

$$\varphi(z) = a_1 z.$$

But for an arbitrary z, if n is sufficiently large, z/n is near 0, so we conclude that for any $z \in \mathbb{C}$ we must have

$$\varphi(z) \equiv a_1 z \pmod{\Lambda}.$$

Now, it is clear that $m\Lambda \subseteq \Lambda'$.

Conversely, if $m\Lambda \subseteq \Lambda'$, multiplication by m induces an analytic group homomorphism from $\mathbb{C}/\Lambda$ to $\mathbb{C}/\Lambda'$. □

Definition 1.5. A non-zero analytic group homomorphism between $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ is called **isogeny**.

Since any analytic map between compact Riemann surfaces is either a surjection or a constant map, an isogeny surjects and has a finite kernel, because it is a discrete subset of a compact set. We will discuss about isogenies in more detail in Section 1.5.

The previous theorem has an important consequence. Let $\Lambda = [\omega_1, \omega_2]$ be a lattice, with $\tau = \omega_1/\omega_2 \in \mathcal{H}$. If Λ_τ denotes the lattice of the form $[\tau, 1]$, then $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda_\tau$. Therefore for every lattice Λ there exists another one of the form $\Lambda_\tau = [\tau, 1]$ for some $\tau \in \mathcal{H}$ such that $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda_\tau$.

Now we wonder when $\mathbb{C}/\Lambda_\tau \cong \mathbb{C}/\Lambda_{\tau'}$. We recall that if $\Lambda = [\omega_1, \omega_2]$ and $\Lambda' = [\omega_1', \omega_2']$ are two normalized lattices (i.e. with $\omega_1/\omega_2 \in \mathcal{H}$ and $\omega_1'/\omega_2' \in \mathcal{H}$), then $\Lambda = \Lambda'$ if and only if

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \quad \text{for some } M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

So, by the previous theorem, if $\mathbb{C}/\Lambda_\tau \cong \mathbb{C}/\Lambda_{\tau'}$ there exist $M \in SL_2(\mathbb{Z})$ and $\alpha \in \mathbb{C}$ such that

$$\begin{pmatrix} \tau' \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha(a\tau + b) \\ \alpha(c\tau + d) \end{pmatrix}.$$

Thus $\alpha = (c\tau + d)^{-1}$ and

$$\tau' = \frac{a\tau + b}{c\tau + d}.$$

So we have just proved the following result.

Corollary 1.3.2. *Let $\Lambda = [\omega_1, \omega_2]$ and $\Lambda' = [\omega_1', \omega_2']$ be two normalized lattices such that there is a analytic group isomorphism $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda'$. Then there exist $M \in SL_2(\mathbb{Z})$ and $\alpha \in \mathbb{C}$ such that*

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \alpha M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}.$$

If Λ_τ denotes the lattice of the form $[\tau, 1]$, then there exists $\tau \in \mathcal{H}$ such that $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda_\tau$.

Let $\tau, \tau' \in \mathcal{H}$. Then $\mathbb{C}/\Lambda_\tau \cong \mathbb{C}/\Lambda_{\tau'}$ if and only if there exists a matrix $M \in SL_2(\mathbb{Z})$ such that $\left(\begin{smallmatrix} \tau' \\ 1 \end{smallmatrix}\right) = M \left(\begin{smallmatrix} \tau \\ 1 \end{smallmatrix}\right)$.

In the next chapter we will define an action of $SL_2(\mathbb{Z})$ on $\mathcal{H}$ (the upper half plane). For any $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ and any $\tau \in \mathcal{H}$ we define $M\tau = \frac{a\tau+b}{c\tau+d}$. Thus we can rewrite the last statement in the corollary by saying that $\mathbb{C}/\Lambda_\tau \cong \mathbb{C}/\Lambda_{\tau'}$ if and only if there exists a matrix $M \in SL_2(\mathbb{Z})$ such that $\tau' = M\tau$.

1.4 The Uniformization Theorem

In this section we prove that there is a bijection between the set of the complex tori $\mathbb{C}/\Lambda$ and the set of the elliptic curves $E(\mathbb{C}) : y^2 = 4x^3 + a_2x + a_3$. This result is known as *The Uniformization Theorem*. We need some results which will be proved in the next chapters. However, we want to show here this theorem to complete the discussion on elliptic curves.

Let $\Lambda = [\omega_1, \omega_2]$ be a lattice and put $\Lambda^* = \Lambda \setminus \{0\}$. Recall that the Weierstrass function (with respect to Λ) satisfies the equation

$$\wp'(z) = 4\wp^3(z) - g_2(\Lambda)\wp(z) - g_3(\Lambda),$$

where

$$G_m(\Lambda) = \sum_{\omega \in \Lambda^*} \frac{1}{\omega^m}, \qquad g_2(\Lambda) = 60G_4(\Lambda), \qquad g_3(\Lambda) = 140G_6(\Lambda).$$

The series G_m is called the *Eisenstein series.* We write $g_2(z)$ for $g_2(\Lambda_z)$ and $g_3(z)$ for $g_3(\Lambda_z)$. We will see that $g_2(z)$ and $g_3(z)$ are two holomorphic functions on $\mathcal{H} = \{z \in \mathbb{C} \mid \Im(z) > 0\}$. We define the *discriminant function* as

$$\Delta : \mathcal{H} \to \mathbb{C}, \qquad \Delta(z) = g_2^3(z) - 27g_3^2(z).$$

Recall that $\Delta(z) \neq 0$ for all $z \in \mathcal{H}$. It follows that the function

$$j : \mathcal{H} \to \mathbb{C}, \qquad j(z) = 1728\frac{g_2^3(z)}{\Delta(z)}$$

is holomorphic on $\mathcal{H}$. This function is called the *modular invariant* for reasons which will be explained later (we will explain also the normalizing term 1728) and it is a surjection.

By using these facts, we can prove the following result.

Theorem 1.4.1. *(Uniformization Theorem).*

1. *Let Λ be a lattice. Then the equation $y^2 = 4x^3 - g_2(\Lambda) - g_3(\Lambda)$ is non-singular (i.e. $g_2^3(\Lambda) - 27g_3^2(\Lambda) \neq 0$) and defines an elliptic curve*

$E(\mathbb{C})$. Moreover, the map 1.4

$$\begin{aligned}\Phi\colon \mathbb{C}/\Lambda &\longrightarrow E(\mathbb{C})\\ 0+\Lambda &\mapsto (0:1:0)\\ z+\Lambda &\mapsto (\wp(z):\wp'(z):1)\end{aligned}$$

is a complex analytic isomorphism of Abelian groups.

2. *Let $E : y^2 = 4x^3 - a_2x - a_3$ be an elliptic curve over $\mathbb{C}$. Then there exists a lattice $\Lambda \subseteq \mathbb{C}$ such that $a_2 = g_2(\Lambda)$, $a_3 = g_3(\Lambda)$ and $\mathbb{C}/\Lambda$ is isomorphic to $E(\mathbb{C})$ via Φ.*

Proof. We have already seen the first statement of the theorem. So we want to prove the second part. We start by considering the case $a_2 \neq 0$ and $a_3 \neq 0$. Since the modular invariant j surject, there exists $z \in \mathcal{H}$ such that

$$j(z) = 1728\frac{a_2^3}{a_2^3 - 27a_3^2}.$$

This gives

$$\frac{a_2^3}{a_2^3 - 27a_3^2} = \frac{g_2^3(z)}{g_2^3(z) - 27g_3^2(z)}$$

and after taking reciprocals

$$\frac{a_2^3}{g_2^3(z)} = \frac{a_3^2}{g_3^2(z)}.$$

For any $\omega_2 \in \mathbb{C}^*$, if $\omega_1 = z\omega_2$ and $\Lambda = [\omega_1, \omega_2]$, then $g_2(\Lambda) = \omega_2^{-4}g_2(z)$ and $g_3(\Lambda) = \omega_2^{-6}g_3(z)$. If we can choose ω_2 such that

$$\omega_2^{-4} = \frac{a_2}{g_2(z)} \qquad \text{and} \qquad \omega_2^{-6} = \frac{a_3}{g_3(z)},$$

we are done. Choose ω_2 to satisfy the first condition. We have

$$\omega_2^{-12} = \frac{a_2^3}{g_2^3(z)} = \frac{a_3^2}{g_3^2(z)}$$

and so

$$\omega_2^{-6} = \pm\frac{a_3}{g_3(z)}.$$

If we have the minus sign, then change ω_2 to $i\omega_2$. This preserves the relation $\omega_2^{-4} = a_2/g_2(z)$ and also yields $\omega_2^{-6} = a_3/g_3(z)$.

Now if $a_2 = 0$, then there exists $z \in \mathcal{H}$ such that $j(z) = 0$ and hence $g_2(z) = 0$. Since $\Delta(z) \neq 0$ it follows that $g_3(z) = \alpha \neq 0$. Let $\Lambda = [\omega_1, \omega_2]$ with $\omega_1 = z\omega_2$ as above. We have $g_3(\Lambda) = \omega_2^{-6} g_3(z) = \omega_2^{-6}\alpha$ and so we can choose ω_2 to satisfy $\omega_2^{-6}\alpha = a_3$ (note that $a_3 \neq 0$ because the curve is non-singular). Then $g_2(\Lambda) = \omega_2^{-4} g_2(z) = 0 = a_2$ as required. Similarly for the case $a_3 = 0$. □

Let $\wp_\Lambda$ denote the Weierstrass function with respect to Λ. An analytic group isomorphism of complex tori takes the form $z+\Lambda \mapsto mz+m\Lambda$, for some $m \in \mathbb{C}$. It is easy to see that $\wp_{m\Lambda}(mz) = m^{-2}\wp_\Lambda(z)$ and $\wp'_{m\Lambda}(mz) = m^{-3}\wp'_\Lambda(z)$, so the corresponding isomorphism of elliptic curves is $(x, y) \mapsto (m^{-2}x, m^{-3}y)$. This last isomorphism goes from the elliptic curve given by the equation $y^2 = 4x^3 + a_2x + a_3$ to one given by the equation $y^2 = 4x^3 + m^{-4}a_2x + m^{-6}a_3$.

The Uniformization Theorem allows us to treat complex analytic objects, like complex tori, and algebraic objects, like elliptic curves, in the same manner. From now on, the term “(complex) elliptic curve” will be used as synonym for “complex torus”.

1.5 Isogenies and Points of Finite Order

Recall from section 1.3 that an isogeny is a non-zero analytic group homomorphism between two elliptic curves. We have also seen that an isogeny always surjects and has a finite kernel.

Let N be a positive integer and let $E = E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ be an elliptic curve. We consider the map given by multiplication by N

$$[N] : E \to E, \qquad z + \Lambda \mapsto Nz + \Lambda.$$

Since $N\Lambda \subset \Lambda$, this is an isogeny (but not an isomorphism). The kernel of this map is denoted by $E[N]$ and it is the subgroup of N-torsion points, i.e.

$$E[N] = \{z + \Lambda \in E \mid Nz + \Lambda = 0 + \Lambda\}.$$

It is easy to see from the representation $E \cong \mathbb{C}/\Lambda$ that

$$E[N] \cong \frac{\mathbb{Z}}{N\mathbb{Z}} \times \frac{\mathbb{Z}}{N\mathbb{Z}}.$$

The inverse image of these points in $\mathbb{C}$ occur as the points of the lattice $\frac{1}{N}\Lambda$, hence their inverse image in $\mathbb{C}/\Lambda$ is the subgroup

$$\frac{1}{N}\Lambda/\Lambda \subseteq \mathbb{C}/\Lambda.$$

If $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ via $\psi : \mathbb{C} \to E(\mathbb{C})$ and $\Lambda = [\omega_1, \omega_2]$, then the elements

$$p_1 = \psi(\frac{\omega_1}{N}) \qquad \text{and} \qquad p_2 = \psi(\frac{\omega_2}{N})$$

form a basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, i.e. $E[N] = \langle p_1 + \Lambda\rangle \oplus \langle p_2 + \Lambda\rangle$.

Now, let K be a cyclic subgroup of $E[N]$ isomorphic to $\mathbb{Z}/N\mathbb{Z}$. Then as a set, K forms a super-lattice of Λ. The *cyclic quotient map*

$$\pi : \mathbb{C}/\Lambda \to \mathbb{C}/K, \qquad z + \Lambda \mapsto z + K$$

is an isogeny with kernel K. One can see that every isogeny is a composition of the examples above. We will not prove this result. For more details about isogenies see [DS05] or [Lan87].

CHAPTER 2

Modular Group and Modular Curves

2.1 Modular Group and Congruence Subgroups

Definition 2.1. The **modular group** is the group of 2×2 matrices with integer entries and determinant 1,

$$SL_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}) \mid ad - bc = 1 \right\}.$$

The modular group acts on $\mathcal{H}$ through fractional linear transformations as follows:

$$Mz = \frac{az+b}{cz+d}, \qquad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}), \quad z \in \mathbb{C}.$$

It is easy to see that

$$Im(Mz) = \frac{Im(z)}{|cz+d|^2},$$

so $SL_2(\mathbb{Z})$ maps the upper half plane back to itself. Moreover $Iz = z$ (where I is the identity matrix) and $(M_1M_2)z = M_1(M_2z)$, for all $M_1, M_2 \in SL_2(\mathbb{Z})$. This shows that $SL_2(\mathbb{Z})$ acts on $\mathcal{H}$.

We will see that the modular group is generated by the two matrices

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

which correspond to the transformations $z \mapsto z+1$ and $z \mapsto -1/z$.

Definition 2.2. Let N be a positive integer. The **principal congruence subgroup of level** N is

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

(The matrix congruence is interpreted entrywise.)

A subgroup Γ of $SL_2(\mathbb{Z})$ is a **congruence subgroup of level** N if $\Gamma(N) \subseteq \Gamma$ for some $N \in \mathbb{Z}^+$.

The subgroup $\Gamma(N)$ is normal in $SL_2(\mathbb{Z})$. In fact it is the kernel of the natural homomorphism $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/N\mathbb{Z})$. Moreover the map is surjective, so induces an isomorphism $SL_2(\mathbb{Z})/\Gamma(N) \cong SL_2(\mathbb{Z}/N\mathbb{Z})$. This show that the index $[SL_2(\mathbb{Z}) : \Gamma(N)]$ is finite for all N. In particular we have the following exact sequence of Abelian groups:

$$0 \longrightarrow \Gamma(N) \longrightarrow SL_2(\mathbb{Z}) \longrightarrow SL_2(\mathbb{Z}/N\mathbb{Z}) \longrightarrow 0$$

The most important congruence subgroups are

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\}$$

and

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

satisfying

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

The map

$$\Gamma_1(N) \to \frac{\mathbb{Z}}{N\mathbb{Z}}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto b \pmod{N}$$

is a surjection with kernel $\Gamma(N)$. Therefore $\Gamma(N) \trianglelefteq \Gamma_1(N)$ and $\Gamma_1(N)/\Gamma(N) \cong \mathbb{Z}/N\mathbb{Z}$. In particular $[\Gamma_1(N) : \Gamma(N)] = N$.

Similarly the map

$$\Gamma_0(N) \to \left(\frac{\mathbb{Z}}{N\mathbb{Z}}\right)^*, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \pmod N$$

is a surjection with kernel $\Gamma_1(N)$. So $\Gamma_1(N) \trianglelefteq \Gamma_0(N)$ and $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^*$. In this case the index is $[\Gamma_0(N) : \Gamma_1(N)] = \varphi(N)$, where φ is the Euler totient function. One can compute that

$$[SL_2(\mathbb{Z}) : \Gamma(N)] = N^3 \prod_{p|N} \left(1 - \frac{1}{p^2}\right)$$

and since

$$\varphi(N) = N \prod_{p|N} \left(1 - \frac{1}{p}\right)$$

it follows that

$$[SL_2(\mathbb{Z}) : \Gamma_0(N)] = N \prod_{p|N} \left(1 + \frac{1}{p}\right).$$

2.2 Modular Group and the Upper Half Plane

The results in the preceding chapter tell us that every $\tau \in \mathcal{H}$ determinates an elliptic curve $E(\mathbb{C}) \cong \mathbb{C}/\Lambda_\tau$. However the choice of τ is not unique, but $E(\mathbb{C}) \cong \mathbb{C}/\Lambda_\tau \cong \mathbb{C}/\Lambda_{\tau'}$ if and only if there exists a matrix $M \in SL_2(\mathbb{Z})$ such that $\tau' = M\tau$. This fact motivates the definition of an equivalence relation between elements on $\mathcal{H}$ under the action of $SL_2(\mathbb{Z})$.

Definition 2.3. We say that two points $\tau,\ \tau' \in \mathcal{H}$ are **equivalent relative to the modular group** $SL_2(\mathbb{Z})$ or $SL_2(\mathbb{Z})$**-equivalent** if there exists a matrix $M \in SL_2(\mathbb{Z})$ such that $\tau' = M\tau$. We write $\tau \sim \tau'$.

"$\sim$" defines an equivalence relation and the set of all equivalence classes is denoted by

$$Y(1) = \mathcal{H}/SL_2(\mathbb{Z}) \quad \text{or} \quad Y(1) = SL_2(\mathbb{Z}) \setminus \mathcal{H},$$

using the second notation to indicate that $SL_2(\mathbb{Z})$ acts on $\mathcal{H}$ on the left. So $Y(1)$ is the quotient space of orbit under $SL_2(\mathbb{Z})$. Since $M\tau = (-M)\tau$, sometimes the equivalence relation is defined with respect to the quotient $PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$. It is easy to see that this group acts faithfully on $\mathcal{H}$.

Next theorem shows what is a *fundamental domain* for $\mathcal{H}/SL_2(\mathbb{Z})$, i.e. a closed subset $\mathcal{F} \subseteq \mathcal{H}$ such that:

- If $z \in \mathcal{H}$, then there exists $z' \in \mathcal{F}$ such that $z \sim z'$;
- If $z, z' \in Int(\mathcal{F})$ are two distinct elements, then $z \nsim z'$.

(Here $Int(\mathcal{F})$ denotes the interior of $\mathcal{F}$.)

Theorem 2.2.1. *A fundamental domain for the quotient $\mathcal{H}/SL_2(\mathbb{Z})$ is*

$$\mathcal{F} = \{z \in \mathbb{C} \mid |z| \geq 1, -\frac{1}{2} \leq Re(z) \leq \frac{1}{2}\}.$$

Moreover two distinct points z_1, z_2 on the boundary of $\mathcal{F}$ are equivalent relative to $SL_2(\mathbb{Z})$ if and only if $Re(z_1) = \pm\frac{1}{2}$ and $z_2 = z_1 \mp 1$ or if $|z_1| = 1$ and $z_2 = -\frac{1}{z_1}$.

Proof. Let Γ be the subgroup of $SL_2(\mathbb{Z})$ generated by

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

Note that $T^{\pm n}z = z \pm n$, so we can translate a given $z \in \mathcal{H}$ in such a way that its real part lies in the interval $[-1/2, 1/2]$. Moreover, since $Im(Sz) = \frac{Im(z)}{|z|^2}$, if $|z| < 1$, then $Im(Sz) > Im(z)$.

Using the formula for the transformation of the imaginary part under $SL_2(\mathbb{Z})$, we can easily show that the imaginary parts of the elements in an orbit of a given $z \in \mathcal{H}$ are bounded from above, and tend to 0 as $\max\{|c|, |d|\} \to \infty$. Therefore, there exists an element w whose imaginary part is maximal in the orbit of z. Note that if $|w| < 1$, then $Im(Sw) > Im(w)$, so we must have $|w| \geq 1$. Now, since $Im(Tw) = Im(w)$, we can obtain an element in $\mathcal{F}$ by applying $T^{\pm n}$ for a suitable choice of n. This shows that if $z \in \mathcal{H}$, then there exists $z' \in \mathcal{F}$ such that $z \sim z'$.

Now we prove that if $z, z' \in \mathcal{F}$ are two distinct elements, then $z \nsim z'$ or z, z' lies on the boundary of $\mathcal{F}$. If $z \sim z'$, then there exists a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}), \qquad M \neq \pm I$$

such that $Mz = z'$. We may assume that $Im(z') \geq Im(z)$ and $c \geq 0$ (after multiplying by $S^2 = -I$, if necessary). We have $|cz + d| \leq 1$ and since $Im(z) \geq \frac{\sqrt{3}}{2}$, it follows that $c = 0$ or $c = 1$.

If $c = 0$, then

$$M = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = T^b \qquad \text{and} \qquad z' = z + b,$$

and it occurs if and only if $Re(z) = \pm\frac{1}{2}$ and $b = \mp 1$.

If $c = 1$, then $d = 0$ or $d = \pm 1$. If $d = 0$, then

$$M = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix} = T^a S \quad \text{and} \quad Mz = a - \frac{1}{z}.$$

In this case, since $|cz + d| = |z| \leq 1$, we have $|z| = 1$ and therefore also $Sz = -\frac{1}{z}$ lies on the unit circle. It follows that $a = 0$ and $z' = Sz$, or if $a \neq 0$, then only two cases occur (and both fix z): $z = \mu_3 = e^{2\pi i/3}$ and $a = -1$ or $z = -\bar{\mu}_3$ and $a = 1$.

If $d = \pm 1$, we have

$$M = \begin{pmatrix} \pm a & a - 1 \\ 1 & \pm 1 \end{pmatrix} = T^{\pm a} S T^{\pm 1} \quad \text{and} \quad Mz = \pm a - \frac{1}{z \pm 1}.$$

Since $|z + d| \leq 1$ and $|z + d|^2 = (Re(z) + d)^2 + Im(z)^2$ it follows that $Re(z) = \mp\frac{1}{2}$ and $Im(z) = \frac{\sqrt{3}}{2}$, because $Im(z) \geq \frac{\sqrt{3}}{2}$ and $-\frac{1}{2} \leq Re(z) \leq \frac{1}{2}$. Therefore we can conclude that if $d = 1$, then $z = \mu_3$ and necessarily $a = 0$ or 1, else if $d = -1$, then $z = -\bar{\mu}_3$ and $a = 0$ or -1.

We have already proven that if $z, z' \in \mathcal{F}$ are two distinct points and $z \sim z'$, then $Re(z) = \pm\frac{1}{2}$ and $z' = z \mp 1$ or $|z| = 1$ and $z' = Sz$. In particular if z, z' lies in the interior of $\mathcal{F}$, then $z = z'$ or $z \nsim z'$. So our theorem is proved. $\square$

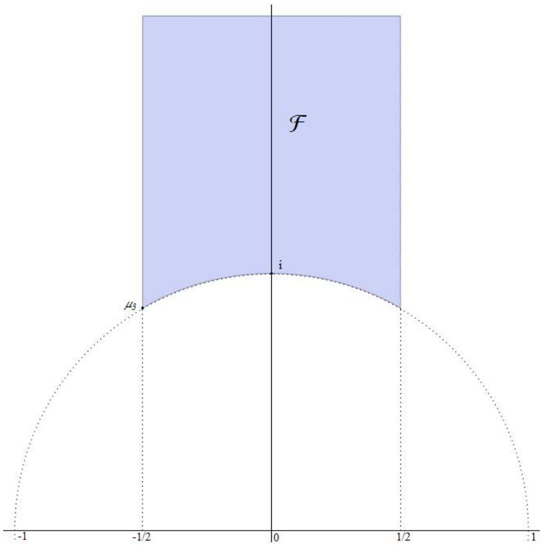

Figure 2.1: Fundamental domain for $SL_2(\mathbb{Z})$

A remarkable fact follows from the proof of the theorem: if z, z' lies in $\mathcal{F}$ and $z \sim z'$ then there exists $M \in SL_2(\mathbb{Z})$ such that $z' = Mz$, but in fact $M \in \Gamma = \langle S, T \rangle$. It follows the useful result

Corollary 2.2.2. *$SL_2(\mathbb{Z})$ is generated by*

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

Proof. Let $A \in SL_2(\mathbb{Z})$ and $z \in Int(\mathcal{F})$. Then there exists $M \in \Gamma = \langle S, T \rangle$ such that $z' = MAz \in \mathcal{F}$. Therefore $z' \sim z$ and since z is in the interior of $\mathcal{F}$, we must have $z = z'$. By the proof of the previous theorem, it follows that $MA = \pm I \in \Gamma$, hence $A \in \Gamma$. □

Now let $\Gamma \leq SL_2(\mathbb{Z})$ be a subgroup of finite index n. Write $SL_2(\mathbb{Z}) = \bigcup_{j=1}^{n} \Gamma A_j$ as a disjoint union of cosets. Then, if $A_j\mathcal{F} = \{A_j z \mid z \in \mathcal{F}\}$ denotes the obvious set, it is not difficult to see that the set

$$\mathcal{F}' = \bigcup_{i=1}^{n} A_j \mathcal{F}$$

is a fundamental domain for the orbit space $\Gamma \setminus \mathcal{H}$.

2.3 Modular Curves and Moduli Spaces

In the previous sections we have shown that each class-isomorphism of elliptic curves corresponds to a point on the fundamental domain $\mathcal{F}$ for $Y(1)$. Our purpose is to generalize this fact: the quotients of the upper half plane by congruence subgroups can be described by the sets of equivalence classes of elliptic curves enhanced by corresponding torsion data.

Definition 2.4. Let N be a positive integer.

An **enhanced elliptic curve for** $\Gamma_0(N)$ is an ordered pair $(E(\mathbb{C}), K)$, where $E(\mathbb{C})$ is an elliptic curve and K is a subgroup of $E[N]$ isomorphic to $\mathbb{Z}/N\mathbb{Z}$.

Two such pairs $(E(\mathbb{C}), K)$ and $(E'(\mathbb{C}), K')$ are equivalent if some isomorphism $E(\mathbb{C}) \to E'(\mathbb{C})$ maps K into K'. An equivalence class is denoted by $[(E(\mathbb{C}), K)]$ and the set of equivalence classes is denoted by $S_0(N)$.

An **enhanced elliptic curve for** $\Gamma_1(N)$ is an ordered pair $(E(\mathbb{C}), Q)$, where $E(\mathbb{C})$ is an elliptic curve and Q is a point of $E[N]$ of order exactly N.

Two such pairs $(E(\mathbb{C}), Q)$ and $(E'(\mathbb{C}), Q')$ are equivalent if some isomorphism $E(\mathbb{C}) \to E'(\mathbb{C})$ maps Q into Q'. An equivalence class is denoted by $[(E(\mathbb{C}), Q)]$ and the set of equivalence classes is denoted by $S_1(N)$.

An **enhanced elliptic curve for** $\Gamma(N)$ is an ordered triple $(E(\mathbb{C}), P, Q)$, where $E(\mathbb{C})$ is an elliptic curve and P and Q are two points that generates $E[N]$.

Two such triples $(E(\mathbb{C}), P, Q)$ and $(E'(\mathbb{C}), P', Q')$ are equivalent if some isomorphism $E(\mathbb{C}) \to E'(\mathbb{C})$ maps P into P' and Q to Q'. An equivalence class is denoted by $[(E(\mathbb{C}), P, Q)]$ and the set of equivalence classes is denoted by $S(N)$.

$S_0(N), S_1(N)$ and $S(N)$ are called **moduli spaces** or **spaces of moduli**.

Note that when $N = 1$, all N-torsion data are trivial and all $S_0(N), S_1(N)$ and $S(N)$ correspond to the isomorphism classes of complex elliptic curves.

Definition 2.5. Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. The **modular**

curve $Y(\Gamma)$ for Γ is defined as

$$Y(\Gamma) = \mathcal{H}/\Gamma = \Gamma \setminus \mathcal{H} = \{\Gamma z \mid z \in \mathcal{H}\}.$$

The modular curves for $\Gamma_0(N), \Gamma_1(N)$ and $\Gamma(N)$ are denoted by

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

Theorem 2.3.1. *Let N be a positive integer and let E_τ denote the elliptic curve $E(\mathbb{C}) \cong \mathbb{C}/\Lambda_\tau$, where $\tau \in \mathcal{H}$.*

1. *The moduli space for $\Gamma_0(N)$ is*

$$S_0(N) = \{[E_\tau, \left\langle \frac{1}{N} + \Lambda_\tau \right\rangle] \mid \tau \in \mathcal{H}\}.$$

Two points $[E_\tau, \langle \frac{1}{N} + \Lambda_\tau \rangle]$ and $[E_{\tau'}, \langle \frac{1}{N} + \Lambda_{\tau'} \rangle]$ are equal if and only if $\Gamma_0(N)\tau = \Gamma_0(N)\tau'$. Thus there is a bijection

$$\psi_0 : S_0(N) \to Y_0(N), \quad [E_\tau, \left\langle \frac{1}{N} + \Lambda_\tau \right\rangle] \mapsto \Gamma_0(N)\tau.$$

2. *The moduli space for $\Gamma_1(N)$ is*

$$S_1(N) = \{[E_\tau, \frac{1}{N} + \Lambda_\tau] \mid \tau \in \mathcal{H}\}.$$

Two points $[E_\tau, \frac{1}{N} + \Lambda_\tau]$ and $[E_{\tau'}, \frac{1}{N} + \Lambda_{\tau'}]$ are equal if and only if $\Gamma_1(N)\tau = \Gamma_1(N)\tau'$. Thus there is a bijection

$$\psi_1 : S_1(N) \to Y_1(N), \quad [E_\tau, \frac{1}{N} + \Lambda_\tau] \mapsto \Gamma_1(N)\tau.$$

3. *The moduli space for $\Gamma(N)$ is*

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

Two points $[E_\tau, \frac{\tau}{N} + \Lambda_\tau, \frac{1}{N} + \Lambda_\tau]$ and $[E_{\tau'}, \frac{\tau'}{N} + \Lambda_{\tau'}, \frac{1}{N} + \Lambda_{\tau'}]$ are equal if and only if $\Gamma(N)\tau = \Gamma(N)\tau'$. Thus there is a bijection

$$\psi : S(N) \to Y(N), \quad [E_\tau, \frac{\tau}{N} + \Lambda_\tau, \frac{1}{N} + \Lambda_\tau] \mapsto \Gamma(N)\tau.$$

Proof. We will prove only part (2). The other two parts can be shown in a similar way. Let $[E, Q]$ be a point of $S_1(N)$. We may assume $E \cong E_{\tau'}$ for some $\tau' \in \mathcal{H}$. Thus $Q = (c\tau' + d)/N$ for some $c, d \in \mathbb{Z}$ with $GCD(c, d, N) = 1$, since Q has order N. So we can write $ad - bc - kN = 1$, for some $a, b, k \in \mathbb{Z}$ and the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is in $SL_2(\mathbb{Z}/N\mathbb{Z})$. Note that modifying the entries of M mod N doesn't affect Q, and hence, since the canonical homomorphism $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/N\mathbb{Z})$ is surjective, we may assume $M \in SL_2(\mathbb{Z})$. Let $\tau = M\tau'$ and $m = c\tau' + d$. Then

$$m\Lambda_\tau = m(\tau\mathbb{Z} \oplus \mathbb{Z}) = (a\tau' + b)\mathbb{Z} \oplus (c\tau' + d)\mathbb{Z} = \tau'\mathbb{Z} \oplus \mathbb{Z} = \Lambda'_\tau.$$

and

$$m(\frac{1}{N} + \Lambda_\tau) = \frac{c\tau' + d}{N} + \Lambda_{\tau'} = Q,$$

so $[E, Q] = [E_\tau, 1/N + \Lambda_\tau]$, where $\tau \in \mathcal{H}$.

Now suppose that $\tau, \tau' \in \mathcal{H}$ are two points such that $M\tau = \tau'$ for some $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$. Let $m = c\tau' + d$. Then $m\Lambda_\tau = \Lambda_{\tau'}$ and

$$m(\frac{1}{N} + \Lambda_\tau) = \frac{c\tau' + d}{N} + \Lambda_{\tau'} = \frac{1}{N} + \Lambda_{\tau'}$$

using $(c, d) \equiv (0, 1)$ mod N for the second equality. So $[E_\tau, 1/N + \Lambda_\tau] = [E_{\tau'}, 1/N + \Lambda_{\tau'}]$.

Conversely, suppose $[E_\tau, 1/N + \Lambda_\tau] = [E_{\tau'}, 1/N + \Lambda_{\tau'}]$, with $\tau, \tau' \in \mathcal{H}$. Then $m\Lambda_\tau = \Lambda_{\tau'}$ for some $m \in \mathbb{C}$ and $m(1/N + \Lambda_\tau) = 1/N + \Lambda_{\tau'}$. So

$$\begin{pmatrix} m\tau \\ m \end{pmatrix} = M \begin{pmatrix} \tau' \\ 1 \end{pmatrix}, \qquad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

and in particular $m = c\tau' + d$. Moreover

$$\frac{c\tau' + d}{N} + \Lambda_{\tau'} = \frac{1}{N} + \Lambda_{\tau'},$$

hence $(c, d) \equiv (0, 1)$ mod N and $M \in \Gamma_1(N)$. It follows that $\Gamma_1(N)\tau = \Gamma_1(N)\tau'$. □

2.4 Cusps

We return to the fundamental domain $\mathcal{F}$ for the modular group $SL_2(\mathbb{Z})$. In order to describe $Y(1)$ it is useful to identify $SL_2(\mathbb{Z})$-equivalent points on the boundary of $\mathcal{F}$. Visually, we fold $\mathcal{F}$ around the imaginary axis, gluing the point $-\frac{1}{2}+iy$ to $\frac{1}{2}+iy$ and the point $e^{2\pi i\theta}$ to $e^{2\pi i(1/2-\theta)}$ for $y \geq \frac{\sqrt{3}}{2}$ and $\frac{1}{6} \leq \theta \leq \frac{1}{3}$. The result is a surface that is homeomorphic to a sphere with one point missing. Our aim is to add this "missing point" to $Y(1)$. Actually, in this section we add several points to $Y(\Gamma)$ for any congruence subgroup $\Gamma \leq SL_2(\mathbb{Z})$. These points are called "**cusps**" for geometric reasons to be explained later. In this section we will discuss only about the formal construction of these points.

We start by defining the *extended half plane* as the union of $\mathcal{H}$ and a copy of the projective line over $\mathbb{Q}$, i.e.

$$\mathcal{H}^* = \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q}).$$

The set $\{(s:1) \mid s \in \mathbb{Q}\}$ can be identified with $\mathbb{Q}$, while the point $(1:0)$ is the point at infinity, which we identify with "$i\infty$". We can extend the action of $SL_2(\mathbb{Z})$ on all $\mathcal{H}^*$. To do this, let $M \in SL_2(\mathbb{Z})$ and $(s:t) \in \mathbb{P}^1(\mathbb{Q})$. We define

$$M(s:t) = (as+bt : cs+dt) \in \mathbb{P}^1(\mathbb{Q}), \qquad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This action is consistent with the previous definition of the action of $SL_2(\mathbb{Z})$ in $\mathcal{H}$. In fact, by the identification $(s:1) = s \in \mathbb{Q}$, if $cs+d \neq 0$ we obtain

$$M(s:1) = Ms = (as+b : cs+d) = (\frac{as+b}{cs+d} : 1) = \frac{as+b}{cs+d} \in \mathbb{Q},$$

while, if $cs+d=0$, we obtain $M(s:1) = (1:0) = i\infty$ and similarly, for the point at infinity we have

$$M(1:0) = (a:c) = \begin{cases} \frac{a}{c} \in \mathbb{Q} & \text{if } c \neq 0 \\ \infty & \text{if } c = 0. \end{cases}$$

Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Two points $z, z' \in \mathcal{H}^*$ are equivalent relative to Γ if there exists a matrix $M \in \Gamma$ such that $z = Mz'$. Clearly this is an equivalence relation and the set of all equivalence classes is denoted by

$$X(\Gamma) = \mathcal{H}^*/\Gamma = \Gamma \setminus \mathcal{H}^*.$$

For the congruence subgroups $\Gamma_0(N),\ \Gamma_1(N)$ and $\Gamma(N)$ we write respectively $X_0(N),\ X_1(N)$ and $X(N)$. In particular $X(1) = X(SL_2(\mathbb{Z}))$.

Definition 2.6. Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. The quotient $X(\Gamma) = \mathcal{H}^*/\Gamma$ is called the **modular curve** for Γ.

The elements in $X(\Gamma)$ that have a representative in $\mathbb{P}^1(\mathbb{Q})$, i.e. the points Γz with $z \in \mathbb{P}^1(\mathbb{Q})$, are called the **cusps**.

As we will see later in this chapter, for any congruence subgroup $\Gamma \leq SL_2(\mathbb{Z})$, the modular curve $X(\Gamma)$ is the *compactification* of $Y(\Gamma)$ and it turns out to be a compact Riemann surface. We note that in order to compactify $Y(\Gamma)$, we just added only finitely many points. In fact we have the following

Lemma 2.4.1. *The modular curve $X(1)$ has one cusp. For any congruence subgroup Γ of $SL_2(\mathbb{Z})$, the modular curve $X(\Gamma)$ has finitely many cusps.*

Proof. For $X(1)$ it suffices to show that if $(s:t) = \mathbb{P}^1(\mathbb{Q})$, then there exists a matrix $M \in SL_2(\mathbb{Z})$ such that $M(1:0) = (s:t)$. We also may assume that $s, t \in \mathbb{Z}$ and $GCD(s,t) = 1$. So we can write $sd - tb = 1$ for some $b, d \in \mathbb{Z}$. Defining $M = \left(\begin{smallmatrix} s & b \\ t & d \end{smallmatrix}\right)$, we obtain $M(1:0) = (s:t)$, as required. This shows that $SL_2(\mathbb{Z})$ acts transitively on $\mathbb{P}^1(\mathbb{Q})$.

Now let $\Gamma \leq SL_2(\mathbb{Z})$ be a congruence subgroup of index m. We can write

$$SL_2(\mathbb{Z}) = \bigcup_{j=i}^{m} \Gamma M_j, \qquad M_j \in SL_2(\mathbb{Z}).$$

Let $\tau \in \mathbb{P}^1(\mathbb{Q})$ be a fixed point. By the previous part, for any $\omega \in \mathbb{P}^1(\mathbb{Q})$ there exist a matrix $M \in \Gamma$ and an index j such that $\omega = MM_j\tau$, hence $\omega \sim_\Gamma M_j\tau$, where $\sim_\Gamma$ means "equivalent relative to Γ" (i.e. in the same orbit under Γ). So, if we put $T_j = \{\omega \in \mathbb{P}^1(\mathbb{Q}) \mid \omega \sim_\Gamma M_j\tau\}$, we obtain:

$$\mathbb{P}^1(\mathbb{Q}) = \bigcup_{j=1}^{m} T_j$$

and the result follows. □

2.5 Elliptic Points

Definition 2.7. Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$ and $z \in \mathcal{H}^*$. The **isotropy subgroup** of z is the subgroup Γ_z of Γ defined by

$$\Gamma_z = \{M \in \Gamma \mid Mz = z\}.$$

A point $z \in \mathcal{H}$ is an **elliptic point** for Γ if the containment $\{\pm I\} \subsetneq \{\pm I\}\Gamma_z$ is proper. The corresponding point $\Gamma z \in X(\Gamma)$ is also called elliptic.

Note that since $-Iz = z$ for all $z \in \mathcal{H}^*$, a point z is elliptic if and only if its isotropy subgroup is non-trivial as a group of transformations. Moreover any element of Γ_z also fixes $\bar{z} \in -\mathcal{H}^*$

Let $z \in \mathcal{H}$ be an elliptic point for $SL_2(\mathbb{Z})$ and $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_z$ be a non-trivial transformation. Then $cz^2 + dz = az + b$ and $c \neq 0$ because $z \notin \mathbb{Q}$. It is easy to see that $|a + d| < 2$, so the characteristic polynomial of M is $X^2 + 1$ or $X^2 \pm X + 1$. Since M satisfies its characteristic polynomial, then M has order 3, 4 or 6 (orders 1 and 2 gives the trivial transformations $\pm I$).

Proposition 2.5.1. *Let $M \in SL_2(\mathbb{Z})$.*

1. *If M has order 3, then M is conjugate to $\left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)^{\pm 1}$ in $SL_2(\mathbb{Z})$;*
2. *If M has order 4, then M is conjugate to $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)^{\pm 1}$ in $SL_2(\mathbb{Z})$;*
3. *If M has order 6, then M is conjugate to $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)^{\pm 1}$ in $SL_2(\mathbb{Z})$.*

Proof. Note that if M has order 3, then $-M$ has order 6, so part (1) follows from (3).

Let M be an element of order 6 and consider the lattice $L = \mathbb{Z} \oplus \mathbb{Z}$. We endow L with a scalar product as follows. For any element $v \in L$ (which is an integral column vector), we define

$$(a + b\mu_6) \cdot v = (aI + bM)v \qquad \forall a, b \in \mathbb{Z}$$

(here $\mu_6 = e^{2\pi i/6}$ as usual). Then L is a module over the ring $\mathbb{Z}[\mu_6]$.

It is well known that $\mathbb{Z}[\mu_6]$ is a PID and by the structure theorem for modules over a PID, L is $\mathbb{Z}[\mu_6]$-isomorphic to a sum $\bigoplus_k \mathbb{Z}[\mu_6]/I_k$, where I_k are ideals. Since every non-zero ideal I has rank 2 as an Abelian group, the quotients $\mathbb{Z}[\mu_6]/I_k$ are torsion groups. But L is free of rank 2 as an Abelian group, so we obtain $\mathbb{Z}[\mu_6] \cong L$ as $\mathbb{Z}[\mu_6]$-modules. Let ϕ_M denotes this isomorphism and put $u = \phi_M(1)$ and $v = \phi_M(\mu_6)$. Then $L = u\mathbb{Z} \oplus v\mathbb{Z}$ and so $det((u\ v)) = \pm 1$. We have

$$Mu = \mu_6 \cdot \phi_M(1) = \phi_M(\mu_6) = v$$

and

$$Mv = \mu_6 \cdot \phi_M(\mu_6) = \phi_M(\mu_6^2) = \phi_M(\mu_6 - 1) = v - u.$$

Thus

$$M(u\ v) = (v\ v - u) = (u\ v)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$$

and

$$M(v\ u) = (v - u\ v) = (v\ u)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)^{-1}.$$

Since one of $(u\ v)$ or $(v\ u)$ is in $SL_2(\mathbb{Z})$, part (3) follows.

If M has order 4, the lattice L is a $\mathbb{Z}[i]$-module and as above there exists a $\mathbb{Z}[i]$-module isomorphism $\phi_M : \mathbb{Z}[i] \to L$. Let $u = \phi_M(1)$ and $v = \phi_M(i)$. This time we have

$$Mu = i \cdot \phi_M(1) = \phi_M(i) = v \quad \text{and} \quad Mv = i \cdot \phi(i) = \phi(-1) = -u$$

so

$$M(u\ v) = (v\ \ -u) = (u\ v)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$$

and

$$M(v\ u) = (-u\ v) = (v\ u)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)^{-1}.$$

Again one of $(u\ v)$ and $(v\ u)$ is in $SL_2(\mathbb{Z})$, proving (2). □

Corollary 2.5.2. *The elliptic points for $SL_2(\mathbb{Z})$ are $SL_2(\mathbb{Z})i$ and $SL_2(\mathbb{Z})\mu_3$, so the modular curve $Y(1)$ has two elliptic points. The isotropy subgroups of i and μ_3 are*

$$SL_2(\mathbb{Z})_i = \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\rangle \quad \textit{and} \quad SL_2(\mathbb{Z})_{\mu_3} = \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \right\rangle.$$

In particular, for each elliptic point z of $SL_2(\mathbb{Z})$, the isotropy subgroup $SL_2(\mathbb{Z})_z$ is finite cyclic.

Proof. Let z be an elliptic point for $SL_2(\mathbb{Z})$ and let $M \in SL_2(\mathbb{Z})_z$ be a non-trivial transformation. Then M is conjugate to T in $SL_2(\mathbb{Z})$, where T is one of the matrices in the previous proposition. Note that T fix i or μ_3. We can write $AM = TA$ for some $A \in SL_2(\mathbb{Z})$, so $T(Az) = (TA)z = AMz = Az$. It follows that T fix Az, hence $SL_2(\mathbb{Z})z = SL_2(\mathbb{Z})i$ or $SL_2(\mathbb{Z})\mu_3$. Since i and μ_3 are not equivalent under $SL_2(\mathbb{Z})$, the first statement follows.

The fact that $SL_2(\mathbb{Z})_i = \left\langle \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \right\rangle$ and $SL_2(\mathbb{Z})_{\mu_3} = \left\langle \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right) \right\rangle$ follows from the calculations in the proof of the Theorem 2.2.1.

The last statement is a consequence of the first part, since any isotropy subgroup of an elliptic point z is conjugate to $SL_2(\mathbb{Z})_i$ or $SL_2(\mathbb{Z})_{\mu_3}$. □

Corollary 2.5.3. *Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. The modular curve $Y(\Gamma)$ has finitely many elliptic points. For each elliptic point z of Γ, the isotropy subgroup Γ_z is finite cyclic.*

Proof. Write $SL_2(\mathbb{Z}) = \bigcup_{j=1}^{n} \Gamma M_j$ as a disjoint union of its cosets. Then the elliptic points of $Y(\Gamma)$ are a subset of the finite set $\{\Gamma M_j i, \Gamma M_j \mu_3\}_{1 \leq j \leq n}$. For the second statement it suffices to note that for any $z \in \mathcal{H}$, Γ_z is a subgroup of $SL_2(\mathbb{Z})_z$. □

Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. We have already seen that for any elliptic point z of Γ, the isotropy subgroup Γ_z is finite cyclic. Thus we can define for each point $z \in \mathcal{H}$ the **period** h_z of z, a positive integer defined as follows

$$h_z = |\{\pm I\}\Gamma_z/\{\pm I\}| = \begin{cases} |\Gamma_z|/2 & \text{if } -I \in \Gamma \\ |\Gamma_z| & \text{if } -I \notin \Gamma. \end{cases}$$

Let $M \in SL_2(\mathbb{Z})$. Since $-I \in \Gamma \Leftrightarrow -I \in M\Gamma M^{-1}$ and $|\Gamma| = |M\Gamma M^{-1}|$, the period of Mz under $M\Gamma M^{-1}$ is the same as the period of z under Γ. So h_z depends only on Γ_z, hence the period of a point on $Y(\Gamma)$ is well defined and counts the z-fixing transformations.

2.6 Modular Curves as Riemann Surfaces

We recall the definition of Riemann surface.

Definition 2.8. A **Riemann Surface** X is a connected Hausdorff space on which there is defined a complex structure S with the following properties:

1. S is a collection of pairs $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$, where $\{U_\alpha\}_{\alpha \in A}$ is an open covering of X and φ_α is a homeomorphism of U_α onto an open subset of $\mathbb{C}$;

2. If $U_\alpha \cap U_\beta \neq \emptyset$, then

$$\varphi_\beta \circ \ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$$

   is holomorphic.

Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. We start to show that $Y(\Gamma)$ is a Riemann surface. The map

$$\pi : \mathcal{H} \to Y(\Gamma), \qquad z \mapsto \Gamma z$$

is an open mapping from $\mathcal{H}$ with the Euclidean topology to $Y(\Gamma)$ endowed with the quotient topology. If $M \in SL_2(\mathbb{Z})$ and $U \subseteq \mathcal{H}$, we write $M(U)$ for the set $\{Mz \mid z \in U\}$ and $\Gamma(U)$ for the set $\{Mz \mid M \in \Gamma, z \in U\}$. It is easy to see that

$$\pi(V) \cap \pi(W) = \emptyset \Longleftrightarrow \Gamma(V) \cap W = \emptyset, \qquad \forall\ V, W \subseteq \mathcal{H}.$$

Since $\mathcal{H}$ is connected and π is continuous, $Y(\Gamma)$ is connected. Moreover $Y(\Gamma)$ is Hausdorff. This is a consequence of the following proposition.

Proposition 2.6.1. *Let z_1, z_2 be two points in $\mathcal{H}$. Then there exist open neighbourhoods U_1 of z_1 and U_2 of z_2 in $\mathcal{H}$ such that for all $M \in SL_2(\mathbb{Z})$*

$$Mz_1 \neq z_2 \Rightarrow M(U_1) \cap U_2 = \emptyset.$$

Proof. Let U_1' and U_2' be neighbourhoods of z_1 and z_2 respectively, with compact closures in $\mathcal{H}$. First we want to show that $M(U_1') \cap U_2' \neq \emptyset$ for only finitely many $M \in SL_2(\mathbb{Z})$. It is not hard to see that the condition

$$\sup\{Im(Mz) \mid M = \left(\begin{smallmatrix} * & * \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z}), z \in U_1'\} < \inf\{Im(z) \mid z \in U_2'\}$$

holds for all but finitely many integer pairs (c, d) with $GCD(c, d) = 1$. Moreover for any fixed pair (c, d) with $GCD(c, d) = 1$, we can write $ad - bc = 1$ for some $a, b \in \mathbb{Z}$ and the matrices $M \in SL_2(\mathbb{Z})$ with bottom row (c, d) are exactly those of the form

$$M = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{for } k \in \mathbb{Z}.$$

Thus $M(U_1') \cap U_2' = \left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(U_1') + k\right) \cap U_2' = \emptyset$ for all but finitely many M with fixed bottom row (c, d).

Now let

$$\Omega = \{M \in SL_2(\mathbb{Z}) \mid M(U_1') \cap U_2 \neq \emptyset, Mz_1 \neq z_2\}.$$

By the previous part Ω is a finite set. For each $M \in \Omega$ there exist disjoint neighbourhoods $U_{1,M}$ of Mz_1 and $U_{2,M}$ of z_2 in $\mathcal{H}$. Define

$$U_1 = U_1' \cap \left(\bigcap_{M \in \Omega} M^{-1}(U_{1,M}) \right)$$

and

$$U_2 = U_2' \cap \left(\bigcap_{M \in \Omega} U_{2,M} \right)$$

which are neighbourhoods of z_1 and z_2 respectively in $\mathcal{H}$. Now we have to prove that if $M \in SL_2(\mathbb{Z})$ is such that $M(U_1) \cap U_2 \neq \emptyset$, then $Mz_1 = z_2$. It suffices to show that $M \notin \Omega$. If $M \in \Omega$, then $U_1 \subset M^{-1}(U_{1,M})$ and $U_2 \subset U_{2,M}$, so $\emptyset \neq M(U_1) \cap U_2 \subset U_{1,M} \cap U_{2,M}$, but this is a contradiction.

This complete the proof. □

Corollary 2.6.2. *Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Then the modular curve $Y(\Gamma)$ is Hausdorff.*

Proof. Let $\pi(z_1)$ and $\pi(z_2)$ be two distinct points in $Y(\Gamma)$ and let U_1, U_2 neighbourhoods of z_1 and z_2 as in the previous proposition. Since $\pi(z_1)$ and $\pi(z_2)$ are distinct, it follows that $\Gamma(U_1) \cap U_2 = \emptyset$ and so $\pi(U_1)$ and $\pi(U_2)$ are disjoint neighbourhoods of $\pi(z_1)$ and $\pi(z_2)$. □

The next step is to define local coordinates on the modular curve $Y(\Gamma)$, i.e. to find for each point $\pi(z) \in Y(\Gamma)$ a neighbourhood U_z and a homeomorphism φ_z of U_z onto an open subset of $\mathbb{C}$ as in the previous definition.

Let's start with the simple case. Let $z \in \mathcal{H}$ be a point fixed only by $\pm I$. Since $Mz \neq z$ for all $M \in \Gamma \setminus \{\pm I\}$, by Proposition 2.6.1 there exists an open neighbourhood U of z such that $M(U) \cap U = \emptyset$ for all $M \in \Gamma \setminus \{\pm I\}$. Then $\pi|_U : U \to \pi(U)$ is an opening continuous surjective map, which is also injective. In fact, let $z_1, z_2 \in U$ such that $\pi(z_1) = \pi(z_2)$. Then there exists $M \in \Gamma$ such that $Mz_1 = z_2 \in M(U) \cap U$. But $M(U) \cap U \neq \emptyset$ if and only if $M = \pm I$, so $z_1 = z_2$. This shows that $\pi(U)$ is homeomorphic to U.

But an elliptic point $\pi(z)$ poses some problems, because any neighbourhood U of z may contain several Γ-equivalent points and thus $\pi|_U$ could not be a homeomorphism. By Proposition 2.6.1, each point $z \in \mathcal{H}$ has an open neighbourhood $U \subset \mathcal{H}$ such that for all $M \in \Gamma$, if $M(U) \cap U \neq \emptyset$ then $M \in \Gamma_z$. Moreover such a neighbourhood can be chosen to have no elliptic points except possibly z. Let $z \in \mathcal{H}$ be given, and $U \subset \mathcal{H}$ an open neighbourhood as above. We define the transformation

$$\delta_z = \begin{pmatrix} 1 & -z \\ 1 & -\bar{z} \end{pmatrix} \in GL_2(\mathbb{C}).$$

So $\delta_z z = 0$ and $\delta_z \bar{z} = \infty$. The isotropy subgroup of 0 in the conjugated transformation group, $(\delta_z \{\pm I\} \Gamma \delta_z^{-1})_0 / \{\pm I\}$, is the conjugate of the isotropy subgroup of z, $\delta_z(\{\pm I\}\Gamma_z / \{\pm I\})\delta_z^{-1}$, and therefore is cyclic of order h_z. Since this group consists of fractional linear transformations fixing 0 and ∞, its elements are of the form $z \mapsto az$, and these must be rotations through angular multiples of $2\pi/h_z$ about the origin. We define a map $\psi = \rho \circ \delta_z : U \to \mathbb{C}$, where ρ is the function $\rho(w) = w^{h_z}$. Let $V = \psi(U)$. Then V is an open subset of $\mathbb{C}$, since ψ is holomorphic. Put $\delta = \delta_z$ and $h = h_z$. For any two points $z_1, z_2 \in U$ we have

$$\pi(z_1) = \pi(z_2) \Longleftrightarrow z_1 \in \Gamma z_2 \Longleftrightarrow z_1 \in \Gamma_z z_2 \Longleftrightarrow$$

$$\delta z_1 \in (\delta\Gamma_z\delta^{-1})\delta z_2 \Longleftrightarrow \delta z_1 = \mu_h^d(\delta z_2) \text{ for some } d,$$

where $\mu_h = e^{2\pi i/h}$, and in the second step we have used the fact that if $z_1 \in \Gamma z_2$, then there exists an element $M \in \Gamma$ such that $M(U) \cap U \neq \emptyset$, hence $M \in \Gamma_z$. So

$$\pi(z_1) = \pi(z_2) \Leftrightarrow (\delta z_1)^h = (\delta z_2)^h \Leftrightarrow \psi(z_1) = \psi(z_2).$$

Thus there exists an injection $\varphi : \pi(U) \to V$ such that $\varphi \circ \pi = \psi$, which is a surjection because ψ surject. Moreover φ is a homeomorphism, because both ψ and π are open and continuous.

In sum, for every point $z \in \mathcal{H}$ there exists a neighbourhood $U = U_z$ with no elliptic points except possibly z such that for all $M \in \Gamma$, if $M(U) \cap U \neq \emptyset$ then $M \in \Gamma_z$. For such neighbourhoods we have the following commutative diagram,

$$\begin{tikzcd}
U \arrow[r, "\delta"] \arrow[d, "\pi"'] \arrow[dr, "\psi"] & \delta(U) \arrow[d, "\rho"] \\
\pi(U) \arrow[r, "\varphi"] & V
\end{tikzcd}$$

where φ is a homeomorphism between $\pi(U)$ and $V = \psi(U)$. Clearly the set $\{\pi(U_z)\}_{z\in\mathcal{H}}$ is an open covering of $Y(\Gamma)$. Now we have to verify that the condition in Definition 2.8 is satisfied.

Let z_1, z_2 be in $\mathcal{H}$. Denote with $U_i, \delta_i, \psi_i, \varphi_i, \rho_i$ and h_i the obvious objects as above. We need to check that the restriction of $\varphi_2 \circ \varphi_1^{-1}$ to $\pi(U_1) \cap \pi(U_2)$, that we denote with $\varphi_{2,1}$, is holomorphic. Put $V_{1,2} = \varphi_1(\pi(U_1) \cap \pi(U_2))$ and

$V_{2,1} = \varphi_2(\pi(U_1) \cap \pi(U_2))$. We have the following diagram

$$\begin{tikzcd}
& \pi(U_1) \cap \pi(U_2) \arrow[dr, "\varphi_2"] & \\
V_{1,2} \arrow[ur, "\varphi_1^{-1}"] \arrow[rr, "\varphi_{2,1}"] & & V_{2,1}.
\end{tikzcd}$$

Let $p = \varphi_1(x)$, where $x \in \pi(U_1) \cap \pi(U_2)$. So we can write $x = \pi(w_1) = \pi(w_2)$, with $w_1 \in U_1$, $w_2 \in U_2$ and $Mw_1 = w_2$ for some $M \in \Gamma$. Let $U_{1,2} = U_1 \cap M^{-1}(U_2)$. Then $\pi(U_{1,2})$ is a neighbourhood of x and so $\varphi_1(\pi(U_{1,2})) \subseteq V_{1,2}$ is a neighbourhood of p. We have to find an explicit representation for $\varphi_{1,2}$ in $\varphi_1(\pi(U_{1,2}))$ (note that it suffices to show the holomorphy in this neighbourhood).

For any fixed $q = \varphi_1(y) \in \varphi_1(\pi(U_{1,2}))$ we have

$$q = \varphi_1(\pi(w)) = \psi(w) = (\delta_1 w)^{h_1}$$

for some $w \in U_{1,2}$. Then we can write

$$\varphi_{2,1}(q) = \varphi_2(\varphi_1^{-1}(q)) = \varphi_2(\pi(w)) = \varphi_2(\pi(Mw)) = \psi_2(Mw) =$$

$$= (\delta_2(Mw))^{h_2} = ((\delta_2 M \delta_1^{-1})(\delta_1 w))^{h_2} = (\delta_2 M \delta_1^{-1}(q^{1/h_1}))^{h_2}$$

where $\psi_2(Mw)$ is defined since $Mw \in U_2$. It shows that $\varphi_{1,2}$ might not be holomorphic only when $h_1 > 1$, i.e. when z_1 is an elliptic point. Suppose that $w_1 = z_1$. Note that it is equivalent to say that $p = 0$, since $p = \varphi_1(x) = \varphi_1(\pi(w_1)) = \psi(w_1) = (\delta_1 w_1)^{h_1}$ and

$$(\delta_1 w_1)^{h_1} = 0 \Leftrightarrow \delta_1 w_1 = 0 \Leftrightarrow \delta_1 = \delta_{w_1} \Leftrightarrow z_1 = w_1.$$

In this case $w_2 = Mw_1$ is also elliptic, and since U_2 contains no elliptic points except z_2, it follows that $z_2 = w_2$ and so $h_1 = h_2$. Moreover

$$\delta_2 M \delta_1^{-1}(0) = 0 \qquad \text{and} \qquad \delta_2 M \delta_1^{-1}(\infty) = \infty$$

hence $\delta_2 M \delta_1^{-1}$ is of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$ for some $a, b \in \mathbb{C}^*$. In this case we obtain

$$\varphi_{1,2}(q) = \left(\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} (q^{1/h_1}) \right)^{h_2} = \left(\frac{a}{b}\right)^{h_2} q$$

and it shows that the map $\varphi_{1,2}$ is holomorphic. Note that this argument also covers the case $\varphi_2(x) = 0$ since the inverse of an holomorphic bijection is again holomorphic. If $p = \varphi_1(x) \neq 0$, consider $z_3 \in \mathcal{H}$ such that $\pi(z_3) = x$. Then $\varphi_3 : \pi(U_3) \to V_3$ maps x to 0 and by the previous part $\varphi_{2,3} = \varphi_2 \circ \varphi_3^{-1}$ and $\varphi_{3,1} = (\varphi_1 \circ \varphi_3^{-1})^{-1}$ are holomorphic. Since $\varphi_{1,2} = \varphi_{2,3} \circ \varphi_{3,1}$, the result follows for all cases.

As we have already said, to compactify the modular curve $Y(\Gamma)$ we need to consider several additional points, the cusps. Recall that $\mathcal{H}^*$ can be identified with $\mathcal{H} \cup \mathbb{Q} \cup \{\infty\}$ and that the modular curve $X(\Gamma)$ is defined as the extended quotient $\Gamma \setminus \mathcal{H}^* = Y(\Gamma) \cup (\Gamma \setminus (\mathbb{Q} \cup \{\infty\}))$. To put an appropriate topology on $X(\Gamma)$ we start by defining a suitable topology on $\mathcal{H}^*$. For any $C > 0$ consider the set

$$\mathcal{N}_C = \{z \in \mathcal{H} \mid Im(z) > C\}.$$

Adjoin to the usual open sets in $\mathcal{H}$, the sets

$$A(\mathcal{N}_C \cup \{\infty\}) \subset \mathcal{H}^*, \qquad C > 0,\ A \in SL_2(\mathbb{Z}).$$

These sets serve as a base of neighbourhoods of the cusps and correspond to disks tangent to the real axis at $A(\infty)$. We endow $\mathcal{H}^*$ with the resulting topology. Clearly under this topology each $M \in SL_2(\mathbb{Z})$ is a continuous map. Now we can consider $X(\Gamma)$ with the quotient topology given by the natural projection $\pi : \mathcal{H}^* \to X(\Gamma)$.

Theorem 2.6.3. *Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Then the modular curve $X(\Gamma)$ is Hausdorff, connected and compact.*

Proof. Let $\mathcal{F}$ denote the fundamental domain for $SL_2(\mathbb{Z})$ as in Theorem 2.2.1, i.e.

$$\mathcal{F} = \{z \in \mathcal{H} \mid -\frac{1}{2} \leq Re(z) \leq \frac{1}{2}, |z| \geq 1\}.$$

Recall that there are the following boundary identifications under $SL_2(\mathbb{Z})$: the translation $z \mapsto z + 1$ identifies the two boundary half-lines and the map $z \mapsto -1/z$ identifies the two halves of the boundary arc. This identifications are the only ones.

(*Compactness*) The set $\mathcal{F}^* = \mathcal{F} \cup \{\infty\}$ is compact in the $\mathcal{H}^*$ topology, since every open covering of $\mathcal{F}^*$ contains ∞ and the complementary of the sets

$\mathcal{N}_C \cup \{\infty\}$ in $\mathcal{F}^*$ are compact. Write $SL_2(\mathbb{Z}) = \bigcup_j \Gamma A_j$ as a disjoint union of its cosets and $\mathcal{H}^* = SL_2(\mathbb{Z})\mathcal{F}^* = \bigcup_j (\Gamma A_j)\mathcal{F}^*$. Then we can write $X(\Gamma)$ as a finite union $X(\Gamma) = \bigcup_j \pi(A_j(\mathcal{F}^*))$ and since A_j and π are continuous, $X(\Gamma)$ is compact.

$(Connection)$ Write $\mathcal{H}^* = U_1 \cup U_2$ as a disjoint union of open subsets, with $U_1 \neq \emptyset$. Since $\mathcal{H}$ is connected we have that $\mathcal{H} \subseteq U_1$ and so $U_2 \subseteq \mathbb{Q} \cup \{\infty\}$. But then, U_2 must be empty and it shows that $\mathcal{H}^*$ is connected, hence so is $X(\Gamma)$.

$(Hausdorff)$ For any two distinct points $x_1, x_2 \in X(\Gamma)$, we have to find disjoint neighbourhoods. The case $x_1 = \pi(z_1)$ and $x_2 = \pi(z_2)$ with $z_i \in \mathcal{H}$ is already been covered.

Suppose $x_1 = \pi(s_1)$ and $x_2 = \pi(z_2)$ with $s_1 \in \mathbb{Q} \cup \{\infty\}$ and $z_2 \in \mathcal{H}$. Write $s_1 = A(\infty)$ for some $A \in SL_2(\mathbb{Z})$. Let U_2 be any neighbourhood of z_2 with compact closure $K \subset \mathcal{H}$. Since for any $z \in \mathcal{H}$ and for any $B \in SL_2(\mathbb{Z})$

$$Im(Bz) \leq \max\left\{Im(z), \frac{1}{Im(z)}\right\},$$

there exists $C > 0$ such that $SL_2(\mathbb{Z})K \cap \mathcal{N}_C = \emptyset$. Let $U_1 = A(\mathcal{N}_C \cup \{\infty\})$. Then $\pi(U_1) \cap \pi(U_2) = \emptyset$. In fact if there exists $x \in \pi(U_1) \cap \pi(U_2)$, then we can write $x = \pi(w_1) = \pi(w_2)$ with $w_i \in U_i$, $Bw_2 = w_1$ for some $B \in \Gamma$ and $w_1 = Au_1$ for some $u_1 \in \mathcal{N}_C \cup \{\infty\}$. It follows that $u_1 = A^{-1}Bw_2 \in SL_2(\mathbb{Z})K \cap \mathcal{N}_C$, a contradiction.

It remains the case $x_1 = \pi(s_1)$ and $x_2 = \pi(s_2)$ with $s_1, s_2 \in \mathbb{Q} \cup \{\infty\}$. We can write $s_1 = A_1(\infty)$ and $s_2 = A_2(\infty)$ for some $A_1, A_2 \in SL_2(\mathbb{Z})$. Let $U_1 = A_1(\mathcal{N}_2 \cup \{\infty\})$ and $U_2 = A_2(\mathcal{N}_2 \cup \{\infty\})$. Then $\pi(U_1) \cap \pi(U_2) = \emptyset$, because otherwise there exist $B \in \Gamma$ and $w_1, w_2 \in \mathcal{N}_2$ such that $BA_1w_1 = w_2$ and so $A_2^{-1}BA_1w_1 = w_2$. It follows that $A_2^{-1}BA_1$ is a translation, since $\mathcal{N}_2$ is tessellated by the integer translates of $\mathcal{F}$ and contains no elliptic points. Thus $A_2^{-1}BA_1$ fixes ∞ and so $Bs_1 = s_2$, but it is impossible if x_1 and x_2 are distinct points. □

Defining charts for $X(\Gamma)$ is similar to the process carried out for $Y(\Gamma)$. We define the dual notion of the period of an elliptic point.

Definition 2.9. Let $s \in \mathbb{Q} \cup \{\infty\}$ and $\delta = \delta_s \in SL_2(\mathbb{Z})$ such that $\delta s = \infty$. The **width** of s under Γ is defined as

$$h_s = |SL_2(\mathbb{Z})_\infty / (\delta\{\pm I\}\Gamma\delta^{-1})_\infty|.$$

Lemma 2.6.4. *Let $s \in \mathbb{Q} \cup \{\infty\}$. Then the width of s is finite and*

$$h_s = |SL_2(\mathbb{Z})_s / \{\pm I\}\Gamma_s|.$$

Moreover for any $M \in SL_2(\mathbb{Z})$, the width of Ms under $M\Gamma M^{-1}$ is the same as the width of s under Γ. In particular h_s depends only on Γs, making the width well defined on $X(\Gamma)$.

Proof. Note that $SL_2(\mathbb{Z})_\infty = \{\left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right) \mid m \in \mathbb{Z}\}$ is the subgroup of translations. Since for any $\delta \in SL_2(\mathbb{Z})$ and for any congruence subgroup Γ, $\delta\Gamma\delta^{-1}$ is again a congruence subgroup, then $(\delta\Gamma\delta^{-1})_\infty = \left\langle \left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \right\rangle$ for some $h \geq 1$. This clearly implies that h_s is finite. For the second statement it suffices to note that $(\delta\{\pm I\}\Gamma\delta^{-1})_\infty = \delta\{\pm I\}\Gamma_s\delta^{-1}$ and $SL_2(\mathbb{Z})_\infty = (\delta SL_2(\mathbb{Z})\delta^{-1})_\infty = \delta SL_2(\mathbb{Z})_s\delta^{-1}$. Now for any $M \in SL_2(\mathbb{Z})$, we have

$$h_{Ms} = |SL_2(\mathbb{Z})_{Ms} / (M\{\pm I\}\Gamma M^{-1})_{Ms}| =$$

$$= |MSL_2(\mathbb{Z})_s M^{-1} / M\{\pm I\}\Gamma_s M^{-1}| =$$

$$= |SL_2(\mathbb{Z})_s / \{\pm I\}\Gamma_s| = h_s,$$

where h_{Ms} is the width of Ms under $M\Gamma M^{-1}$ and h_s is the width of s under Γ. □

Now define $U = U_s = \delta^{-1}(\mathcal{N}_2 \cup \{\infty\})$ and ρ as the h-periodic map $\rho(z) = e^{2\pi i z/h}$, where $h = h_s$ is the width of s under Γ. As before let $\psi = \rho \circ \delta$ and $V = \psi(U)$, which is an open subset of $\mathbb{C}$. For $z_1, z_2 \in U$ we have

$$\pi(z_1) = \pi(z_2) \iff z_1 = Mz_2 \iff \delta z_1 = (\delta M \delta^{-1})\delta z_2$$

for some $M \in \Gamma$. But since δz_1 and δz_2 both lie in $\mathcal{N}_2 \cup \{\infty\}$, $\delta M \delta^{-1}$ is a translation, hence $\delta M \delta^{-1} \in (\delta\Gamma\delta^{-1})_\infty \subseteq \pm\left\langle \left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \right\rangle$. It follows that

$$\pi(z_1) = \pi(z_2) \iff \delta z_1 = \delta z_2 + mh \text{ for some } m \in \mathbb{Z} \iff \psi(z_1) = \psi(z_2).$$

As before there exists a bijection $\varphi : \pi(U) \to V$ such that the following diagram is commutative

$$\begin{tikzcd}
U \arrow[r, "\delta"] \arrow[d, "\pi"] \arrow[dr, "\psi"] & \delta(U) \arrow[d, "\rho"] \\
\pi(U) \arrow[r, "\varphi"] & V.
\end{tikzcd}$$

Since π and ψ are open and continuous maps, φ is a homeomorphism. We have the following situation:

- $\pi : \mathcal{H}^* \to X(\Gamma)$ is the natural projection;
- The neighbourhoods $U \subset \mathcal{H}^*$ we consider contain at most one elliptic point or cusp;
- For $z \in \mathcal{H}$, the map $\delta = \delta_z$ is given by $\delta = \left(\begin{smallmatrix} 1 & -z \\ 1 & -\bar{z} \end{smallmatrix}\right)$, so $\delta z = 0$ and $\delta(U)$ is a neighbourhood of 0. We also define the map $\rho(w) = w^h$, where $h = h_z$ is the period of z (which is the only possible elliptic point in U);
- For $s \in \mathbb{Q} \cup \{\infty\}$, the map $\delta = \delta_s \in SL_2(\mathbb{Z})$ takes s to ∞ and so $\delta(U)$ is a neighbourhood of ∞. We also define the map $\rho(w) = e^{2\pi i w/h}$, where $h = h_z$ is the width of s under Γ;
- The homeomorphism $\varphi : \pi(U) \to V$ satisfies $\psi = \varphi \circ \pi$, where $\psi = \rho \circ \delta$.

It remains to verify that the transition maps are holomorphic. Let $U_1, U_2 \in \mathcal{H}^*$ be open neighbourhoods as above such that $\pi(U_1) \cap \pi(U_2) \neq \emptyset$, let $z_i, s_i, \delta_i, \psi_i, \varphi_i, \rho_i, h_i$ denote the corresponding objects and let $\varphi_{2,1} = \varphi_2 \circ \varphi_1^{-1}$ denote the transition map.

First suppose $U_1 \subset \mathcal{H}$ and $U_2 = \delta_2^{-1}(\mathcal{N}_2 \cup \{\infty\})$. For each $x \in \pi(U_1) \cap \pi(U_2)$ we can write $x = \pi(w_1) = \pi(w_2)$ with $w_1 \in U_1$, $w_2 \in U_2$ and $w_2 = Mw_1$ for some $M \in \Gamma$. Let $U_{1,2} = U_1 \cap M^{-1}(U_2)$. Then $\pi(U_{1,2})$ is a neighbourhood of x and so $\varphi_1(\pi(U_{1,2}))$ is a neighbourhood of $\varphi_1(x)$ in $V_{1,2} = \varphi_1(\pi(U_1) \cap \pi(U_2))$. Note that if $h_1 > 1$, then $z_1 \notin U_{1,2}$, because otherwise $\delta_2(Mz_1) \in \mathcal{N}_2$ is an elliptic point for Γ, but $\mathcal{N}_2$ has no elliptic points for $SL_2(\mathbb{Z})$. Therefore if

$h_1 > 1$, then $z_1 \notin U_{1,2}$ and so $0 = \varphi_1(\pi(z_1)) \notin \varphi_1(\pi(U_{1,2}))$. For any fixed $q = \varphi_1(y) \in U_{1,2}$ we have $q = \varphi_1(\pi(w)) = \psi(w) = (\delta_1 w)^{h_1}$ for some $w \in U_{1,2}$. Then we can write

$$\varphi_{2,1}(q) = \varphi_2(\varphi_1^{-1}(q)) = \varphi_2(\pi(w)) = \varphi_2(\pi(Mw)) = \psi_2(Mw) =$$

$$= \exp\{2\pi i \delta_2 Mw/h_2\} = \exp\{2\pi i \delta_2 M\delta_1^{-1}(q^{1/h_1})/h_2\}.$$

Thus $\varphi_{2,1}$ might not be holomorphic only when $h_1 > 1$ and $0 \in \varphi_1(\pi(U_{1,2}))$, but we have already seen that this cannot happen. Clearly we have also covered the case with the roles of U_1 and U_2 exchanged, because the inverse of an holomorphic function is again holomorphic.

Now suppose $U_1 = \delta_1^{-1}(\mathcal{N}_2 \cup \{\infty\})$ and $U_2 = \delta_2^{-1}(\mathcal{N}_2 \cup \{\infty\})$. Since $\pi(U_1) \cap \pi(U_2) \neq \emptyset$, $M\delta_1^{-1}(\mathcal{N}_2 \cup \{\infty\}) \cap \delta_2^{-1}(\mathcal{N}_2 \cup \{\infty\}) \neq \emptyset$ for some $M \in \Gamma$. This means that $\delta_2 M \delta_1^{-1}$ moves some point in $\mathcal{N}_2 \cup \{\infty\}$ to another and therefore must be a translation $z \mapsto z + m$. So $Ms_1 = M\delta_1^{-1}(\infty) = \delta_2^{-1}\delta_2 M\delta_1^{-1}(\infty) = \delta_2^{-1}(\infty) = s_2$, hence $h_1 = h_2 = h$. For a point $q \in \varphi_1(\pi(U_{1,2}))$ we can write

$$q = \psi_1(w) = exp\{2\pi i(\delta_1 w)/h\}$$

and this time

$$\varphi_{2,1}(q)\varphi_{2,1}(q) = \psi_2(Mw) = exp\{2\pi i \delta_2 M\delta_1^{-1}(\delta_1 w)/h\} =$$

$$= exp\{2\pi i(\delta_1 w + m)/h\} = e^{2\pi i m/h} q$$

and it is clearly holomorphic.

This completes the purpose of this section: we have shown that for every congruence subgroup $\Gamma \leq SL_2(\mathbb{Z})$, the modular curve $Y(\Gamma)$ is a Riemann surface, that can be compactified to obtain a compact Riemann surface $X(\Gamma)$.

CHAPTER 3

Modular Forms

3.1 Functions of Lattices

Definition 3.1. A **function of lattices** is a complex-valued function G defined over the set of lattices in the complex plane $\mathcal{L}$, which satisfies

$$G(\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}) = G(M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix})$$

for any $M \in SL_2(\mathbb{Z})$ and for any $\Lambda = [\omega_1, \omega_2] \in \mathcal{L}$. We can also write $G(\Lambda)$ or $G([\omega_1, \omega_2])$ to indicate that G is evaluated at the lattice $\Lambda = [\omega_1, \omega_2] \in \mathcal{L}$.

Let k be an integer. We say that G is homogeneous of degree k if for any lattice Λ and for any $c \in \mathbb{C}^*$

$$G(c\Lambda) = c^{-k} G(\Lambda).$$

Let G be a function of lattices, homogeneous of degree k. We define a function $g : \mathcal{H} \to \mathbb{C}$ by setting

$$g(z) = G(\Lambda_z) = G(\begin{pmatrix} z \\ 1 \end{pmatrix}).$$

Then for any $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$

$$g(Mz) = G([\frac{az+b}{cz+d}, 1]) = (cz+d)^k G([az+b, cz+d]) =$$

$$= (cz+d)^k G(M \begin{pmatrix} z \\ 1 \end{pmatrix}) = (cz+d)^k G(\begin{pmatrix} z \\ 1 \end{pmatrix}) = (cz+d)^k g(z).$$

Conversely , given a function $g : \mathcal{H} \to \mathbb{C}$ satisfying the condition

$$g(Mz) = (cz+d)^k g(z) \qquad \forall M \in SL_2(\mathbb{Z}),\ z \in \mathcal{H}$$

we can set for any lattice $\Lambda = [\omega_1, \omega_2] \in \mathcal{L}$

$$G(\Lambda) = \omega_2^{-k} g(\frac{\omega_1}{\omega_2}).$$

and since $G(c\Lambda) = c^{-k}G(\Lambda)$, G defines a function of lattices. Now the following result immediately follows.

Proposition 3.1.1. *There is a bijection between functions of lattices, homogeneous of degree k and functions $g : \mathcal{H} \to \mathbb{C}$ satisfying the condition*

$$g(Mz) = (cz+d)^k g(z) \qquad \forall M \in SL_2(\mathbb{Z}),\ z \in \mathcal{H}.$$

Given a function of lattices G, homogeneous of degree k, we can also define a corresponding complex-valued function $\tilde{G}$ of elliptic curves in the obvious way

$$\tilde{G}(\mathbb{C}/\Lambda) = G(\Lambda)$$

for any elliptic curve $E = \mathbb{C}/\Lambda$. Note that $\tilde{G}$ satisfies the homogeneity condition

$$\tilde{G}(\mathbb{C}/c\Lambda) = c^{-k}\tilde{G}(\mathbb{C}/\Lambda) \quad \text{for any } c \in \mathbb{C}.$$

Recall the enhanced elliptic curves for the congruence subgroups $\Gamma_0(N), \Gamma_1(N)$ and $\Gamma(N)$ in Section 2.3. We can define the corresponding concepts in terms of lattices.

Definition 3.2. Let N be an integer and let Γ be one of $\Gamma_0(N), \Gamma_1(N)$, $\Gamma(N)$ or $SL_2(\mathbb{Z})$. A **modular point** for Γ is

i. a lattice $\Lambda \in \mathcal{L}$ if $\Gamma = SL_2(\mathbb{Z})$;

ii. a pair (Λ, u), where $\Lambda \in \mathcal{L}$ and $u \in \mathbb{C}/\Lambda$ is a point of exact order N, if $\Gamma = \Gamma_1(N)$;

iii. a pair (Λ, S), where $\Lambda \in \mathcal{L}$ and $S \subset \mathbb{C}/\Lambda$ is a cyclic subgroup of order N, if $\Gamma = \Gamma_0(N)$;

iv. a triple (Λ, u, v), where $\Lambda \in \mathcal{L}$ and $u, v \in \mathbb{C}/\Lambda$ generates the subgroup of N-torsion points, if $\Gamma = \Gamma(N)$.

Then we can extend the definition of functions of lattices to functions of modular points $G(\Lambda)$, $G(\Lambda, u)$, $G(\Lambda, S)$ and $G(\Lambda, u, v)$ respectively for $\Gamma = SL_2(\mathbb{Z}), \Gamma_0(N), \Gamma_1(N)$ or $\Gamma(N)$, and in an obvious way, we can obtain the corresponding functions $\tilde{G}$ of enhanced elliptic curves. The homogeneity conditions of degree k become

i. $G(c\Lambda) = c^{-k}G(\Lambda) \qquad \forall c \in \mathbb{C}^*;$

ii. $G(c\Lambda, cu) = c^{-k}G(\Lambda, u) \qquad \forall c \in \mathbb{C}^*;$

iii. $G(c\Lambda, cS) = c^{-k}G(\Lambda, u) \qquad \forall c \in \mathbb{C}^*;$

iv. $G(c\Lambda, cu, cv) = c^{-k}G(\Lambda, u, v) \qquad \forall c \in \mathbb{C}^*.$

We also define

$$g(z) = \begin{cases} G(\Lambda_z) & \text{for modular points for } SL_2(\mathbb{Z}) \\ G(\Lambda_z, u) & \text{for modular points for } \Gamma_1(N) \\ G(\Lambda_z, S) & \text{for modular points for } \Gamma_0(N) \\ G(\Lambda_z, u, v) & \text{for modular points for } \Gamma(N), \end{cases}$$

which satisfies $g(Mz) = (cz + d)^k g(z)$ for all $M \in \Gamma$, $z \in \mathcal{H}$. Now it is easy to see that the following result holds.

Proposition 3.1.2. *Let N and k be integers and let Γ be one of $\Gamma_0(N)$, $\Gamma_1(N)$, $\Gamma(N)$ or $SL_2(\mathbb{Z})$. The above association of G with $\tilde{G}$ and g gives a bijection between the following sets:*

- *complex-valued functions G on modular points which are homogeneous of degree k;*
- *complex-valued functions $\tilde{G}$ on enhanced elliptic curves which are homogeneous of degree k;*

- *complex-valued functions g on $\mathcal{H}$ which satisfy $g(Mz) = (cz+d)^k g(z)$ for all $M \in SL_2(\mathbb{Z})$, $z \in \mathcal{H}$.*

In the following sections and in the next chapter, we focus our attention to a class of holomorphic functions f on $\mathcal{H}$ which satisfy $g(Mz) = (cz+d)^k g(z)$ for any $M \in \Gamma$, $z \in \mathcal{H}$.

3.2 Modular Forms and Cusp Forms

We have already shown how to obtain functions f on $\mathcal{H}$ which satisfy a certain homogeneity condition by starting from complex-valued function of lattices. In this section we will requesting that such functions have some other properties.

Definition 3.3. Let k be an integer. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is **weakly modular of weight k** if

$$f(Mz) = (cz+d)^k f(z), \quad \forall\ M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z}),\ z \in \mathcal{H}.$$

Necessary and sufficient condition for f to be weakly modular of weight k is that this transformation law holds when M is each of the generators of the modular group. In other words, f is weakly modular of weight k if and only if $f(z+1) = f(z)$ and $f(-1/z) = z^k f(z)$. In particular f is $\mathbb{Z}$-periodic. Moreover, since the factor $cz + d \neq 0$ if $z \in \mathcal{H}$, $f(z)$ and $f(Mz)$ always have the same zeros and poles. We also observe that letting $M = -I$ we obtain $f = (-1)^k f$ and so the only weakly modular function of any odd weight k is the zero function. However, there exists a more general context (to be developed soon) in which we can find non-zero odd weight examples.

Now we want to introduce the concept of "*holomorphy at $i\infty$*". We have just seen that a weakly modular function f is $\mathbb{Z}$-periodic. Let D be the open complex punctured unit disk, i.e. $D = \{z \in \mathbb{C} \mid |z| < 1\} \setminus \{0\}$. Recall that the $\mathbb{Z}$-periodic holomorphic map $z \mapsto q = e^{2\pi i z}$ takes $\mathcal{H}$ to D. Thus, corresponding to f, the function $g : D \to \mathbb{C}$ where $g(q) = f(log(q)/2\pi i)$ is well defined even though the logarithm is only determinated up to $2\pi i\mathbb{Z}$, and $f(z) = g(e^{2\pi i z})$. If f is holomorphic on $\mathcal{H}$ then g is holomorphic on D, and so g has a Laurent expansion $g(q) = \sum_{n \in \mathbb{Z}} a_n q^n$ for $q \in D$. Define f to be **holomorphic at** $i\infty$ if g can be extended holomorphically to the unit disk, i.e. the Laurent series sums over $n \in \mathbb{N}$. This means that f has a *Fourier expansion*

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

Since $q \to 0$ if and only if $Im(z) \to \infty$, to show that a weakly modular form f is holomorphic at $i\infty$ it suffices to verify that $\lim_{Im(z) \to \infty} f(z)$ exists or even just that $f(z)$ is bounded as $Im(z) \to \infty$.

Definition 3.4. Let k be an integer. A function $f : \mathcal{H} \to \mathbb{C}$ is a **modular form of weight k** if

1. f is holomorphic on $\mathcal{H}$;
2. f is weakly modular of weight k;
3. f is holomorphic at $i\infty$.

The set of modular forms of weight k is denoted by $\mathcal{M}_k(SL_2(\mathbb{Z}))$.

For our purposes, particular attention should be paid to modular forms that have constant term equal to zero.

Definition 3.5. A **cusp form** of weight k is a modular form of weight k whose Fourier expansion has leading coefficient $a_0 = 0$, i.e.

$$f(z) = \sum_{n=1}^{\infty} a_n q^n, \qquad q = e^{2\pi i z}.$$

The set of cusp forms is denoted by $\mathcal{S}_k(SL_2(\mathbb{Z}))$.

$\mathcal{M}_k(SL_2(\mathbb{Z}))$ is a vector space over $\mathbb{C}$ and we will see that it is finite-dimensional. If f and g are two modular forms of weight k and l respectively, then their product is a modular form of weight $k + l$. Thus the sum

$$\mathcal{M}(SL_2(\mathbb{Z})) = \bigoplus_{k \in \mathbb{Z}} \mathcal{M}_k(SL_2(\mathbb{Z}))$$

forms a graded ring because of its structure as a sum. The zero function on $\mathcal{H}$ is a modular form of every weight, and every constant function on $\mathcal{H}$ is a modular form of weight zero.

The cusp forms $\mathcal{S}_k(SL_2(\mathbb{Z}))$ form a vector subspace of the modular forms $\mathcal{M}_k(SL_2(\mathbb{Z}))$, and the graded ring

$$\mathcal{S}(SL_2(\mathbb{Z})) = \bigoplus_{k \in \mathbb{Z}} \mathcal{S}_k(SL_2(\mathbb{Z}))$$

is an ideal in $\mathcal{M}(SL_2(\mathbb{Z}))$.

A modular form is a cusp form when $\lim_{Im(z)\to\infty} f(z) = 0$. So a cusp form can be viewed as vanishing at the limit point $i\infty$ of $\mathcal{H}$. This point is called the *cusp* of $SL_2(\mathbb{Z})$ (we also say that *a cusp form vanishes at the cusp*).

By the following definition we introduce a shorthand notation.

Definition 3.6. For any matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ define the **factor of automorphy** $j(M, z) \in \mathbb{C}$ for $z \in \mathcal{H}$ to be $j(M, z) = cz + d$, and for any integer k define the **weight-k operator** $[M]_k$ on functions $f : \mathcal{H} \to \mathbb{C}$ by

$$(f[M]_k)(z) = j(M, z)^{-k} f(Mz), \qquad z \in \mathcal{H}.$$

Since the factor of automorphy is never zero or infinity, f and $f[M]_k$ have the same zeros and poles. In view of this notation, a function f is weakly modular of weight k if $f[M]_k = f$ for all $M \in SL_2(\mathbb{Z})$ (sometimes we express this condition by saying that f is weight-k invariant under $SL_2(\mathbb{Z})$).

Lemma 3.2.1. *For all $M_1, M_2 \in SL_2(\mathbb{Z})$ and $z \in \mathcal{H}$,*

1. $j(M_1M_2, z) = j(M_1, M_2(z))j(M_2, z)$;
2. $(M_1M_2)z = M_1(M_2z)$;
3. $[M_1M_2]_k = [M_1]_k[M_2]_k$;
4. $Im(M_1(z)) = \frac{Im(z)}{|j(M_1,z)|^2}$;
5. $\frac{dM_1z}{dz} = \frac{1}{j(M_1,z)^2}$.

Proof. This lemma can be shown by a straightforward calculation. □

Now we want to extend this concept in a more general context. We have the following definition.

Definition 3.7. Let k be an integer and let Γ be a subgroup of $SL_2(\mathbb{Z})$. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is **weakly modular of weight k with respect to** Γ if $f[M]_k = f$ for all $M \in \Gamma$.

An immediate consequence of the lemma is that if f is weakly modular of weight k with respect to some subset $S \subseteq SL_2(\mathbb{Z})$, then f is weakly modular of weight k with respect to the subgroup generated by S.

We now develop the definition of a modular form with respect to a congruence subgroup similar to what we saw before for modular forms. Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Since Γ contains $\Gamma(N)$ for some N, Γ contains a translation $z \mapsto z + h$, i.e. an element of the form $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$, for some minimal $h \in \mathbb{Z}$ (note that h may properly divide N). If $f : \mathcal{H} \to \mathbb{C}$ is weakly modular with respect to Γ, then f is $h\mathbb{Z}$-periodic and thus has a corresponding function $g : D \to \mathbb{C}$ where again D is the open complex punctured unit disk, but now we have $f(z) = g(q_h)$, with $q_h = e^{2\pi i z/h}$. As before, if f is also holomorphic on the upper half plane then g is holomorphic on D and so it has a Laurent expansion. Define such f to be **holomorphic at** $i\infty$ if g extends holomorphically to $q = 0$. Thus f has a **Fourier expansion**

$$f(z) = \sum_{n=0}^{\infty} a_n q_h^n, \qquad q_h = e^{2\pi i z/h}.$$

We are now able to give the following definition.

Definition 3.8. Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$ and let k be an integer. A function $f : \mathcal{H} \to \mathbb{C}$ is a **modular form of weight k with respect to** Γ if

1. f is holomorphic on $\mathcal{H}$;
2. f is weakly modular of weight k with respect to Γ;
3. $f[M]_k$ is holomorphic at $i\infty$ for all $M \in SL_2(\mathbb{Z})$.

If in addition

4. $a_0 = 0$ in the Fourier expansion of $f[M]_k$ for all $M \in SL_2(\mathbb{Z})$,

then we say that f is a **cusp form of weight k with respect to** Γ. The set of modular forms of weight k with respect to Γ is denoted by $\mathcal{M}_k(\Gamma)$, while the subset of the cusp forms is denoted by $\mathcal{S}_k(\Gamma)$.

Note that if f is weakly modular of weight k with respect to a congruence subgroup Γ, then for any $M \in SL_2(\mathbb{Z})$, $f[M]_k$ is weakly modular of weight k with respect to $M^{-1}\Gamma M$, which is again a congruence subgroup, and the condition (3) make sense. Moreover conditions (3) and (4) give us the holomorphy at the cusps in terms of holomorphy at ∞ in a natural way via the $[M]_k$ operators. For this reason we sometimes say that f is *holomorphic at the cusps* and *vanishes at the cusps* respectively. Then note that since any congruence subgroup has finite index in $SL_2(\mathbb{Z})$, conditions (3) and (4) need to be checked only for the finitely many coset representatives M_j for $SL_2(\mathbb{Z})/\Gamma$.

3.3 Examples and Main Results

In this section we will discuss about the most relevant examples and results of modular forms.

Let Λ be a lattice and let k be an even integer greater that 2. Recall from Chapter 1 the Eisenstein series

$$G_k(\Lambda) = \sum_{\omega \in \Lambda^*} \frac{1}{\omega^k}.$$

If the lattice is of the form $\Lambda_z = [z, 1]$ for some $z \in \mathcal{H}$, then we can write

$$G_k(z) = G_k(\Lambda_z) = \sum_{(c,d) \in \mathbb{Z}^2 \setminus \{(0,0)\}} \frac{1}{(cz+d)^k}. \tag{3.1}$$

Proposition 3.3.1. *The Eisenstein series $G_k(z)$ defines a modular form of weight k for $SL_2(\mathbb{Z})$.*

Proof. We have seen that the double sum (3.1) is absolutely convergent and uniformly convergent in any compact subset of $\mathcal{H}$ and hence it defines an holomorphic function on $\mathcal{H}$. Furthermore it is easy to see that

$$G_k(z+1) = G_k(z) \qquad \text{and} \qquad G_k(-\frac{1}{z}) = z^k G_k(z).$$

Finally the Fourier expansion of G_k has no negative terms, since

$$\lim_{Im(z)\to\infty} G_k(z) = \sum_{n\neq 0} \frac{1}{n^k} = 2\zeta(k)$$

where $\zeta(k)$ denotes the well-known Riemann zeta function. □

Now we are interested to compute the Fourier expansion for G_k. We use the identities

$$\frac{1}{z} + \sum_{d=1}^{\infty} \left(\frac{1}{z+d} + \frac{1}{z-d} \right) = \pi \cot(\pi z) = \pi i - 2\pi i \sum_{m=0}^{\infty} q^m$$

where $q = e^{2\pi i z}$. Differentiating $k-1$ times with respect to z we obtain

$$(-1)^{k-1}(k-1)! \sum_{d\in\mathbb{Z}} \frac{1}{(z+d)^k} = -(2\pi i)^k \sum_{m=0}^{\infty} m^{k-1} q^m$$

and so

$$\sum_{d\in\mathbb{Z}} \frac{1}{(z+d)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{m=1}^{\infty} m^{k-1} q^m.$$

At the same time, for even k greater than 2 we have

$$\sum_{(0,0)\neq(c,d)\in\mathbb{Z}^2} \frac{1}{(cz+d)^k} = \sum_{d\neq 0} \frac{1}{d^k} + 2 \sum_{c=1}^{\infty} \sum_{d\in\mathbb{Z}} \frac{1}{(cz+d)^k} =$$

$$= 2\zeta(k) + 2 \sum_{c=1}^{\infty} \left[\frac{(-2\pi i)^k}{(k-1)!} \sum_{m=1}^{\infty} m^{k-1} q^{cm} \right] =$$

$$= 2\zeta(k) + 2 \frac{(-2\pi i)^k}{(k-1)!} \sum_{c=1}^{\infty} \sum_{m=1}^{\infty} m^{k-1} q^{cm}.$$

By setting $n = cm$ and rearranging the last expression we have

$$G_k(z) = 2\zeta(k) + 2\frac{(-2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n$$

where the n^{th}-coefficient is the arithmetic function

$$\sigma_{k-1}(n) = \sum_{m|n, m>0} m^{k-1}.$$

We can normalize the Eisenstein series by dividing by $2\zeta(k)$. We obtain the **normalized Eisenstein series** $E_k(z)$, whose Fourier expansion at infinity has the constant term equal to one and the other coefficients equal to rational numbers with a common denominator. In fact, since the Bernoulli numbers B_k, for positive even k, satisfy

$$\zeta(k) = -\frac{(2\pi i)^k}{2}\frac{B_k}{k!},$$

we obtain

$$E_k(z) = 1 - \frac{2k}{B_k}\sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n.$$

Recall the *discriminant function* from Section 1.4. It is defined as

$$\Delta : \mathcal{H} \to \mathbb{C}, \qquad \Delta(z) = g_2^3(z) - 27g_3^2(z).$$

where

$$g_2(z) = 60G_4(z), \qquad \text{and} \qquad g_3(z) = 140G_6(z).$$

Since $\zeta(4) = \frac{\pi^4}{90}$ and $\zeta(6) = \frac{\pi^6}{945}$, we can express the discriminant function in terms of the normalized Eisenstein series E_4 and E_6:

$$\Delta(z) = \frac{(2\pi)^{12}}{1728}(E_4(z)^3 - E_6(z)^2).$$

Note that $\Delta \in \mathcal{M}_{12}(SL_2(\mathbb{Z}))$ and actually it is a cusp form, since both E_4 and E_6 have constant term equal to 1 in their Fourier expansion.

Let $\mu_3 = e^{\frac{2}{3}\pi i}$ be the complex cube root of unity and put $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$. Since $S\mu_3 = \mu_3 + 1$, by periodicity of $E_4(z)$ we have $E_4(S\mu_3) = E_4(\mu_3)$ and, by modularity, we also have that $E_4(S\mu_3) = \mu_3^4 E_4(\mu_3)$. Therefore

$E_4(\mu_3) = 0$ and $E_6(\mu_3) \neq 0$ because $\Delta(z)$ does not vanish on $\mathcal{H}$. Similarly, we can show that $E_6(i) = 0$ and $E_4(i) \neq 0$. We use this fact later in the section.

The next useful result allow us to determine the spaces $\mathcal{M}_k(SL_2(\mathbb{Z}))$ and $\mathcal{S}_k(SL_2(\mathbb{Z}))$, and it can be used to establish when two modular forms are actually the same. We need two lemmas.

Lemma 3.3.2. *Let f be a weakly modular function of weight k for $SL_2(\mathbb{Z})$. Then the imaginary part of the zeros and poles of f is bounded from above.*

Proof. We have seen that the change of variables $z \mapsto q = e^{2\pi i z}$ makes $f(z)$ into a meromorphic function $g(q)$ in a disc around $q = 0$. Since g is meromorphic, there exists a disc around $q = 0$ on which g has no zeros or poles, except possibly at $q = 0$. It follows that the imaginary part of the zeros and poles of f is bounded from above. □

Lemma 3.3.3. *Let f be a weakly modular function of weight k for $SL_2(\mathbb{Z})$, with no zeros or poles on a contour $\gamma \subset \mathcal{H}$, and let $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$, with $c \neq 0$. Then*

$$\int_\gamma \frac{f'(z)}{f(z)}dz - \int_{M\gamma} \frac{f'(z)}{f(z)}dz = -k \int_\gamma \frac{1}{z + \frac{d}{c}}dz$$

Proof. Differentiating $f(Mz) = (cz+d)^k f(z)$ we obtain

$$f'(Mz)\frac{dMz}{dz} = (cz+d)^k f'(z) + kc(cz+d)^{k-1} f(z)$$

and dividing the second equality by the first we have

$$\frac{f'(Mz)}{f(Mz)}dMz = \frac{f'(z)}{f(z)}dz + k\frac{cdz}{cz+d}.$$

Now it suffices to integrate over γ to obtain

$$\int_{M\gamma} \frac{f'(z)}{f(z)}dz = \int_\gamma \frac{f'(z)}{f(z)}dz + k\int_\gamma \frac{dz}{z + \frac{d}{c}},$$

so the lemma is proved. □

Theorem 3.3.4. *Let f be a weakly modular function of weight k for $SL_2(\mathbb{Z})$. For any $\tau \in \mathcal{H}$, let $v_\tau(f)$ denote the order of zero or minus the order of pole of f at the point τ. Let $v_\infty(f)$ denote the index of the first non-vanishing term in the q-expansion of f. Then*

$$v_\infty(f) + \frac{1}{2}v_i(f) + \frac{1}{3}v_{\mu_3}(f) + \sum_{\tau \in SL_2(\mathbb{Z})\backslash\mathcal{H},\ \tau \neq i,\mu_3} v_\tau(f) = \frac{k}{12}. \tag{3.2}$$

(With $\tau \in SL_2(\mathbb{Z}) \setminus \mathcal{H}$, $\tau \neq i, \mu_3$, we means that the sum is taken over all points $\tau \in \mathcal{H}$ modulo $SL_2(\mathbb{Z})$, not in the orbit of i or μ_3.)

Proof. First note that $v_\tau(f) = v_{M\tau}(f)$ for any $M \in SL_2(\mathbb{Z})$, so the sum is well-defined.

Let L be a positive number such that $Im(\tau) < L$ for all $\tau \in \mathcal{H}$ such that f has a pole or zero in τ. To prove our theorem we want to integrate the logarithmic derivative of f on the boundary of the fundamental domain $\mathcal{F}$ of $SL_2(\mathbb{Z})$ (Theorem 2.2.1) intersected with the horizontal stripe $\{z \in \mathcal{H} \mid Im(z) \leq L\}$, but modified by taking small arcs with small radius ϵ around any pole or zero on the boundary, as on the Figure (3.3). Let γ denote such contour. Note that we include every $SL_2(\mathbb{Z})$-equivalence class of zero or pole exactly once inside γ, except for i and μ_3 that are kept outside of γ if they are zeros or poles.

By the Residue Theorem we have

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz = \sum_{\tau \in SL_2(\mathbb{Z})\backslash\mathcal{H},\tau\neq i,\mu_3} v_\tau(f).$$

Now we evaluate this integral section by section. For the integral over the top segment AH, make the change of coordinates $z \mapsto q = e^{2\pi i z}$. Remembering that f' denotes derivative with respect to z, we obtain that the integral over this segment is equal to

$$-v_0(g) = -v_\infty(f)$$

where $g(q) = f(z)$. Then, the integral over the left vertical segment AB cancels the integral over the right vertical segment GH, because of the periodicity of f.

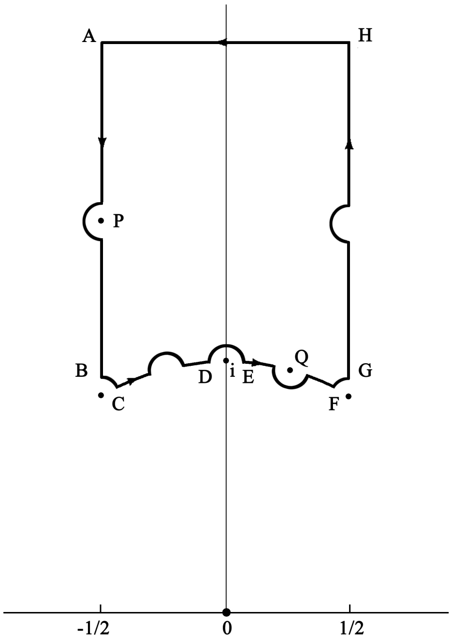

Figure 3.1: Contour for the proof of Theorem (3.3.4)

To evaluate the integral over the arcs BC, DE and FG recall that if $h(z)$ is a meromorphic function with a Laurent expansion $h(z) = \sum_{n=m}^{\infty} c_m(z-a)^m$ near a and if β is a counter-clockwise oriented circular arc of angle θ centred at a with small radius ϵ, then

$$\int_\beta \frac{h'(z)}{h(z)}dz = mi\theta.$$

We can apply this to the arcs BC and FG. For $\epsilon \to 0$ the angles approach to $\pi/3$ and so we obtain respectively

$$-\frac{1}{2\pi i}(\frac{i\pi}{3}v_{\mu_3}(f)) = -\frac{1}{6}v_{\mu_3}(f) \quad \text{and} \quad -\frac{1}{2\pi i}(\frac{i\pi}{3}v_{-\bar{\mu}_3}(f)) = -\frac{1}{6}v_{-\bar{\mu}_3}(f).$$

Here the minus sign is because the arcs goes clockwise. Note that $v_{\mu_3}(f) = v_{-\bar{\mu}_3}(f)$, so the integral over BC and FG gives us a contribute of $-\frac{1}{3}v_{\mu_3}(f)$.

To prove our theorem it remains to evaluate the integrals from C to D and from E to F. We need to show that

$$\frac{1}{2\pi i}\int_{CD}\frac{f'(z)}{f(z)}dz + \frac{1}{2\pi i}\int_{EF}\frac{f'(z)}{f(z)}dz \to \frac{k}{12} \qquad \text{as} \quad \epsilon \to 0.$$

But this follows from the previous lemma by setting $M = S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. In fact, S takes CD to FE and

$$\frac{1}{2\pi i}\int_{CD}\frac{dz}{z} \to \int_{\frac{1}{3}}^{\frac{1}{4}} d\theta = -\frac{1}{12}, \quad \text{where} \quad z = e^{2\pi i\theta}.$$

Now, adding all contributes together, the theorem follows. □

There are several important consequences Theorem (3.3.4), which are listed in the following proposition.

Proposition 3.3.5. *Let k be an even integer.*

1. *The only modular forms of weight 0 for $SL_2(\mathbb{Z})$ are constants, i.e. $\mathcal{M}_0(SL_2(\mathbb{Z})) = \mathbb{C}$;*
2. *$\mathcal{M}_k(SL_2(\mathbb{Z})) = 0$ if $k < 0$ or $k = 2$;*
3. *$\mathcal{M}_k(SL_2(\mathbb{Z}))$ is one-dimensional, generated by E_k (i.e. $\mathcal{M}_k(SL_2(\mathbb{Z})) = \mathbb{C}E_k$), if k =4,6,8,10 or 14;*
4. *$\mathcal{S}_k(SL_2(\mathbb{Z})) = 0$ if $k < 12$ or $k = 14$; $\mathcal{S}_{12}(SL_2(\mathbb{Z})) = \mathbb{C}\Delta$ and for $k > 14$, $\mathcal{S}_k(SL_2(\mathbb{Z})) = \Delta\mathcal{M}_{k-12}(SL_2(\mathbb{Z}))$;*
5. *$\mathcal{M}_k(SL_2(\mathbb{Z})) = \mathcal{S}_k(SL_2(\mathbb{Z})) \oplus \mathbb{C}E_k$ for $k > 2$.*

Proof. All the results follows from the fact that for a modular form all terms on the left in (3.2) are non-negative.

1. Let $f \in \mathcal{M}_0(SL_2(\mathbb{Z}))$ and let c any value taken by $f(z)$. Then $f(z) - c$ has a zero and so the sum on the left in (3.2) is strictly positive. Since the right side is equal to 0, it follows that $f(z) - c$ must be the zero function, i.e. $f(z)$ is constant.
2. If $k \leq 2$ there is no way to obtain the value $\frac{k}{12}$ on the left side in 3.2.
3. It is easy to see that
   - for $k = 4$ we must have $v_{\mu_3}(f) = 1$ and all other $v_\tau = 0$;
   - for $k = 6$ we must have $v_i(f) = 1$ and all other $v_\tau = 0$;

- for $k = 8$ we must have $v_{\mu_3}(f) = 2$ and all other $v_\tau = 0$;
- for $k = 10$ we must have $v_{\mu_3}(f) = v_i(f) = 1$ and all other $v_\tau = 0$;
- for $k = 14$ we must have $v_{\mu_3}(f) = 2$, $v_i(f) = 1$ and all other $v_\tau = 0$.

Now let f, g be two non-zero modular form for those values of k. Since f and g have the same zeros, f/g is a modular form of weight zero and by part (1), f and g are proportional. Choosing $f = E_k$, the result follows.

4. Since $v_\infty(f) > 0$ if $f \in \mathcal{S}_k(SL_2(\mathbb{Z}))$, by the previous part we must have $\mathcal{S}_k(SL_2(\mathbb{Z})) = 0$ for $k < 12$ or $k = 14$. Then recall that Δ has no zeros on $\mathcal{H}$ and hence for any $k > 14$ or $k = 12$ and any $f \in \mathcal{S}_k(SL_2(\mathbb{Z}))$, $f/\Delta \in \mathcal{M}_{k-12}(SL_2(\mathbb{Z}))$.

5. Since E_k does not vanish at infinity, for any given $f \in \mathcal{M}_k(SL_2(\mathbb{Z}))$, we can find $c \in \mathbb{C}$ such that $f - cE_k$ vanishes at infinity, i.e. $f - cE_k \in \mathcal{S}_k(SL_2(\mathbb{Z}))$.

□

Corollary 3.3.6. *Let k be an even integer. Then any $f \in \mathcal{M}_k(SL_2(\mathbb{Z}))$ is a polynomial in E_4 and E_6, i.e. f can be written in the form*

$$f(z) = \sum_{4i+6j=k} c_{ij} E_4(z)^i E_6(z)^j, \qquad c_{i,j} \in \mathbb{C}.$$

Moreover E_4 and E_6 are algebraically independent, and so the polynomial above is unique.

Proof. We prove the first statement by induction on k. For $k = 4, 6, 8, 10, 14$ note that $E_4, E_6, E_4^2, E_4E_6, E_4^2E_6$ respectively, is an element of $\mathcal{M}_k(SL_2(\mathbb{Z}))$, and the statement follows by the previous proposition. Now suppose that $k = 12$ or $k > 14$. Let i, j be such that $4i + 6j = k$ (note that it is always possible to find such integers). Clearly $E_4^i E_6^j \in \mathcal{M}_k(SL_2(\mathbb{Z}))$. Given $f \in \mathcal{M}_k(SL_2(\mathbb{Z}))$ we can always subtract a suitable multiple of $E_4^i E_6^j$ so that the resulting $f - cE_4^i E_6^j \in \mathcal{S}_k(SL_2(\mathbb{Z}))$. By the previous proposition we can write

$$f = E_4^i E_6^j + \Delta f_1 = E_4^i E_6^j + \frac{(2\pi)^{12}}{1728}(E_4^3 - E_6^2) f_1$$

for some $f_1 \in \mathcal{M}_{k-12}(SL_2(\mathbb{Z}))$. Now the result follows by the induction assumption, with k replaced by $k-12$.

For the second statement, if E_4 and E_6 are not algebraically independent, then there exists an irreducible polynomial $F(X,Y) \in \mathbb{C}[X,Y]$ such that $F(E_4, E_6) = 0$. Note that $F(X,Y)$ must be homogeneous because a non-trivial linear relation among elements of distinct weights cannot exist. It follows that $F(X,Y)$ can be write in the form $X^h + YG(X,Y)$ for some $h > 0$ and $G(X,Y) \in \mathbb{C}[X,Y]$. Since $E_4(i) \neq 0$ and $E_6(i) = 0$, by substituting $z = i$ in $F(E_4(z), E_6(z))$, we obtain a contradiction. □

The previous proposition allow us to compute inductively the dimension of spaces of modular forms for $SL_2(\mathbb{Z})$.

Corollary 3.3.7. *The dimension of $\mathcal{M}_k(SL_2(\mathbb{Z}))$ as a $\mathbb{C}$-vector space is finite and it is given by*

$$\dim_{\mathbb{C}}(\mathcal{M}_k(SL_2(\mathbb{Z}))) = \begin{cases} 0 & \textit{if } k < 0 \textit{ or } k \textit{ is odd;} \\ \left[\frac{k}{12}\right] & \textit{if } k \geq 0,\ k \equiv 2\ (\textit{mod } 12) \\ \left[\frac{k}{12}\right] + 1 & \textit{otherwise} \end{cases}$$

where $[x]$ denotes the floor function. Moreover, for all integers k there exists an isomorphism of $\mathbb{C}$-vector spaces

$$\mathcal{M}_{2k-12}(SL_2(\mathbb{Z})) \xrightarrow{\sim} \mathcal{S}_{2k}(SL_2(\mathbb{Z})) \qquad f(z) \mapsto \Delta(z) f(z)$$

and in particular

$$\dim_{\mathbb{C}}(\mathcal{S}_{2k}(SL_2(\mathbb{Z}))) = \dim_{\mathbb{C}}(\mathcal{M}_{2k-12}(SL_2(\mathbb{Z}))).$$

One can prove that for any integer k and any congruence subgroup $\Gamma \leq SL_2(\mathbb{Z})$, the $\mathbb{C}$-vector space $\mathcal{M}_k(\Gamma)$ is finite dimensional. The formulas for $\dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma))$ and $\dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma))$ involve the genus and the number of cusps of the modular curve $X(\Gamma)$. Below, we provide this formulas. For the proof, see [DS05].

Theorem 3.3.8. *Let k be an integer and let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Then let g denote the genus of $X(\Gamma)$, e_2 the number of elliptic points with period 2, e_3 the number of elliptic points with period 3, e_∞ the number of cusps and e'_∞, e''_∞ two integers depending on the cusps of $X(\Gamma)$. Then for k even*

$$\dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \left[\frac{k}{4}\right] e_2 + \left[\frac{k}{3}\right] e_3 + \frac{k}{2} e_\infty & \textit{if } k \geq 2 \\ 1 & \textit{if } k = 0 \\ 0 & \textit{if } k < 0 \end{cases}$$

and

$$\dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \left[\frac{k}{4}\right] e_2 + \left[\frac{k}{3}\right] e_3 + \left(\frac{k}{2} - 1\right) e_\infty & \textit{if } k \geq 4 \\ 1 & \textit{if } k = 2 \\ 0 & \textit{if } k \leq 0. \end{cases}$$

For k odd, if $-I \in \Gamma$, then $\mathcal{M}_k(\Gamma) = \mathcal{S}_k(\Gamma) = \{0\}$, else

$$\dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \left[\frac{k}{3}\right] e_3 + \frac{k}{2} e'_\infty + \frac{k-1}{2} e''_\infty & \textit{if } k \geq 3 \\ \frac{1}{2} e'_\infty & \textit{if } k = 1 \\ 0 & \textit{if } k < 0 \end{cases}$$

and

$$\dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \left[\frac{k}{3}\right] e_3 + \frac{k-2}{2} e'_\infty + \frac{k-1}{2} e''_\infty & \textit{if } k \geq 3 \\ 0 \quad \textit{or} \quad \dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma)) - \frac{1}{2} e'_\infty & \textit{if } k = 1 \\ 0 & \textit{if } k < 0 \end{cases}$$

where for $k = 1$, $\dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma)) = 0$ if $e'_\infty > 2g - 2$ and $\dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma)) = \dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma)) - e'_\infty/2$ otherwise.

3.4 A Loose End From Chapter 1: The j-invariant

In Section 1.4 we have claimed that the *modular invariant*

$$j(z) = 1728\frac{g_2^3(z)}{\Delta(z)} = 1728\frac{E_4(z)^3}{E_4(z)^3 - E_6(z)^2}$$

is a surjective function. Now we want to prove this fact. Recall that since $\Delta(z) \neq 0$ for all $z \in \mathcal{H}$, $j(z)$ is an holomorphic function on $\mathcal{H}$. It is also easy to see that j is $SL_2(\mathbb{Z})$-invariant, i.e. $j(Mz) = j(z)$ for all $M \in SL_2(\mathbb{Z}),\ z \in \mathcal{H}$. For this reason, this function is called the *modular invariant.* The Fourier expansion

$$j(z) = \frac{(2\pi)^{12} + ...}{(2\pi)^{12}q + ...} = \frac{1}{q} + ..., \qquad q = e^{2\pi iz}$$

shows that j has a simple pole at ∞, so it is not a modular form. We note that j is normalized to have residue 1 at the pole at ∞.

Proposition 3.4.1. *The modular function j induces a bijection from $X(1)$ to the Riemann sphere $\mathbb{P}^1_{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$*

$$J : SL_2(\mathbb{Z})z \mapsto j(z).$$

In particular, the modular invariant $j : \mathcal{H} \to \mathbb{C}$ is surjective.

Proof. Since j has a simple pole at infinity and it is holomorphic on $\mathcal{H}$, it suffices to show that for any $c \in \mathbb{C}$, the function $j(z) - c$ has exactly one zero modulo $SL_2(\mathbb{Z})$. We apply the relation 3.2 to the modular form $1728g_2^3 - c\Delta \in \mathcal{M}_{12}(SL_2(\mathbb{Z}))$. In this case the term on the left is equal to 1 and hence $1728g_2^3 - c\Delta \in \mathcal{M}_{12}(SL_2(\mathbb{Z}))$ has exactly one zero modulo $SL_2(\mathbb{Z})$. Dividing by Δ we obtain that $j(z) - c = 0$ for exactly one value of $z \in SL_2(\mathbb{Z}) \setminus \mathcal{H}$, as required. □

We can view j as a function of lattices. Since j has weight 0, we can write

$$j(\Lambda) = j(z)$$

for any lattice $\Lambda = [\omega_1, \omega_2]$, with $\omega_1/\omega_2 = z$. Moreover if $\Lambda = c\Lambda'$ for some $c \in \mathbb{C}^*$, then $j(\Lambda) = j(\Lambda')$. Clearly, the converse holds as well, because if

$\Lambda = [\omega_1, \omega_2]$, with $\omega_1/\omega_2 = z$ and $\Lambda' = [\omega_1', \omega_2']$, with $\omega_1'/\omega_2' = z'$, then

$$j(\Lambda) = j(\Lambda') \iff j(z) = j(z') \iff SL_2(\mathbb{Z})z = SL_2(\mathbb{Z})z'.$$

So we have the following

Corollary 3.4.2. *Let Λ and Λ' be two lattices in $\mathbb{C}$. Then $j(\Lambda) = j(\Lambda')$ if and only if $\mathbb{C}/\Lambda \cong \mathbb{C}/\Lambda'$.*

We have seen that for any elliptic curve

$$E : Y^2 = 4X^3 - a_2X - a_3$$

we can always select a lattice Λ such that $a_2 = g_2(\Lambda)$ and $c_3 = g_3(\Lambda)$. We denote by j_E the value of $j(\Lambda)$. Clearly, it is independent from the choice of Λ and it is called the j-**invariant** of the curve. Note that it is defined in term of the coefficients a_2 and a_3 of the equation of E:

$$j_E = 1728\frac{c_2^3}{c_2^3 - 27c_3^2}.$$

We have the following result.

Corollary 3.4.3. *Two elliptic curves E and E' are isomorphic if and only if $j_E = j_{E'}$.*

CHAPTER 4

Hecke Operators

4.1 The Double Coset Operators

Let Γ_1 and Γ_2 be two congruence subgroups of $SL_2(\mathbb{Z})$. Our aim is to transform modular forms with respect to Γ_1 into modular forms with respect to Γ_2. For each $A \in GL_2^+(\mathbb{Q})$ we define the *double coset* as the set

$$\Gamma_1 A \Gamma_2 = \{M_1 A M_2 \mid M_i \in \Gamma_i\}.$$

The group Γ_1 acts on the double coset by left multiplication, so we can write $\Gamma_1 A \Gamma_2$ as a disjoint union of its orbits $\Gamma_1 B_j$, for some representatives $B_j = M_{1,j} A M_{2,j}$. Moreover for each $B \in GL_2^+(\mathbb{Q})$ and $k \in \mathbb{Z}$ we can define the weight-k operator on functions $f : \mathcal{H} \to \mathbb{C}$ by

$$(f[B]_k)(z) = det(B)^{k-1} j(B,z)^{-k} f(Bz), \qquad z \in \mathbb{H}$$

where the factor of automorphy $j(B,z)$ and the action of $GL_2^+(\mathbb{Q})$ on $\mathcal{H}$ are defined as in the case of the modular group $SL_2(\mathbb{Z})$, i.e.

$$Bz = \frac{az+b}{cz+d}, \quad j(B,z) = cz+d, \quad \text{for all } B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbb{Q}).$$

Definition 4.1. Let Γ_1 and Γ_2 be congruence subgroups of $SL_2(\mathbb{Z})$. Let $A \in GL_2^+(\mathbb{Q})$ and $k \in \mathbb{Z}$. The **weight-k $\Gamma_1 A \Gamma_2$ double coset operator** on $\mathcal{M}_k(\Gamma_1(N))$ is given by

$$f[\Gamma_1 A \Gamma_2]_k = \sum_j f[B_j]_k$$

where $\{B_j\}$ are orbit representatives for $\Gamma_1 A \Gamma_2$, as above.

We must show that the double coset operator is well defined. First we will show that if $\Gamma_1 A \Gamma_2 = \bigcup_j \Gamma_1 B_j$ is a disjoint union of its orbits, then this union is finite. We need the following lemmas.

Lemma 4.1.1. *Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$ and let $A \in GL_2^+(\mathbb{Q})$. Then $A^{-1}\Gamma A \cap SL_2(\mathbb{Z})$ is again a congruence subgroup of $SL_2(\mathbb{Z})$.*

Proof. Let N' be a positive integer such that $\Gamma(N') \subseteq \Gamma$, $N'A \in M_2(\mathbb{Z})$ and $N'A^{-1} \in M_2(\mathbb{Z})$. Set $N = N'^3$. We have

$$A\Gamma(N)A^{-1} \subseteq A(I + NM_2(\mathbb{Z}))A^{-1} =$$

$$= I + N'(N'A)M_2(\mathbb{Z})(N'A^{-1}) \subseteq I + N'M_2(\mathbb{Z}).$$

So $A\Gamma(N)A^{-1} \subseteq \Gamma(N')$, since any element in $A\Gamma(N)A^{-1}$ has determinant equal to 1. Thus $\Gamma(N) \subseteq A^{-1}\Gamma(N')A \subseteq A^{-1}\Gamma A$, and so $\Gamma(N) \subseteq A^{-1}\Gamma A \cap SL_2(\mathbb{Z})$. □

Lemma 4.1.2. *Let Γ_1 and Γ_2 be congruence subgroups of $SL_2(\mathbb{Z})$ and let $A \in GL_2^+(\mathbb{Q})$. Set $\Gamma_3 = A^{-1}\Gamma_1 A \cap \Gamma_2$. Then the map*

$$\Gamma_2 \to A\Gamma_2, \qquad M_2 \mapsto AM_2$$

induces a bijection

$$\Gamma_2/\Gamma_3 \to \Gamma_1 A \Gamma_2/\Gamma_1, \qquad M_2 \mapsto \Gamma_1 A M_2.$$

Proof. Clearly the map $\psi : \Gamma_2 \to \Gamma_1 A \Gamma_2/\Gamma_1$ given by $M_2 \mapsto \Gamma_1 A M_2$ is a surjection. We have $\psi(M_2) = \psi(M_2')$ if and only if $M_2^{-1}M_2' \in A^{-1}\Gamma_1 A$. Since $M_2^{-1}M_2' \in \Gamma_2$, the result follows. □

Lemma 4.1.3. *Let G_1 and G_2 be congruence subgroups of $SL_2(\mathbb{Z})$. Then G_1 and G_2 are commensurable, i.e.*

$$[G_1 : G_1 \cap G_2] < +\infty \quad \textit{and} \quad [G_2 : G_1 \cap G_2] < +\infty.$$

Proof. Since G_i are congruence subgroups, there exist two positive integers N_i such that $\Gamma(N_i) \subseteq G_i$. Set $N = lcm(N_1, N_2)$. Then $\Gamma(N) \trianglelefteq G_1 \cap G_2 \trianglelefteq G_i$ for $i = 1, 2$. The map

$$G_i \to SL_2(\mathbb{Z}/N\mathbb{Z})), \quad M \mapsto M \pmod N$$

is a group homomorphism with kernel $\Gamma(N)$, hence the index $[G_i : \Gamma(N)]$ is finite. It follows that $[G_i : G_1 \cap G_2]$ is also finite. □

Now if we put $G_1 = A^{-1}\Gamma_1 A \cap SL_2(\mathbb{Z})$ and $G_2 = \Gamma_2$, by the previous lemmas $|\Gamma_1 A \Gamma_2/\Gamma_1| = |\Gamma_2/\Gamma_3| = [G_2 : G_1 \cap G_2] < +\infty$, so the union $\Gamma_1 A \Gamma_2 = \bigcup_j \Gamma_1 B_j$ is finite.

It remains to prove that the double coset operator is independent from the choice of this set. Let B and B' be two representatives of the same orbit in $\Gamma_1 \setminus \Gamma_1 A \Gamma_2$. Then $B = \bar{M} B'$ for some $\bar{M} \in \Gamma_1$. For each $f \in \mathcal{M}_k(\Gamma_1)$ we have

$$f[B]_k = f[\bar{M}B']_k = f[\bar{M}]_k[B']_k = f[B']_k.$$

It follows that the double coset operator is well defined.

The double coset operator takes modular forms with respect to Γ_1 to modular forms with respect to Γ_2 and takes cusp forms into cusp forms. Let $f \in \mathcal{M}_k(\Gamma_1)$. Since any $M_2 \in \Gamma_2$ permutes the orbit space by right multiplication, if $\{B_j\}$ is a set of representatives so is $\{B_j M_2\}$. It follows that

$$f[\Gamma_1 A \Gamma_2]_k[M_2]_k = \left(\sum_j f[B_j]_k\right)[M_2]_k = \sum_j f[B_j M_2]_k = f[\Gamma_1 A \Gamma_2]_k$$

so the transformed $f[\Gamma_1 A \Gamma_2]_k$ is weight-k invariant under Γ_2. To show holomorphy at the cusps, first note that if functions $g_1, ..., g_m : \mathcal{H} \to \mathbb{C}$ are holomorphic at infinity, i.e. each g_j has a Fourier expansion

$$g_j(z) = \sum_{n \geq o} a_n(g_j) e^{2\pi i n z/h_j},$$

then so is their sum $g_1 + ... + g_m$. In particular its period is $h = lcm(h_1, ..., h_m)$. The following lemma shows us that for any $f \in \mathcal{M}_k(\Gamma_1)$ and for any $M \in GL_2^+(\mathbb{Q})$, the function $g = f[M]_k$ is holomorphic at infinity and if $f \in \mathcal{S}_k(\Gamma_1)$, then g vanishes at infinity.

Lemma 4.1.4. *Any $A \in GL_2^+(\mathbb{Q})$ can be written in the form $A = MA'$, where $M \in SL_2(\mathbb{Z})$ and $A' = r\left(\begin{smallmatrix} u & v \\ 0 & w \end{smallmatrix}\right)$ with $r \in \mathbb{Q}^+$ and $u, v, w \in \mathbb{Z}$ relatively prime. Moreover if Γ is a congruence subgroup and $f \in \mathcal{M}_k(\Gamma)$, then $f[A]_k$ is holomorphic at infinity and if the Fourier expansion for $f[M]_k$ has constant term equal to 0, then so does the Fourier expansion for $f[A]_k$.*

Proof. Let $A = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in GL_2^+(\mathbb{Q})$. If $c' \neq 0$ then we can write $a'/c' = a/c$, where $a, c \in \mathbb{Z}$ are relatively prime. So there exist $\lambda, \mu \in \mathbb{Z}$ such that $\lambda a + \mu c = 1$. If we set $M^{-1} = \left(\begin{smallmatrix} -\lambda & -\mu \\ c & -a \end{smallmatrix}\right)$, then $M^{-1}A = A'$ has the lower left entry equal to zero. Now the first part easily follows.

If $f \in \mathcal{M}_k(\Gamma)$, then $f[M]_k$ has a Fourier expansion for all $M \in SL_2(\mathbb{Z})$. Let h be the period of $f[M]_k$. By using the previous part it is easy to see that $(f[A]_k)(z + wh) = (f[A]_k)(z)$, where w is the lower right entry of M' as in the statement, so $f[A]_k$ is periodic. Moreover $(f[A]_k)(z)$ is bounded for $Im(z) \to \infty$ (because it is $f[M]_k$), so $f[A]_k$ has a Fourier expansion at infinity. Note that

$$\lim_{Im(z)\to\infty} (f[A]_k)(z) = 0 \Longleftrightarrow \lim_{Im(z)\to\infty} (f[M]_k)(z) = 0$$

so $f[A]_k$ vanishes at infinity if and only if $f[M]_k$ vanishes at infinity. So if f is a cusp form, then $f[A]_k$ has a Fourier expansion with the constant term equal to zero. □

Now it is clear that $f[\Gamma_1 A \Gamma_2]_k$ is holomorphic at the cusps, since for any $M \in SL_2(\mathbb{Z})$, $f[\Gamma_1 A \Gamma_2]_k[M]_k = \sum_j f[B_j M]_k$. So

$$[\Gamma_1 A \Gamma_2]_k : \mathcal{M}_k(\Gamma_1) \to \mathcal{M}_k(\Gamma_2)$$

and

$$[\Gamma_1 A \Gamma_2]_k : \mathcal{S}_k(\Gamma_1) \to \mathcal{S}_k(\Gamma_2).$$

There are three special cases for the double coset operator $[\Gamma_1 A \Gamma_2]_k$, and actually any double coset operator is a composition of these:

- when $\Gamma_2 \subseteq \Gamma_1$, if $A = I$, then $f[\Gamma_1 A \Gamma_2]_k = f$ and the double coset operator is the natural inclusion $\mathcal{M}_k(\Gamma_1) \hookrightarrow \mathcal{M}_k(\Gamma_2)$.
- when $A^{-1}\Gamma_1 A = \Gamma_2$, the double coset operator is $f[\Gamma_1 A \Gamma_2]_k = f[A]_k$ and it is a vector-space isomorphism from $\mathcal{M}_k(\Gamma_1)$ to $\mathcal{M}_k(\Gamma_2)$.
- when $\Gamma_1 \subseteq \Gamma_2$, if $A = I$ and $\{M_{2,j}\}$ is a set of representatives for Γ_1/Γ_2, then $f[\Gamma_1 A \Gamma_2]_k = \sum_j f[M_{2,j}]_k$ is a surjection. In fact it is the natural projection of $\mathcal{M}_k(\Gamma_1)$ into $\mathcal{M}_k(\Gamma_2)$.

Let Γ_1, Γ_2 and $A \in GL_2^+(\mathbb{Q})$ be given. We set $\Gamma_3 = A^{-1}\Gamma_1 A \cap \Gamma_2$ as above and $\Gamma_3' = A\Gamma_3 A^{-1} = \Gamma_1 \cap A\Gamma_2 A^{-1}$. Then $\Gamma_3' \subseteq \Gamma_1$, $\Gamma_3' = A\Gamma_3 A^{-1}$ and $\Gamma_3 \subseteq \Gamma_2$ covering the three cases. The corresponding composition of double coset operators is

$$\mathcal{M}_k(\Gamma_1) \hookrightarrow \mathcal{M}_k(\Gamma_3') \xrightarrow{\sim} \mathcal{M}_k(\Gamma_3) \twoheadrightarrow \mathcal{M}_k(\Gamma_2)$$

given by

$$f \longmapsto f \longmapsto f[A]_k \longmapsto \sum_j f[AM_{2,j}]_k$$

which is the coset operator $[\Gamma_1 A \Gamma_2]_k$, since $\{AM_{2,j}\}$ are orbit representatives for $\Gamma_1 A \Gamma_2$ if and only if $\{M_{2,j}\}$ is a set of representatives for the quotient Γ_2/Γ_3.

4.2 The Diamond Operator

In this section we introduce the first type of Hecke operator. For any positive integer N, we denote by G_N the multiplicative group $(\mathbb{Z}/N\mathbb{Z})^*$. We consider the map

$$\Gamma_0(N) \to G_N, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \pmod{N}.$$

It is clearly a surjective group homomorphism with kernel $\Gamma_1(N)$, so it induces an isomorphism $\Gamma_0(N)/\Gamma_1(N) \cong G_N$. Since $\Gamma_0(N)$ acts on $\mathcal{M}_k(\Gamma_1(N))$ and its subgroup $\Gamma_1(N)$ acts trivially, it is well defined an action of G_N on $\mathcal{M}_k(\Gamma_1(N))$, as follows.

Definition 4.2. Let $d \in G_N$. The **diamond operator**, denoted by $\langle d \rangle$, is an operator on $\mathcal{M}_k(\Gamma_1(N))$,

$$\langle d \rangle : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N))$$

given by

$$\langle d \rangle \, f = f[M]_k \quad \text{for any } M = \begin{pmatrix} a & b \\ c & \delta \end{pmatrix} \in \Gamma_0(N) \text{ with } \delta \equiv d \pmod N$$

By using the fact that the operator $\langle d \rangle$ is independent from the choice of $M \in \Gamma_0(N)$, one can easily see that for any $d, e \in G_N$ the relation $\langle d \rangle \langle e \rangle = \langle e \rangle \langle d \rangle = \langle de \rangle$ holds.

Let $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$. We can see the diamond operator as a weight-k double coset operator. Since $\Gamma_1(N)$ is normal in $\Gamma_0(N)$, for each $f \in \mathcal{M}_k(\Gamma_1(N))$ we have

$$f[\Gamma_1 M \Gamma_2]_k = f[M]_k, \qquad \forall M \in \Gamma_0(N).$$

The definition of the diamond operator $\langle d \rangle$ can be extended to $\langle n \rangle$ for all $n \in \mathbb{Z}$. If $GCD(N, n)) = 1$, then $\langle n \rangle$ is determined by n mod N. If $GCD(N, n) > 1$, then we define $\langle n \rangle$ to be the zero operator on $\mathcal{M}_k(\Gamma_1(N))$. For all $n \in \mathbb{Z}$ we have $\langle mn \rangle = \langle m \rangle \langle n \rangle$.

We continue defining the Dirichlet characters, which will allow us to give a decomposition of $\mathcal{M}_k(\Gamma_1(N))$ into a direct sum of its subspaces.

Definition 4.3. Let N be a positive integer. A **Dirichlet character modulo** N is a homomorphism of multiplicative groups

$$\chi : G_N \to \mathbb{C}^*.$$

The set of Dirichlet characters form a group, which is called the **dual group of** G_N and it is denoted by $\hat{G}_N$. The group product in $\hat{G}_N$ is defined by the rule

$$(\chi\psi)(n) = \chi(n)\psi(n), \; \forall \; n \in G_N, \forall \; \chi, \psi \in \hat{G}_N$$

and the identity element is the **trivial character modulo** N, denoted by 1_N, which maps every element of G_N to 1.

Clearly for any Dirichlet character χ, $\chi(n)$ is a complex root of unity for all $n \in G_N$ and the inverse of χ is its complex conjugate $\bar{\chi}(n) = \overline{\chi(n)}$. Moreover the groups G_N and $\hat{G}_N$ are isomorphic in a non-canonical way (see [Ser73]) and in particular the number of Dirichlet characters modulo N is $\varphi(N)$.

Lemma 4.2.1. *Let N be a positive integer. The groups G_N and $\hat{G}_N$ satisfy the orthogonality relations*

$$\sum_{n \in G_N} \chi(n) = \begin{cases} \varphi(N) & \text{if } \chi = 1_N \\ 0 & \text{if } \chi \neq 1_N \end{cases} \quad , \quad \sum_{\chi \in \hat{G}_N} \chi(n) = \begin{cases} \varphi(N) & \text{if } n = 1 \\ 0 & \text{if } n \neq 1 \end{cases}$$

Proof. Clearly we may assume $N \geq 2$. For the first relation, if $\chi = 1_N$ the result immediately follows. If χ is non-trivial then there exists an element $m \in G_N$ such that $\chi(m) \neq 1$. Replacing n by nm in the sum we have

$$\sum_{n \in G_N} \chi(n) = \sum_{n \in G_N} \chi(nm) = \sum_{n \in G_N} \chi(n)\chi(m) = \chi(m) \sum_{n \in G_N} \chi(n).$$

Since $\chi(m) \neq 0, 1$, the result follows.

For the second relation, the case $n = 1$ is trivial. If $n \neq 1$, then there exists a Dirichlet character $\chi' \in \hat{G}_N$ such that $\chi'(n) \neq 1$. Similarly to the previous case, we can conclude after writing

$$\sum_{\chi \in \hat{G}_N} \chi(n) = \sum_{\chi \in \hat{G}_N} \chi\chi'(n) = \sum_{\chi \in \hat{G}_N} \chi(n)\chi'(n) = \chi'(n) \sum_{\chi \in \hat{G}_N} \chi(n).$$

□

Now we introduce some subspaces of $\mathcal{M}_k(\Gamma_1(N))$, on which the diamond operator $\langle d \rangle$ acts as multiplication by $\chi(d)$.

Definition 4.4. Let χ be a Dirichlet character modulo N. The χ-eigenspace of $\mathcal{M}_k(\Gamma_1(N))$ is

$$\mathcal{M}_k(N, \chi) = \{f \in \mathcal{M}_k(\Gamma_1(N)) \mid \langle d \rangle f = \chi(d) f \quad \forall\, d \in G_N\}.$$

As we have already said, we have introduced the Dirichlet characters because they decompose the vector space $\mathcal{M}_k(\Gamma_1(N))$ into a direct sum of its χ-eigenspaces.

Theorem 4.2.2. *Let N be a positive integer. Then*

$$\mathcal{M}_k(\Gamma_1(N)) = \bigoplus_{\chi\in\hat{G}_N} \mathcal{M}_k(N,\chi)$$

and the same result holds for cusp forms.

Proof. For each Dirichlet character $\chi \in \hat{G}_N$ we define an operator on $\mathcal{M}_k(\Gamma_1(N))$:

$$\pi_\chi = \frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1} \langle d\rangle .$$

We need to proceed in several steps. We will show that:

1. π_χ is a projection, i.e. $\pi_\chi^2 = \pi_\chi$;
2. $\pi_\chi(\mathcal{M}_k(\Gamma_1(N))) \subseteq \mathcal{M}_k(N,\chi)$ and $\pi_\chi = Id$ on $\mathcal{M}_k(N,\chi)$, so π_χ is a projection on $\mathcal{M}_k(N,\chi)$;
3. $\sum_{\chi\in\hat{G}_N} \pi_\chi = Id$, i.e. the subspaces $\mathcal{M}_k(N,\chi)$ span $\mathcal{M}_k(\Gamma_1(N))$;
4. $\pi_\chi \circ \pi_{\chi'} = 0$ if $\chi \neq \chi'$, i.e. the subspaces $\mathcal{M}_k(N,\chi)$ are linearly disjoint.

Let $f \in \mathcal{M}_k(\Gamma_1(N))$ and $\chi \in \hat{G}_N$. First we want to compute $\pi_\chi^2(f)$.

$$\pi_\chi^2(f) = \pi_\chi\left(\frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1} \langle d\rangle f\right) =$$

$$= \frac{1}{\varphi(N)} \sum_{e\in G_N} \chi(e)^{-1} \langle d\rangle \left(\frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1} \langle d\rangle f\right) =$$

$$= \frac{1}{\varphi(N)^2} \sum_{(d,e)\in G_N^2} \chi(e)^{-1}\chi(d)^{-1} \langle e\rangle \langle d\rangle f = \frac{1}{\varphi(N)^2} \sum_{(d,e)\in G_N^2} \chi(de)^{-1} \langle de\rangle f =$$

$$= \frac{1}{\varphi(N)^2}\left(\varphi(N) \sum_{c\in G_N} \chi(c)^{-1} \langle c\rangle f\right) = \frac{1}{\varphi(N)} \sum_{c\in G_N} \chi(c)^{-1} \langle c\rangle f = \pi_\chi(f),$$

where for the fifth equality we have used the fact that every product $de \in G_N$ can be written in exactly $\varphi(N)$ different ways:

$$c_1(c_1^{-1}de),\ \ c_2(c_2^{-1}de),\ \dots\ ,\ \ c_{\varphi(N)}(c_{\varphi(N)}^{-1}de), \qquad c_i \in G_N.$$

Now we denote by M_e a matrix in $\Gamma_0(N)$ with lower right entry equal to e. We have

$$\langle e\rangle\,(\pi_\chi(f)) = \langle e\rangle \left(\frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1}\,\langle d\rangle\, f\right) =$$

$$= \frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1}\,\langle de\rangle\, f = \frac{1}{\varphi(N)}\chi(e) \sum_{d\in G_N} \chi(d)^{-1}\chi(e)^{-1}\,\langle de\rangle\, f =$$

$$= \frac{\chi(e)}{\varphi(N)} \sum_{de\in G_N} \chi(de)^{-1}\,\langle de\rangle\, f = \chi(e)\pi_\chi(f)$$

hence $\pi_\chi(\mathcal{M}_k(\Gamma_1(N))) \subseteq \mathcal{M}_k(N,\chi)$. Moreover if $f \in \mathcal{M}_k(N,\chi)$, then

$$\pi_\chi(f) = \frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1}\,\langle de\rangle\, f = \frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1}\,\langle d\rangle\, f =$$

$$= \frac{1}{\varphi(N)} \sum_{d\in G_N} \chi(d)^{-1}\chi(d) f = \frac{1}{\varphi(N)}\varphi(N) f = f$$

so π_χ is a projection on $\mathcal{M}_k(N,\chi)$.

Now, using the orthogonality relations, we obtain

$$\left(\sum_{\chi\in\hat{G}_N} \pi_\chi\right) f = \frac{1}{\varphi(N)} \sum_{\chi\in\hat{G}_N} \sum_{d\in G_N} \chi(d)^{-1}\,\langle d\rangle\, f =$$

$$= \frac{1}{\varphi(N)} \sum_{d\in G_N} \left(\sum_{\chi\in\hat{G}_N} \chi(d)^{-1}\right) \langle d\rangle\, f = \langle 1\rangle\, f = f.$$

Finally, if $\chi, \chi' \in \hat{G}_N$ are two different Dirichlet characters we have

$$\pi_\chi(\pi_{\chi'}(f)) = \pi_\chi \left(\frac{1}{\varphi(N)} \sum_{d\in G_N} \chi'(d)^{-1}\,\langle d\rangle\right) =$$

$$= \frac{1}{\varphi(N)^2} \sum_{(d,e) \in G_N^2} \chi(e)^{-1} \chi'(d)^{-1} \langle de \rangle f =$$

$$= \frac{1}{\varphi(N)^2} \sum_{de \in G_N} \left(\sum_{i=1}^{\varphi(N)} \chi(c_i)^{-1} \chi'(c_i^{-1} de)^{-1} \right) \langle de \rangle f =$$

$$= \frac{1}{\varphi(N)^2} \sum_{de \in G_N} \left(\sum_{i=1}^{\varphi(N)} (\chi^{-1}\chi')(c_i) \right) \chi'(de)^{-1} \langle de \rangle f = 0$$

again by the orthogonality relations. This completes the proof. □

We can lift every Dirichlet character χ modulo N to a function $\chi : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ (which is not a group homomorphism) by setting $\chi(n) = 0$ if $n \in \mathbb{Z}/N\mathbb{Z}$ is a non invertible element. So we can extend it to a function $\chi : \mathbb{Z} \to \mathbb{C}$ defined (with abuse of notation) by $\chi(n) = \chi(n \pmod N)$ for all $n \in \mathbb{Z}$.

4.3 The T_p Operator

The second type of Hecke operator is also a weight-k double coset operator.

Definition 4.5. Let N, p be positive integers, p prime. Put $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$. The T_p operator is defined as

$$T_p = [\Gamma_1 \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_2]_k : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N)).$$

In this case the double coset is

$$\Gamma_1 \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_2 = \{M \in M_2(\mathbb{Z}) \mid M \equiv \begin{pmatrix} 1 & * \\ 0 & p \end{pmatrix} \pmod N,\ det(M) = p\}$$

and in fact in the definition of T_p, the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ can be replaced by any other element in this double coset. The following proposition gives us an explicit representation of T_p.

Proposition 4.3.1. *Let N, p be positive integers, p prime. Put $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$ and $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. The operator $T_p = [\Gamma_1 A \Gamma_2]_k$ is given by*

$$T_p f = \begin{cases} \sum_{j=0}^{p-1} f[B_j]_k & \text{if } p \mid N \\ \sum_{j=0}^{p-1} f[B_j]_k + f[B_\infty]_k & \text{if } p \nmid N \end{cases},$$

where

$$B_j = \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} \quad \text{and} \quad B_\infty = \begin{pmatrix} mp & n \\ Np & 1 \end{pmatrix}, \quad mp - nN = 1.$$

Proof. We need to find a set of orbit representatives for $\Gamma_1 A \Gamma_2$, or equivalently, a set of representatives for the quotient Γ_2/Γ_3, where $\Gamma_3 = A^{-1}\Gamma_1 A \cap \Gamma_2$. Let

$$\Gamma^0(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \pmod{p} \right\}.$$

Then $\Gamma_3 = \Gamma_1(N) \cap \Gamma^0(p)$ (it is just a straightforward calculation). If $p \mid N$, we want to show that the representatives for the quotient Γ_2/Γ_3 are

$$M_{2,j} = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix} \quad \text{for} \quad 0 \le j < p.$$

Let $M_2 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_2$ be given. Since a is of the form $kN + 1$ and $p \mid N$, it follows that $p \nmid a$. By setting $j \equiv ba^{-1} \pmod{p}$ we have $b - ja \equiv 0 \pmod{p}$. So it is easy to check that $M_2 M_{2,j}^{-1} \in \Gamma_3$, i.e. $M_2 \in \Gamma_3 M_{2,j}$. Since the representatives $M_{2,j}$ are all distinct, if $p \mid N$ we are done.

If $p \nmid N$ then p may divide a (the upper left entry of M_2). In this case $b - aj$ cannot be $0 \pmod{p}$ for any j, because if $p \mid b$, then $p \mid ad - bc = 1$ (note that $p \mid a$ only if $p \nmid N$). To complete the set of representatives, we put

$$M_{2,\infty} = \begin{pmatrix} mp & n \\ N & 1 \end{pmatrix}$$

where $mp - nN = 1$. It is easy to see that if $p \mid a$, then $M_2 M_{2,\infty}^{-1} \in \Gamma_3$, i.e. $M_2 \in \Gamma_3 M_{2,\infty}$, as needed. Once again the representatives are all distinct.

Therefore, if $p \mid N$ the corresponding orbit representatives are

$$B_j = AM_{2,j} = \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}, \text{ for } 0 \leq j < p$$

and if $p \nmid N$, we have to add

$$B_\infty = AM_{2,\infty} = \begin{pmatrix} mp & n \\ Np & p \end{pmatrix}, \text{ where } mp - nN = 1.$$

□

The T_p operator commutes with the diamond operator. To see this, let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ and $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$ as above. Note that $MAM^{-1} \in \Gamma_1 A \Gamma_2$ for any $M \in \Gamma_0(N)$. If $\{B_j\}$ is a set of orbit representatives for the double coset, since $\Gamma_1(N)$ is normal in $\Gamma_0(N)$, we have

$$\bigcup_j \Gamma_1 B_j = \Gamma_1 A \Gamma_2 = \Gamma_1 MAM^{-1} \Gamma_2 = M \Gamma_1 A \Gamma_2 M^{-1} =$$

$$= M \left(\bigcup_j \Gamma_1 B_j \right) M^{-1} = \bigcup_j \Gamma_1 M B_j M^{-1},$$

hence

$$\bigcup_j \Gamma_1 M B_j = \bigcup_j \Gamma_1 B_j M.$$

Thus for any $f \in \mathcal{M}_k(\Gamma_1(N))$ and for any $M \in \Gamma_0(N)$ with lower right entry $\delta \equiv d \bmod N$, we can write

$$\langle d \rangle T_p f = \sum_j f[B_j M]_k = \sum_j f[M B_j]_k = T_p \langle d \rangle f.$$

The next result describes the effect of T_p on Fourier coefficients.

Proposition 4.3.2. *Let $f \in \mathcal{M}_k(\Gamma_1(N))$. Since f has a Fourier expansion, we can write*

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

1. *If 1_N is the trivial character modulo N, then T_pf has a Fourier expansion*

$$(T_pf)(z) = \sum_{n=0}^{\infty} a_{np}(f)q^n + 1_N(p)p^{k-1}\sum_{n=0}^{\infty} a_n(\langle p\rangle f)q^{np} =$$

$$= \sum_{n=0}^{\infty}(a_{np}(f) + 1_N(p)p^{k-1}a_{n/p}(\langle p\rangle f))q^n$$

where $a_{n/p} = 0$ when $p \nmid n$ and

$$a_n(T_pf) = a_{np}(f) + 1_N(p)p^{k-1}a_{n/p}(\langle p\rangle f).$$

2. *Let χ be a Dirichlet character and $f \in \mathcal{M}_k(N,\chi)$. Then also $T_pf \in \mathcal{M}_k(N,\chi)$ and its Fourier expansion is*

$$(T_pf)(z) = \sum_{n=0}^{\infty} a_{np}(f)q^n + \chi(p)p^{k-1}\sum_{n=0}^{\infty} a_n(f)q^{np} =$$

$$= \sum_{n=0}^{\infty}\left(a_{np}(f) + \chi(p)p^{k-1}a_{n/p}(f)\right)q^n.$$

In particular

$$a_n(T_pf) = a_{np}(f) + \chi(p)p^{k-1}a_{n/p}(f).$$

Proof. For part (a), let B_j be as in *Proposition* 4.3.1 for $0 \le j < p$. We compute

$$f[B_j]_k(z) = \frac{1}{p}f\left(\frac{z+j}{p}\right) = \frac{1}{p}\sum_{n=0}^{\infty} a_n(f)e^{2\pi i n(z+j)/p} = \frac{1}{p}\sum_{n=0}^{\infty} a_n(f)q_p^n\mu_p^{nj}$$

where $q_p = e^{2\pi i z/p}$ and $\mu_p = e^{2\pi i/p}$. Since the geometric sum

$$\sum_{j=0}^{p-1}\mu_p^{nj} = \begin{cases} p & \text{if } p \mid n \\ 0 & \text{if } p \nmid n \end{cases},$$

we obtain

$$\sum_{j=0}^{p-1} f[B_j]_k(z) = \sum_{n\equiv_p 0} a_n(f)q_p^n = \sum_{n=0}^{\infty} a_{np}(f)q^n.$$

This is $(T_pf)(z)$ when $p \nmid N$. When $p \mid N$ we must include the term

$$f[B_\infty]_k(z) = p^{k-1}(\langle p \rangle f)(pz) = p^{k-1}\sum_{n=0}^{\infty} a_n(\langle p \rangle f)q^{np},$$

where we have used the fact that

$$B_\infty = \begin{pmatrix} m & n \\ N & p \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Now part (b) follows from the relation $\langle d \rangle (T_pf) = T_p(\langle d \rangle f)$. □

Corollary 4.3.3. *Let p and q be prime. Then $T_pT_q = T_qT_p$.*

Proof. Since $\mathcal{M}_k(\Gamma_1(N)) = \bigoplus \mathcal{M}_k(N,\chi)$, it suffices to verify the relation for $f \in \mathcal{M}_k(N,\chi)$ (note that $\langle d \rangle$ and T_p preserve this decomposition). By the previous proposition we have

$$a_n(T_p(T_qf)) = a_{np}(T_qf) + \chi(p)p^{k-1}a_{n/p}(T_qf) =$$

$$= a_{npq}(f) + \chi(q)q^{k-1}a_{np/q}(f) + \chi(p)p^{k-1}(a_{nq/p}(f) + \chi(q)q^{k-1}a_{n/pq}(f)) =$$

$$= a_{npq}(f) + \chi(q)q^{k-1}a_{np/q}(f) + \chi(p)p^{k-1}a_{nq/p}(f) + \chi(pq)(pq)^{k-1}a_{n/pq}(f),$$

and this is symmetric in p and q. □

As in the case of the diamond operator, we can extend the definition of T_p to T_n for all $n \in \mathbb{Z}$. First set $T_1 = Id$. Then for any prime powers, we define inductively

$$T_{p^r} = T_pT_{p^{r-1}} - p^{k-1}\langle p \rangle T_{p^{r-2}}, \quad \text{for } r \geq 2.$$

One can shows inductively that $T_{p^r}T_{p^s} = T_{p^s}T_{p^r}$ for distinct primes and for any $r, s \geq 1$. So it is well defined for all $n \in \mathbb{Z}$

$$T_n = \prod_{i=1}^{m} T_{p_i^{e_i}} \quad \text{where } n = p_1^{e_1} \cdot ... \cdot p_m^{e_m}.$$

Clearly $T_mT_n = T_nT_m$ and in particular, $T_{nm} = T_nT_m$ if $GCD(m,n) = 1$.

The Fourier coefficient formulas generalize to

Proposition 4.3.4. *Let $f \in \mathcal{M}_k(\Gamma_1(N))$ have a Fourier expansion*

$$f(z) = \sum_{m=0}^{\infty} a_m(f) q^m, \qquad q = e^{2\pi i z}.$$

Then for all $n \in \mathbb{Z}^+$, $T_n f$ has a Fourier expansion

$$(T_n f)(z) = \sum_{m=0}^{\infty} a_m(T_n f) q^m$$

where

$$a_m(T_n f) = \sum_{d|(n,m)} d^{k-1} a_{mn/d^2}(\langle d \rangle f). \tag{4.1}$$

In particular, if $f \in \mathcal{M}_k(N, \chi)$ then

$$a_m(T_n f) = \sum_{d|(n,m)} \chi(d) d^{k-1} a_{mn/d^2}(f). \tag{4.2}$$

4.4 The Petersson Scalar Product

In order to study the space of cusp forms $\mathcal{S}_k(\Gamma_1(N))$ further, we want to endow it with an inner product, which will be defined as an integral. Before giving the definition, we want to preface some remarks.

In the definition we will replace the Euclidean measure $dxdy$ on $\mathbb{C}$ by the *hyperbolic measure*

$$d\mu(z) = \frac{dxdy}{y^2}, \quad \text{where } z = x + iy \in \mathcal{H}.$$

This makes sense because the hyperbolic measure is invariant under the action of $GL_2^+(\mathbb{Q})$, i.e. under a change of variable of the form

$$z \mapsto Mz = \frac{az+b}{cz+d}, \qquad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbb{Q}).$$

In general, if we make a differentiable change of complex variable $z' = u(z)$, then the area element $dxdy$ near z is multiplied by $|u'(z)|^2$. By using the

relations
$$\frac{dMz}{dz} = \frac{det(M)}{(cz+d)^2}, \qquad \frac{Im(Mz)}{Im(z)} = \frac{det(M)}{|cz+d|^2},$$
it follows that $d\mu(Mz) = d\mu(z)$ for all $M \in GL_2^+(\mathbb{Q})$ and $z \in \mathcal{H}$.

Let $\mathcal{F}$ be the fundamental domain of $\mathcal{H}/SL_2(\mathbb{Z})$ as in *Theorem* 2.2.1 and let $\mathcal{F}^* = \mathcal{F} \cup \{\infty\}$ denote the fundamental domain of $\mathcal{H}^*/SL_2(\mathbb{Z})$. Then for any continuous and bounded function $\psi : \mathcal{H} \to \mathbb{C}$ and any $M \in SL_2(\mathbb{Z})$, the integral
$$\int_{\mathcal{F}^*} \psi(Mz) d\mu(z)$$
converges. In fact
$$\int_{\mathcal{F}^*} \psi(Mz) d\mu(z) \leq \sup_{z \in \mathcal{H}}(\psi) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\frac{\sqrt{3}}{2}}^{\infty} \frac{dxdy}{y^2} = \frac{2}{\sqrt{3}} \sup_{z \in \mathcal{H}}(\psi).$$

Let $\Gamma \subseteq SL_2(\mathbb{Z})$ be a congruence subgroup and set $\bar{\Gamma} = \{\pm I\}\Gamma$. Then we can write $SL_2(\mathbb{Z})$ as a disjoint union of the form
$$SL_2(\mathbb{Z}) = \bigcup_j \bar{\Gamma} M_j$$
where $\{M_j\} \subseteq SL_2(\mathbb{Z})$ is a set of representatives for the orbit space $\bar{\Gamma} \backslash SL_2(\mathbb{Z})$. Note that $\bigcup_j M_j(\mathcal{F}^*)$ is a fundamental domain for Γ, and so it can be identified with the modular curve $X(\Gamma)$. Moreover if the function ψ is Γ-invariant, then the sum $\sum_j \int_{\mathcal{F}^*} \psi(M_j z) d\mu(z)$ is independent of the choice of the set $\{M_j\}$ and it is equal to $\int_{\cup_j M_j(\mathcal{F}^*)} \psi(z) d\mu(z)$. Thus we can define
$$\int_{X(\Gamma)} \psi(z) d\mu(z) = \int_{\cup_j M_j(\mathcal{F}^*)} \psi(z) d\mu(z) = \sum_j \int_{\mathcal{F}^*} \psi(M_j z) d\mu(z) \qquad (4.3)$$
for every continuous, bounded and Γ-invariant function $\psi : \mathcal{H} \to \mathbb{C}$. In particular, if $\psi = 1$ we define the **volume** of $X(\Gamma)$ as
$$V_\Gamma = \int_{X(\Gamma)} d\mu(z).$$
Clearly $V_\Gamma = [SL_2(\mathbb{Z}) : \bar{\Gamma}]\ V_{SL_2(\mathbb{Z})}$ and one can easily compute that $V_{SL_2(\mathbb{Z})} = \pi/3$.

For any $f, g \in \mathcal{S}_k(\Gamma)$ consider the function $\psi(z) = f(z)\overline{g(z)}Im(z)^k$. It is clearly continuous and for any $M \in \Gamma$ we have

$$\psi(Mz) = f(Mz)\overline{g(Mz)}Im(Mz)^k =$$

$$= (f[M]_k)(z)j(M,z)^k\overline{(g[M]_k)(z)j(M,z)^k}Im(z)^k|j(M,z)|^{-2k} =$$

$$= (f[M]_k)(z)\overline{(g[M]_k)(z)}Im(z)^k = f(z)\overline{g(z)}Im(z)^k = \psi(z)$$

so ψ is Γ-invariant. To show that ψ is bounded on $\mathcal{H}$, it suffices to verify that for any $M \in SL_2(\mathbb{Z})$, $\psi(Mz)$ is bounded on $\mathcal{F}$. By writing the Fourier expansions

$$(f[M]_k)(z) = \sum_{n=1}^{\infty} a_n(f[M]_k)q_h^n, \quad \text{and} \quad (g[M]_k)(z) = \sum_{n=1}^{\infty} a_n(g[M]_k)q_h^n$$

where, as usual, $q_h = e^{2\pi i z/h}$ for some $h \in \mathbb{Z}$, we obtain that as $Im(z) \to \infty$

$$\psi(Mz) = \mathcal{O}(q_h)^2 Im(z)^k.$$

Since $|q_h| = e^{-2\pi Im(z)/h}$, $\psi(Mz) \to 0$ as $Im(z) \to \infty$ and so $\psi(M\cdot)$ is bounded on $\mathcal{F}$ as desired. Note that in fact ψ is bounded on $\mathcal{H}$ whenever the product fg vanishes at each cusp. We will not work with this additional generality.

Now we can endow the space of cusp forms with an inner product.

Definition 4.6. Let $\Gamma \subseteq SL_2(\mathbb{Z})$ be a congruence subgroup. The **Petersson inner product**

$$\langle \cdot\, , \cdot \rangle_\Gamma : \mathcal{S}_k(\Gamma) \times \mathcal{S}_k(\Gamma) \to \mathbb{C}$$

is given by

$$\langle f, g \rangle_\Gamma = \frac{1}{V_\Gamma} \int_{X(\Gamma)} f(z)\overline{g(z)}Im(z)^k d\mu(z).$$

The Petersson inner product is an *Hermitian inner product* on $\mathcal{S}_k(\Gamma)$. In fact

- it is linear in f and anti-linear in g: for any $\lambda_1, \lambda_2 \in \mathbb{C}$ $\langle \lambda_1 f_1 + \lambda_2 f_2, g \rangle_\Gamma = \lambda_1 \langle f_1, g \rangle_\Gamma + \lambda_2 \langle f_2, g \rangle_\Gamma$ and $\langle f, \lambda_1 g_1 + \lambda_2 g_2 \rangle_\Gamma = \bar{\lambda_1} \langle f, g_1 \rangle_\Gamma + \bar{\lambda_2} \langle f, g_2 \rangle_\Gamma$;
- it is antisymmetric: $\langle f, g \rangle = \overline{\langle g, f \rangle}$;

- it is positive definite: $\langle f, f\rangle > 0$ for $f \neq 0$.

Now suppose that $\Gamma' \leq \Gamma$ is another congruence subgroup. Any two modular forms f, g in $\mathcal{S}_k(\Gamma)$ may also be considered as modular forms in $\mathcal{S}_k(\Gamma')$. Then $\langle \cdot\ , \cdot\rangle_\Gamma = \langle \cdot\ , \cdot\rangle_{\Gamma'}$ on $\mathcal{S}_k(\Gamma)$. To see this write $SL_2(\mathbb{Z}) = \bigcup_i \bar{\Gamma} A_i$ and $\bar{\Gamma} = \bigcup_j \bar{\Gamma}' B_j$ as disjoint unions of cosets, with $\{A_i\} \subseteq SL_2(\mathbb{Z})$ and $\{B_j\} \subseteq \Gamma$. Then $SL_2(\mathbb{Z}) = \bigcup_j \bigcup_i \bar{\Gamma}' B_j A_i$ and for any $f, g \in \mathcal{S}_k(\Gamma)$ we have

$$\langle f, g\rangle_{\Gamma'} = \frac{1}{V_{\Gamma'}} \int_{X(\Gamma')} f(z)\overline{g(z)} Im(z)^k d\mu(z) =$$

$$= \frac{1}{V_{\Gamma'}} \sum_j \sum_i \int_{\mathcal{F}} f(B_j A_i z)\overline{g(B_j A_i z)} Im(B_j A_i z)^k d\mu(z) =$$

$$= \frac{1}{V_{\Gamma'}} \sum_j \sum_i \int_{\mathcal{F}} f(A_i z)\overline{g(A_i z)} Im(A_i z)^k d\mu(z) =$$

$$= \frac{[\bar{\Gamma} : \bar{\Gamma}']}{V_{\Gamma'}} \sum_i \int_{\mathcal{F}} f(A_i z)\overline{g(zA_i)} Im(A_i z)^k d\mu(z) =$$

$$= \frac{[\bar{\Gamma} : \bar{\Gamma}']}{V_{\Gamma'}} \int_{X(\Gamma)} f(z)\overline{g(z)} Im(z)^k d\mu(z) =$$

$$= \frac{1}{V_\Gamma} \int_{X(\Gamma)} f(z)\overline{g(z)} Im(z)^k d\mu(z) = \langle f, g\rangle_\Gamma$$

as required.

Now we can consider the vector space $\mathcal{S}_k(\Gamma_1(N))$ as an *inner product space*. Recall that if T is a linear operator on an inner product space V, then its adjoin T^* is a linear operator on V such that

$$\langle Tu, v\rangle = \langle u, T^* v\rangle \qquad \forall u, v \in V.$$

The operator T is called *normal* when $TT^* = T^*T$. One can compute the *adjoins* of the Hecke operators in $\mathcal{S}_k(\Gamma_1(N))$ (see [DS05]). For p prime, with $p \nmid N$, we have

$$\langle p\rangle^* = \langle p\rangle^{-1} \quad \text{and} \quad T_p^* = \langle p\rangle^{-1} T_p.$$

It follows that the Hecke operators $\langle n\rangle$ and T_n are normal for $GCD(n, N) = 1$.

Definition 4.7. An **eigenform for an Hecke operator** T is a modular form $f \in \mathcal{S}_k(\Gamma_1(N))$ such that f is an eigenvector for T.

From the Spectral Theorem of linear algebra, given a commuting family of normal operators on a finite-dimensional inner product space, the space has a orthogonal basis of simultaneous eigenvectors for the operators. Thus we have the following result.

Theorem 4.4.1. *The space $\mathcal{S}_k(\Gamma_1(N))$ has an orthogonal basis of simultaneous eigenforms for the family of Hecke operators $\{\langle n \rangle, T_n \mid GCD(n, N) = 1\}$.*

4.5 Old Forms and New Forms

In this section we will explain how we can move between levels, lifting the level from $\frac{N}{d}$ to N. We will also distinguish between the part of $\mathcal{S}_k(\Gamma_1(N))$ "coming" from lower levels and the "new" one.

Lemma 4.5.1. *Let $M \in GL_2^+(\mathbb{Q})$ and let Γ_1, Γ_2 be two congruence subgroups.*

1. *Suppose that $M\Gamma_2M^{-1} \subseteq \Gamma_1$. If $f \in \mathcal{M}_k(\Gamma_1)$, then $f[M]_k \in \mathcal{M}_k(\Gamma_2)$. The same result holds for cusp forms.*
2. *Let N, d be two positive integers, with $d \mid N$, and let $f \in \mathcal{M}_k(\Gamma_1(\frac{N}{d}))$. Then $g(z) = f(dz) \in \mathcal{M}_k(\Gamma_1(N))$. The same result holds for cusp forms.*

In particular, if we set

$$M_d = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} \in GL_2^+(\mathbb{Q}),$$

then we have an injective linear map from $\mathcal{S}_k(\Gamma_1(\frac{N}{d}))$ to $\mathcal{S}_k(\Gamma_1(N))$:

$$[M_d]_k : \mathcal{S}_k(\Gamma_1(\frac{N}{d})) \to \mathcal{S}_k(\Gamma_1(N)), \qquad f(z) \mapsto (f[M_d]_k)(z) = d^{k-1}f(dz).$$

Proof. By *Lemma* 4.1.4, if $f \in \mathcal{M}_k(\Gamma_1)$, then $(f[M]_k)[A]_k$ is holomorphic at infinity for all $A \in SL_2(\mathbb{Z})$ and if f is a cusp form, then $(f[M]_k)[A]_k$ has a

Fourier expansion with the constant term equal to 0 for all $A \in SL_2(\mathbb{Z})$ (to see this replace M in the lemma by MA).

To show the first statement it remains to prove that if $f \in \mathcal{M}_k(\Gamma_1)$, then $f[M]_k$ is weight-k invariant under Γ_2. But it is easy, because for any $M_2 \in \Gamma_2$ there exists $M_1 \in \Gamma_1$ such that $MM_2 = M_1M$. So we have

$$f[M]_k[M_2]_k = f[MM_2]_k = f[M_1M]_k = f[M_1]_k[M]_k = f[M]_k.$$

Now set $\Gamma_1 = \Gamma_1(\frac{N}{d})$ and $\Gamma_2 = \Gamma_1(N)$. Since

$$\begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} \Gamma_2 \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix}^{-1} \subseteq \Gamma_1$$

the second part follows by setting $M = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. □

Let N, N' be two integers, with $N' \mid N$. Clearly $\mathcal{S}_k(\Gamma_1(N')) \subseteq \mathcal{S}_k(\Gamma_1(N))$, but the previous lemma gives us another way to embed $\mathcal{S}_k(\Gamma_1(N'))$ into $\mathcal{S}_k(\Gamma_1(N))$. In fact we can use the map $[M_d]_k$, where d is any factor of N/N'. Now we can distinguish the part of $\mathcal{S}_k(\Gamma_1(N))$ coming from lower levels.

Definition 4.8. For each divisor d of N, let i_d be the map

$$\mathcal{S}_k(\Gamma_1(\frac{N}{d})) \times \mathcal{S}_k(\Gamma_1(\frac{N}{d})) \longrightarrow \mathcal{S}_k(\Gamma_1(N))$$

given by

$$(f, g) \longmapsto f + g[M_d]_k.$$

The **subspace of old forms at level** N is

$$\mathcal{S}_k(\Gamma_1(N))^{old} = \sum_{p|N} i_p(\mathcal{S}_k(\Gamma_1(\frac{N}{p})) \times \mathcal{S}_k(\Gamma_1(\frac{N}{p})))$$

and the **subspace of new forms at level** N is defined as

$$\mathcal{S}_k(\Gamma_1(N))^{new} = (\mathcal{S}_k(\Gamma_1(N))^{old})^{\perp},$$

i.e. the orthogonal complement of the subspace of old forms with respect to the Petersson inner product.

In the definition of old forms, we can extend the sum over all divisors of N, without making difference. In fact, if $d \mid N$ and $p \mid d$, then $\mathcal{S}_k(\Gamma_1(\frac{N}{d})) \subseteq \mathcal{S}_k(\Gamma_1(\frac{N}{p}))$. So changing the definition of $\mathcal{S}_k(\Gamma_1(N))^{old}$ by extending the sum over all divisors of N does not add anything to the space of old forms.

The subspaces $\mathcal{S}_k(\Gamma_1(N))^{old}$ and $\mathcal{S}_k(\Gamma_1(N))^{new}$ are stable under the Hecke operators and in particular they have orthogonal bases of eigenforms for the Hecke operators $\{\langle n \rangle, T_n \mid GCD(n, N) = 1\}$.

We can normalize the map $i_d : f(z) \mapsto d^{k-1} f(dz)$ by defining $\iota_d = d^{1-k} i_d$. Let $f \in Sk(\Gamma_1(N))$ be an element of the form $\sum_{p|N} \iota_p f_p$, with each $f_p \in \mathcal{S}_k(\Gamma(\frac{N}{p}))$. If f has a Fourier expansion $f(z) = \sum_{n\geq 1} a_n(f) q^n$, then $a_n(f) = 0$ for all n such that $GCD(n, N) = 1$, because the map ι_d acts on Fourier expansions as

$$\iota_d : \sum_{n=1}^{\infty} a_n q^n \longmapsto \sum_{n=1}^{\infty} a_n q^{dn}.$$

An important result due to Atkin and Lehner (see [AL70]), known as the *Main Lemma*, says us that the converse holds as well:

Theorem 4.5.2 (Main Lemma). *If $f \in \mathcal{S}_k(\Gamma_1(N))$ has a Fourier expansion $f(z) = \sum_{n\geq 1} a_n(f) q^n$ with $a_n(f) = 0$ whenever $GCD(n, N) = 1$, then f takes the form $\sum_{p|N} \iota_p f_p$, with each $f_p \in \mathcal{S}_k(\Gamma(\frac{N}{p}))$.*

We don't prove it, but we use the Main Lemma in the following.

4.6 Eigenforms

Recall that the space $\mathcal{S}_k(\Gamma_1(N))$ has an orthogonal basis of simultaneous eigenforms for the family of Hecke operators $\{\langle n \rangle, T_n \mid GCD(n, N) = 1\}$. In this section we will eliminate the restriction $GCD(n, N) = 1$ for $\mathcal{S}_k(\Gamma_1(N))^{new}$.

Definition 4.9. A **Hecke eigenform** or simply an **eigenform** is a non-zero modular form $f \in \mathcal{M}_k(\Gamma_1(N))$ that is an eigenform for the operators T_n and $\langle n \rangle$ for all $n \in \mathbb{Z}^+$. We say that $f = \sum_{n\geq 0} a_n(f) q^n$ is **normalized** when $a_1(f) = 1$. A **newform** is a normalized eigenform in $\mathcal{S}_k(\Gamma_1(N))^{new}$.

Theorem 4.6.1. *Let $f \in \mathcal{S}_k(\Gamma_1(N))^{new}$ be a non-zero eigenform for the Hecke operators T_n and $\langle n \rangle$ for all n with $GCD(n, N) = 1$. Then*

1. *f is a Hecke eigenform and it is a newform up to a suitable scalar multiple;*

2. *if g satisfies the same conditions as f and has the same T_n-eigenvalues, then $g = cf$ for some $c \in \mathbb{C}$.*

The set of newforms is an orthogonal basis for the space $\mathcal{S}_k(\Gamma_1(N))^{new}$. Each such newform lies in an eigenspace $\mathcal{S}_k(N, \chi)$ for some Dirichlet character χ and its Fourier coefficients are its T_n-eigenvalues. That is, every newform satisfies $T_n f = a_n(f) f$ for all $n \in \mathbb{Z}^+$.

Proof. First recall that when $GCD(n, N) \geq 2$, the diamond operator $\langle n \rangle$ is defined to be the zero operator, so if $f \in \mathcal{S}_k(\Gamma_1(N))$ is an eigenform for each operator $\langle n \rangle$ with $GCD(n, N) = 1$, then f is an eigenform for the operators $\langle n \rangle$ for all $n \in \mathbb{Z}^+$. For such f there exist eigenvalues $d_n \in \mathbb{C}$ such that $\langle n \rangle f = d_n f$. By using $d_{nm} f = \langle nm \rangle f = \langle n \rangle (\langle m \rangle f) = d_n d_m f$, it is clear that the map $n \mapsto d_n$ defines a Dirichlet character χ and $f \in \mathcal{S}_k(N, \chi)$. Conversely if $\chi \in \hat{G}_N$ is a Dirichlet character and $f \in \mathcal{S}_k(N, \chi)$, then f is an eigenform for all the operators $\langle n \rangle$, since $\langle n \rangle f = \chi(n) f$ for all $n \in \mathbb{Z}^+$.

Now let $f \in \mathcal{S}_k(\Gamma_1(N))$ be an eigenform for the Hecke operators T_n and $\langle n \rangle$ with $GCD(n, N) = 1$ and let $c_n \in \mathbb{C}$ such that $T_n f = c_n f$. We have

$$c_n a_1(f) = a_1(T_n f) = a_n(f)$$

when $GCD(n, N) = 1$ (the second equality follows from (4.2)). Thus if $a_1(f) = 0$, then $f \in \mathcal{S}_k(\Gamma_1(N))^{old}$ by the Main Lemma. So if f is a non-zero form in $\mathcal{S}_k(\Gamma_1(N))^{new}$, then $a_1(f) \neq 0$ and we may assume that f is normalized to have $a_1(f) = 1$. For any $m \in \mathbb{Z}^+$ define

$$g_m = T_m f - a_m(f) f \in \mathcal{S}_k(\Gamma_1(N))^{new}.$$

Clearly it is an eigenform for the Hecke operators T_n and $\langle n \rangle$ for all $n \in \mathbb{Z}^+$

such that $GCD(n, N) = 1$. We have

$$a_1(g_m) = a_1(T_m f) - a_1(a_m(f)f) = a_m(f) - a_m(f) = 0$$

so g_m is also an old form, hence $g_m = 0$. This shows that $T_m f = a_m(f)f$ for all $m \in \mathbb{Z}^+$.

To conclude the proof we must prove that the set of newforms is linearly independent. Suppose that there exists a non-trivial linear relation

$$\sum_{i=1}^{n} b_i f_i = 0, \qquad b_i \in \mathbb{C}^*$$

with n minimal, necessarily greater than 2. Let m be a positive integer. Applying $T_m - a_m(f_1)$ we obtain

$$\sum_{i=2}^{n} b_i(a_m(f_i) - a_m(f_1))f_i = 0.$$

This relation must be trivial, so $a_m(f_i) = a_m(f_1)$. Since m is arbitrary, $f_i = f_1$ for all i, giving a contradiction. □

Corollary 4.6.2. *If $f \in \mathcal{S}_k(\Gamma_1(N))$ is an eigenform, then f is old or new.*

Proof. Let $f = g + h$ be an eigenform, with $g \in \mathcal{S}_k(\Gamma_1(N))^{old}$ and $h \in \mathcal{S}_k(\Gamma_1(N))^{new}$. If $h = 0$, then $f = g$ lies in the space of old forms and we are done. So assume $h \neq 0$. By the Main Lemma $a_1(g) = 0$ and so $a_1(f) = a_1(h)$. By the methods seen above $a_1(f)$ must be non-zero and we may assume $a_1(f) = a_1(h) = 1$. So $T_n f = a_n(f)f$ for all $n \in \mathbb{Z}^+$ (again by the methods seen before). Recall that T_n and $\langle n \rangle$ preserves the decomposition of $\mathcal{S}_k(\Gamma_1(N))$ as a direct sum of old and new spaces, hence applying T_n and $\langle n \rangle$, we have that g and h are two eigenforms and in particular

$$T_n h = a_n(f)h \qquad \text{and} \qquad T_n g = a_n(f)g.$$

But since $h \neq 0$ is a newform, $T_n h = a_n(h)h$ for all $n \in \mathbb{Z}^+$. It follows that $a_n(f) = a_n(h)$ and so $f = h$ is a newform (note that we have normalized f to have $a_1(f) = 1$, so we have just proved that f lies in $\mathcal{S}_k(\Gamma_1(N))^{new}$). □

We conclude this chapter by stating a result which is the converse of what we have seen above.

Theorem 4.6.3. *Let $g \in \mathcal{S}_k(\Gamma_1(N))$ be a normalized eigenform. Then there exists a newform $f \in \mathcal{S}_k(\Gamma_1(M))^{new}$ for some $M \mid N$ such that $a_p(f) = a_p(g)$ for all $p \nmid N$.*

CHAPTER 5

Applications and Outstanding Problems

5.1 The Ramanujan Function

The Ramanujan τ function is an arithmetic function that, in the early twentieth century, evoked the curiosity of the great Indian mathematician Srinivasa Ramanujan. He proved or conjectured many of its properties and that is why this function is named after him. In this section we define this function and we show some of the most important results concerning it.

Definition 5.1. The **Ramanujan τ function** is the function $\tau : \mathbb{N} \to \mathbb{Z}$ defined by the following identity:

$$\sum_{n \geq 1} \tau(n) q^n = (2\pi)^{-12} \Delta(z).$$

In other words, $\tau(n)$ is defined to be the n^{th}-coefficient in the Fourier expansion of $(2\pi)^{-12}\Delta(z)$.

The first property that we are going to prove is a famous congruence due to Ramanujan in 1916.

Theorem 5.1.1.

$$\tau(n) \equiv \sigma_{11}(n) \quad (mod\, 691)$$

Proof. We shall deduce the statement from the following identity:

$$E_{12} - E_6^2 = c\Delta, \qquad \text{where } c = \frac{2^6 \cdot 3^5 \cdot 7^2}{(2\pi)^{12} \cdot 691}.$$

Since both sides of the equality lies in the one-dimensional space $\mathcal{S}_{12}(SL_2(\mathbb{Z}))$ it suffices to check the equality for the first non-constant terms in their Fourier expansions. For the Eisenstein series we have

$$E_4 = 1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n$$

$$E_6 = 1 - 504\sum_{n=1}^{\infty}\sigma_5(n)q^n$$

$$E_{12} = 1 + \frac{65520}{691}\sum_{n=1}^{\infty}\sigma_{11}(n)q^n$$

and after doing a little algebra, we obtain

$$E_{12} - E_6^2 = 0 + \frac{2^6 3^5 7^2}{691}q + \text{highter terms...}$$

and

$$\Delta = 0 + (2\pi)^{12}q + \text{highter terms...},$$

so the equality follows. Now, by expanding $E_{12} - E_6^2 = c\Delta$, we find an expression for $\tau(n)$ in terms of $\sigma_{11}(n)$ and $\sigma_5(n)$:

$$\tau(n) = \frac{65}{756}\sigma_{11}(n) + \frac{691}{756}\sigma_5(n) - \frac{691}{3}\sum_{d|n}\sigma_5(d)\sigma_5(\frac{n}{d}).$$

To conclude the proof, it suffices to reduce this equality modulo 691. □

There is another important property of the τ function, which was proved by Mordell in 1917.

Theorem 5.1.2. *The τ function is a multiplicative function, i.e.*

$$\tau(mn) = \tau(m)\tau(n) \quad \textit{if } GCD(m,n) = 1.$$

Proof. The theorem easily follows by using the theory of Hecke operators. Recall that $\mathcal{S}_{12}(SL_2(\mathbb{Z})) = \mathbb{C}\Delta$ and the operators T_m preserve $\mathcal{S}_{12}(SL_2(\mathbb{Z}))$ for all $m \in \mathbb{Z}^+$. Since $\tau(1) = 1$, $f = \sum_{n=1}^{\infty}\tau(n)q^n$ is a normalized eigenform for T_m for all $m \in \mathbb{Z}^+$. So we have $T_m f = \tau(m)f$ and using the fact that $T_{mn} = T_m T_n$ for m, n relatively prime, we obtain $\tau(mn) = \tau(m)\tau(n)$. □

There are some outstanding problems concerning the τ function. For example, in 1947, D.H.Lehner conjectured that $\tau(n) \neq 0$ for all $n \in \mathbb{N}$. This assertion, also known as *Lehmer's conjecture*, was verified by Derickx, van Hoeij, and Zeng for all $n < 816212624008487344127999$.

5.2 Fermat's Last Theorem

In 1637 the Frenchman Pierre de Fermat, a lawyer and amateur mathematician, obtained a copy of the Greek work by Diophantus, Arithmetica. While he was studying the second volume, Fermat read remarks, problems and solutions about Pythagorean theorem and Pythagorean triples. Playing with Pythagorean equation, he created a variant of this one. Instead of considering $X^2 + Y^2 = Z^2$, he considered the equation $X^n + Y^n = Z^n$ for $n \in \mathbb{Z}$, $n \geq 3$.

In the margin of his copy of Arithmetica, close to the 8^{th} problem, Fermat wrote " It is impossible to separate a cube into two cubes, a fourth power into two fourth powers or generally any power above the second into two powers of the same degree. I have discovered a truly marvellous demonstration which this margin is too narrow to contain ". In modern notation, Fermat states that the equation $X^n + Y^n = Z^n$ has no positive integer solutions for all integers n greater than 2.

Fermat left no details of his proof apart from the special case $n = 4$. He proved this case using a particular method known as "infinite descent". With the special case $n = 4$ proven, the problem was to prove the theorem for exponents n that are prime numbers. Around 1753, Leonhard Euler adapted the Fermat's method to the case $n = 3$. In 1825, Gustav Lejeune Dirichlet and Adrien-Marie Legendre proved (independently) the case $n = 5$ and fourteen years later, by adapting a new technique due to Sophie Germain, Gabriel Lamè showed the case $n = 7$. In 1846, Ernst Eduard Kummer proved the theorem for all regular primes, but a proof for all primes seemed to be inaccessible.

The proof of Fermat's Last Theorem for all n, was finally accomplished by Andrew Wiles in 1994 by using the theory of elliptic curves and modular forms (a paper [Wil95] was published in May 1995). The most significant steps of the proof are outlined below.

Around 1955 Japanese mathematicians Goro Shimura and Yutaka Taniyama conjectured a strong link between elliptic curves and modular forms. This result is known as the Taniyama–Shimura-Weil conjecture or the modularity theorem and apparently, it has no connection to Fermat's Last Theorem. A statement of the conjecture is the following.

Theorem 5.2.1 (Taniyama-Shimura-Weil conjecture)**.** *A series of the form $L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ with $a_n \in \mathbb{Z}$ is the L-function $L_E(s)$ of an elliptic curve $E(\mathbb{Q})$ of conductor N if and only if $L(s) = L_f(s)$ is the L-function of a newform of weight 2 for $\Gamma_0(N)$.*

Let's see what an L-function and the conductor of an elliptic curve are. For cusp forms we have a simple definition.

Definition 5.2. Let $f \in \mathcal{S}_k(\Gamma_1(N))$ and let $f(z) = \sum_{n=1}^{\infty} a_n(f)q^n$ be its Fourier expansion (with $q = e^{2\pi iz}$ as usual). The **L-function attached to** f is defined as

$$L_f(s) = \sum_{n=1}^{\infty} a_n(f)\frac{1}{n^s},$$

where $s \in \mathbb{C}$ is a complex variable.

To define the L-function of an elliptic curve is a little more complicated. Consider an elliptic curve $E(\mathbb{Q})$ defined over $\mathbb{Q}$ by an equation

$$E : Y^2 = X^3 + aX + b, \qquad a, b \in \mathbb{Q}$$

and let p be a prime. We can reduce the equation modulo p and consider the curve $E(\mathbb{F}_p)$ but, after reduction, the condition $\Delta = 4a^3 + 27b^2 \neq 0$ could fail. If it happens, then there exist singular points on the reduced curve $E(\mathbb{F}_p)$, i.e. the elliptic curve $E(\mathbb{F}_p)$ is singular. We have the following definition.

Definition 5.3. Let $E(\mathbb{Q})$ be an elliptic curve defined by an equation $E : Y^2 = X^3 + aX + b$ $a, b \in \mathbb{Q}$ and let p be a prime. We say that

- $E(\mathbb{Q})$ has a **good reduction** at p if $X^3 + aX + b$ has three distinct roots mod p;

- $E(\mathbb{Q})$ has an **additive reduction** at p if $X^3 + aX + b$ has a triple root mod p;
- $E(\mathbb{Q})$ has a **split multiplicative reduction** at p if $X^3 + aX + b$ has a double root α mod p and $\alpha \in \mathbb{F}_p$;
- $E(\mathbb{Q})$ has a **non-split multiplicative reduction** at p if $X^3 + aX + b$ has a double root α mod p and $\alpha \notin \mathbb{F}_p$;

We say that the elliptic curve $E(\mathbb{Q})$ is **semistable** if it has a good or multiplicative reduction at all primes.

We are now able to give the following definition.

Definition 5.4. Let $E(\mathbb{Q})$ be an elliptic curve over $\mathbb{Q}$. We define coefficients a_n for all $n \geq 1$ as follows. For p prime, we define

$$a_p := \begin{cases} p+1-\#E(\mathbb{F}_p) & \text{if } E \text{ has a good reduction at } p \\ 0 & \text{if } E \text{ has an additive reduction at } p \\ 1 & \text{if } E \text{ has a split multiplicative reduction at } p \\ -1 & \text{if } E \text{ has a non-split multiplicative reduction at } p \end{cases}$$

where $\#E(\mathbb{F}_p)$ denotes the number of points of the elliptic curve $E(\mathbb{F}_p)$. If $n = p^r$, we define a_{p^r} recursively by using the relation

$$a_p a_{p^r} = a_{p^{r+1}} + p a_p^{r-1}$$

if $E(\mathbb{Q})$ has a good reduction at p and

$$a_{p^r} = (a_p)^r$$

otherwise. Finally we put $a_1 = 1$ and if $GCD(n, m) = 1$, we define

$$a_{mn} = a_m a_n.$$

The **L-function attached to** E is the series

$$L_E(s) = \sum_{n=1}^{\infty} a_n \frac{1}{n^s},$$

where $s \in \mathbb{C}$ is a complex variable.

We spend a few words about the convergence of these series. For the L-function attached to an elliptic curve E, by using the recursive definition of a_n, we can rewrite $L_E(s)$ as an Euler product

$$L_E(s) := \prod_{bad\ p} (1 - a_p p^{-s})^{-1} \prod_{good\ p} (1 - a_p p^{-s} + p^{1-2s})^{-1}.$$

The Hasse's Theorem (see [Was03]) states that $|a_p| \leq 2\sqrt{p}$ and it easly implies that $L_E(s)$ converges for $Re(s) > 3/2$.

The study of the convergence of the L-function attached to a cusp form $f \in \mathcal{S}_k(\Gamma_1(N))$ requires some estimates for the Fourier coefficients of f. If $g(q) = f(z) = \sum_{n=1}^{\infty} a_n(f) q^n$ is the holomorphic function on the open unit disk $\{q \in \mathbb{C} \mid |q| < 1\}$, then by Cauchy's formula

$$a_n(f) = \int_{|q|=r} \frac{g(q)}{q^{n+1}} dq$$

for any $r \in (0,1)$. If we write $q = e^{2\pi i(x+iy)}$, we have for any $y > 0$

$$a_n(f) = \int_0^1 f(x+iy) e^{-2\pi i(x+iy)} dx$$

and after the change of variable $y = 1/n$ we obtain

$$a_n(f) = e^{2\pi} \int_0^1 f(x + \frac{i}{n}) e^{-2\pi i n x} dx.$$

Since $Im(z)^{k/2}|f(z)|$ is bounded on $\mathcal{H}$, by estimating the last integral we obtain that $|a_n(f)| \leq C n^{k/2}$ and hence $L_f(s)$ converges for $Re(s) > k/2 + 1$, since $|a_n(f) n^{-s}| = \mathcal{O}(n^{k/2 - Re(s)})$.

Definition 5.5. We say that an elliptic curve $E(\mathbb{Q})$ over $\mathbb{Q}$ is **modular** if there exists a cusp form f such that

$$L_E(s) = L_f(s).$$

It remains to define the conductor of an elliptic curve. Since Wiles proved the Taniyama-Shimura-Weil conjecture for semistable elliptic curves, we define the conductor only in this context. For the general case we need a technical invariant, which is not easy to describe.

Definition 5.6. Let $E(\mathbb{Q})$ be a semistable elliptic curve. The conductor of E is defined to be

$$N_E = \prod_{p|\Delta} p$$

where Δ is the discriminant of the elliptic curve.

In 1986 Frey pointed out a connection between Fermat's Last Theorem and the Taniyama-Shimura-Weil conjecture. Let ℓ be a prime and suppose that there exist integers a, b, c such that

$$a^\ell + b^\ell = c^\ell, \qquad abc \neq 0.$$

Frey suggested that the semistable elliptic curve

$$E_{Frey} : Y^2 = X(X - a^\ell)(X + b^\ell)$$

has such restrictive properties that it cannot exist, and therefore there cannot exists integers a, b, c such that $a^\ell + b^\ell = c^\ell$ and $abc \neq 0$.

Two years later, recurring to the theory of Galois representations, Ribet showed that E_{Frey} cannot be modular ([Rib90a]) and therefore the Taniyama-Shimura-Weil conjecture implies Fermat's Last Theorem. In 1994 Wiles proved Taniyama-Shimura-Wiles conjecture for semistable elliptic curves and in particular he showed what Frey had conjectured: E_{Frey} cannot exist and so Fermat's Last Theorem is true.

List of Symbols

$a_n(f)$	n^{th} Fourier coefficient of f
B_k	Bernoulli number
$\langle d \rangle$	diamond Hecke operator
Δ	discriminant function
$E[N]$	N-torsion subgroup of E
$E(\mathbb{K})$	the group of $\mathbb{K}$-rationals point of E
$E_m(z)$	normalized Eisenstein series of weight m
$\mathcal{F}$	fundamental domain
$\langle f, g \rangle_\Gamma$	Petersson inner product with respect to Γ
$f[\Gamma_1 A \Gamma_2]_k$	weight-k double coset operator
$G_m(z)$	Eisenstein series of weight m
G_N	multiplicative group $(\mathbb{Z}/N\mathbb{Z})^*$
$\Gamma_0(N)$	congruence subgroup of $SL_2(\mathbb{Z})$
$\Gamma_1(N)$	congruence subgroup of $SL_2(\mathbb{Z})$
$\Gamma(N)$	principal congruence subgroup of level N
$GL_2(\mathbb{Q})^+$	general linear group of 2-by-2 matrices with positive determinant and rational entries
$\mathcal{H}$	upper half plane
h_s	width of a cusp
h_z	period of an elliptic point
$j(z)$	modular invariant
Λ	lattice in $\mathbb{C}$

$\mathcal{M}_k(\Gamma)$	modular forms of weight k with respect to Γ
$\mathcal{M}_k(N,\chi)$	χ-eigenspace of $\mathcal{M}_k(\Gamma_1(N))$
$\mathcal{M}_k(SL_2(\mathbb{Z}))$	modular forms of weight k
μ_n	n^{th} root of unity $e^{2\pi i/n}$
$\mathbb{P}^n_{\mathbb{K}}$	n-dimensional projective space over $\mathbb{K}$
$\wp(z)$	Weierstrass $\wp$-function
q	$e^{2\pi iz}$, point of the punctured unit disk
q_h	$e^{2\pi iz/h}$, point of the punctured unit disk
$S_0(N)$	moduli space for $\Gamma_0(N)$
$S_1(N)$	moduli space for $\Gamma_1(N)$
$S(N)$	moduli space for $\Gamma(N)$
$\mathcal{S}_k(\Gamma)$	cusp forms of weight k with respect to Γ
$\mathcal{S}_k(SL_2(\mathbb{Z}))$	cusp forms of weight k
$\mathcal{S}_k(\Gamma_1(N))^{new}$	new forms at level N
$\mathcal{S}_k(\Gamma_1(N))^{old}$	old forms at level N
$\sigma_k(n)$	arithmetic function
T_n	Hecke operator
T^*	adjoin of T
V_Γ	volume of $X(\Gamma)$
$Y(\Gamma)$	modular curve for Γ
$Y_0(N)$	modular curve for $\Gamma_0(N)$
$Y_1(N)$	modular curve for $\Gamma_1(N)$
$Y(N)$	modular curve for $\Gamma(N)$
$X(\Gamma)$	compact modular curve for Γ
$X_0(N)$	compact modular curve for $\Gamma_0(N)$
$X_1(N)$	compact modular curve for $\Gamma_1(N)$
$X(N)$	compact modular curve for $\Gamma(N)$

Bibliography

[AL70] A.O.L.Atkin and J.Lehner, “Hecke operators on $\Gamma_0(m)$”, *Mathematische Annalen*, **185:134-160** (1970).

[Apo76] T.M.Apostol, “Introduction to Analytic Number Theory”, *Undergraduate Texts in Mathematics, Springer*, (1976).

[AZ95] A. N. Andrianov and V. G. Zhuravlev, “Modular Forms and Hecke Operators”, *Translations of Mathematical Monographs, American Mathematical Society*, **145** (1995).

[BK07] B.C. Berndt and M.I.Knopp, “Hecke's Theory of Modular Forms and Dirichlet Series”, *Monographs in Number Theory, World Scientific Pub Co Inc.*, (2007).

[CSS97] G.Cornell, J.H.Silverman and G.Stevens, “Modular Forms and Fermat's Last Theorem”, *Papers from a conference held Aug.9-18,1995, at Boston University, Spingler-Verlag, New York*, (1997).

[DS05] F.Diamond and J.Shurman, “A First Course in Modular Forms”, *Graduate Texts in Mathematics, Spingler, New York*, **228** (2005).

[Kil08] L.J.P.Kilford, “Modular Forms: A Classical and Computational Introduction”, *Imperial College Press*, (2008).

[Kob93] N.Koblitz, “Introduction to Elliptic Curves and Modular Forms”, *Graduate Texts in Mathematics, Spingler-Verlag, New York*, **97** (1993).

[Lan76] S.Lang, “Introduction to Modular Forms”, *Grundl. Math. Wiss., Spingler-Verlag, Berlin, New York*, **222** (1976).

[Lan87] S.Lang, “Elliptic Functions”, *Graduate Texts in Mathematics, Spingler-Verlag, Berlin, New York*, **112** (1987).

[Lan93] S.LANG, “Complex Analysis”, *Graduate Texts in Mathematics, Spingler-Verlag, New York*, **103** (1993).

[Loz11] Á.LOZANO-ROBLEDO, “Elliptic Curves, Modular Forms, and Their L-functions”, *Student Mathematical Library, American Mathematical Society*, (2011).

[Mil90] J.S.MILNE, “Modular Functions and Modular Forms”, (1990) **http://www.jmilne.org/math** .

[Miy89] T.MIYAKE, “Modular Forms”, *Springer Monographs in Mathematics, Springer Berlin Heidelberg*, (1989).

[Mur93] M.RAM MURTY, “Topics in Number Theory”, *Lectures at Mehta Research Institute, Allahabad*, **211 002** (1993).

[Rib90] KENNETH RIBET, “ From the Taniyama-Shimura conjecture to Fermat’s last theorem”, *Annales de la faculté des sciences de Toulouse* 5^e *Série*, **11 (1): 116–139** (1990).

[Rib90a] KENNETH RIBET, “On modular representation of $Gal(\bar{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms”, *Inven.math.*, **100: 431-476** (1990).

[Ser73] J.P.SERRE, “A Course in Arithmetic”, *Graduate Texts in Mathematics, Springer-Verlag*, **7** (1973).

[Shi11] G.SHIMURA, “Modular Forms: Basics and Beyond”, *Springer Monographs in Mathematics, Springer*, (2011).

[Shi71] G.SHIMURA, “Introduction to the Arithmetic Theory of Automorphic Functions”, *Iwanami Shoten and Princeton University Press*, (1971).

[Sil09] J.H.SILVERMAN, “The Arithmetic of Elliptic Curves”, *Graduate Texts in Mathematics, Springer*, **106** (2009).

[Was03] L.C.WASHINGTON, “Elliptic Curves, Number Theory and Cryptography”, *CRC Press Series on Discrete Mathematics and its Applications, Chapman &Hall/CRC, New York*, **114** (2003).

[Wil95] ANDREW WILES, “Modular Elliptic Curves and Fermat’s Last Theorem”, *Annals of Mathematics*, **142 (3): 443–551** (1995).