Elliptic Curves and Modular Forms

by

Dongho Tommy Kim

A thesis submitted in partial satisfaction of the

requirements for the degree of

Honors in Mathematics

in the

Undergraduate Division

of the

University of California, Berkeley

Advisor:

Dr. Zeyu Liu

Fall 2025

Acknowledgments

I would like to express my special thanks of gratitude to my thesis advisor, Dr. Zeyu Liu, who offered me the opportunity to do this as well as his invaluable guidance and support through both my thesis work and my developement as a student of mathematics. I would like to thank Seewoo Lee for enlightening discussions and for his comments that greatly improved this paper. I would like to thank Professor Martin Olsson for serving as my faculty sponsor.

I would also like to thank my parents, my sister, and my grandparents for their constant support throughout my years of study.

ELLIPTIC CURVES AND MODULAR FORMS

DONGHO KIM

Abstract. The modularity theorem states that every elliptic curve defined over $\mathbb{Q}$ is modular (specifically, arises from a weight 2 modular form of a given level). This paper provides an expository overview of the theorem, with an emphasis on eigenforms and L-functions. After reviewing the necessary background on elliptic curves and modular forms, we build up the theory of Hecke operators and eigenforms as well as the theory of L-functions associated to both elliptic curves and modular forms. We then discuss the connections between these objects and finally, state a version of the modularity theorem.

Contents

1. Introduction 2
2. Elliptic Curves 4
3. Modular Forms 10
4. Hecke Operators, Eigenforms, and Newforms 14
5. Modularity Theorem 19
References 20

1. Introduction

1.1. **Overview.** This paper serves an overview of the modularity theorem, and our ultimate goal is to state two versions of the theorem. We start with a brief overview of the theory of elliptic curves and modular forms, with the aim of understanding the correspondence between elliptic curves and modular forms. However, we will be omitting some important details in the theory of elliptic curves (such as isogenies, the Tate module, and Weil pairing) and the theory of modular forms (such as the dimension formulas), as our singular aim will be to state the modularity theorem. For more details, the reader can refer to [Sil09] and [DS05], respectively. The heart of the correspondence between elliptic curves and modular forms lies in the theory of Hecke operators, eigenforms, and L-functions. Ultimately, we will be building towards the following statements:

Theorem 1.1.1 (Modularity Theorem, Version 1)**.** *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E. Then, for some newform $f \in \mathcal{S}_2(\Gamma_0(N_E))$,*

$$a_p(f) \;=\; a_p(E) \quad \textit{for all primes } p.$$

Theorem 1.1.2 (Modularity Theorem, Version 2)**.** *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E. Then, for some newform $f \in \mathcal{S}_2(\Gamma_0(N_E))$,*

$$L(s, f) \;=\; L(s, E).$$

The motivation for this thesis is to develop a clear understanding of the statement of the modularity theorem and its role in modern number theory. This interest is partly motivated by the proof of Fermat's Last Theorem, which relied on early cases of modularity for elliptic curves (specifically for semi-stable elliptic curves) and was later generalized to the full modularity theorem.

1.2. **History.** The original statement of the modularity theorem can be traced back to the 1955 International Symposium on Algebraic Number Theory held in Tokyo and Nikkō, Japan, where Yutaka Taniyama posed a set of open problems. The twelfth problem, specifically, was a problem that related the L-functions of elliptic curves to the L-functions of certain modular forms [Lan95]. The problem was later refined jointly by Taniyama and Goro Shimura, leading to the Taniyama-Shimura conjecture (also known as the modularity conjecture): "every elliptic curve over $\mathbb{Q}$ is modular".

In 1967, André Weil separately explored the relationship between the zeta functions of elliptic curves and Mellin transform of modular forms [Wei67]; in the same paper, he introduced the notion of the conductor of an elliptic curve, which he suggested was equal to the level of a modular form corresponding to the elliptic curve, and he also restated the question of the modularity of elliptic curves over $\mathbb{Q}$. Weil's contribution lent credence to the Taniyama-Shimura conjecture. That is to say, his work provided additional evidence supporting the conjecture by establishing a connection between elliptic curves and modular forms.

The modularity conjecture gained significant attention in the 1980s when Gerhard Frey suggested in 1986 that it implies Fermat's Last Theorem [Fre86]. Frey, expanding on work done by Yves Hellegouarch [Hel72], associated hypothetical solutions of the Fermat equation with an elliptic curve. Specifically, if p is a odd prime and a, b, c are positive integers such that $a^p + b^p = c^p$, then the corresponding Frey curve (or Frey-Hellegouarch curve) is the elliptic curve

$$y^2 = x(x - a^p)(x + b^p),$$

or equivalently

$$y^2 = x(x - a^p)(x - c^p).$$

This nonsingular algebraic curve of genus one defined over $\mathbb{Q}$ has the property that its projective completion is an elliptic curve over $\mathbb{Q}$. In 1985, Jean-Pierre Serre formalized Frey's observations into what is now known as Serre's conjecture [Ser87], where he used the theory of Galois representations to prove that if a statement he called ϵ (or the ϵ-conjecture) were true, then the Frey curve could not be modular. Ken Ribet subsequently proved Serre's epsilon conjecture (a result now known as Ribet's theorem) in 1986 [Rib90], thereby establishing that the Taniyama-Shimura conjecture implies Fermat's Last Theorem.

While this result was significant, it was largely believed that the Taniyama-Shimura conjecture was inaccessible. However, in 1993, Andrew Wiles announced a proof of Fermat's Last Theorem by proving the Taniyama-Shimura conjecture for a special class of elliptic curves known as semistable elliptic curves. While Wiles's initial proof contained a gap, he (along with Richard Taylor) was able to fix this gap by late 1994 and published a complete proof through [Wil95] and [TW95]. Afterwards, Christophe Breuil, Brian Conrad,

Fred Diamond, and Richard Taylor extended Wiles's techniques to prove the full modularity theorem in 2001 [BCDT01].

2. Elliptic Curves

This section is largely based on [Sil09], [CSS97], and [Mil21]. We begin with an introduction to the theory of elliptic curves, as they are one of the two objects of study in the modularity theorem. Accordingly, our discussion will be framed around elliptic curves defined over the rational numbers $\mathbb{Q}$; however, we note that many of the fundamental concepts remain valid over arbitrary fields. Let us now begin with the basic definition of an elliptic curve.

2.1. Definition and Basic Properties.

Definition 2.1.1. An *elliptic curve* is a pair (E, O), where E is a nonsingular curve of genus one and $O \in E$. In general, when the point O is understood, we simply refer to the elliptic curve as E. The curve E is said to be *defined over the field* K, denoted E/K, if E is defined over K as a curve and $O \in E(K)$, where $E(K)$ is the set of K-rational points of E.

While this is the formal definition of an elliptic curve, we introduce an equivalent definition that is more practical for our purposes. Let us first state a result that connects elliptic curves to plane cubic curves via Weierstrass equations. This result establishes the fact that every elliptic curve can be represented as a plane cubic and that every smooth Weierstrass plane cubic curve is an elliptic curve.

Proposition 2.1.2. *Let E/K be an elliptic curve. Then, we have the following results.*

(a) There exist functions $x, y \in K(E)$ such that the map

$$\varphi : E \to \mathbb{P}^2, \quad \varphi \;=\; [x, y, 1]$$

gives an isomorphism of E/K onto a curve given by a Weierstrass equation of the form

$$C \;:\; Y^2 + a_1XY + a_3Y \;=\; X^3 + a_2X^2 + a_4X + a_6$$

with coefficients $a_1, \ldots, a_6 \in K$ and satisfying $\varphi(O) = [0 : 1 : 0]$. The functions x and y are called **Weierstrass coordinates** *for the elliptic curve E.*

(b) Any two Weierstrass equations for E as in **(a)** *are related by a linear change of variables of the form*

$$X \;=\; u^2X' + r, \quad Y \;=\; u^3Y' + su^2X' + t,$$

where $u \in K^*$ and $r, s, t \in K$.

(c) Conversely, every smooth cubic curve C given by a Weierstrass equation of the form in **(a)** *is an elliptic curve defined over K with base point $O = [0 : 1 : 0]$, which is the point at infinity in the projective plane $\mathbb{P}^2$.*

Remark 2.1.3. A detailed proof for Proposition 2.1.2 can be found at [Sil09, Proposition III.3.1]. The main idea behind the proof is to use the Riemann-Roch theorem to construct functions $x, y \in K(E)$ with poles of order 2 and 3 at O, respectively, and to observe that $\mathcal{L}(6(O))$ has dimension 6 while containing the seven functions

$$1,\; x,\; y,\; x^2,\; xy,\; y^2,\; x^3.$$

We also use the fact the invariant differential associated to a Weierstrass equation for a curve is holomorphic and nowhere vanishing, together with the Riemann-Roch theorem, to establish that the curve, along with the base point $[0 : 1 : 0]$ is an elliptic curve.

From Proposition 2.1.2, we have a practical definition of an elliptic curve as follows.

Definition 2.1.4. An *elliptic curve* E/K is a nonsingular projective plane curve over a field K of the form

$$E \;:\; y^2 + a_1xy + a_3y \;=\; x^3 + a_2x^2 + a_4x + a_6, \tag{2.1.1}$$

with coefficients $a_1, \ldots, a_6 \in K$, together with the point at infinity $O = [0 : 1 : 0]$.

Given that $\operatorname{char} K \neq 2$, a linear change of variables can be performed to obtain an equivalent Weierstrass equation of the form

$$E \;:\; y^2 \;=\; 4x^3 + b_2x^2 + b_4x + b_6,$$

where

$$b_2 \;=\; a_1^2 + 4a_4, \quad b_4 \;=\; 2a_4 + a_1a_3, \quad b_6 \;=\; a_3^2 + 4a_6.$$

Remark 2.1.5. We can also define the following quantities associated to the elliptic curve E:

$$\begin{aligned} b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2, \\ c_4 &= b_2^2 - 24b_4, \\ c_6 &= -b_2^3 + 36b_2b_4 - 216b_6, \\ \Delta &= -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6, \\ j(E) &= c_4^3/\Delta, \\ \omega &= \frac{dx}{2y + a_1x + a_3} = \frac{dy}{3x^2 + 2a_2x + a_4 - a_1y}. \end{aligned}$$

We note that $4b_8 = b_2b_6 - b_4^2$ and that $1728\Delta = c_4^3 - b_6^2$.

Definition 2.1.6. We denote Δ as the *discriminant* of the elliptic curve E, j as the *j-invariant* of E, and ω as the *invariant differential* of E.

Remark 2.1.7. Although the quantities Δ, j, and ω are defined here in the context of an elliptic curve E, the formulas above make sense for *any* plane curve given by a Weierstrass equation of the form (2.1.1). In this more general setting, the curve may be singular; nevertheless, these invariants remain well-defined in the coefficients a_i and encodes geometric information about the curve.

We further observe that when $\operatorname{char} K$ is not equal to 2 or 3, we can remove the a_1xy, a_3y, x^2 terms by performing a linear change of variables, resulting in a simplified Weierstrass equation of the form

$$E : y^2 = x^3 + Ax + B. \tag{2.1.2}$$

Then, it is clear from Remark 2.1.5 that, in this case, we have

$$\Delta = -16(4a^3 + 27B^2) \quad \text{and} \quad j(E) = -1728 \cdot \frac{64A^3}{\Delta} = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2}.$$

Recall from Definition 2.1.1 that an elliptic curve is nonsingular. Geometrically, this means that the curve has no cusps, self-intersections, or isolated points. Algebraically, this nonsingularity condition holds if and only if the discriminant of the elliptic curve is nonzero, i.e. $4A^3 + 27B^2 \neq 0$. We can, in fact, use the discriminant to prove a more general result regarding the nonsingularity of elliptic curves.

Proposition 2.1.8. *Let E/K be a curve given by a Weierstrass equation of the form in* (2.1.1) *or* (2.1.2) *(note that E need not be an elliptic curve as we are missing the condition that E is nonsingular). Then, the curve E satisfies the following:*

(a) E is nonsingular if and only if $\Delta \neq 0$,
(b) E has a node if and only if $\Delta = 0$ and $c_4 \neq 0$,
(c) E has a cusp if and only if $\Delta = c_4 = 0$,

where c_4 is as defined in Remark 2.1.5.

2.2. The Group Law on Elliptic Curves.

We now describe the group structure on the set of points of an elliptic curve E/K, where K is a field such that $\operatorname{char} K$ is not 2 or 3. The existence of this group structure can be characterized in various ways, and we will first present a geometric description of the group law, which more intuitively illustrates the addition of points on E.

2.2.1. *Geometric Description.* Let $P = (x_P, y_P), Q = (x_Q, y_Q)$ be points on E.

(a) If $P \neq Q$, let ℓ be the line passing through P and Q.
(b) If $P = Q$, let ℓ be the tangent line to E at P, defined by the derivative

$$\frac{dy}{dx} = \frac{3x_P^2 + A}{2y_P}.$$

In either case (a) or (b), we observe through Bézout's theorem that the line ℓ and curve E must intersect in exactly three points (counting multiplicity). As P and Q are two of these points, there necessarily exists a third point of intersection, which we denote by $R = (x_R, y_R)$. Let $P + Q$ be defined as the reflection of R across the x-axis, i.e.

$$P + Q \coloneqq (x_R, -y_R).$$

When adding a point $P = (x_P, y_P)$ with O, we first obtain the straight vertical line crossing the point P, which intersects E at the points (x_P, y_P) and $(x_P, -y_P)$. Reflecting across the x-axis, it is clear that $P + O = P$. As the group operation is clearly commutative, we conclude that O is the identity element of the group $E(K)$.

We also observe that three collinear points on E sum to zero:

$$P + Q + R = O \quad \Longleftrightarrow \quad P, Q, R \text{ are collinear on } E.$$

Example illustrations of the group law on elliptic curves can be seen at Figure 1. This geometric construction

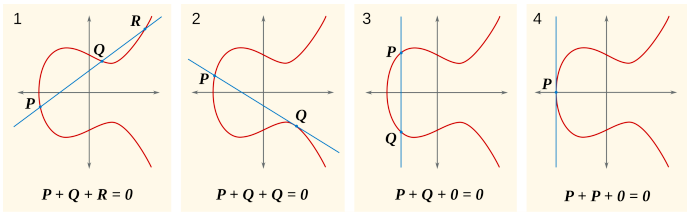

FIGURE 1. Examples of the group law on elliptic curves [Wik07].

is helpful not only for visualizing the group law, but also for writing down explicit formulas for the elliptic curves.

2.2.2. *Algebraic Formulas.* Let us now present explicit algebraic formulas for the group law on our curve E. Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ with $P, Q \neq O$. We then consider three cases:

(a) Case 1: $P \neq Q$. The line ℓ through P and Q has slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Then, the coordinates of $P + Q = (x_3, y_3)$ are given by

$$x_3 = m^2 - x_1 - x_2 \quad \text{and} \quad y_3 = -y_1 + m(x_1 - x_3).$$

(b) Case 2: $P = Q$. Let ℓ be the tangent line at P, whose slope is

$$m = \frac{3x_1^2 + A}{2y_1}.$$

Then

$$x_3 = m^2 - 2x_1, \qquad y_3 = -y_1 + m(x_1 - x_3).$$

(c) Case 3: $P + (-P)$. If $P = (x, y)$, then

$$-P = (x, -y),$$

since $P + (-P) = O$. Thus, the group operation is compatible with the geometric reflection law:

$$R = P + Q \quad \Longleftrightarrow \quad P, Q, -R \text{ are collinear.}$$

Remark 2.2.1. A question that may arise is why we care about the group structure and group law on elliptic curves. The answer lies in the arithmetic data encoded by elliptic curves. Note that we will be working with elliptic curves over $\mathbb{Q}$ in the coming sections and that elliptic curves are, by definition, nonsingular. However, it is often useful to reduce the coefficients of E modulo p for various primes p and to consider E as a curve defined over the finite field $\mathbb{F}_p$.

Let us recall Δ, the discriminant of E. For almost all primes p, the reduced curve $E \mod p$ remains nonsingular, and thus, E modulo p will be an elliptic curve over $\mathbb{F}_p$. On the other hand, if p divides Δ, then the reduced curve E modulo p will have discriminant zero, and thus will be singular. As such, even when we are working with nonsingular elliptic curves over $\mathbb{Q}$, singular curves naturally arise. What we will observe is that the group law on elliptic curves can be extended to study singular curves as well, which will be crucial in understanding certain arithmetic invariants of elliptic curves (namely the conductor, which will be defined later).

2.3. **Singular Curves.** Let E now be a (possibly) singular curve given a Weierstrass equation, either of the form (2.1.1) or (2.1.2).

Proposition 2.3.1. *If the discriminant Δ of E is zero, then E is singular and thus has a singular point, denoted by S, and we know this singular point is unique.*

Proof. We know that if $\Delta = 0$, then E is singular by Proposition 2.1.8. To see that the singular point is unique, let us suppose for contradiction there exist two distinct singular points S_1 and S_2 on E. Let ℓ be the line that passes through both S_1 and S_2. It is clear that ℓ would intersect E in at least 4 points, when counting with multiplicity, but this contradicts Bézout's theorem, which states that a line and a cubic curve can intersect in exactly 3 points (counting multiplicity). We thus have the uniqueness of the singular point on E. □

As noted in Proposition 2.1.8, the singular point S on E is either a node or a cusp. We can, in fact, separately study the nonsingular points on E.

Definition 2.3.2. The nonsingular locus of E is denoted

$$E_{\text{ns}} \coloneqq \{P \in E : P \text{ is a nonsingular point of } E\}.$$

If E is defined over K, then $E_{\text{ns}}(K)$ denotes the set of nonsingular points of $E(K)$.

We can now see the importance of the group law, as we can extend it to the set E_{ns} of nonsingular points on E.

Proposition 2.3.3. *Let E/K be a curve given by a Weierstrass equation with $\Delta = 0$. Then, the group law on elliptic curves makes E_{ns} into an abelian group.*

(1) Suppose that E has a node. Then, E_{ns} is isomorphic to the multiplicative group $\overline{K}^{\times}$ or $\mathbb{G}_m$.
(2) Suppose that E has a cusp. Then, E_{ns} is isomorphic to the additive group $\overline{K}^{+}$ or $\mathbb{G}_a$.

The details for this map can be found in [Sil09, Proposition III.2.5]. The signficance of Proposition 2.3.3 is that it describes the group structure on the nonsingular locus of a singular cubic curve. This will be used when we study the reduction of elliptic curves modulo primes.

2.4. **Reductions of Elliptic Curves.** We now turn our attention to the reduction of elliptic curves defined over $\mathbb{Q}$. Before we define the different types of reductions, however, we first need to establish the minimal Weierstrass equation for an elliptic curve defined over $\mathbb{Q}$.

Definition 2.4.1. Recall the Weierstrass equation of the form (2.1.1) for an elliptic curve $E/\mathbb{Q}$:

$$E \ : \ y^2 + a_1xy + a_3y \ = \ x^3 + a_2x^2 + a_4x + a_6.$$

We will make a change of variables with following maps:

$$x \ \mapsto u^2x \quad \text{and} \quad y \mapsto u^3y + su^2x + t,$$

with $u, r, s, t \in \mathbb{Q}$ and $u \neq 0$ chosen so that the new a_i are all in $\mathbb{Z}$ and $|\Delta|$ is minimized. This equation is said to be *minimal* or, equivalently, a *minimal Weierstrass equation* for $E/\mathbb{Q}$.

Remark 2.4.2. We note that the choice of u, r, s, t is not unique, and thus, most of the theory is independent of the choice of the minimal Weierstrass equation. However, the minimal Weierstrass equation is unique up to a change of variables with $r, s, t \in \mathbb{Z}$ and $u \in \mathbb{Z}^{\times}$. A more generalized approach to the minimal Weierstrass equation can be found in [Sil09, Chapter VII].

Definition 2.4.3. The curve $\overline{E}$ obtained by reducing a minimal equation for E modulo a prime p is called the *reduction of E modulo p*.

Remark 2.4.4. As with the case in Equation (2.1.2), when we are interested only in reductions modulo primes $p \neq 2, 3$, we can work with the simplified Weierstrass equation

$$E \ : \ y^2 \ = \ x^3 + Ax + B,$$

and we can make a change of variables $x \mapsto x/c^2$ and $y \mapsto y/c^3$ with $c \in \mathbb{Q}^*$ chosen so that the new coefficients A, B are in $\mathbb{Z}$ and $|\Delta|$ is minimal.

We are now prepared to study the reduction of elliptic curves defined over $\mathbb{Q}$. As the reduction of an elliptic curve modulo a prime p is given by a Weierstrass equation, we know by Proposition 2.1.8 that the reduced curve is one of three types. We classify E according to these possibilities.

Definition 2.4.5. Let $E/\mathbb{Q}$ be an elliptic curve given by a minimal Weierstrass equation, and let $\overline{E}$ be its reduction modulo a prime p, given by the equation

$$y^2 + \overline{a}_1xy + \overline{a}_3y \;=\; x^3 + \overline{a}_2x^2 + \overline{a}_4x + \overline{a}_6,$$

where $\overline{a}_i = a_i \mod p$. Then, we say that

(a) E has *good* (or *stable*) *reduction* if $\overline{E}$ is nonsingular.
(b) E has *additive* (or *unstable*, *cuspidal*) *reduction* if $\overline{E}$ has a cusp.
(c) E has *multiplicative* (or *semi-stable*, *nodal*) *reduction* if $\overline{E}$ has a node.

In cases (b) and (c), we say that E has *bad reduction*. If E has multiplicative reduction, then the reduction is said to be *split* if the slopes of the tangent lines at the node are in $\mathbb{Q}$, and otherwise it is said to be *nonsplit*.

We can provide further details on the types of reduction based on the quantities associated to the elliptic curve E, as mentioned in Remark 2.1.5.

2.4.1. *Good Reduction.* If p does not divide Δ, then $\overline{E}$ is nonsingular, and $\overline{E}$ is an elliptic curve over $\mathbb{F}_p$. In this case, we say that E has good reduction at p.

2.4.2. *Additive Reduction.* This is the case when $\overline{E}$ has a cusp, and so $\overline{E}_{\text{ns}} \cong \mathbb{G}_a$. This occurs exactly when p divides both Δ and c_4. When p is not 2 or 3, this is equivalent to p dividing both $4A^3 + 27B^2$ and $-48A$ in the simplified Weierstrass equation. In fact, for $p \neq 2, 3$, this is equivalent to p dividing both A and B, or p dividing $4A^3 + 27B^2$ and $-2AB$.

2.4.3. *Multiplicative Reduction.* This is the case when $\overline{E}$ has a node. This occurs exactly when p divides Δ but not c_4. When p is not 2 or 3, this is equivalent to p dividing $4A^3 + 27B^2$ but not $-48A$ in the simplified Weierstrass equation. For $p \neq 2, 3$, this is equivalent to p dividing $4A^3 + 27B^2$ but not A, or p dividing $4A^3 + 27B^2$ but not $-2AB$.

The choice of $-2AB$ is made to identify when the reduction is split or nonsplit. The tangents of the node are rational over $\mathbb{F}_p$ if and only if $-2AB$ is a square in $\mathbb{F}_p$, in which case $\overline{E}_{\text{ns}} \cong \mathbb{G}_m$ and $\overline{E}$ is said to have *split multiplicative reduction*. Otherwise, $\overline{E}_{\text{ns}}$ is isomorphic to $\mathbb{G}_m[-2\overline{ab}]$ and E is said to have *nonsplit multiplicative reduction*.

2.5. **Conductor.** Having established the reduction of an elliptic curve, we are now suited to define its conductor, which can be seen as a measure of the bad reduction of the elliptic curve across all primes. Let us start by defining the quantity f_p for each prime $p \in \mathbb{Z}$.

Definition 2.5.1. Let E be an elliptic curve defined over $\mathbb{Q}$. For each prime $p \in \mathbb{Z}$, let f_p be defined by

$$f_p \;=\; \begin{cases} 0, & \text{if } E \text{ has a good reduction at } p, \\ 1, & \text{if } E \text{ has a multiplicative reduction at } p, \\ 2, & \text{if } E \text{ has an additive reduction at } p \text{ and } p \notin \{2,3\}, \\ 2+\delta_p, & \text{if } E \text{ has an additive reduction at } p \text{ and } p \in \{2,3\}, \end{cases}$$

where δ_p depends on wild ramification in the action of the inertia group at p of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the Tate module $T_p(E)$. This quantity can be computed using Tate's algorithm (see [Sil94, Section IV.9.4]).

With this, we can now define the conductor of an elliptic curve.

Definition 2.5.2. Let E be an elliptic curve defined over $\mathbb{Q}$. The *conductor* of E is defined as the integer

$$N_E \;=\; \prod_{p \text{ prime}} p^{f_p},$$

where f_p is as defined in Definition 2.5.1.

2.6. **The L-Function of an Elliptic Curve.** As mentioned earlier in Section 2.4 and Section 2.5, we note that the global arithmetic of an object can be understood by studying its local behavior at each prime p. This philosophy, central to number theory, can again be observed through the study of an L-function associated to an elliptic curve. Let us first define the following functions that will be used in the definition of the L-function. We will assume any $E/\mathbb{Q}$ is given by a minimal Weierstrass equation.

Definition 2.6.1. Let E be an elliptic curve defined over $\mathbb{Q}$. Let $\mathbf{1}_{N_E}$ be the trivial character modulo the conductor N_E of E. That is,

$$1_{N_E}(m) \;=\; \begin{cases} 1, & \text{if } \gcd(m, N_E) = 1, \\ 0, & \text{otherwise.} \end{cases}$$

When E is clear from context, we simply write 1_N.

Remark 2.6.2. In particular, we observe that for a prime p,

$$1_N(p) \;=\; \begin{cases} 1, & \text{if } E \text{ has a good reduction at } p, \\ 0, & \text{if } E \text{ has a bad reduction at } p. \end{cases}$$

Let us also define the integer $a_p(E)$ associated to the elliptic curve E at each prime p.

Definition 2.6.3. Let E be an elliptic curve defined over $\mathbb{Q}$. For each prime p, we define the integer

$$a_p(E) \;=\; p + 1 - \left|\overline{E}(\mathbb{F}_p)\right|,$$

where $\overline{E}$ is the reduction of E modulo p.

Remark 2.6.4. What we observe is that the value of $a_p(E)$ for a prime p of bad reduction is

$$a_p(E) \;=\; \begin{cases} 1, & \text{if } E \text{ has a split multiplicative reduction at } p, \\ -1, & \text{if } E \text{ has a nonsplit multiplicative reduction at } p, \\ 0, & \text{if } E \text{ has an additive reduction at } p. \end{cases}$$

For primes p of good reduction, there does not exist such a simple formula for $a_p(E)$, but Hasse's theorem on elliptic curves (or the Hasse bound) provides an estimate of the number of points on an elliptic curve over a finite field, bounding the value both above and below. Specifically, Hasse's theorem states that for a prime p of good reduction,

$$|a_p(E)| \;\leq\; 2\sqrt{p}.$$

Even without going into the details of the proof, it is apparent that there is a deeper structure that governs the values of $a_p(E)$ for primes of good reduction. We will explore this further in Section 5, when we discuss the modularity theorem.

Remark 2.6.5. For primes p of good reduction, the value $a_p(E)$ is defined in terms of the number of points on the reduced curve $\overline{E}$ over the finite field $\mathbb{F}_p$. Specifically, we define

$$a_p(E) \;\coloneqq\; p + 1 - \#E(\mathbb{F}_p),$$

where $\#E(\mathbb{F}_p)$ denotes the number of $\mathbb{F}_p$-rational points on the reduced curve $\overline{E}$. This definition "captures" the deviation of the number of points on the elliptic curve from the expected value of $p+1$, which is the number of points on a projective line over $\mathbb{F}_p$. The quantity $a_p(E)$ thus encodes important arithmetic information about the elliptic curve at the prime p.

We are now prepared to define the L-function associated to an elliptic curve defined over the rationals.

Definition 2.6.6. Let E be an elliptic curve defined over $\mathbb{Q}$, and let N be its conductor. The Hasse-Weil L-function associated to $E/\mathbb{Q}$ is defined as

$$L(s, E) \;=\; \prod_p (1 - a_p(E)p^{-s} + \mathbf{1}_N(p)p^{1-2s})^{-1}.$$

3. Modular Forms

Having established the necessary background for elliptic curves, we now turn our attention to modular forms. Modular forms are complex analytic functions defined on the upper half plane that satisfy specific transformation properties under the action of the modular group $\mathrm{SL}_2(\mathbb{Z})$. They will be necessary to develop the theory of Hecke operators, eigenforms, and newforms, which will ultimately lead us to our statement of the Modularity Theorem. We note that this section is largely based on [DS05] and [CSS97]

3.1. Basic Definitions.

Definition 3.1.1. Let k be an integer. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k* if

$$f(\gamma(z)) \;=\; (cz+d)^k f(z) \text{ for } \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \text{ and } z \in \mathcal{H}.$$

Example 3.1.2. An important example of a weakly modular function is Felix Klein's *j-invariant* or *j function*. Let us first define

$$G_k(z) \;\coloneqq\; \sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}} \frac{1}{(cz+d)^k}, \quad z \in \mathcal{H}.$$

We note that G_k is holomorphic and of weight k. We then define $g_2(z) = 60G_4(z)$ and $g_3(z) = 140G_6(z)$ as well as the discriminant function $\Delta : \mathcal{H} \to \mathbb{C}$ given by

$$\Delta(z) \;\coloneqq\; (g_2(z))^3 - 27(g_3(z))^2,$$

and it is clear that Δ is weakly modular of weight 12. What we can also observe is that the only zero of Δ at ∞ and that Δ is a cusp form, which is defined later at Definition 3.1.8. We then define the *modular function* $j : \mathcal{H} \to \mathbb{C}$ given by

$$j(z) \;\coloneqq\; 1728\frac{(g_2(z))^3}{\Delta(z)},$$

and it is the case that j is holomorphic on $\mathcal{H}$. Because the numerator and denominator of j are both of weight 12, j is $\mathrm{SL}_2(\mathbb{Z})$-invariant, i.e.

$$j(\gamma(z)) \;=\; j(z), \quad \text{for } \gamma \in \mathrm{SL}_2(\mathbb{Z}),\ z \in \mathcal{H},$$

and j is also called the *modular invariant*.

Remark 3.1.3. Though not proven in this paper, if the transformation law for Definition 3.1.1 holds when γ is each of the generators

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

then it holds for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. As such, we have an equivalent definition for Definition 3.1.1: a meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is weakly modular of weight k if

$$f(z+1) \;=\; f \quad \text{and} \quad f(-1/z) \;=\; z^k f(z).$$

Recall that a meromorphic function on an open subset U of the complex plane $\mathbb{C}$ is a function that is holomorphic on all of U except for a set of isolated points, which are called the poles of the function. Given that a function is weakly modular, we can then define a modular form.

Definition 3.1.4. 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 ∞.

The set of modular forms of weight k is denoted $\mathcal{M}_k(\mathrm{SL}_2(\mathbb{Z}))$.

Remark 3.1.5. We note that there are no nonzero modular forms of odd weight. Because of the invariance of a modular form f under the action of $-\mathrm{id}$, where id is the identity matrix in $\mathrm{SL}_2(\mathbb{Z})$, we observe that for $z = x + iy \in \mathcal{H}$,

$$f(-\mathrm{id}(z)) \;=\; f\left(\frac{-x-iy}{-1}\right) \;=\; f(z),$$

so for the relation $f(-\mathrm{id}(z)) = (-1)^k f(z)$ to hold, we necessarily have $(-1)^k = 1$, which implies that k is even.

We further note there are no nonzero modular forms of weight 2 or modular forms of negative (with respect to $\mathrm{SL}_2(\mathbb{Z})$). This fact can be proven through the dimension formulas for modular forms of even weight (see [DS05, Section 3.5] for details).

Remark 3.1.6. We note from Remark 3.1.3 that if f is a modular form of weight k, then $f(z+1) = f(z)$ for all $z \in \mathcal{H}$. This periodicity, along with the fact that f is holomorphic at ∞, implies that f has a Fourier expansion of the form

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

Example 3.1.7. A trivial example of a modular form is the zero function on $\mathcal{H}$, which we can observe is a modular form of every weight. Similarly, we note that very constant function on $\mathcal{H}$ is a modular form of weight 0. The standard example of a nontrivial modular form is the Eisenstein series of weight k for even integers $k \geq 4$. The Eisenstein series of weight k, denoted as G_k, is defined by

$$G_k(z) \;=\; \sum_{(m,n) \in \mathbb{Z}^2 \setminus \{(0,0)} \frac{1}{(m+nz)^k}.$$

Having defined a modular form, we can further define what a cusp form is:

Definition 3.1.8. A *cusp form of weight k* is a modular form of weight k whose Fourier expansion has leading coefficient $a_0 = 0$, i.e.,

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

The set of cusp forms is denoted $\mathcal{S}_k(\mathrm{SL}_2(\mathbb{Z}))$.

Defining modular forms and cusp forms for the group $\mathrm{SL}_2(\mathbb{Z})$ naturally leads to the question of whether we can define modular forms and cusp forms for other groups. What we can, in fact, define are modular forms for specific subgroups of $\mathrm{SL}_2(\mathbb{Z})$, specifically congruence subgroups. We first introduce the principal congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$.

Definition 3.1.9. Let N be a positive integer. The *principal congruence subgroup of level N* is

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

Using the principal congruence subgroup, we define the congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Definition 3.1.10. A subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is a *congruence subgroup* if $\Gamma(N) \subset \Gamma$ for some $N \in \mathbb{Z}^+$, in which case Γ is a congruence subgroup of *level N*.

Lemma 3.1.11. *Let M and N be positive integers such that $M \mid N$. Then, we have the inclusion $\Gamma(N) \subset \Gamma(M)$. Furthermore, if Γ is a congruence subgroup of level N, then Γ is also a congruence subgroup of level M.*

Proof. Suppose M divides N. Let $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma \in \Gamma(N)$. Then, $a \equiv d \equiv 1 \mod N$ and $b \equiv c \equiv 0 \mod N$. Since $M \mid N$, it follows that $a \equiv d \equiv 1 \mod M$ and $b \equiv c \equiv 0 \mod M$, so $\gamma \in \Gamma(M)$. The second part of the lemma follows directly. □

From Definition 3.1.10, it is clear that every congruence subgroup Γ has finite index in $\mathrm{SL}_2(\mathbb{Z})$. While there are various congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$, the most important congruence subgroups that we will focus on are

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

(where "$*$" means "unspecified") and

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

Remark 3.1.12. From quick inspection, we can see that the above congruence subgroups satisfy the following relation:

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

For any congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$, we can also consider the space of modular forms of weight k with respect to Γ, instead of $\mathrm{SL}_2(\mathbb{Z})$. We can similarly consider the space of cusp forms of weight k with respect to Γ. To do so, we will first generalize Definition 3.1.1.

Definition 3.1.13. Let k be an integer. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k with respect to* Γ if it is weight-k invariant under Γ. Equivalently, f satisfies

$$f[\gamma]_k \ = \ f \quad \text{for all } \gamma \in \Gamma,$$

where

$$(f[\gamma]_k)(z) \ = \ j(\gamma, z)^{-k} f(\gamma(z)), \quad z \in \mathcal{H},$$

with the *factor of automorphy* $j(\gamma, z) \in \mathbb{C}$ defined by $j(\gamma, z) = cz + d$, where $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \Gamma$.

We can now generalize Definition 3.1.4 and Definition 3.1.8.

Definition 3.1.14. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ 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,
(2) f is weight-k invariant under Γ,
(3) $f[\alpha]_k$ is holomorphic at ∞ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

If in addition,

(4) $a_0 = 0$ in the Fourier expansion of $f[\alpha]_k$ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$,

then f is a *cusp form of weight k with respect to* Γ. The modular forms of weight k with respect to Γ are denoted $\mathcal{M}_k(\Gamma)$, and the cusp forms are denoted $\mathcal{S}_k(\Gamma)$.

From Definition 3.1.14, it is clear that if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, then we recover Definition 3.1.4 and Definition 3.1.8. We also have the following inclusion lemma.

Lemma 3.1.15. *Let Γ and Γ' be congruence subgroups of $SL_2(\mathbb{Z})$ such that $\Gamma \subset \Gamma'$. Then, for any integer $k \geq 0$, we have*

$$\mathcal{M}_k(\Gamma') \subset \mathcal{M}_k(\Gamma),$$

and similarly

$$\mathcal{S}_k(\Gamma') \subset \mathcal{S}_k(\Gamma).$$

Proof. Let $f \in \mathcal{M}_k(\Gamma')$. By definition, f satisfies the transformation law

$$f(\gamma z) \ = \ (cz + d)^k f(z) \quad \text{for all } \gamma \in \Gamma'.$$

Since $\Gamma \subset \Gamma'$, this identity holds in particular for all $\gamma \in \Gamma$, so $f \in \mathcal{M}_k(\Gamma)$. □

Corollary 3.1.16. *A simple but important consequence of Lemma 3.1.15 is that for any positive integer N and integer $k \geq 0$, we have the following relations:*

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

and

$$\mathcal{S}_k(SL_2(\mathbb{Z})) \ \subset \ \mathcal{S}_k(\Gamma_0(N)) \ \subset \ \mathcal{S}_k(\Gamma_1(N)) \ \subset \ \mathcal{S}_k(\Gamma(N)).$$

Remark 3.1.17. We note that $\gamma = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right] \in \Gamma_1(N)$ for any N, so any $f \in \mathcal{M}_k(\Gamma_1(N))$ is 1-periodic. Then, by Definition 3.1.14, it follows by the same reasoning as in Remark 3.1.6 that f has a Fourier expansion of the form

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

We now present an example of Definition 3.1.14. This function will be our running example through the rest of this paper, as its explicit form makes it convenient for understanding and demonstrating the general theory that will be developed.

Example 3.1.18. Consider the function $f : \mathcal{H} \to \mathbb{C}$ defined by

$$f(z) \;=\; q\prod_{n\geq 1}(1-q^n)^2(1-q^{11n})^2, \quad q = e^{2\pi iz}.$$

We state without proof that f is a modular form of weight 2 with respect to the congruence subgroup $\Gamma_0(11)$, or $f \in \mathcal{M}_k(\Gamma_0(11))$. Furthermore, since the Fourier expansion of f has leading coefficient 0, i.e. $a_0 = 0$, we have that f is a cusp form of weight 2 with respect to $\Gamma_0(11)$. That is, $f \in \mathcal{S}_2(\Gamma_0(11))$.

Up until now, we have been working with only holomorphic functions from $\mathcal{H}$ to $\mathbb{C}$. What is, in fact, possible is to extend our discussion to meromorphic functions from $\mathcal{H}$ to $\hat{\mathbb{C}}$.

Definition 3.1.19. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and let k be an integer. A function $f : \mathcal{H} \to \hat{\mathbb{C}}$ is an *automorphic form of weight k with respect to* Γ if

(1) f is meromorphic,
(2) f is weight-k invariant under Γ,
(3) $f[\alpha]_k$ is meromorphic at ∞ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

The set of automorphic forms of weight k with respect to Γ is denoted $\mathcal{A}_k(\Gamma)$.

Example 3.1.20. Recall the j function from Example 3.1.2. Even though it is not a modular form, the j function is an automorphic form of weight 0.

Remark 3.1.21. Even though the theory of automorphic forms will not be directly used in this paper, it provides a broader context for modular forms. From this viewpoint, classical modular forms (i.e. the modular forms defined above) can be seen as holomorphic automorphic forms for $\mathrm{SL}_2(\mathbb{Z})$. This framework is particularly useful when extending the theory to more general settings, such as Hilbert modular forms and Siegel modular forms, which are associated to higher-dimensional spaces.

4. Hecke Operators, Eigenforms, and Newforms

We are now prepared to study the theory of Hecke operators, eigenforms, and newforms. Building on our previous discussion of modular forms and cusp forms, we will define Hecke operators acting on these spaces, explore the concept of Hecke eigenforms, and finally introduce the notion of newforms. This framework will culminate in the next section with the statement of a version of the modularity theorem. We note that this section is largely based on [DS05] and [Ste07].

4.1. **Vector Space Structure of Modular Forms.** We have defined the space of modular forms $\mathcal{M}_k(\Gamma(N))$ and cusp forms $\mathcal{S}_k(\Gamma)$ of weight k with respect to the congruence subgroup Γ. What we can, in fact, do is view these spaces as vector spaces over $\mathbb{C}$.

Proposition 4.1.1. *Let Γ be a fixed congruence subgroup of $SL_2(\mathbb{Z})$, and let $k \in \mathbb{Z}_{>0}$. Then, both $\mathcal{M}_k(\Gamma)$ and $\mathcal{S}_k(\Gamma)$ are vector spaces over $\mathbb{C}$ under the usual addition and scalar multiplication of functions.*

Remark 4.1.2. Though we will not prove this proposition, the key point is that the sum of modular forms (or cusp forms) of weight k with respect to Γ is again a modular form (or cusp form) remain within the same space. We can similarly show that scalar multiplication also preserves these spaces.

In addition, it can be shown that these vector spaces are finite-dimensional (over $\mathbb{C}$). This can be proven using the Riemann-Roch theorem and the theory of meromorphic differentials. For more details, see [DS05, Chapter 3].

We have now established the vector space structure of modular forms and cusp forms. This allows us to consider possible linear operators acting on these spaces, leading us to the theory of Hecke operators.

4.2. **Double Coset Operator.** To define Hecke operators, we first introduce the concept of double cosets and double coset operators.

Let Γ_A and Γ_B be congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$. Let $\mathrm{GL}_2^+(\mathbb{Q})$ denote the group of 2×2 matrices with positive determinant and rational entries. Clearly, $\mathrm{SL}_2(\mathbb{Z}) \subset \mathrm{GL}_2^+(\mathbb{Q})$, so Γ_A and Γ_B are naturally subgroups of $\mathrm{GL}_2^+(\mathbb{Q})$.

Definition 4.2.1. Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. The *double coset* of α with respect to Γ_A and Γ_B is defined as

$$\Gamma_A \alpha \Gamma_B \;=\; \{\gamma_A \alpha \gamma_B : \gamma_A \in \Gamma_A, \gamma_B \in \Gamma_B\}.$$

Remark 4.2.2. While we will not be providing a detailed proof here, we note that the group Γ_A acts on the double coset $\Gamma_A \alpha \Gamma_B$ by left multiplication. This action partitions the double coset into distinct orbits, where the typical orbit is $\Gamma_A \beta$ with representative $\beta = \gamma_A \alpha \gamma_B$.

The orbit space $\Gamma_A \backslash \Gamma_A \alpha \Gamma_B$ is therefore a disjoint union $\bigcup \Gamma_A \beta_j$ for some choice of representatives β_j. This union is, in fact, finite (see [DS05, Lemma 5.1.1, Lemma 5.1.2]).

Definition 4.2.3. For congruence subgroups Γ_A and Γ_B of $\mathrm{SL}_2(\mathbb{Z})$ and for $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, the *weight-$k$ $\Gamma_A \alpha \Gamma_B$ operator* takes a modular form $f \in \mathcal{M}_k(\Gamma_B)$ to

$$f[\Gamma_A \alpha \Gamma_B] \;=\; \sum_j f[\beta_j]_k,$$

where $\{\beta_j\}$ are orbit representatives, i.e. $\Gamma_A \alpha \Gamma_B = \bigcup_j \Gamma_A \beta_j$ is a disjoint union. This operator is known as the *double coset operator* associated to $\Gamma_A \alpha \Gamma_B$.

Remark 4.2.4. We can observe that the double coset operator takes modular forms with respect to Γ_A to modular forms with respect to Γ_B,

$$[\Gamma_A \alpha \Gamma_B] \;:\; \mathcal{M}_k(\Gamma_A) \;\longrightarrow\; \mathcal{M}_k(\Gamma_B).$$

Similarly, cusp forms with respect to Γ_A are sent to cusp forms with respect to Γ_B,

$$[\Gamma_A \alpha \Gamma_B] \;:\; \mathcal{S}_k(\Gamma_A) \;\longrightarrow\; \mathcal{S}_k(\Gamma_B).$$

The well-definedness of the double coset operator and the above maps can be seen in [DS05, Section 5.1].

4.3. **The $\langle d \rangle$ and T_p Operators.** We can now define two important types of double coset operators: the diamond operator $\langle d \rangle$ and the operator T_p. These two operators will be generalized to define the Hecke operators $\langle n \rangle$ and T_n for any positive integer n.

4.3.1. *Diamond Operators and Decompositions.* Let us first recall $\Gamma_0(N)$ and $\Gamma_1(N)$, as defined in Section 3.1. As previously noted, $\Gamma_1(N) \subset \Gamma_0(N)$ and $\mathcal{M}_k(\Gamma_0(N)) \subset \mathcal{M}_k(\Gamma_1(N))$. We now define the diamond operator as follows:

Definition 4.3.1. Let $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. The diamond operator $\langle d\rangle : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N))$ is given by

$$\langle d\rangle f \;=\; f[\Gamma_1(N)\alpha\Gamma_1(N)]_k,$$

where $\alpha = \left[\begin{smallmatrix} a & b \\ c & \delta \end{smallmatrix}\right] \in \Gamma_0(N)$ with $\delta \equiv d \pmod N$.

Let us now explain the significance of the diamond operator by introducing the concept of χ-eigenspaces. For any Dirichlet character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$, we can decompose the vector space $\mathcal{M}_k(\Gamma_1(N))$ into a direct sum of subspaces that we can analyze independently.

Definition 4.3.2. For each Dirichlet character χ modulo N, we define the *χ-eigenspace* of $\mathcal{M}_k(\Gamma_1(N))$ as

$$\mathcal{M}_k(N,\chi) \;=\; \{f \in \mathcal{M}_k(\Gamma_1(N)) : f[\gamma]_k = \chi(d_\gamma)f \text{ for all } \gamma \in \Gamma_0(N)\},$$

where d_γ is the lower-right entry of γ.

Remark 4.3.3. The decomposition of $\mathcal{M}_k(\Gamma_1(N))$ into χ-eigenspaces is given by:

$$\mathcal{M}_k(\Gamma_1(N)) \;=\; \bigoplus_\chi \mathcal{M}_k(N,\chi).$$

One important χ-eigenspace is the space $\mathcal{M}_k(N,\mathbf{1})$, where $\mathbf{1}$ is the trivial character modulo N. This space is precisely the space of modular forms with respect to $\Gamma_0(N)$, i.e. $\mathcal{M}_k(\Gamma_0(N))$.

What we can observe about this χ-eigenspace is that it is precisely the χ-eigenspace of the diamond operators:

$$\mathcal{M}_k(N,\chi) \;=\; \{f \in \mathcal{M}_k(\Gamma_1(N)) : \langle d\rangle f = \chi(d)f \text{ for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times\},$$

i.e. the diamond operator $\langle d\rangle$ respects the above decomposition of $\mathcal{M}_k(\Gamma_1(N))$. As we noted above, $\mathcal{M}_k(N,\mathbf{1})$ is the $\mathbf{1}$-eigenspace of the diamond operators, corresponding to the trivial Dirichlet character modulo N.

From now on and throughout the remainder of this paper, we restrict our space to the trivial character. That is, we work in the subspace

$$\mathcal{M}_k(\Gamma_0(N)) \;\subset\; \mathcal{M}_k(\Gamma_1(N))$$

consisting of modular forms fixed by all diamond operators. As will later state, the Hecke operators T_p commute with the diamond operators, so this subspace is preserved under the action of the Hecke operators. We will similarly restrict $\mathcal{S}_k(\Gamma_1(N))$ to $\mathcal{S}_k(\Gamma_0(N))$. As the diamond operators act trivially, once we restrict to $\mathcal{M}_k(\Gamma_0(N))$, we no longer have to consider the diamond operators explicitly.

4.3.2. *The T_p Operators.* We also define the operator T_p for primes $p \nmid N$ as follows:

Definition 4.3.4. Let $\alpha = \left[\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right]$. For prime $p \nmid N$, the operator $T_p : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N))$ is given by

$$T_p f \;=\; f[\Gamma_1(N)\alpha\Gamma_1(N)]_k.$$

We present one important property of the operator T_p.

Proposition 4.3.5. *Let $f \in \mathcal{M}_k(\Gamma_1(N))$. We note from Remark 3.1.17 that f has a Fourier expansion $f(z) = \sum_{n\geq 0} a_n q^n$, where $q = e^{2\pi i z}$. Then, for $\mathbf{1}_N : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ (the trivial character modulo N), we have*

$$\begin{aligned}(T_p f)(z) \;&=\; \sum_{n\geq 0} a_{np}(f)q^n + \mathbf{1}_N(p)p^{k-1}\sum_{n\geq 0} a_n(\langle p\rangle f)q^{np} \\ &=\; \sum_{n\geq 0}\left(a_{np}(f) + \mathbf{1}_N(p)p^{k-1}a_{n/p}(\langle p\rangle f)\right)q^n.\end{aligned}$$

That is,

$$a_n(T_p f) \;=\; a_{np}(f) + \mathbf{1}_N(p)p^{k-1}a_{n/p}(\langle p\rangle f) \quad \text{for } f \in \mathcal{M}_k(\Gamma_1(N)),$$

where $a_{n/p} = 0$ when $n/p \notin \mathbb{Z}$.

Note that there is also the case when $f \in \mathcal{M}_k(N, \chi)$ for some nontrivial character χ. However, we note again that we will be focusing only on the case where $f \in \mathcal{M}_k(\Gamma_0(N)) = \mathcal{M}_k(N, \mathbf{1})$.

In general, we also have the following results.

Proposition 4.3.6. *Let d and e be elements of $(\mathbb{Z}/N\mathbb{Z})^\times$, and let p and q be primes not dividing N. Then*

(a) $\langle d\rangle T_p = T_p\langle d\rangle$,
(b) $\langle d\rangle\langle e\rangle = \langle e\rangle\langle d\rangle$,
(c) $T_pT_q = T_qT_p$.

4.4. **Hecke Operators $\langle n\rangle$ and T_n.** We now generalize the operators $\langle d\rangle$ and T_p to define the Hecke operators $\langle n\rangle$ and T_n for any positive integer n.

Definition 4.4.1. Let $n \in \mathbb{Z}_{>0}$. The *Hecke operator* $\langle n\rangle : \mathcal{M}_k(\Gamma_0(N)) \to \mathcal{M}_k(\Gamma_0(N))$ is given by

(a) $\langle n\rangle = \langle d\rangle$ if $n \equiv d \mod N$ for some $d \in (\mathbb{Z}/N\mathbb{Z})^\times$,
(b) $\langle n\rangle = 0$ if $\gcd(n, N) \neq 1$.

Remark 4.4.2. However, we again note that by focusing on the space $\mathcal{M}_k(\Gamma_0(N))$, the diamond operators act trivially. Therefore, for any n coprime to N, the Hecke operator $\langle n\rangle$ acts trivially on $\mathcal{M}_k(\Gamma_0(N))$.

To define the Hecke operator T_n, we first define T_{p^r} for prime powers:

$$T_{p^r} = T_pT_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}}, \quad \text{for } r \geq 2.$$

We can then extend this definition to all positive integers n using the prime factorization of n.

Definition 4.4.3. Let $n \in \mathbb{Z}_{>0}$. The *Hecke operator* $T_n : \mathcal{M}_k(\Gamma_0(N)) \to \mathcal{M}_k(\Gamma_0(N))$ is given by

$$T_n = \prod_i T_{p_i^{e_i}}, \quad \text{where } n = \prod_i p_i^{e_i}.$$

Remark 4.4.4. While the definition of T_n may come naturally, the definition of T_{p^r} does not seem to be as intuitive. To motivate this definition, we can consider the generating function g of T_n, as given by

$$g(s) = \sum_{n\geq 1} T_n n^{-s},$$

which can be expressed as an Euler product:

$$g(s) = \prod_p (1 - T_p p^{-s} + \langle p\rangle p^{k-1-2s})^{-1},$$

where the product is over all primes p. This expression has a connection with the L-function of modular forms, which can be used to study the connection between modular forms and elliptic curves.

One important property of the Hecke operators is extended from Proposition 4.3.6.

Proposition 4.4.5. *Let $m, n \in \mathbb{Z}_{>0}$. Then,*

(a) $\langle m\rangle\langle n\rangle = \langle n\rangle\langle m\rangle = \langle mn\rangle$,
(b) $T_mT_n = T_nT_m$.

4.5. **Eigenforms and the Petersson Inner Product.** We now introduce eigenforms and the Petersson inner product, which will be useful in studying the properties of $\mathcal{M}_k(\Gamma_0(N))$. We begin with the definition of eigenforms.

Definition 4.5.1. Let $f \in \mathcal{M}_k(\Gamma_0(N))$ be a nonzero modular form. If f is an eigenvector for a Hecke operator, then we say f is an *eigenform* for that operator. If f is an eigenform for the Hecke operators T_n and $\langle n\rangle$ for all $n \in \mathbb{Z}_{>0}$, then f is called a *Hecke eigenform* or simply an *eigenform*. The eigenform f is said to be *normalized* if $a_1(f) = 1$ for the Fourier expansion.

We now turn to the Petersson inner product, which is crucial in the study of modular forms. While we will state its definition, what is more important for our purposes are its properties.

Definition 4.5.2. Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup. The *Petersson inner product* $\langle \cdot, \cdot \rangle : \mathcal{S}_k(\Gamma) \times \mathcal{S}_k(\Gamma) \to \mathbb{C}$ is given by

$$\langle f, g \rangle_\Gamma \;=\; \frac{1}{V_\Gamma} \int_{X(\Gamma)} f(z)\overline{g(z)}(\mathrm{Im}(z))^k \; d\mu(z),$$

where $X(\Gamma)$ is the modular curve associated to Γ and V_Γ is the volume of $X(\Gamma)$.

Remark 4.5.3. Note that $X(\Gamma)$ is not defined in this paper, as we will not be directly using its properties. What is notable about $X(\Gamma)$ is that it is a compact Riemann surface, which ensures the convergence of the integral in the Petersson inner product.

What is important about the Petersson inner product is that the Hecke operators are self-adjoint with respect to this inner product. This means that for any $f, g \in \mathcal{S}_k(\Gamma)$, we have

$$\langle T_n f, g \rangle_\Gamma \;=\; \langle f, T_n g \rangle_\Gamma.$$

This allows us to conclude the orthogonality about eigenforms in the space of cusp forms.

One important result is the following:

Theorem 4.5.4. *The space $\mathcal{S}_k(\Gamma_0(N))$ has an orthogonal basis of simultaenous eigenforms for the Hecke operators $\{T_n : (n, N) = 1\}$.*

4.6. **Oldforms and Newforms.** Finally, we introduce the concepts of oldforms and newforms, which are essential in understanding the structure of cusp forms. As stated in Remark 4.5.3, the Petersson inner product plays a crucial role in this discussion.

Definition 4.6.1. For each divisor d of N, let $i_d : (\mathcal{S}_k(\Gamma_0(Nd^{-1})))^2 \to \mathcal{S}_k(\Gamma_0(N))$ be the map given by

$$(f, g) \;\mapsto\; f + g[\alpha_d]_k.$$

The subspace of *oldforms of level N* is defined as

$$\mathcal{S}_k(\Gamma_0(N))^{\mathrm{old}} \;=\; \sum_{p|N} i_p((\mathcal{S}_k(\Gamma_0(Np^{-1})))^2),$$

where p runs over all primes, and the subspace of *newforms of level N* is defined as the orthogonal complement with respect to the Petersson inner product:

$$\mathcal{S}_k(\Gamma_0(N))^{\mathrm{new}} \;=\; (\mathcal{S}_k(\Gamma_0(N))^{\mathrm{old}})^\perp.$$

Remark 4.6.2. Clearly, we again see the importance of the Petersson inner product in defining newforms. The space of cusp forms can thus be decomposed as

$$\mathcal{S}_k(\Gamma_0(N)) \;=\; \mathcal{S}_k(\Gamma_0(N))^{\mathrm{old}} \oplus \mathcal{S}_k(\Gamma_0(N))^{\mathrm{new}}.$$

We now present the canonical definition of newforms.

Definition 4.6.3. A *newform* is a normalized Hecke eigenform that lies in the space of newforms $\mathcal{S}_k(\Gamma_0(N))^{\mathrm{new}}$.

This leads us to our final important result regarding Hecke operators.

Theorem 4.6.4. *Let $f \in \mathcal{S}_k(\Gamma_0(N))^{new}$ be a nonzero eigenform for the Hecke operators T_n with $(n, N) = 1$. Then*

- *(a) f is a Hecke eigenform, i.e. an eigenform for T_n for all $n \in \mathbb{Z}_{>0}$, and a suitable scalar multiple of f is a newform;*
- *(b) if $\tilde{f}$ satisfies the same conditions as f and has the same T_n-eigenvalues, then $\tilde{f} = cf$ for some constant $c \in \mathbb{C}$.*

Remark 4.6.5. The significance of Theorem 4.6.4 is that it shows newforms are the fundamental building blocks of cusp forms of level $\Gamma_0(N)$. Part (a) implies that a newform is automatically an eigenform for all Hecke operators, so its Hecke eigenvalues encode all its data. From part (b), we can reliably conclude the uniqueness of newforms up to scaling.

This uniqueness is crucial in arithmetic applications. As we will see later, it ensures that when an arithmetic object, such as an elliptic curve over $\mathbb{Q}$, gives rise to a system of Hecke eigenvalues, there exists unique associated newforms. Thus, newforms provide a canonical way to isolate modular forms of minimal level and weight that capture essential arithmetic information.

Example 4.6.6. We now again consider our modular form of weight 2 and level 11, as in Example 3.1.18:

$$f(z) \;=\; q\prod_{n\geq 1}(1-q^n)^2(1-q^{11n})^2, \quad q=e^{2\pi iz}.$$

As we can see in [LMF25a] and in [DS05, Theorem 3.5.1], $\dim(\mathcal{S}_2(\Gamma_0(11))) = 1$. Furthermore, the dimension of the old subspace is 0, so the new subspace necessarily has dimension 1 by 4.6.2. Therefore, any nonzero cusp form in $\mathcal{S}_2(\Gamma_0(11))$ must lie in $\mathcal{S}_2(\Gamma_0(11))^{\text{new}}$, so f spans the entire space of cusp forms of weight 2 and level 11.

That is, f must be a nonzero cusp form, and since $a_1 = 1$ for f, f is a newform by definition. In addition, as the space is one-dimensional, f is automatically an eigenform for all Hecke operators.

We will present a further observation about the function in the following section.

4.7. **Connection with L-Functions.** We conclude this section by briefly discussing the connection between modular forms and L-functions. We note that each cusp form $f \in \mathcal{S}_k(\Gamma_0(N))$ has an associated L-function, and it is defined as follows.

Definition 4.7.1. Let $f(z) = \sum_{n\geq 1} a_n q^n \in \mathcal{S}_k(\Gamma_0(N))$ be a modular form. The **L-function** associated to f is defined as

$$L(s,f) \;=\; \sum_{n\geq 1} a_n n^{-s},$$

where $s \in \mathbb{C}$.

It is clear that the L-function is a Dirichlet series and is derived from the Fourier coefficients of the modular form. Thus, the L-function encodes information about the modular form. We note as a remark that the L-function converges uniformly to a holomorphic function on the half-plane $\text{Re}(s) > k/2 + 1$. Furthermore, the L-function has an analytic continuaation to a holomorphic function on the entire complex plane $\mathbb{C}$ (for more details, refer to [DS05]).

The discussion of Hecke operators leads us to an important result regarding the L-function of eigenforms.

Theorem 4.7.2. *Let $f \in \mathcal{M}_k(\Gamma_0(N))$ such that $f(z) = \sum_{n\geq 1} a_n q^n$. The following are equivalent:*

(a) f is a normalized eigenform.

(b) $L(s,f)$ has an Euler product expansion:

$$L(s,f) \;=\; \prod_{p}(1-a_p p^{-s}+\chi(p)p^{k-1-2s})^{-1},$$

where the product is taken over all primes p.

5. Modularity Theorem

5.1. **Statement of the Modularity Theorem.** Recall the statement of the modularity theorem from Theorem 1.1.1 in Section 1:

Theorem 5.1.1. *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E. Then, for some newform $f \in \mathcal{S}_2(\Gamma_0(N_E))$,*

$$a_p(f) \;=\; a_p(E) \quad \textit{for all primes } p.$$

Remark 5.1.2. Let us again recall our running example from Example 3.1.18:

$$f(z) \;=\; q\prod_{n\geq 1}(1-q^n)^2(1-q^{11n})^2, \quad q = e^{2\pi iz}.$$

We know the values $a_p(f)$ for various primes p from the Fourier expansion of f:

$$f(q) \;=\; q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + q^{11} + \cdots.$$

We recall that for primes p of good reduction, the value $a_p(E)$ for an elliptic curve E over $\mathbb{Q}$ is defined as

$$a_p(E) \;\coloneqq\; p + 1 - \#E(\mathbb{F}_p),$$

where $\#E(\mathbb{F}_p)$ denotes the number of $\mathbb{F}_p$-rational points on the reduced curve modulo p. For primes of bad reduction, the value $a_p(E)$ is defined differently based on the type of reduction (additive or multiplicative), as in Remark 2.6.4. For more details, one can refer to [Sil09].

We now consider the elliptic curve E over $\mathbb{Q}$ defined by the Weierstrass equation

$$E \;:\; y^2 + y \;=\; x^3 - x^2.$$

Using Sagemath, we can compute the conductor of E, which is $N_E = 11$, and the values of $a_p(E)$ for various primes p:

$$\begin{aligned}
a_2(E) &= -2,\\
a_3(E) &= -1,\\
a_5(E) &= 1,\\
a_7(E) &= -2,\\
a_{11}(E) &= 1,\\
a_{13}(E) &= 4,\\
a_{17}(E) &= -2,\\
a_{19}(E) &= 0,
\end{aligned}$$

and so on. What we can observe is that the values $a_p(f)$ and $a_p(E)$ match for all primes p. Thus, the modularity theorem guarantees that there exists a newform f in $\mathcal{S}_2(\Gamma_0(11))$ such that $a_p(f) = a_p(E)$ for all primes p, and in this case, our running example f is precisely that newform.

This example also demonstrates Theorem 1.1.2 from Section 1, which states that the L-function of the elliptic curve E is equal to the L-function of the newform f:

Theorem 5.1.3. *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E. Then, for some newform $f \in \mathcal{S}_2(\Gamma_0(N_E))$,*

$$L(s, f) \;=\; L(s, E).$$

As we have established that $a_p(f) = a_p(E)$ for all primes p, it follows directly from the definitions of the L-functions that $L(s, f) = L(s, E)$. Thus, our running example f also satisfies this version of the modularity theorem for the elliptic curve E. From this, we see that the modularity theorem establishes a deep connection between elliptic curves over $\mathbb{Q}$ and modular forms, linking their arithmetic properties through the equality of their Fourier coefficients and L-functions.

References

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor. On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises. *J. Amer. Math. Soc.*, 14(4):843–939, 2001.

[Con25] Keith Conrad. The local-global principle. `https://kconrad.math.uconn.edu/blurbs/gradnumthy/localglobal.pdf`, 2025.

[CSS97] Gary Cornell, Joseph H. Silverman, and Glenn Stevens, editors. *Modular forms and Fermat's last theorem*. Springer-Verlag, New York, 1997. Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995.

[DS05] Fred Diamond and Jerry Shurman. *A first course in modular forms*, volume 228 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005.

[Fre86] Gerhard Frey. Links between stable elliptic curves and certain Diophantine equations. *Ann. Univ. Sarav. Ser. Math.*, 1(1):iv+40, 1986.

[Hel72] Yves Hellegouarch. *Courbes elliptiques et équation de Fermat*. Thése, 1972.

[Kob93] Neal Koblitz. *Introduction to elliptic curves and modular forms*, volume 97 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1993.

[Lan95] Serge Lang. Some history of the Shimura-Taniyama conjecture. *Notices Amer. Math. Soc.*, 42(11):1301–1307, 1995.

[LMF25a] The LMFDB Collaboration. The L-functions and modular forms database. `https://www.lmfdb.org/ModularForm/GL2/Q/holomorphic/11/2/a/a/`, 2025. [Online; accessed 2 October 2025].

[LMF25b] The LMFDB Collaboration. The L-functions and modular forms database. `https://www.lmfdb.org`, 2025. [Online; accessed 2 October 2025].

[LR11] Álvaro Lozano-Robledo. *Elliptic curves, modular forms, and their L-functions*, volume 58 of *Student Mathematical Library*. American Mathematical Society, Providence, RI; Institute for Advanced Study (IAS), Princeton, NJ, 2011. IAS/Park City Mathematical Subseries.

[Mil21] James S. Milne. *Elliptic curves*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, [2021] ©2021. Second edition [of 2267743].

[Rib90] K. A. Ribet. On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms. *Invent. Math.*, 100(2):431–476, 1990.

[Ser87] Jean-Pierre Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987.

[Sil94] Joseph H. Silverman. *Advanced topics in the arithmetic of elliptic curves*, volume 151 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[Sil09] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[ST15] Joseph H. Silverman and John T. Tate. *Rational points on elliptic curves*. Undergraduate Texts in Mathematics. Springer, Cham, second edition, 2015.

[Ste07] William Stein. *Modular forms, a computational approach*, volume 79 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2007. With an appendix by Paul E. Gunnells.

[Tok56] *Proceedings of the international symposium on algebraic number theory, Tokyo & Nikko, 1955*. Science Council of Japan, Tokyo, 1956.

[TW95] Richard Taylor and Andrew Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995.

[Wei67] André Weil. Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. *Math. Ann.*, 168:149–156, 1967.

[Wik07] Wikimedia. Example elliptic curves. ECClines.svg, 2007.

[Wil95] Andrew Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995.

University of California Berkeley, Department of Mathematics, Berkeley, CA 94720, USA
Email address: `dtkim25@berkeley.edu`