GALOIS REPRESENTATIONS AND MODULARITY THEOREM

WENSHI ZHAO

Abstract. This paper serves as an introduction to the theory of modular forms and Galois representations. We begin by defining the Galois representation associated with an elliptic curve. Next, we describe modular curves as moduli spaces of elliptic curves equipped with enhanced torsion data. We then outline the theory of Hecke operators and their actions on various objects, including the modular curves and the space of cusp forms. Using these preliminaries, we construct the Galois representation associated with a newform in the weight-2 case and state the Modularity Theorem. Finally, we apply the Eichler—Shimura relation to connect this formulation of the Modularity Theorem with another version, which predicts the relationship between the number of points on elliptic curves modulo p and the Fourier coefficients of their associated newforms.

Contents

Date: 10 September 2025.

1. Introduction

The Modularity Theorem is a theorem roughly of the following form: every elliptic curve is associated with a modular form. In this paper, we will introduce two formulations of the Modularity Theorem. The first is through Galois representations, i.e., representations of the absolute Galois group $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. For every prime l and every elliptic curve E over $\mathbb{Q}$, there is an associated Galois representation $\rho_{E,l}$. For a newform f (defined in subsection 4.3), we may associate an abelian variety A_f over $\mathbb{Q}$, and through this geometric object we may also define a representation, denoted $\rho_{f,\lambda}$. The Modularity Theorem states that every $\rho_{E,l}$ is equivalent to some $\rho_{f,\lambda}$ for some newform f. The central achievement of Andrew Wiles' famous 1995 paper [1] proving Fermat's Last Theorem was to establish this formulation of the Modularity Theorem in the case of *semistable* elliptic curves.[1]

The second formulation of the Modularity Theorem is as follows. In many cases (see Definition 6.1), one can reduce an elliptic curve E over $\mathbb{Q}$ at a prime $p \in \mathbb{Z}$. Let $a_p(E) = p + 1 - |\widetilde{E}(\mathbb{F}_p)|$, where $\widetilde{E}$ is the reduction of E at p and $\widetilde{E}(\mathbb{F}_p)$ denotes its $\mathbb{F}_p$-points. For a modular form f, let $f(\tau) = \sum_{n=0}^{\infty} a_n(f) q^n$ be its Fourier expansion. The second version of the Modularity Theorem states that, for each elliptic curve E over $\mathbb{Q}$, there is a newform f such that $a_p(E) = a_p(f)$. Thus, the Fourier coefficients of f encode the number of points of the various reductions of E.

The main objective of this paper is to introduce these two formulations of the Modularity Theorem and show that they are equivalent. In doing so, we will touch on Galois representations in general, modular forms, modular curves, Hecke operators, Jacobians, and related topics. In section 2, we review some basic facts about elliptic curves, define the Galois representation associated with elliptic curves, and compute the characteristic polynomial of the image of crucial elements $\mathrm{Frob}_{\mathfrak{p}} \in G_{\mathbb{Q}}$. This characteristic polynomial will serve as the link between the two versions of the Modularity Theorem. In section 3 and section 4, we outline parts of the classical theory of modular curves, modular forms, and Hecke operators. We will repeatedly emphasize the perspective of the modular curve as a moduli space (i.e., a space of solutions to classification problems) of elliptic curves. In section 5, we define the Galois representation associated with a newform and state the first version of the Modularity Theorem. Finally, section 6 introduces reductions of algebraic curves (in particular, modular curves), sketches the proof of the Eichler—Shimura relation, and shows the equivalence of the two formulations of the Modularity Theorem.

Readers for this paper should be familiar with basic algebra, complex analysis, and topology. Familiarity with Riemann surface, basic algebraic geometry (Hartshorne 1.1-1.6 and 2.2), and elliptic curves will be very helpful.

This paper is intended as a short introduction, so we will sketch many results without going into too much detail. We want to give readers a glimpse into these beautiful theories without requiring them to work through excessive technicalities. References are included for those who wish to study this material in greater depth.

[1] Roughly speaking, the proof of Fermat's Last Theorem reduces to showing the nonexistence of certain elliptic curves. This nonexistence follows from properties of modular forms, together with the connections established by the Modularity Theorem. See [2] for a nice exposition.

2. Elliptic Curves and Galois Representations

2.1. **Basic facts of elliptic curves.** We begin by recording the basic facts about elliptic curves over a field. These facts will be used freely in the rest of the paper. For details, see [3].

(1) An **elliptic curve** over a field k is a pair $(E, 0_E)$, where E is a nonsingular curve of genus one over k and $0_E \in E$. For the rest of this subsection, assume that $\mathrm{Char}(k) \neq 2, 3$. Every elliptic curve is isomorphic to the plane curve in $\mathbb{P}^2_k$ given by a **Weierstrass equation**

$$E : y^2 = 4x^3 - g_2 x - g_3,$$

where the isomorphism sends 0_E to the point of infinity $[0 : 1 : 0]$. Now let

$$\Delta = g_2^3 - 27g_3^2 \in k, \qquad j = 1728 g_2^3/\Delta \in k.$$

The quantity Δ (resp. j) is called the **discriminant** (resp. **invariant**) of the Weierstrass equation, respectively.

(2) Any two Weierstrass equations for isomorphic elliptic curves E are related by an **admissible change of variables**, meaning one of the form

$$x = u^2 x', \qquad y = u^3 y' \qquad (u \in k^*).$$

We say that two Weierstrass equations are **equivalent** if they are related by an admissible change of variables. Under this change of variables, we have

$$\Delta = u^{12}\Delta', \qquad j = j'.$$

In particular, the invariant depends only on the elliptic curve E and not on the choice of Weierstrass equation.

(3) For any extension K/k, let $E(K)$ denote the set of K-points of E. If E is given by a Weierstrass equation $E(x, y) \in k(x, y)$, then $E(K) = \{(x, y) \in K : E(x, y) = 0\} \cup \{0_E\}$.

(4) The $\overline{k}$-points of an elliptic curve have a **group law** defined by the property that if P, Q, and R are collinear, then $P + Q + R = 0$. If we denote the coordinate $P = (x_P, y_P) \in \mathbb{A}^2_{\overline{k}}$, then

$$(x_{P+Q}, y_{P+Q}) = (r(x_P, x_Q, y_P, y_Q), s(x_P, x_Q, y_P, y_Q))$$

for some $r, s \in k(x_P, x_Q, y_P, y_Q, g_2, g_3)$. For $K \subset \overline{k}$, $E(K)$ is a subgroup of $E = E(\overline{k})$.

(5) Let $[N] : E \to E$ be the multiplication by N map, and let $E[N] = \ker([N])$ be the **N-torsion points**. Then

$$E[N] \cong \prod E[p^{e_p}] \qquad \text{where } N = \prod p^{e_p}.$$

If $p \neq \mathrm{Char}(k)$, then $E[p^e] \cong (\mathbb{Z}/p^e\mathbb{Z})^2$. If $p = \mathrm{Char}(k)$, then $E[p^e] \cong \mathbb{Z}/p^e\mathbb{Z}$ for all $e \geq 1$, or $E[p^e] = 0$ for all $e \geq 1$. If $p = \mathrm{Char}(k)$, and $E[p] \cong \mathbb{Z}/p\mathbb{Z}$, then E is called **ordinary**; if $E[p] = 0$, then E is **supersingular**.

(6) Let X be an algebraic curve. The divisor group is the group of finite integer combinations of points of X:

$$\mathrm{Div}(X) = \bigoplus_{x \text{ a closed point of } X} \mathbb{Z}.$$

For a regular function f on X, let $\mathrm{div}(f)$ be the divisor whose value at x is the order of vanishing of f at x. A divisor of the form $\mathrm{div}(f)$ is called a

principal divisor. The subgroup of principal divisors of X is denoted as $\mathrm{Div}^l(X)$.

(7) The **degree** of a divisor $D = (n_x) \in \mathrm{Div}(X)$ is

$$\deg D = \sum_{x \in X} n_x \deg x,$$

where $\deg x$ is the degree of its residue field[2]. Principal divisors on algebraic curves have degree 0. Let $\mathrm{Div}^0(X) \subset \mathrm{Div}(X)$ be the subgroup of degree-zero divisors. Define the **Picard group** of C as

$$\mathrm{Pic}^0(X) = \mathrm{Div}^0(X)/\mathrm{Div}^l(X).$$

(8) We denote the point x in the Picard group as $[x]$. If $h : X \to Y$ is a morphism of algebraic curves, then it induces forward backward maps on Picard groups:

$$h_* : \mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y) \qquad h_*\left(\sum_{x \in X} n_x[x]\right) = \sum_{x \in X} n_x[h(x)],$$

$$h^* : \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X) \qquad h^*\left(\sum_{y \in Y} n_y[y]\right) = \sum_{y \in Y} n_y \sum_{x \in h^{-1}(y)} e_h(x)[x],$$

where $e_h(x)$ is the degree of ramification at x. These induced maps are functorial, namely, if $h : X \to Y$ and $g : Y \to Z$ are morphisms, then $(g \circ h)_* = g_* \circ h_*$, and $(g \circ h)^* = g^* \circ h^*$. Moreover, they satisfy

$$h_* \circ h^* = \deg h,$$

where $\deg h$ denotes multiplication by $\deg h$.

(9) Let E be an elliptic curve. Then the map

$$\mathrm{Div}(E) \to E\colon \qquad \sum n_P(P) \to \sum n_P P$$

(where (P) is the point in $\mathrm{Div}(E)$ corresponding to $P \in E$) induces a group isomorphism

$$\mathrm{Pic}^0(E) \xrightarrow{\sim} E.$$

In particular, this gives a characterization of principal divisors:

$$\sum n_P(P) \in \mathrm{Div}^l(E) \iff \sum n_P = 0 \text{ and } \sum n_P P = 0_E.$$

(10) Let μ_N denote the group of N^{th} roots of unity in $\overline{k}$. If $\mathrm{Char}(k) = 0$, then there is a bilinear, alternating, and non-degenerate pairing $e_N : E[N] \times E[N] \to \mu_N$ that also satisfies the following properties:

- For any $\sigma \in \mathrm{Gal}(\overline{k}/k)$, $e_N(P,Q)^\sigma = e_N(P^\sigma, Q^\sigma)$.
- e_N is compatible with isomorphism, i.e., if $\varphi : E \to E'$ is an isomorphism of elliptic curves taking P (resp. Q) to P' (resp. Q'), then $e_N(P,Q) = e_N(P',Q')$.
- If P, Q forms an ordered basis for $E[N]$, then $e_N(P,Q)$ is a primitive Nth root of unity.

This is called the **Weil pairing**.

[2]If X is an algebraic variety, then the residue field of $x \in X$ is defined to be $k(x) = \mathscr{O}_{X,x}/\mathfrak{m}_{X,x}$, where $\mathscr{O}_{X,x}$ is the local ring at x and $\mathfrak{m}_{X,x}$ is its maximal ideal.

2.2. **Galois representation.** One of the fundamental objects of study for number theorists is finite extensions of $\mathbb{Q}$, and in particular, the Galois extensions. This naturally leads to investigating the structure of their Galois groups, and hence, the absolute Galois group

$$G_{\mathbb{Q}} := \mathrm{Aut}(\overline{\mathbb{Q}}/\mathbb{Q}).$$

Understanding $G_{\mathbb{Q}}$ amounts to understanding its representations. This subsection defines l-adic Galois representations and discusses the one-dimensional case.

Because the group structure of the infinite group $G_{\mathbb{Q}}$ is extremely complicated, it is desirable to impose a continuity condition on the representations of $G_{\mathbb{Q}}$, and to this end we first need to define a topology on $G_{\mathbb{Q}}$. We would like this topology to "restrict to the discrete topology" when we project to the quotient $\mathrm{Gal}(K/\mathbb{Q}) = G_{\mathbb{Q}}/G_K$ for any finite extension $K/\mathbb{Q}$. This motivates:

Definition 2.1. The **Krull topology** on $G_{\mathbb{Q}}$ is the topology on $G_{\mathbb{Q}}$ generated by

$$U_\sigma(K) := \sigma \cdot \ker(G_{\mathbb{Q}} \to \mathrm{Gal}(K/\mathbb{Q}))$$

for K a number field Galois over $\mathbb{Q}$ and $\sigma \in G_{\mathbb{Q}}$, where $G_{\mathbb{Q}} \to \mathrm{Gal}(K/\mathbb{Q})$ is restriction from $\overline{\mathbb{Q}}$ to K. Let $U(K) = U_1(K)$.

Definition 2.2. Let l be a prime of $\mathbb{Z}$. An **l-adic Galois representation** is a continuous group homomorphism

$$\rho : G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{L}),$$

where $\mathbb{L}$ is a finite extension of $\mathbb{Q}_l$.

Before proceeding, we shall explain why we consider l-adic representations instead of complex representations. Consider a continuous representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{C})$. Choose an open neighborhood W of the identity in $\mathrm{GL}_n(\mathbb{C})$ that contains no non-trivial subgroups. Since $\rho^{-1}(W)$ is a neighborhood of 1 in $G_{\mathbb{Q}}$, it must contain some $U(K)$ for some number field K. Since $U(K)$ is a subgroup, so is $\rho(U(K)) \subset W$, and by the choice of W, $\rho(U(K)) = 1$. Thus, ρ factors as a representation $\mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$, which means that ρ fails to capture the infinite structure of $G_{\mathbb{Q}}$.

On the other hand, we now construct a 1-dimensional l-adic Galois representation that does keep track of some infinite structure of $G_{\mathbb{Q}}$. For $\sigma \in G_{\mathbb{Q}}$, let $\chi_l(\sigma, n)$ be defined by

$$\mu_{l^n}^{\sigma} = \mu_{l^n}^{\chi_l(\sigma,n)} \qquad \chi_l(\sigma, n) \in \mathbb{Z}.$$

The identification $\mathrm{Gal}(\mathbb{Q}(\mu_{l^n})/\mathbb{Q}) \cong (\mathbb{Z}/l^n\mathbb{Z})^*$ takes $\sigma|_{\mathbb{Q}(\mu_{l^n})}$ to $\chi_l(\sigma, n)$. Furthermore, under this identification, the restriction $\mathrm{Gal}(\mathbb{Q}(\mu_{l^n})/\mathbb{Q}) \to \mathrm{Gal}(\mathbb{Q}(\mu_{l^{n-1}})/\mathbb{Q})$ is compatible with the natural restriction $(\mathbb{Z}/l^n\mathbb{Z})^* \to (\mathbb{Z}/l^{n-1}\mathbb{Z})^*$. Hence, there is a map

$$\chi_l \colon G_{\mathbb{Q}} \to \varprojlim_n \{(\mathbb{Z}/l^n\mathbb{Z})^*\} = \mathbb{Z}_l^* \subset \mathbb{Q}_l^* \cong \mathrm{GL}_1(\mathbb{Q}_l) \colon \quad \sigma \mapsto (\chi_l(\sigma,1), \chi_l(\sigma,2), \ldots).$$

This is clearly a group homomorphism. To check continuity, it suffices to how that $\chi_l^{-1}(1 + l^n\mathbb{Z}_l)$ is open in $G_{\mathbb{Q}}$. To see this, observe that

$$\begin{aligned} \sigma \in \chi_l^{-1}(1 + l^n\mathbb{Z}_l) &\iff \chi_l(\sigma, m) = 1 \text{for each } m \leq n \\ &\iff \sigma|_{\mathbb{Q}(\mu_{l^n})} = 1 \\ &\iff \sigma \in U(\mathbb{Q}(\mu_{l^n})). \end{aligned}$$

Definition 2.3. The l-adic Galois representation $\chi_l : G_{\mathbb{Q}} \to \mathrm{GL}_1(\mathbb{Q}_l)$ constructed above is called the **l-adic cyclotomic character**.

Example 2.4. Recall the following notions from algebraic number theory. Let $\mathfrak{p} \subset \overline{\mathbb{Z}}$ be a prime that lies over a prime $p \in \mathbb{Z}$. Then the decomposition group is defined to be

$$D_{\mathfrak{p}} := \{\sigma \in G_{\mathbb{Q}} : \mathfrak{p}^{\sigma} = \mathfrak{p}\}.$$

Now, reduction mod $\mathfrak{p}$ defines a surjective homomorphism

$$D_{\mathfrak{p}} \to \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p),$$

whose kernel is

$$I_{\mathfrak{p}} := \{\sigma \in G_{\mathbb{Q}} : x^{\sigma} \equiv x \mod \mathfrak{p} \text{ for all } x \in G_{\mathbb{Q}}\}.$$

The Frobenius automorphism in $\overline{\mathbb{F}}_p/\mathbb{F}_p$ which takes $x \in \overline{\mathbb{F}}_p$ to x^p pulls back to an element $\mathrm{Frob}_{\mathfrak{p}} \in D_{\mathfrak{p}}/I_{\mathfrak{p}}$, which satisfies

$$x^{\mathrm{Frob}_{\mathfrak{p}}} \equiv x^p \mod \mathfrak{p} \qquad x \in \overline{\mathbb{Q}}.$$

The notation $\mathrm{Frob}_{\mathfrak{p}}$ may be used to denote any element in the coset $D_{\mathfrak{p}}/I_{\mathfrak{p}}$. We now claim that $\chi_l(\mathrm{Frob}_{\mathfrak{p}}) = p$ for $p \neq l$. Indeed, $\mathbb{Q}(\mu_{l^n})/\mathbb{Q}$ is unramified over p and the Galois group is abelian, hence $\mathrm{Frob}_{\mathfrak{p}}|_{\mathbb{Q}(\mu_{l^n})}$ is uniquely determined. It is precisely the map $\mu_{l^n} \mapsto \mu_{l^n}^p$. Thus, $\chi_l(\mathrm{Frob}_{\mathfrak{p}}, n) = p$ for all n, so $\chi_l(\mathrm{Frob}_{\mathfrak{p}}) = p \in \mathrm{GL}_1(\mathbb{Q}_l)$.

Definition 2.5. Let ρ be a Galois representation and let p be a prime. Then ρ is unramified at p if $I_{\mathfrak{p}} \subset \ker \rho$ for any nonzero prime $\mathfrak{p} \subset \overline{\mathbb{Z}}$ lying over p.

For example, χ_l is unramified at p for $p \neq l$.

For ρ unramified, $\rho(\mathrm{Frob}_{\mathfrak{p}})$ is well-defined for each $\mathfrak{p}$ and changing $\mathfrak{p}$ conjugates it, and so the conjugacy class of $\rho(\mathrm{Frob}_{\mathfrak{p}})$ only depends on p.

2.3. **Tate module.** In this subsection we define the Tate module and construct the Galois representation associated to an elliptic curve. Each Galois action is sent to an automorphism of the Tate module.

Definition 2.6. Let E be an elliptic curve over $\mathbb{Q}$. Consider the inverse system

$$E[l] \xleftarrow{\cdot l} E[l^2] \xleftarrow{\cdot l} E[l^3] \leftarrow \cdots$$

The **Tate module** of E is defined to be the inverse limit

$$\mathrm{Ta}_l(E) := \varprojlim_n \{E[l^n]\}.$$

By choosing a basis for all n, compatibly with multiplication by l, we have

$$\mathrm{Ta}_l(E) \cong \mathbb{Z}_l^2.$$

The Galois group $G_{\mathbb{Q}}$ acts on the points of an elliptic curve E over $\mathbb{Q}$ coordinate-wise: if $P = (x_P, y_P) \in \overline{\mathbb{Q}}(E)$, then $P^{\sigma} = (x_P^{\sigma}, y_P^{\sigma})$. Since $(x_{P+Q}, y_{P+Q}) = (r_1(x_P, y_P, x_Q, y_Q), r_2(x_P, y_P, x_Q, y_Q))$, where r_i are rational functions over $\mathbb{Q}$, the Galois action commutes with r_i, hence $(P+Q)^{\sigma} = P^{\sigma} + Q^{\sigma}$. In other words, any $\sigma \in G_{\mathbb{Q}}$ gives a group homomorphism on the points of E. If we take the inverse limit of the following diagram

$$\begin{array}{ccc} E[l^{n+1}] & \xrightarrow{\cdot l} & E[l^n] \\ \sigma\downarrow & & \downarrow\sigma \\ E[l^{n+1}] & \xrightarrow[\cdot l]{} & E[l^n] \end{array}$$

we obtain an automorphism σ of $\mathrm{Ta}_l(E)$. We thus get a group homomorphism

$$\rho_{E,l} : G_{\mathbb{Q}} \to \mathrm{Aut}(\mathrm{Ta}_l(E)) \cong \mathrm{GL}_2(\mathbb{Z}_l) \subset \mathrm{GL}_2(\mathbb{Q}_l).$$

This is the desired representation. This is continuous. Indeed, a distinguished base near $\mathrm{id} \in \mathrm{GL}_2(\mathbb{Q}_l)$ is the set of $(\mathrm{id} + M_{2\times 2}(l^n\mathbb{Z}_l)) \cap \mathrm{GL}_2(\mathbb{Q}_l)$, and

$$\begin{aligned} \rho_{E,l}^{-1}\left((\mathrm{id} + M_{2\times 2}(l^n\mathbb{Z}_l)) \cap \mathrm{GL}_2(\mathbb{Q}_l)\right) &= \{\sigma \in G_{\mathbb{Q}} : \sigma|_{E[l^n]} = 1\} \\ &= \{\sigma \in G_{\mathbb{Q}} : \sigma|_{\mathbb{Q}(E[l^n])} = 1\}, \end{aligned}$$

where $\mathbb{Q}(E[l^n])$ is the finite extension over $\mathbb{Q}$ obtained by adjoining the coordinates of all $E[l^n]$-torsion points. This is open in the Krull topology, hence proving the continuity.

2.4. **Reduction of elliptic curves.** In this subsection, we first state some preliminaries on reductions of elliptic curves (for details, see [4] or [3]), then record some important facts about the Frobenius morphism.

For elliptic curves in general characteristic (particularly 2 or 3), one has to consider a more general form of Weierstrass equation, and express the discriminant and the invariant using the general form. We will skip this detail and refer the interested readers to Section 8.1 of [4].

Definition 2.7. For an elliptic curve E over $\mathbb{Q}$, let

$$v_p(E) = \min\{v_p(\Delta(E')) : E' \text{ integer coefficient and equivalent to } E\}.$$

Considering possible admissible changes of variables, if $v_p(\Delta) < 12$, $v_p(E) = v_p(\Delta)$.

Define the **global minimal discriminant of** E to be

$$\Delta_{\min}(E) = \prod_p p^{v_p(E)}.$$

In fact, the p-adic valuation of the discriminant can be minimized to $v_p(E)$ simultaneously for all p by an admissible change of variables. Thus, there is a **global minimal Weierstrass equation**. Assume E is global minimal. Define a Weierstrass equation $\widetilde{E}$ over $\mathbb{F}_p$ by reducing the coefficients of E mod p.

Definition 2.8. An elliptic curve E has reduction at p which is

(1) **good** (*nonsingular, stable*) if $\widetilde{E}$ is again an elliptic curve. Equivalently, $p \nmid \Delta_{\min}(E)$. Moreover, the reduction is
   (a) **ordinary** if $\widetilde{E}[p] \cong \mathbb{Z}/p\mathbb{Z}$.
   (b) **supersingular** if $\widetilde{E}[p] = 0$.

(2) **bad** (*singular*) if $\widetilde{E}$ is not an elliptic curve. Moreover, putting $\widetilde{E}$ in the form $(y - m_1x)(y - m_2x) = x^3$, the reduction is
   (a) **multiplicative** (*semistable*) if the singular point is a node ($m_1 \neq m_2$);
      (i) **split** if $m_1, m_2 \in \mathbb{F}_p$.
      (ii) **non-split** if $m_1, m_2 \notin \mathbb{F}_p$.
   (b) **additive** (*unstable*) if the singular point is a cusp ($m_1 = m_2$).

These reduction types are all independent of the choice of global minimal Weierstrass equation.

The following notion tracks the primes where an elliptic curve has bad reduction. This will be used in section 5 and section 6.

Definition 2.9. Let E be an elliptic curve over $\mathbb{Q}$. Then define the **algebraic conductor** of E to be $N_E = \prod_p p^{f_p}$, where

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

for certain $\delta_2 \leq 6$ and $\delta_3 \leq 3$, which measures "wild ramification" in the action of the inertia group on $\mathrm{Ta}_l(E)$ (see Appendix C in [3]). In particular, $p \mid \Delta_{\min}(E) \iff p \mid N_E$.

In the positive characteristic setting, there is an important map, called the Frobenius endomorphism. We state preliminary results on the Frobenius map, which will be used freely later in the paper.

(1) Recall the **Frobenius endomorphism**

$$\sigma_p \colon k \to k \colon x \mapsto x^p$$

for any field k of characteristic p, which is an automorphism for $k = \overline{\mathbb{F}}_p$. It induces for any projective curve $C \subset \mathbb{P}^n$ over $\overline{\mathbb{F}}_p$ a morphism

$$\sigma_p \colon C \to C^{\sigma_p} \colon \qquad [x_0 : \cdots : x_n] \mapsto [x_0^p : \cdots : x_n^p],$$

where C^{σ_p} is defined by applying σ_p to the coefficients of the equations defining C.

(2) We say that a morphism of curves is separable (resp. inseparable, purely inseparable) if the induced map of function fields is separable (resp. inseparable, purely inseparable). The Frobenius morphism σ_p is purely inseparable with degree p [3, p. 25].

(3) Since any field extension can be factored as a separable extension and a purely inseparable extension, any morphism $h : C \to C'$ of curves can be factored as

$$h : C \xrightarrow{h_{\mathrm{ins}}} C_{\mathrm{sep}} \xrightarrow{h_{\mathrm{sep}}} C'.$$

Any purely inseparable extension is obtained by successively adjoining p-th roots of multiplicity p. Geometrically, this means (for curves) that $h_{\mathrm{ins}} = \sigma_p^e$ for some $e \geq 0$ up to isomorphism. Thus, $h = h_{\mathrm{sep}} \circ \sigma_p^e$. For all but finitely many points $y \in C'$, $|h^{-1}(y)| = \deg_{\mathrm{sep}}(h)$. In the special case where $\varphi : C \to C'$ is an isogeny, all fibers have the same cardinality (since φ is a group homomorphism), so $|h^{-1}(y)| = \deg_{\mathrm{sep}}(h)$ for all y. In particular, $|\ker \varphi| = \deg_{\mathrm{sep}}(h)$.

(4) The separable (resp. inseparable) degree of a morphism $h : C \to C'$ is defined to be the degree of h_{sep} (resp. h_{ins}). These degrees are multiplicative: if $h' : C' \to C''$ is another morphism, then $\deg_{\mathrm{sep}}(h' \circ h) = \deg_{\mathrm{sep}}(h') \cdot \deg_{\mathrm{sep}}(h)$ and similarly for inseparable degrees.

(5) The induced forward and backward maps of σ_p on the Picard group are easily computed to be

$$\sigma_{p,*}\colon (P) \mapsto (\sigma_p(P)), \qquad \sigma_p^*\colon (P) \mapsto p(\sigma_p^{-1}(P)).$$

For $h : C \to C'$, the following equalities hold:

- $h \circ \sigma_{p,C} = \sigma_{p,C'} \circ h$.
- $h_* \circ \sigma_{p,C}^* = \sigma_{p,C'}^* \circ h_*$.

2.5. **The characteristic polynomial.** In this subsection, we will show that the characteristic polynomial of $\rho_{E,l}(\mathrm{Frob}_\mathfrak{p})$ is related to the solution counts of E modulo p. The characteristic polynomial is especially important, since by the Chebotarev density theorem, the Frobenius elements (possibly excluding those over finitely many integer primes for which $\rho_{E,l}$ is ramified) are dense in $G_\mathbb{Q}$. Since $\rho_{E,l}$ is continuous, these $\rho_{E,l}(\mathrm{Frob}_\mathfrak{p})$ determine the representation $\rho_{E,l}$ completely.

Definition 2.10. Let E be an elliptic curve over $\mathbb{Q}$ with good reduction at p and $\widetilde{E}$ be its reduction mod p. Then define

$$a_p(E) := p + 1 - |\widetilde{E}(\mathbb{F}_p)|.$$

This definition captures how much the number of $\mathbb{F}_p$-points of $\widetilde{E}$ differs from the estimated number, which is $p + 1$. (Intuitively, for p not equal to 2 and 3, the Weierstrass equation can be written as $y^2 = f(x)$ for some cubic polynomial f. For roughly half the x's in $\mathbb{F}_p$, $f(x)$ is a quadratic residue, yielding two solutions for y, while for roughly half the x's, it is a non-quadratic residue, yielding no solutions for y. Adding in the point at infinity, there should be about $p+1$ solutions in $\mathbb{F}_p$.)

Lemma 2.11. *With the same setup as in the above definition,*

$$a_p(E) = \sigma_{p,*} + \sigma_p^* \qquad \text{on } Pic^0(\widetilde{E}).$$

Proof. Note that

$$\widetilde{E}(\mathbb{F}_p) = \{P \in \widetilde{E} : P^{\sigma_p} = P\} = \ker(\sigma_p - 1).$$

And observe that $\sigma_p - 1$ is separable: indeed, if not, then $\sigma_p - 1 = f \circ \sigma_p$ for some morphism f, so that $(1 - f) \circ \sigma_p = 1$, showing that σ_p is an isomorphism; this is not the case, contradiction. Thus,

$$|\widetilde{E}(\mathbb{F}_p)| = |\ker(\sigma_p - 1)| = \deg(\sigma_p - 1) = (\sigma_p - 1)_* \circ (\sigma_p - 1)^* = p + 1 - \sigma_{p,*} - \sigma_p^*,$$

from which the result follows. □

Theorem 2.12. *Let l be a prime and E an elliptic curve over $\mathbb{Q}$ with conductor N. The Galois representation $\rho_{E,l}$ is unramified at every prime $p \nmid lN$. For any such p, let $\mathfrak{p} \subset \overline{\mathbb{Z}}$ be any maximal ideal over p. Then the characteristic polynomial of $\rho_{E,l}(Frob_\mathfrak{p})$ is*

$$x^2 - a_p(E)x + p = 0.$$

Moreover, $\rho_{E,l}$ is irreducible.

Proof. Observe that the following diagram commutes:

$$\begin{array}{ccc} D_\mathfrak{p} & \xrightarrow{\rho_n} & \mathrm{Aut}(E[l^n]) \\ {\scriptstyle \pi_\mathfrak{p}}\downarrow & & \downarrow{\scriptstyle \pi_n} \\ G_{\mathbb{F}_p} & \longrightarrow & \mathrm{Aut}(\widetilde{E}[l^n]) \end{array}$$

From Chapter VII Proposition 3.1 of [3], one knows that, if E has good reduction at p and $p \neq l$ (i.e., if $p \nmid lN$), $E[l^n] \to \widetilde{E}[l^n]$ is injective.[3] Using the known structure of the torsion subgroups of elliptic curves and comparing cardinalities, it follows that $E[l^n] \to \widetilde{E}[l^n]$ is an isomorphism. Hence, π_n is an isomorphism. By definition, $\ker \pi_{\mathfrak{p}} = I_{\mathfrak{p}}$, so $I_{\mathfrak{p}} = \ker \pi_{\mathfrak{p}} \subset \ker \pi_n \circ \rho_n = \ker \rho_n$. This proves that the representation is unramified at such primes.

Consider the two diagrams below:

$$\begin{array}{ccc} E[l^n] & \xrightarrow{a_p(E)} & E[l^n] \\ \downarrow & & \downarrow \\ \widetilde{E}[l^n] & \xrightarrow[\sigma_p + p\sigma_p^{-1}]{} & \widetilde{E}[l^n] \end{array} \qquad \begin{array}{ccc} E[l^n] & \xrightarrow{\mathrm{Frob}_{\mathfrak{p}} + p\mathrm{Frob}_{\mathfrak{p}}^{-1}} & E[l^n] \\ \downarrow & & \downarrow \\ \widetilde{E}[l^n] & \xrightarrow[\sigma_p + p\sigma_p^{-1}]{} & \widetilde{E}[l^n] \end{array}$$

The second diagram commutes by facts from the last subsection; the first diagram commutes by the lemma and the identification of E with $\mathrm{Pic}^0(E)$. Since the vertical maps are isomorphism, this means $a_p(E) = \mathrm{Frob}_{\mathfrak{p}} + p\mathrm{Frob}_{\mathfrak{p}}^{-1}$. Thus, $\mathrm{Frob}_{\mathfrak{p}}^2 - a_p(E)\mathrm{Frob}_{\mathfrak{p}} + p = 0$.

To show that this is indeed the characteristic polynomial, it suffices to show that $\det \rho_{E,l}(\mathrm{Frob}_{\mathfrak{p}}) = p$. To see this, pick an ordered basis P, Q of $E[l^n]$, and pick a primitive root of unity $\mu_{l^n} = e_N(P, Q)$. Then

$$\mu_{l^n}^{\sigma} = e_N(P, Q)^{\sigma} = e_N(P^{\sigma}, Q^{\sigma}) = \mu_{l^n}^{\det \rho_n(\sigma)}$$

where e_N is the Weil pairing. But by definition of χ_l, $\mu_{l^n}^{\sigma} = \mu_{l^n}^{\chi_l(\sigma,n)}$. Thus, $\det \rho_n(\sigma) = \chi_l(\sigma, n)$ in $\mathbb{Z}/l^n\mathbb{Z}$. Taking inverse limits, this shows that $\det \rho_{E,l} = \chi_l$. But $\chi_l(\mathrm{Frob}_{\mathfrak{p}}) = p$, so $\det \rho_{E,l}(\mathrm{Frob}_{\mathfrak{p}}) = p$, as desired.

For the last statement, Theorem IV.2.1(a) of [5] proves that $\rho_{E,l}$ is irreducible if E has no complex multiplication over $\mathbb{Q}$. But $E/\mathbb{Q}$ cannot have complex multiplication over $\mathbb{Q}$ (cf. discussions in [6]). □

We will come back to use this characteristic polynomial when we show that the two formulations of the Modularity Theorem discussed in this paper are equivalent.

3. Modular Curves as Moduli Spaces of Elliptic Curves

Whenever we are given a mathematical object, it is natural to ask: can we classify these objects up to isomorphism? In algebraic geometry, such classification problems often have a set of solutions that itself carries a geometric structure, such as a Riemann surface, a variety, or a scheme. Roughly speaking, these "parameter spaces" are called **moduli spaces**. In this section, we introduce modular curves as moduli spaces of elliptic curves equipped with additional data, and modular forms as differentials on modular curves. We begin by examining the relationship between complex tori and complex elliptic curves. Next, we define three important families of modular curves: $X_0(N)$, $X_1(N)$, and $X(N)$. We then briefly discuss modular forms, and finally, we present a model of these modular curves over $\mathbb{Q}$.

[3] In [3], E is assumed to be an elliptic curve over a local field. The Proposition applies to our case because there is an injection $E[m] \to E(\mathbb{Q}_p)[m]$, which allows us to pass to the local field case.

3.1. **Complex tori and complex elliptic curves.** In this subsection, we prove some elementary properties of complex tori, and we will show that complex tori are the same as complex elliptic curves.

Definition 3.1. A **complex torus** is the space $\mathbb{C}/\Lambda$, for a lattice Λ in $\mathbb{C}$.

A complex torus has a natural Riemann surface structure induced from that of $\mathbb{C}$. It is also an abelian group under addition.

Definition 3.2. An **isogeny** between complex tori is a non-constant holomorphic group homomorphism between complex tori $\varphi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$. An **isomorphism of complex tori** is a biholomorphic isogeny.

We now characterize different maps of complex tori.

Proposition 3.3. *Let $\varphi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ be a non-constant map of complex tori.*

1. *φ is holomorphic if and only if φ is of the form $\varphi(z+\Lambda) = mz + b + \Lambda'$, where $m, b \in \mathbb{C}$ and $m\Lambda \subset \Lambda'$.*
2. *φ is an isogeny if and only if φ is of the form $\varphi(z+\Lambda) = mz + \Lambda'$, where $m \in \mathbb{C}$ and $m\Lambda \subset \Lambda'$.*
3. *φ is an isomorphism of complex tori if and only if φ is of the form $\varphi(z+\Lambda) = mz + \Lambda'$, where $m \in \mathbb{C}$ and $m\Lambda = \Lambda'$.*

Proof. All the "if" directions are clear. We will only prove (1) since the rest easily follows. Suppose φ is holomorphic. Since $\mathbb{C}$ is the universal cover for both complex tori, φ can be lifted to a map $\widetilde{\varphi} : \mathbb{C} \to \mathbb{C}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\widetilde{\varphi}} & \mathbb{C} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi'} \\ \mathbb{C}/\Lambda & \xrightarrow[\varphi]{} & \mathbb{C}/\Lambda' \end{array}$$

Now for each $\lambda \in \Lambda$, let $f_\lambda(z) := \widetilde{\varphi}(z+\lambda) - \widetilde{\varphi}(z)$. Then $\pi' \circ f_\lambda = \varphi(z+\lambda+\Lambda) - \varphi(z+\Lambda) = 0 + \Lambda'$, so $\mathrm{im}(f_\lambda) \subseteq \Lambda'$. But f_λ is continuous and Λ' is discrete, forcing f_λ to be a constant. Differentiating f_λ, we obtain $\widetilde{\varphi}'(z+\lambda) = \widetilde{\varphi}'(z)$, so $\widetilde{\varphi}'$ is Λ-periodic. But this implies that $\widetilde{\varphi}'$ is bounded, hence constant by Liouville's Theorem. Thus, $\widetilde{\varphi}$ is linear, and φ is of the desired form. This proves (1). □

Let $\mathcal{H}$ denote the complex upper half-plane $\{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathcal{H}$ by fractional linear transformations:

$$\gamma(z) = \frac{az+b}{cz+d}, \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ z \in \mathcal{H}.$$

Proposition 3.4.

1. *Let $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ and $\Lambda' = \mathbb{Z}\omega_1' \oplus \mathbb{Z}\omega_2'$ such that $\omega_1/\omega_2, \omega_1'/\omega_2' \in \mathcal{H}$. Then*
$$\Lambda = \Lambda' \iff \begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \gamma \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \quad \text{for some } \gamma \in SL_2(\mathbb{Z}).$$
2. *Let $\Lambda_\tau = \mathbb{Z}\tau \oplus \mathbb{Z}$ for $\tau \in \mathcal{H}$. Then every complex torus is isomorphic to some $\mathbb{C}/\Lambda_\tau$.*
3. *$\mathbb{C}/\Lambda_{\tau_1}$ is isomorphic to $\mathbb{C}/\Lambda_{\tau_2}$ if and only if $\tau_1 = \gamma\tau_2$ for some $\gamma \in SL_2(\mathbb{Z})$.*

Proof. Simple exercise. □

Now we state the relationship between complex tori and elliptic curves.

Definition 3.5. The **Weierstrass $\wp$-function** is

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda - \{0\}} \Big(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \Big), \qquad z \in \mathbb{C} - \Lambda.$$

Definition 3.6. Let $\Lambda \subset \mathbb{C}$ be a lattice. The **weight-k Eisenstein series** is

$$G_k(\Lambda) = \sum_{\omega \in \Lambda - \{0\}} \frac{1}{\omega^k}, \qquad k > 2 \text{ even}.$$

Also define

$$g_2(\Lambda) = 60 G_4(\Lambda), \qquad g_3(\Lambda) = 140 G_6(\Lambda).$$

Theorem 3.7. *Let $\Lambda \subset \mathbb{C}$ be a lattice, and $g_2 = g_2(\Lambda)$, $g_3 = g_3(\Lambda)$.*

(1) *The functions $\wp$ and its derivative $\wp'$ satisfy*

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

(2) *The polynomial $4x^3 - g_2(\Lambda) - g_3(\Lambda)$ has distinct roots, hence the discriminant $\Delta(\Lambda) = g_2(\Lambda)^3 - 27 g_3(\Lambda)^2$ is non-zero.*

(3) *The curve $E : y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$ is an elliptic curve, and the map*

$$\phi \colon \mathbb{C}/\Lambda \to E(\mathbb{C}) \subset \mathbb{P}^2(\mathbb{C}) \colon \qquad z \mapsto [\wp(z) : \wp'(z) : 1]$$

is an isomorphism of complex Lie groups (i.e., a biholomorphic group isomorphism).

(4) *Every pair of $g_2, g_3 \in \mathbb{C}$ satisfying $g_2^3 - 27 g_3^2 \neq 0$ can be expressed as $g_2 = g_2(\Lambda)$, $g_3 = g_3(\Lambda)$ for some lattice $\Lambda \subset \mathbb{C}$.*

(5) *Let E_1, E_2 be the complex elliptic curves associated to lattices Λ_1, Λ_2 respectively. Then there is a functorial bijection*

$$\{\text{isogenies } \phi \colon \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2\} \to \{\text{isogenies } \phi \colon E_1 \to E_2\}.$$

defining an equivalence between the categories of tori and elliptic curves.

Proof. See Chapter VI of [3] or Chapter 2 of [4]. □

3.2. **$X_0(N)$, $X_1(N)$, and $X(N)$.** By Proposition 3.4, we see that each complex torus can be represented by $\mathbb{C}/\Lambda_\tau$ where $\tau \in \mathcal{H}$, and such a representation is unique up to action by $\mathrm{SL}_2(\mathbb{Z})$. Thus, the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ is a **moduli space** of complex tori, i.e., in bijection with the set of complex tori up to isomorphism. By Theorem 3.7, there is a bijective correspondence between complex tori and elliptic curves. Thus, the space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ also classifies complex elliptic curves. The space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ is an example of a modular curve over $\mathbb{C}$. The group $\mathrm{SL}_2(\mathbb{Z})$ is generated by

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

and the space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ can be visually represented by:

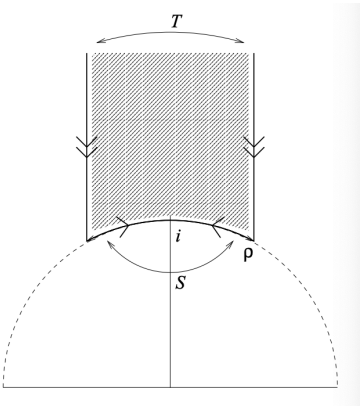

FIGURE 1. Fundamental Domain of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ [7].

The (open) shaded region is called a **fundamental domain** for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. The arrows on the boundary lines indicate that these lines are identified by the given transformations.

By modifying the classification problem, we can obtain other spaces of solutions. For example, we could keep track of not just the isomorphism class of an elliptic curve but also additional torsion data.

Definition 3.8. Let

$$S_0(N) = \{\text{isomorphism classes of pairs } (E, C), \text{ where } E \text{ is a complex elliptic curve and } C \text{ is a cyclic subgroup of order } N\},$$

$$S_1(N) = \{\text{isomorphism classes of pairs } (E, Q), \text{ where } E \text{ is a complex elliptic curve and } Q \text{ is a point of order } N\},$$

$$S(N) = \{\text{isomorphism classes of pairs } (E, (P, Q)), \text{ where } E \text{ is a complex elliptic curve and } (P, Q) \text{ is a pair of points that generate } E[N]\}.$$

As a special case, the space of elliptic curves over $\mathbb{C}$ can be viewed as $S_1(1)(\mathbb{C})$. As in this special case, there are subgroups $\Gamma_0(N)$, $\Gamma_1(N)$, and $\Gamma(N)$ that describe algebraically the parameter space for these elliptic curves with additional data. These groups are defined by

$$\Gamma(N) = \left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N\right\},$$

$$\Gamma_1(N) = \left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod N\right\},$$

$$\Gamma_0(N) = \left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod N\right\}.$$

Note that $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$, and $\Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$. Now also define

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

Theorem 3.9. *Let N be a positive integer. For $\tau \in \mathcal{H}$, denote by E_τ the complex elliptic curve corresponding to the complex torus $\mathbb{C}/\Lambda_\tau$. Then*

(1) $S_0(N) = \{[E_\tau, \langle 1/N + \Lambda_\tau\rangle] : \tau \in \mathcal{H}\}$, *and there is a bijection*

$$\psi_0 \colon S_0(N) \xrightarrow{\sim} Y_0(N) \colon \qquad [\mathbb{C}/\Lambda_\tau, \langle 1/N + \Lambda_\tau\rangle] \mapsto \Gamma_0(N)\tau.$$

(2) $S_1(N) = \{[E_\tau, 1/N + \Lambda_\tau] : \tau \in \mathcal{H}\}$, *and there is a bijection*

$$\psi_1 \colon S_1(N) \xrightarrow{\sim} Y_1(N) \colon \qquad [\mathbb{C}/\Lambda_\tau, 1/N + \Lambda_\tau] \mapsto \Gamma_1(N)\tau.$$

(3) $S(N) = \{[\mathbb{C}/\Lambda_\tau, (\tau/N + \Lambda_\tau, 1/N + \Lambda_\tau)] : \tau \in \mathcal{H}\}$, *and there is a bijection*

$$\psi \colon S(N) \xrightarrow{\sim} Y(N) \colon \qquad [\mathbb{C}/\Lambda_\tau, (\tau/N + \Lambda_\tau, 1/N + \Lambda_\tau)] \mapsto \Gamma(N)\tau.$$

Proof sketch. (For details, see section 1.5 in [4].) We will prove (2), the other parts being similar. We may take $E = \mathbb{C}/\Lambda_{\tau'}$ for some $\tau' \in \mathcal{H}$, so that $Q = (c\tau' + d)/N + \Lambda_{\tau'}$ for some $c, d \in \mathbb{Z}$. For the first statement, it suffices to find $m, \tau \in \mathbb{C}$ such that $m\Lambda_\tau = \Lambda_{\tau'}$ and $m(1/N + \Lambda_\tau) = Q$. Towards this, since Q has order N, $\gcd(c, d, N) = 1$, so we can find $a, b, k \in \mathbb{Z}$ such that $ad - bc - kN = 1$. The matrix $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ reduces mod N to $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Observe that changing its entries mod N does not change Q. So, since $\mathrm{SL}_2(\mathbb{Z})$ surjects onto $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, we can take $\gamma \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Take $m = c\tau' + d$ and $\tau = \gamma(\tau')$, and this yields what we wanted following a direct computation. We can also show that $[E_\tau, 1/N + \Lambda_\tau] = [E_{\tau'}, 1/N + \Lambda_{\tau'}]$ if and only if $\Gamma_1(N)\tau = \Gamma_1(N)\tau'$. Hence, ψ_1 is a bijection. □

Definition 3.10. A subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma(N)$ is called a **congruence subgroup** of level N. A space of the form $Y(\Gamma) := \Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$ is called a **modular curve**. By adding finitely many points, called **cusps**, one can compactify $Y(\Gamma)$, and the resulting compact Riemann surface is denoted $X(\Gamma)$. The spaces $X(\Gamma)$ are also called modular curves. In particular, denote by $X_0(N)$, $X_1(N)$, and $X(N)$ the compactifications of $Y_0(N), Y_1(N)$, and $Y(N)$, respectively.

3.3. **Modular forms.** Modular forms naturally arise as differentials on the modular curve. In this subsection, we will first define modular forms as functions on $\mathcal{H}$, then interpret them as differentials on $X(\Gamma)$.

Definition 3.11. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and k 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, i.e.,

$$f[\alpha]_k(\tau) := \frac{f(\alpha(\tau))}{(c\tau + d)^k} = f(\tau) \quad \text{for all } \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

(3) $f[\alpha]_k$ is holomorphic at ∞ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

We explain condition (3) here. Any congruence subgroup $\Gamma(N) \subset \Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ contains a translation matrix of the form

$$\begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} : \tau \mapsto \tau + h,$$

for example for $h = N$. Thus, if f satisfies (1) and (2) for Γ, then $g = f[\alpha]_k$ satisfies them for $\alpha^{-1}\Gamma\alpha \supset \Gamma(N)$ and can be written as $g' \circ q_h$, where $q_h(\tau) = e^{2\pi i\tau/h}$. Define

$f[\alpha]_k$ to be holomorphic at ∞ if g' extends holomorphically to $q = 0$. In particular, this means that f has a Fourier expansion

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

Definition 3.12. The space of modular forms of weight k with respect to Γ is denoted $\mathcal{M}_k(\Gamma)$. If $a_0 = 0$ in the Fourier expansion for $f[\alpha]_k$ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, then f is said to be a **cusp form**. The space of cusp forms of weight k with respect to Γ is denoted $\mathcal{S}_k(\Gamma)$.

The spaces $\mathcal{M}_k(\Gamma)$ are all *finite-dimensional* vector spaces [4, Sections 3.5, 3.6].

Example 3.13.

(a) The weight-k Eisenstein series defined before is a weight-k modular form in $\mathcal{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ for even $k \geq 4$. In particular, $g_2(\tau) = 60G_4(\tau)$ and $g_3 = 140G_6(\tau)$ are modular forms of weight 4 and 6, respectively.

(b) The **discriminant function** Δ is defined by

$$\Delta\colon \mathcal{H} \to \mathbb{C}\colon \qquad \tau \mapsto (g_2(\tau))^3 - 27(g_3(\tau))^2.$$

This is a cusp form of weight 12.

(c) Any weight-0 modular form is constant by compactness of $X(\Gamma)$.

(d) Consider

$$j\colon \mathcal{H} \to \mathbb{C}\colon \qquad \tau \mapsto 1728\frac{g_2(\tau)^3}{\Delta(\tau)}.$$

It is almost a modular form, except that it is not holomorphic at ∞. In fact, it defines an isomorphism of Riemann surfaces from $X(1)$ to $\mathbb{P}^1$.

Introduce the notation

$$j(\gamma, \tau) = c\tau + d \qquad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Observe that

$$d\gamma(\tau) = j(\gamma, \tau)^{-2} d\tau.$$

It follows that the differential $f(\tau)(d\tau)^{k/2}$ is Γ-invariant. The most important case for us occurs when $k = 2$. In this case, a weight-2 cusp form $f \in \mathcal{S}_2(\Gamma)$ corresponds to a holomorphic 1-form ω on $Y(\Gamma) = X(\Gamma) - \{\text{cusps}\}$, and with a bit of work one can show that this extends to a holomorphic 1-form on all of $X(\Gamma)$. For instance, at infinity, $f(\tau) = \sum_{n=1}^{\infty} a_n q_h^n$, where $q_h = e^{2\pi i\tau/h}$. It follows that $d\tau = \frac{h}{2\pi i}\frac{dq_h}{q_h}$, hence

$$f(\tau)d\tau = \left(\sum_{i=1}^{\infty} a_n q_h^n\right) \cdot \frac{h}{2\pi i}\frac{dq_h}{q_h} = \left(\frac{h}{2\pi i}\sum_{i=0}^{\infty} a_{n+1} q_h^n\right) dq_h$$

is holomorphic in q_h at 0, which is a Riemann surface coordinate at ∞. Applying this to $f[\alpha]_k$ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, we see that ω extends to all the cusps.

Conversely, if we are given a holomorphic 1-form ω on $X(\Gamma)$, then using the projection $\mathcal{H} \to Y(\Gamma)$, ω pulls back to a form $f(\tau)d\tau$ on $\mathcal{H}$ (every holomorphic 1-form on $\mathcal{H}$ can be expressed in this way since $\mathcal{H}$ is simply connected). Since ω is defined at $Y(\Gamma)$, one easily shows that $f[\alpha]_2 = f$ for all $\alpha \in \Gamma$, so one may write $f(\tau) = \sum_{n=0}^{\infty} a_n q_h^n$, where $q_h = e^{2\pi i\tau/h}$ for suitable h. By a similar calculation as above, one shows that $f \in \mathcal{S}_2(\Gamma)$. This (cf. section 3.3 of [4] for the full details) shows that:

Proposition 3.14. *The complex vector space of weight-2 cusp forms $\mathcal{S}_2(\Gamma)$ is isomorphic to the space of holomorphic differentials $\Omega^1_{hol}(X(\Gamma))$.*

3.4. **Modular curves over $\mathbb{Q}$.** In order to extract the arithmetic information from the modular curves, we will need an algebro-geometric description of the modular curve. In this subsection, we will sketch such a description. The main tool is the following:

Definition 3.15. A field K/k is called a function field over k if $K \cap \overline{k} = k$ and K is a finite extension of $k(t)$ for some transcendental element t over k.

Theorem 3.16. *There is an equivalence of categories between the category of non-singular projective curves over k (with non-constant morphisms) and the category of fields of transcendence degree 1 over k (with injective field homomorphism). In this correspondence:*

1. *C corresponds to $k(C)$, the function field of C*
2. *To pass from a function field K to a curve C, first write $K = k(x, f)$, where x is transcendental over k and f is algebraic over $k(x)$. Then $K \cong k[x, y]/(p(x, y))$ for some irreducible polynomial p. Normalizing the affine curve defined by $p(x, y)$ gives the desired curve C.*
3. *Non-constant morphisms $C \to C'$ bijectively correspond to injective homomorphisms $k(C') \to k(C)$ by pullback.*

Proof. See Chapter 7 of [8] or [9, Tag 0BY1]. □

Remark 3.1. An analogous correspondence holds for Riemann surfaces and their fields of meromorphic functions. There is an equivalence of categories between the category of Riemann surfaces (with non-constant holomorphic maps) and the category of function fields over $\mathbb{C}$ [10]. Thus, algebraic curves over $\mathbb{C}$ may be viewed analytically as complex Riemann surfaces, and vice versa.

The first step in defining a model of the modular curve over $\mathbb{Q}$ is to compute the function fields over $\mathbb{C}$. Let

$$f_0^{\overline{v}} = \frac{g_2(\tau)}{g_3(\tau)} \wp_\tau \left(\frac{c_v \tau + d_v}{N} \right) \qquad v = (c_v, d_v) \in \mathbb{Z}^2.$$

where the overline $\overline{v}$ denote reduction mod N. This function lies in $\mathbb{C}(X(N))$. Let

$$f_0^{\overline{d}}(\tau) = f_0^{\overline{(0,d)}}(\tau) \quad \text{for } d \not\equiv 0 \mod N, \qquad f_0(\tau) = \sum_{d=1}^{N-1} f_0^{\overline{d}}(\tau),$$

$$f_{1,0} = f_0^{\pm\overline{(1,0)}}, \qquad f_{0,1} = f_1 = f_0^{\pm\overline{(0,1)}},$$

and finally let $j_N(\tau) = j(N\tau)$.

Proposition 3.17. *The fields of meromorphic functions on $X_1(N)$ and $X_0(N)$ are*

$$\mathbb{C}(X_1(N)) = \mathbb{C}(j, f_1),$$

$$\mathbb{C}(X_0(N)) = \mathbb{C}(j, f_0) = \mathbb{C}(j, j_N).$$

Proof. See Section 7.5 of [4]. □

Now we define the "same" function fields for $X_1(N)$ and $X_0(N)$ but over $\mathbb{Q}$. Let

$$K_0 = \mathbb{Q}(j, f_0), \qquad K_1 = \mathbb{Q}(j, f_1).$$

It turns out that K_0 and K_1 are actually *function fields over* $\mathbb{Q}$, i.e., for $i = 0, 1$, f_i is algebraic over $\mathbb{Q}(j)$, and $K_i \cap \overline{\mathbb{Q}} = \mathbb{Q}$.

Definition 3.18. The curves that correspond (by Theorem 3.16) to K_0 and K_1 are denoted $X_0(N)_{\mathbb{Q}}$ and $X_1(N)_{\mathbb{Q}}$, respectively.

Example 3.19. The curve $X_0(11)_{\mathbb{Q}}$ is the elliptic curve with Weierstrass equation $y^2+y = x^3-x^2-10x-20$. The curve $X_1(11)_{\mathbb{Q}}$ is the elliptic curve with Weierstrass equation $y^2 + y = x^3 - x^2$. Obtaining these equations is a non-trivial task, and we refer the interested readers to [7].

The most difficult part of this argument is in proving that $K_i \cap \overline{\mathbb{Q}} = \mathbb{Q}$. For details, see Section 7.5-7.6 of [4]. The condition that $K_i \cap \overline{\mathbb{Q}} = \mathbb{Q}$ for $i = 0, 1$ is important for the following reasons.

Lemma 3.20. *Let X be a variety over a field k. Then X is geometrically irreducible over k (i.e., for any field extension k' over k, the base change $X(k')$ is irreducible) if and only if the separable algebraic closure of k in $K(X)$ is k.*

Proof. [9, Tag o54Q]. □

Thus, the fact that K_i is a function field ensures that the base change $X_i(N)_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{C}$ is a variety, and in fact, a non-singular projective curve over $\mathbb{C}$ by [9, Tag 0BY4]. Now take an open affine subvariety $X' = \mathrm{Spec}(\mathbb{Q}[j, y]/(p(j, y)))$ of $X_i(N)_{\mathbb{Q}}$. Then $k(X) = k(X') = \mathrm{Frac}(\mathbb{Q}[j, y]/(p))$. The base change $X' \times_{\mathbb{Q}} \mathbb{C}$ is an open subvariety of $X_i(N)_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{C}$. Thus, $k(X_i(N)_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{C}) = k(X' \times_{\mathbb{Q}} \mathbb{C}) = \mathrm{Frac}(\mathbb{C}[j, y]/(p))$, which is also the function field of $X_i(N)_{\mathbb{C}}$. Thus, it follows that

$$X_i(N)_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{C} = X_i(N)_{\mathbb{C}}.$$

Remark 3.2. There are also more intrinsic ways of defining the modular curve over $\mathbb{Q}$, using the general theory of representable functors. See [2].

4. Hecke operators

In this section, we sketch the theory of Hecke operators. We begin by defining them in several equivalent forms and then specialize the definitions to the case over $\mathbb{Q}$. Next, we focus on the interpretation of Hecke operators as endomorphisms on the space of cusp forms and state some classical results. The proofs of these results will not be given here, as they would take us too far afield. Finally, we introduce the Jacobian of a curve and describe the action of Hecke operators on the Jacobians of modular curves.

4.1. Hecke operators over $\mathbb{C}$.

We will define Hecke operators in this subsection. There are two kinds of Hecke operators, and each can be interpreted as operators on spaces of modular forms, homomorphisms of divisor groups, or maps between moduli spaces. See [4] for a detailed treatment. For an interpretation of Hecke operators as functions on the space of lattice in the level 1 case, see Chapter VII of [11].

Recall that there is a map $\psi_1 : S_1(N) \xrightarrow{\sim} Y_1(N) \subset X_1(N)$, where $S_1(N)$ is the moduli space of elliptic curves over $\mathbb{C}$ together with a torsion point of order N. Now define the two types of Hecke operators on $S_1(N)$ as follows:

(1) (Diamond operator) For $d \in \mathbb{Z}_+$ such that $(d, N) = 1$, let

$$\langle d \rangle \colon \mathrm{Div}(S_1(N)) \to \mathrm{Div}(S_1(N)) \colon \qquad [E, Q] \mapsto [E, d \cdot Q].$$

(2) For a prime $p \in \mathbb{Z}_+$, define

$$T_p \colon \mathrm{Div}(S_1(N)) \to \mathrm{Div}(S_1(N)) \colon \qquad [E, Q] \mapsto \sum_C [E/C, Q + C]$$

where the sum is taken over all order p subgroups $C \subset E$ such that C intersects $\langle Q \rangle$ (the cyclic subgroup generated by Q) trivially. Using the isomorphism $Y_1(N) \xrightarrow{\sim} S_1(N)$ to obtain an endomorphism of $\mathrm{Div}(Y_1(N))$, and extending over the finite number of cusps in a canonical way, one can obtain a corresponding endomorphism of $\mathrm{Div}(X_1(N))$:

$$\langle d \rangle \colon \mathrm{Div}(X_1(N)) \to \mathrm{Div}(X_1(N)) \colon \qquad x \mapsto \alpha(x),$$
$$T_p \colon \mathrm{Div}(X_1(N)) \to \mathrm{Div}(X_1(N)) \colon \qquad \Gamma_1(N)\tau \mapsto \sum_j \Gamma_1(N)\beta_j(\tau),$$

where α is any matrix $\begin{pmatrix} a & b \\ c & \delta \end{pmatrix} \in \Gamma_0$ with $\delta \equiv d \mod N$, and β_j are representatives for cosets of $\Gamma_1(N)$ in the double-coset $\Gamma_1(N) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_1(N)$.

Since modular forms behave as "differentials" on $X_1(N)$, one can pull back these maps $\mathrm{Div}(X_1(N)) \to \mathrm{Div}(X_1(N))$ and obtain operators on the space of modular forms (cusp forms, in particular). One obtains

$$\langle d \rangle \colon \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N)) \colon \qquad f \mapsto f[\alpha]_k$$

for any $\alpha = \begin{pmatrix} a & b \\ c & \delta \end{pmatrix} \in \Gamma_0(N)$ with $\delta \equiv d \mod N$. It can be shown that the space $\mathcal{M}_k(\Gamma_1(N))$ decomposes into eigenspaces

$$\mathcal{M}_k(\Gamma_1(N)) = \bigoplus_{\text{Dirichlet characters } \chi \bmod N} \mathcal{M}_k(N, \chi),$$

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

The second type of Hecke operators becomes the following map. Write the following double coset as a union of cosets:

$$\Gamma_1(N) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_1(N) = \bigcup_j \Gamma_1(N)\beta_j.$$

Then

$$T_p \colon \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N)) \colon \qquad f \mapsto \sum_j f[\beta_j]_k.$$

Explicitly, if f has Fourier expansion

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

then

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

This Hecke operator is also compatible with the decomposition of $\mathcal{M}_k(\Gamma_1(N))$ into $\mathcal{M}_k(N,\chi)$ above. So

$$a_n(T_p f) = a_{np}(f) + \chi(p)p^{k-1}a_{n/p}(f), \qquad \text{if } f \in \mathcal{M}_k(N,\chi).$$

For general n, first define

$$T_{p^r} = T_p T_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}} \qquad (r \geq 2).$$

Then take

$$T_n = \prod T_{p_i}^{r_i} \qquad \text{for } n = \prod p_i^{r_i}.$$

Finally, set $T_1 = 1$. All the above Hecke operators restrict to endomorphisms of $\mathcal{S}_k(\Gamma_1(N))$.

Remark 4.1. We can define T_n in the moduli space definition directly by having

$$T_n([E,Q]) = \sum_C [E/C, Q+C]$$

where the sum runs over all cyclic subgroups of order n that intersect $\langle Q\rangle$ trivially, and show the recurrence above for T_{p^r} and T_n.

Proposition 4.1. *All Hecke operators commute with each other.*

Proof. See section 5.2-5.3 of [4]. □

4.2. **Hecke operators over $\mathbb{Q}$.** In this subsection we will sketch a definition of Hecke operators as correspondences of varieties *over* $\mathbb{Q}$. We will provide intuitions to the fact that the "same" definition of the Hecke operators as maps between moduli spaces carries over to a moduli space over $\overline{\mathbb{Q}}$. First define

$$S_1(N)_{\mathbb{Q}} = \{\text{isomorphism classes of pairs } (E,C), \text{ where } E \text{ is an elliptic curve over } \overline{\mathbb{Q}} \text{ and } Q \text{ is a point of order } N\}.$$

Our previously defined $S_1(N)$ will be written as $S_1(N)_{\mathbb{C}}$. If E, E' are elliptic curves over $\overline{\mathbb{Q}}$, and if $[E,Q]$ and $[E',Q']$ are isomorphic over $\mathbb{C}$, then they are isomorphic over $\overline{\mathbb{Q}}$ (this can be seen using Weierstrass equations). Thus, $S_1(N)_{\mathbb{Q}} \subset \mathcal{S}_1(N)_{\mathbb{C}}$.

To connect the moduli space with modular curves, we will need a version of $\psi_1 : S_1(N)_{\mathbb{C}} \to X_1(N)_{\mathbb{C}}$ (cf. Theorem 3.9), but over $\overline{\mathbb{Q}}$. First, consider the commutative diagram

$$\begin{array}{ccc} S_1(N)_{\mathbb{C}} & \longrightarrow & S_1(1)_{\mathbb{C}} \\ {\scriptstyle\psi_1}\downarrow & & \downarrow \\ X_1(N)_{\mathbb{Q}}(\mathbb{C}) & \xrightarrow[\alpha]{} & X_1(1)_{\mathbb{C}} \end{array} \qquad \begin{array}{ccc} [E,Q] & \longrightarrow & [E] \\ \downarrow & & \downarrow \\ P & \longrightarrow & j(E) \end{array}$$

(The point P simply denotes the image of $[E,Q]$ under ψ_1.) Now $[E,Q] \in S_1(N)_{\mathbb{C}}$ is an element of $S_1(N)_{\mathbb{Q}}$ if and only if $j(E) \in \overline{\mathbb{Q}}$ (the forward direction is clear; for converse, if $j \neq 0, 1728$, E has a Weierstrass model $y^2 = 4x^3 - \frac{27j}{j-1728}\cdot x - \frac{27j}{j-1728}$; if $j = 0$, take $y^2 = x^3 + B$ for $B \in \mathbb{Q}^*$; and if $j = 1728$, take $y^2 = x^3 + Ax$ for $A \in \mathbb{Q}^*$). The morphism α is defined over $\mathbb{Q}$, so the preimage of the $\overline{\mathbb{Q}}$-points of $X_1(1)_{\mathbb{Q}}(\mathbb{C})$ under α also consists of $\overline{\mathbb{Q}}$-points; in other words,

$\psi_1(S_1(N)_{\mathbb{Q}}) \subset \alpha^{-1}(X_1(1)_{\mathbb{Q}}(\overline{\mathbb{Q}})) \subset X_1(N)_{\mathbb{Q}}(\overline{\mathbb{Q}})$. Thus, the restriction of ψ_1 to $S_1(N)_{\mathbb{Q}}$ defines

$$\psi_{1,\mathbb{Q}} : S_1(N)_{\mathbb{Q}} \to X_1(N)_{\mathbb{Q}}.$$

Now, on $S_1(N)_{\mathbb{Q}}$, the Hecke operators can be defined in the same way:

$$\langle d\rangle \colon \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \to \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \colon \qquad [E,Q] \mapsto [E, d\cdot Q] \qquad (\text{for } (d,N)=1),$$

$$T_p \colon \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \to \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \colon \qquad [E,Q] \mapsto \sum_C [E/C, Q+C].$$

Then the diagrams

$$\begin{array}{ccc} \mathrm{Div}(S_1(N)_{\mathbb{C}}) & \xrightarrow{T_p} & \mathrm{Div}(S_1(N)_{\mathbb{C}}) \\ {\scriptstyle\psi_1}\downarrow & & \downarrow{\scriptstyle\psi_1} \\ \mathrm{Div}(X_1(N)_{\mathbb{C}}) & \xrightarrow[T_p]{} & \mathrm{Div}(X_1(N)_{\mathbb{C}}) \end{array} \qquad \begin{array}{ccc} \mathrm{Div}(S_1(N)_{\mathbb{C}}) & \xrightarrow{\langle d\rangle} & \mathrm{Div}(S_1(N)_{\mathbb{C}}) \\ {\scriptstyle\psi_1}\downarrow & & \downarrow{\scriptstyle\psi_1} \\ \mathrm{Div}(X_1(N)\mathbb{C}) & \xrightarrow[\langle d\rangle]{} & \mathrm{Div}(X_1(N)_{\mathbb{C}}) \end{array}$$

restrict to the diagrams

$$\begin{array}{ccc} \mathrm{Div}(S_1(N)_{\mathbb{Q}}) & \xrightarrow{T_p} & \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \\ {\scriptstyle\psi_{1,\mathbb{Q}}}\downarrow & & \downarrow{\scriptstyle\psi_{1,\mathbb{Q}}} \\ \mathrm{Div}(X_1(N)_{\mathbb{Q}}) & \xrightarrow[T_p]{} & \mathrm{Div}(X_1(N)_{\mathbb{Q}}) \end{array} \qquad \begin{array}{ccc} \mathrm{Div}(S_1(N)_{\mathbb{Q}}) & \xrightarrow{\langle d\rangle} & \mathrm{Div}(S_1(N)_{\mathbb{Q}}) \\ {\scriptstyle\psi_{1,\mathbb{Q}}}\downarrow & & \downarrow{\scriptstyle\psi_{1,\mathbb{Q}}} \\ \mathrm{Div}(X_1(N)\mathbb{Q}) & \xrightarrow[\langle d\rangle]{} & \mathrm{Div}(X_1(N)_{\mathbb{Q}}) \end{array}$$

which means that the moduli space interpretation of Hecke operators carries over when one defines them over $\mathbb{Q}$.

Another way to define Hecke operators over $\mathbb{Q}$ is by passing to function fields.

Let $(d,N)=1$. The corresponding map of function fields is the pullback $\langle d\rangle^* : \mathbb{C}(X_1(N)) \to \mathbb{C}(X_1(N))$. Recall that $\mathbb{C}(X_1(N)) = \mathbb{C}(j, f_1)$ and $\mathbb{Q}(X_1(N)) = \mathbb{Q}(j, f_1)$. Thus, we shall show that $\langle d\rangle^*(\mathbb{Q}(j,f_1)) \subset \mathbb{Q}(j,f_1)$, which would correspond to a morphism $\langle d\rangle \colon X_1(N)_{\mathbb{Q}} \to X_1(N)_{\mathbb{Q}}$ by curve-field correspondence. Now

$$\langle d\rangle \Gamma_1(N)\tau = \Gamma_1(N)\gamma(\tau) \qquad \gamma = \begin{pmatrix} a & b \\ c & \delta \end{pmatrix} \quad \delta \equiv d \mod N.$$

Thus, $(\langle d\rangle^* j)(\tau) = j\circ\gamma(\tau)$. Since j is of weight 0, $j\circ\gamma(\tau)=\tau$, hence $\langle d\rangle^* j = j$. For f_1, compute that

$$f_1 \circ \gamma = f_0^{\pm\overline{(0,1)}} \circ \gamma = f_0^{\pm\overline{(0,1)\gamma}} = f_0^{\pm\overline{(0,d)}}.$$

Recall that $f_0^{\pm\overline{(0,d)}}(\tau) = \wp(d/N)g_2(\tau)/g_3(\tau)$, and $f_1 = \wp(1/N)g_2(\tau)/g_3(\tau)$. Fix $\tau \in \mathcal{H}$ such that $j(\tau) \notin \{0, 1728\}$. This means that $g_2(\tau)$ and $g_3(\tau)$ are nonzero since $j = 1728 g_2^3/(g_2^3 - 27 g_3^2)$. Consider the map

$$(\wp_\tau g_2(\tau)/g_3(\tau), \wp'_\tau(g_2(\tau)/g_3(\tau))^{3/2}\wp'_\tau) : \mathbb{C}/\Lambda_\tau \to \mathbb{C}^2 \cup \{\infty\}$$

This differs from $(\wp_\tau, \wp'_\tau)$ by an admissible change of variables $(x,y) = (u^2x', u^3y')$ where $u = (g_3(\tau)/g_2(\tau))^{1/2}$, and defines a bijective correspondence between the complex torus $\mathbb{C}/\Lambda_\tau$ and the elliptic curve

$$E : y^2 = 4x^3 - \left(\frac{27j(\tau)}{j(\tau)-1728}\right)x - \left(\frac{27j(\tau)}{j(\tau)-1728}\right)$$

Then f_1 (resp. $f_0^{\pm\overline{(0,d)}}$ is the function that takes τ to the x-coordinate of torsion point Q (resp. $d\cdot Q$) in $E_{j(\tau)}$ that corresponds to the point $1/N$ in the complex

torus $\mathbb{C}/\Lambda_\tau$. By Exercise 3.7(a) and (d) in [3], the point $x(d \cdot Q)$ can be expressed as $g(27j/(j-1728), x(Q))$, where g is a rational function in two variables over $\mathbb{Q}$. Thus, for some $g \in \mathbb{Q}(j, f_1)$, $\langle d \rangle^* f_1 = f_0^{\pm(0,d)}$ and $g(j, f_1)$ agree at all but finitely many $\tau \in \mathbb{C}(X_1(N))$ (excluding the cusps and the points with j-invariant 0 or 1728), which means they must agree completely. Thus, $\langle d \rangle^* f_1 \in \mathbb{Q}(j, f_1)$.

The proof that T_p is defined over $\mathbb{Q}$ follows roughly the same structure, but it is a bit more technical, and we omit it here. See Section 7.9 in [4].

4.3. **Hecke eigenforms.** We have defined the Hecke operators as correspondences over $\mathbb{C}$ and over $\mathbb{Q}$. In this subsection we will consider them as endomorphisms of $\mathcal{S}_k(\Gamma_1(N))$, which leads to many beautiful results and reveals a lot about modular forms and the Hecke operators. We only state the main results. The details can be found in Chapter 5 of [4].

We begin by giving $\mathcal{S}_k(\Gamma)$ the structure of an inner product space. The **Petersson inner product** on $\mathcal{S}_k(\Gamma)$ is defined as:

$$\langle , \rangle_\Gamma : \mathcal{S}_k(\Gamma) \times \mathcal{S}_k(\Gamma) \to \mathbb{C} \qquad \langle f, g \rangle_\Gamma = \frac{1}{V_\Gamma} \int_{X(\Gamma)} f(\tau)\overline{g(\tau)}(\mathrm{Im}(\tau))^k d\mu(\tau).$$

where $d\mu(\tau) = \frac{dxdy}{y^2}$ for $\tau = x + iy \in \mathcal{H}$ is the hyperbolic measure. (Intuitively, this is the familiar inner product on function spaces, but adding the factor of $(\mathrm{Im}(\tau))^k$ to account for the factors of automorphy in the definition of modular forms.) Then

Theorem 4.2. *The Hecke operators $\langle n \rangle$ and T_n for $(n, N) = 1$ are normal (i.e. commutes with their Hermitian adjoints). This is a commuting family of normal operators on a finite-dimensional inner product space. Thus, by linear algebra, $\mathcal{S}_k(\Gamma_1(N))$ has an orthogonal basis of simultaneous eigenforms for the Hecke operators $\{\langle n \rangle, T_n : (n, N) = 1\}$.*

Definition 4.3. Let

$$\mathbb{T} = \mathbb{Z}[\{T_n, \langle n \rangle : n \in \mathbb{N}\}] \qquad \mathbb{T}_0 = \mathbb{Z}[\{T_n, \langle n \rangle : (n, N) = 1\}] \subset \mathbb{T}$$

i.e., the *algebra* generated by the Hecke operators. The algebra $\mathbb{T}$ is called the **Hecke algebra**, and $\mathbb{T}_0$ is called the **anemic Hecke algebra**.

Now we introduce the notion of newforms. Some modular forms in $\mathcal{S}_k(\Gamma_1(N))$ more naturally belongs to lower levels. These are the oldforms, defined as follows.

Definition 4.4. Let $d \mid N$. Let

$$\alpha_d = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} \qquad f[\alpha_d]_k(\tau) = d^{k-1} f(d \cdot \tau).$$

Let

$$i_d \colon \mathcal{S}_k(\Gamma_1(N/d))^2 \to \mathcal{S}_k(\Gamma_1(N)) \colon \qquad (f, g) \mapsto f + g[\alpha_d]_k.$$

The subspace of **oldforms at level** N consists of

$$\mathcal{S}_k(\Gamma_1(N))^{\mathrm{old}} = \sum_{p \mid N} i_p(\mathcal{S}_k(\Gamma_1(N/p))^2),$$

where the p's are primes. The subspace $\mathcal{S}_k(\Gamma_1(N))^{\mathrm{new}}$ is defined to be the orthogonal complement of the subspace of oldforms at level N with respect to the Petersson inner product. The oldforms and newforms are stable under the Hecke operators T_n and $\langle n \rangle$. Thus, they also have orthogonal bases of eigenforms for $\mathbb{T}_0$.

A modular form $f \in \mathcal{M}_k(\Gamma_1(N))$ that is an eigenform for $\mathbb{T}$ (not just $\mathbb{T}_0$) is called an **Hecke eigenform**, or simply an **eigenform**. An eigenform is **normalized** if $a_1(f) = 1$. A **newform** is a normalized eigenform in $\mathcal{S}_k(\Gamma_1(N))^{\text{new}}$. Amazingly, the eigenvalues for the newforms are precisely their Fourier coefficients:

Theorem 4.5. *Let $f \in \mathcal{S}_k(\Gamma_1(N))^{new}$ be a nonzero eigenform for $\mathbb{T}_0$. Then*

1. *f is a Hecke eigenform, i.e., f is an eigenform for $\mathbb{T}$. A suitable scalar multiple of f is a newform.*
2. *If f is a newform, then $T_n f = a_n(f) f$ for all $n \in \mathbb{Z}^+$.*
3. ***(Multiplicity One)** If $\widetilde{f}$ is also an eigenform with the same T_n-eigenvalues, then $\widetilde{f} = cf$ for some constant c.*

Moreover, the set of newforms in $\mathcal{S}_k(\Gamma_1(N))^{new}$ is an orthogonal basis of the space.

This theory will allow one to give an explicit basis for $\mathcal{S}_k(\Gamma_1(N))$.

Theorem 4.6. *The set*

$$\mathcal{B}_k = \{f(n\tau) : M \mid N, f \text{ is a newform of level } M, n \mid N/M\}$$

is a basis for $\mathcal{S}_k(\Gamma_1(N))$.

The following proposition says that every normalized eigenform is almost a newform in some potentially lower levels.

Proposition 4.7. *Let $g \in \mathcal{S}_k(\Gamma_1(N))$ be a normalized eigenform. Then there is 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 $p \nmid N/M$. If $g \in \mathcal{S}_k(N, \chi)$, then $f \in \mathcal{S}_k(M, \chi_M)$ where χ_M lifts to χ mod N.*

Proposition 4.8. *Let $f \in \mathcal{M}_k(N, \chi)$. Then f is a normalized eigenform if and only if the following conditions are satisfied:*

1. *$a_1(f) = 1$.*
2. *$a_{p^r} = a_p(f) a_{p^{r-1}}(f) - \chi(p) p^{k-1} a_{p^{r-2}}(f)$ for all prime p and $r \geq 2$.*
3. *$a_{mn}(f) = a_m(f) a_n(f)$ when $(m, n) = 1$.*

Note that the forward direction of this Proposition easily follows from definitions.

Example 4.9. In [12], Ramanujan studied the following function

$$(2\pi)^{-12}\Delta = q \prod_{n=1} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n$$

and conjectured that $\tau(nm) = \tau(n)\tau(m)$ if $(n, m) = 1$ and $\tau(p)\tau(p^n) = \tau(p^{n+1}) + p^{11}\tau(p^{n-1})$ if p is a prime and $n \geq 1$. This conjecture can be proved using the theory sketched above. It is known (cf. subsection 3.3) that Δ is a cusp form of weight 12 and that the first Fourier coefficient $\Delta(1) = (2\pi)^{12}$. Moreover, one has $\dim \mathcal{S}_{12}(\Gamma_1(1)) = 1$ (cf. [11] or Chapter 3 of [4] for a more complete dimension computation), so the space spanned by Δ must be stable under the Hecke operators. Hence, $(2\pi)^{-12}\Delta$ is a normalized eigenform, and Ramanujan's conjecture follows from Proposition 4.8.

4.4. **Jacobian and Picard groups.** In algebraic geometry, there is an important association of an algebraic curve over a field k with their Jacobians. In this subsection, we will consider the action of Hecke operators on the Jacobian and the Picard groups.

Definition 4.10. Let X be a curve of genus g over a field k. The **Jacobian** of X is a certain abelian variety of dimension g over k whose underlying group is functorially isomorphic to $\mathrm{Pic}^0(X)$.

Definition 4.11. Let $J_1(N) = \mathrm{Jac}(X_1(N)_{\mathbb{Q}})$, and $J_0(N) = \mathrm{Jac}(X_0(N)_{\mathbb{Q}})$.

The Jacobians of curves over $\mathbb{C}$ have a more explicit model. By the theory of Riemann surfaces, there is an embedding

$$\iota\colon H_1(X,\mathbb{Z}) \to \Omega^1(X)^\wedge\colon \qquad \gamma \mapsto \int_\gamma \cdot$$

where X is a Riemann surface (equivalently, a complex non-singular projective curve) and $\Omega^1(X)$ is the space of holomorphic 1-forms on X. If X has genus g, then $H_1(X,\mathbb{Z}) \cong \mathbb{Z}^{2g}$. Then we realize $\mathrm{Jac}(X)$ as a quotient

$$\mathrm{Jac}(X) = \Omega^1(X)^\wedge / \iota(H_1(X,\mathbb{Z})).$$

This is a complex torus of complex dimension g. Now consider the modular curve $X_1(N)_{\mathbb{Q}}$. We have

$$J_1(N)(\mathbb{C}) = \mathrm{Jac}(X_1(N)_{\mathbb{Q}}(\mathbb{C})) = \mathrm{Jac}(X_1(N)_{\mathbb{C}}) = \Omega^1(X_1(N)_{\mathbb{C}})^\wedge / \iota(H_1(X_1(N)_{\mathbb{C}},\mathbb{Z})).$$

The most important case is when $k = 2$. Then by Proposition 3.14, $H_0(X_1(N),\Omega^1) \cong \mathcal{S}_2(\Gamma_1(N))$. Thus,

$$J_1(N)(\mathbb{C}) \cong \mathcal{S}_2(\Gamma_1(N))^\wedge / \iota(H_1(X_1(N)_{\mathbb{Q}},\mathbb{Z})).$$

For simplicity of notation, we may as well write $J_1(N)(\mathbb{C}) \cong \mathcal{S}_2(\Gamma_1(N))^\wedge / H_1(X_1(N),\mathbb{Z})$.

Proposition 4.12. *The dual of the Hecke operator*

$$T\colon \mathcal{S}_2(\Gamma_1(N))^\wedge \to \mathcal{S}_2(\Gamma_1(N))^\wedge\colon \qquad \varphi \mapsto \varphi \circ T$$

for $T = T_p$ or $T = \langle d\rangle$ descends to a map $T : J_1(N)(\mathbb{C}) \to J_1(N)(\mathbb{C})$.

To prove this, recall that a map $h : X \to Y$ of curves induces a forward map h_* and a backward map h^* of Picard group (cf. subsection 2.1). This functorially transfers to a forward map and a backward map of the Jacobians.

Proof sketch. (Details can be found in Section 6.3 of [4]) Denote $\Gamma_1 = \Gamma_1(N)$ and $\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$. We first decompose T_p into a composition of forward and backward maps on Picard groups. Consider the configuration

$$\Gamma_1 \leftarrow \Gamma_3 \xrightarrow{\sim} \Gamma_3 \to \Gamma_1,$$

where $\Gamma_3 = \alpha^{-1}\Gamma_1\alpha \cap \Gamma_1$ and $\Gamma_3' = \alpha\Gamma_3\alpha^{-1} = \alpha\Gamma_1\alpha^{-1} \cap \Gamma_1$. This induces a configuration on modular curves

$$X_1 \xleftarrow{\pi_1} X_3 \xrightarrow{\sim} X_3' \xrightarrow{\pi_2} X_1,$$

where the isomorphism is $\Gamma_3\tau \to \Gamma_3'\alpha(\tau)$. The configuration produces a map of divisor groups given by

$$(\Gamma_1\tau) \xrightarrow{\pi_1^{-1}} \sum_j (\Gamma_3\gamma_j(\tau)) \xrightarrow{\alpha} \sum_j (\Gamma_3'\beta_j(\tau)) \xrightarrow{\pi_2} \sum_j \Gamma_1\beta_j(\tau),$$

where $\Gamma_3\backslash\Gamma_1 = \bigcup_j \Gamma_3\gamma_j$ and $\beta_j = \alpha\gamma_j$. One can check that this composition is exactly T_p on $\mathrm{Div}(X_1(N))$. One also checks from the definition of forward and backward maps of Picard groups (cf. subsection 2.1) that this composition is exactly

$\pi_{1,*} \circ \alpha_* \circ \pi_2^* = (\pi_1 \circ \alpha)_* \circ \pi_2^*$. The forward and backward maps are defined on Picard groups, hence also on the Jacobians.

For the Diamond operator $\langle d \rangle$, take $\alpha = \begin{pmatrix} a & b \\ c & \delta \end{pmatrix}$ as before, and observe that $\langle d \rangle$ is exactly the forward map α_*. □

Remark 4.2. It also follows from the fact that T is a morphism over $\mathbb{Q}$ that the map $T : J_1(N)(\mathbb{C}) \to J_1(N)(\mathbb{C})$ restricts to a map $T : J_1(N) \to J_1(N)$. We will skip this detials. See Corollary 5.22 of [13].

Theorem 4.13. *Let $f \in \mathcal{S}_2(\Gamma_1(N))$ be a normalized eigenform. Then the field $K_f = \mathbb{Q}(\{a_n : n \in \mathbb{Z}^+\})$ is a finite extension over $\mathbb{Q}$. This is called the* ***number field of*** *f.*

Proof. Let $T = T_p$ or $T = \langle d \rangle$ for $(d, N) = 1$. Dualizing, we have

$$T \colon \mathcal{S}_2(\Gamma_1(N))^\wedge \to \mathcal{S}_2(\Gamma_1(N))^\wedge \colon \qquad \varphi \mapsto \varphi \circ T.$$

This descends to the Jacobian $J_1(N)(\mathbb{C})$. Thus, the operators act as endomorphisms on the kernel $H_1(X_1(N)_{\mathbb{C}}, \mathbb{Z}) \cong \mathbb{Z}^{2g}$, so $\mathbb{T}_{\mathbb{Z}} \subset \mathrm{End}(H_1(X_1(N)_{\mathbb{C}}, \mathbb{Z}))$, which implies that $\mathbb{T}_{\mathbb{Z}}$ is a finitely generated $\mathbb{Z}$-module. Let

$$\lambda_f : \mathbb{T}_{\mathbb{Z}} \to \mathbb{C} \qquad Tf = \lambda_f(T) f.$$

The image of this map is

$$\mathcal{O}_f := \mathbb{Z}[\{a_n(f) : n \in \mathbb{Z}^+\}],$$

which must also be finitely generated over $\mathbb{Z}$. Thus, $K_f = \mathbb{Q}(\{a_n(f)\})$ is finitely generated as a vector space over $\mathbb{Q}$, hence a finite extension. □

Theorem 4.14. *Let $f \in \mathcal{S}_2(N, \chi)$ be a normalized eigenform. Let K_f be its number field. For any embedding $\sigma : K_f \to \mathbb{C}$, the conjugated f^σ (defined by conjugating the coefficients in the Fourier expansion) is also a normalized eigenform in $\mathcal{S}_2(N, \chi^\sigma)$, where $\chi^\sigma(n) = \chi(n)^\sigma$.*

Proof. See Section 6.5 of [4]. □

5. Modular Galois Representation

In this section, we will construct the Galois representation associated to a modular form. We will then state the Modularity Theorem using this construction. The construction of Galois representations for elliptic curves is relatively simple. To associate a Galois representation to normalized eigenforms, however, is much more difficult. For $k \neq 2$, the construction involves étale cohomology. If $k = 2$, there is luckily a geometric object associated to normalized eigenform, namely the abelian variety A_f. We will describe this construction for a newform $f \in \mathcal{S}_2(\Gamma_1(N))$. Throughout the rest of the subsection, we will fix a positive integer N, and every modular form considered will be of weight 2.

5.1. Abelian varieties.

The main tool in the construction of A_f is the following theorem:

Theorem 5.1. *Let k be a field and A be an abelian variety over k. Suppose that B is an abelian subvariety of A. Then there exists an abelian variety C over k and a surjective morphism $A \to C$ with kernel exactly B.*

The abelian variety C may be considered to be the quotient A/B of A by B.

Proof. See Section 9.5 of [14] or the discussion in [15]. □

Let $f \in \mathcal{S}_2(N,\chi)$ be a newform, and let $\lambda_f\colon \mathbb{T} \to \mathbb{C}$ be defined as in the proof of Theorem 4.13. Because the Hecke operators are morphisms of abelian varieties over $\mathbb{Q}$ on $J_1(N)$, $I_f J_1(N)$ is an abelian subvariety of $J_1(N)$, where $I_f = \ker(\lambda_f)$. By Theorem 5.1, we may define an abelian variety A_f over $\mathbb{Q}$ as the quotient $J_1(N)/I_f J_1(N)$. Over $\mathbb{C}$, the same theorem shows the existence of a surjective morphism $J_1(N)(\mathbb{C}) \to A_f(\mathbb{C})$ with kernel exactly $I_f J_1(N)(\mathbb{C})$. Moreover, $A_f(\mathbb{C}) = A_{f,\mathbb{C}}$. (See Chapter 14 of [16] for an alternative definition of A_f.)

Definition 5.2. Let $f \in \mathcal{S}_2(\Gamma_1(N))$ be a newform. Let $V_f = \mathrm{Span}(f^\sigma : \sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})) \subset \mathcal{S}_2(\Gamma_1(N))$. Restricting the subgroup $H_1(X_1(N),\mathbb{Z})$ of $\mathcal{S}_2(\Gamma_1(N))^\wedge$ to functions on V_f gives a subgroup $\Lambda_f = H_1(X_1(N),\mathbb{Z})|_{V_f} \subset V_f^\wedge$.

Proposition 5.3. *Let $f \in \mathcal{S}_2(\Gamma_1(N))$ be a newform with number field K_f. Then there is a group isomorphism*

$$A_{f,\mathbb{C}} \xrightarrow{\sim} V_f^\wedge/\Lambda_f \qquad \varphi + I_f J_1(N) \to \varphi|_{V_f} + \Lambda_f \text{ for } \varphi \in \mathcal{S}_2(\Gamma_1(N))^\wedge.$$

The right side is a complex torus of dimension $[K_f : \mathbb{Q}]$. It follows that $\dim A_f = \dim_{\mathbb{C}} A_{f,\mathbb{C}} = [K_f : \mathbb{Q}]$.

Proof. Let $\mathcal{S}_2 = \mathcal{S}_2(\Gamma_1(N))$, $H_1 = H_1(X_1(N),\mathbb{Z})$, $A_f = A_{f,\mathbb{C}}$. Then

$$A_f = J_1(N)/I_f J_1(N) = (\mathcal{S}_2^\wedge/H_1)/I_f(\mathcal{S}_2^\wedge/H_1) \cong (\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge)/\overline{H_1},$$

where $\overline{H_1}$ is the image of H_1 in $\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge$. By some linear algebra, $\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge \cong \mathcal{S}_2[I_f]^\wedge$, where $\mathcal{S}_2[I_f]$ consists of the elements of $\mathcal{S}_2$ annihilated by I_f and the isomorphism is given by restriction $\varphi + I_f\mathcal{S}_2^\wedge \to \varphi|_{\mathcal{S}_2[I_f]}$. Thus, $A_f \xrightarrow{\sim} \mathcal{S}_2[I_f]^\wedge/H_1|_{\mathcal{S}_2[I_f]}$.

We need to show that $\mathcal{S}_2[I_f] = V_f$ and $\Lambda_f = H_1|_{V_f}$ is a lattice. Clearly $V_f \subset \mathcal{S}_2[I_f]$. The converse is shown by a dimension argument. First, observe that we have a perfect pairing, invariant under the action of the Hecke operators,

$$\mathbb{T}_{\mathbb{C}} \times \mathcal{S}_2 \to \mathbb{C}\colon \qquad (T,f) \mapsto a_1(Tf),$$

where $\mathbb{T}_{\mathbb{C}} = \mathbb{C}[\{T_n, \langle n\rangle : n \in \mathbb{Z}_+\}]$, so $\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge \cong \mathbb{T}_{\mathbb{C}}/I_f\mathbb{T}_{\mathbb{C}}$. Hence

$$\dim(\mathcal{S}_2[I_f]) = \dim(\mathcal{S}_2[I_f]^\wedge) = \dim(\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge) = \dim(\mathbb{T}_{\mathbb{C}}/I_f\mathbb{T}_{\mathbb{C}}).$$

We also know that the natural surjection $\mathbb{T}_{\mathbb{Z}} \otimes \mathbb{C} \to \mathbb{T}_{\mathbb{C}}$ descends to a surjection $(\mathbb{T}_{\mathbb{Z}} \otimes \mathbb{C})/(I_f \otimes \mathbb{C}) \to \mathbb{T}_{\mathbb{C}}/I_f\mathbb{T}_{\mathbb{C}}$. So

$$\begin{aligned}\dim \mathcal{S}_2[I_f] &\le \dim((\mathbb{T}_{\mathbb{Z}} \otimes \mathbb{C})/(I_f \otimes \mathbb{C})) = \dim((\mathbb{T}_{\mathbb{Z}}/I_f) \otimes \mathbb{C}) \\ &= \mathrm{rank}(\mathbb{T}_{\mathbb{Z}}/I_f) = [K_f : \mathbb{Q}] = \dim V_f.\end{aligned}$$

Since $V_f \subset \mathcal{S}_2[I_f]$, the dimension argument shows that $\mathcal{S}_2[I_f] = V_f$. Lastly, to show that Λ_f is a lattice in $V_f^\wedge$, one needs to show that $\mathrm{Span}_{\mathbb{R}}(\Lambda_f) = V_f^\wedge$ and $\mathrm{rank}(\Lambda_f) \le \dim_{\mathbb{R}}(V_f^\wedge)$. For the first, note that the inclusion $V_f \subset \mathcal{S}_2$ gives a surjective restriction map $\pi : \mathcal{S}_2^\wedge \to V_f^\wedge$. Since $\mathrm{Span}_{\mathbb{R}}(H_1) = \mathcal{S}_2^\wedge$, $\mathrm{Span}_{\mathbb{R}}(\Lambda_f) = \mathrm{Span}_{\mathbb{R}}(\pi(H_1)) = \pi(\mathcal{S}_2^\wedge) = V_f^\wedge$. For the second, take dimensions over $\mathbb{R}$:

$$\begin{aligned}\dim_{\mathbb{R}} V_f^\wedge &= \dim_{\mathbb{R}}(\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge) = \dim_{\mathbb{R}}((H_1 \otimes \mathbb{R})/I_f(H_1 \otimes \mathbb{R})) \\ &= \dim_{\mathbb{R}}((H_1/I_fH_1) \otimes \mathbb{R}) = \mathrm{rank}(H_1/I_fH_1).\end{aligned}$$

[4]Since by Theorem 4.6, the f^σ's for $\sigma\colon K_f \to \mathbb{C}$ are linearly independent.

But $\Lambda_f = \pi(H_1) \cong H_1/(H_1 \cap \ker \pi)$. Since $V_f = \mathcal{S}_2[I_f]$, $I_f H_1 \subset H_1 \cap \ker \pi$. Thus, there is a surjection $H_1/I_f H_1 \to \Lambda_f$, so $\operatorname{rank}(\Lambda_f) \leq \operatorname{rank}(H_1/I_f H_1) = \dim_{\mathbb{R}}(V_f^{\wedge})$. This completes the proof. □

5.2. **Modular Galois Representations.** We will use the following general fact about abelian varieties.

Theorem 5.4. *Let $n \in \mathbb{Z}_+$. Let k be a field of characteristic 0 or p with $(p, n) = 1$. If A is an abelian variety of dimension d, then $A[n] \cong (\mathbb{Z}/n\mathbb{Z})^{2d}$.*

Proof. See section I.7 of [17]. □

Let f be a newform of level N. Write $A = A_f$, $K = K_f$, and $d = [K : \mathbb{Q}]$ for simplicity. Let l be a prime. Define

$$\operatorname{Ta}_l(A) := \varprojlim_n A[l^n], \qquad V_l(A) := \mathbb{Q}_l \otimes_{\mathbb{Z}_l} \operatorname{Ta}_l(A).$$

As a $\mathbb{Q}_l$-vector space, $V_l(A) \cong \mathbb{Q}_l^{2d}$ by Theorem 5.4. Now note that $\mathcal{O}_f := \mathbb{T}_{\mathbb{Z}}/I_f$ is isomorphic as a $\mathbb{Z}$-module to $\mathbb{Z}[\{a_n(f) : n \in \mathbb{Z}_+\}]$, which has rank $[K_f : \mathbb{Q}]$, and it acts on the group $A_f = J_1(N)/I_f J_1(N)$. Hence, there is a natural action of $\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q}_l = \mathcal{O}_f \otimes_{\mathbb{Z}} (\mathbb{Q} \otimes_{\mathbb{Q}} \mathbb{Q}_l) = (\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_l = K \otimes \mathbb{Q}_l$ on $V_l(A)$. By algebraic number theory,

$$K \otimes_{\mathbb{Q}} \mathbb{Q}_l = \prod_{\lambda | l} K_\lambda,$$

where λ are primes lying over l and K_λ are the λ-adic completion of K. Thus, there is a decomposition

$$V_l(A) = \prod_{\lambda | l} V_\lambda(f).$$

Lemma 5.5. *For all λ lying over l, $\dim_{K_\lambda} V_\lambda(f) = 2$.*

Proof. Write $A(\mathbb{C}) = V/\mathcal{L}$, where $V = V_f$ and $\mathcal{L}$ is a lattice. We have $A[l^n] = l^{-n}\mathcal{L}/\mathcal{L} \cong \mathcal{L}/l^n\mathcal{L}$. Thus, $V_l(A) \cong \mathcal{L} \otimes \mathbb{Q}_l$ as a $K \otimes \mathbb{Q}_l$-module (using the Hecke action on $V/\mathcal{L}$). Also, $\mathcal{L}$ is a free $\mathbb{Z}$-module of rank $2d$, and since $\mathcal{L}$ is an $\mathcal{O}_f$-module, $\mathcal{L} \otimes \mathbb{Q}$ is a vector space over K. Then $\mathcal{L} \otimes \mathbb{Q} \cong K^2$ as a K-vector space by dimension counting over $\mathbb{Q}$. This shows that

$$V_l(A) \cong (\mathcal{L} \otimes \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong (K^2) \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong (K \otimes_{\mathbb{Q}} \mathbb{Q}_l)^2$$

It follows that $V_\lambda(A) \cong K_\lambda^2$ when we decompose it using $K \otimes_{\mathbb{Q}} \mathbb{Q}_l = \prod K_\lambda$. □

Since the Hecke operators are defined over $\mathbb{Q}$, as morphisms of varieties, they can be locally expressed as rational functions over $\mathbb{Q}$, hence the Galois action commutes with the Hecke operators, so $G_{\mathbb{Q}}$ acts compatibly on $V_l(A)$ with the action of $K \otimes_{\mathbb{Q}} \mathbb{Q}_l$. Thus, we can define

$$\rho_l = \rho_{f,l} : G_{\mathbb{Q}} \to \operatorname{Aut}_{K \otimes \mathbb{Q}_l} V_l(A) \cong \operatorname{GL}_2(K \otimes \mathbb{Q}_l) \cong \prod_{\lambda | l} \operatorname{GL}_2(K_\lambda),$$

so ρ_l decomposes into

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to \operatorname{GL}_2(K_\lambda).$$

This is the desired Galois representation associated to f.

5.3. **Modularity Theorem.** We are finally in a position to state the Modularity Theorem.

Definition 5.6. An irreducible Galois representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_l)$ such that $\det \rho = \chi_l$ is **modular** if there exists a newform $f \in \mathcal{S}_2(\Gamma_0(M_f))$ such that $K_{f,\lambda} = \mathbb{Q}_l$ for some maximal ideal λ of $\mathcal{O}_{K_f}$ lying over l and such that $\rho_{f,\lambda} \sim \rho$.

Theorem 5.7. ***(Modularity Theorem, Version** R)* *Let E be an elliptic curve ovre $\mathbb{Q}$. Then $\rho_{E,l}$ is modular for some l.*

This is the version of Modularity Theorem proved in Wiles' paper [1].

Theorem 5.8. ***(Modularity Theorem, Strong Version** R)* *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N (cf. Definition 2.9). Then for some newform $f \in \mathcal{S}_2(\Gamma_0(N))$ with number field $K_f = \mathbb{Q}$, $\rho_{f,l} \sim \rho_{E,l}$ for all l.*

These two versions are in fact equivalent, as we will see in the next section.

6. Eichler–Shimura Relation

In this section, we will exhibit the relationship between the Modularity Theorem stated in the previous section and explicit problem of counting points on elliptic curves mod p. We will almost never work on the complex analytic modular curve, hence the notations $X_1(N)$, $S_1(N)$, and so on, will replace the previous notations $X_1(N)_{\mathbb{Q}}$, $S_1(N)_{\mathbb{Q}}$, and so on.

6.1. **Reduction of elliptic curves over $\overline{\mathbb{Q}}$.** We've defined reduction of elliptic curves over $\mathbb{Q}$ in section 2. Here, we will define the reduction of elliptic curves over $\overline{\mathbb{Q}}$. This is important because the moduli space interpretation of $S_1(N) \subset X_1(N)$ is in terms of isomorphisms classes of elliptic curves (with torsion data) over $\overline{\mathbb{Q}}$.

Let E be an elliptic curve over $\overline{\mathbb{Q}}$. Let $\mathfrak{p}$ be a prime in $\overline{\mathbb{Z}}$. Take a Weierstrass equation for E. Using an admissible change of variables, we may assume that the Weierstrass equation has coefficients in $\overline{\mathbb{Z}}_{(\mathfrak{p})}$. We say that such a Weierstrass equation is $\mathfrak{p}$-integral. Any $\mathfrak{p}$-integral Weierstrass equation reduces coefficient-wise to a Weierstrass equation $\widetilde{E}$ over $\overline{\mathbb{F}}_p$.

Definition 6.1. The reduction types of E at $\mathfrak{p}$ (good, bad, ordinary, etc.) are defined for $\mathfrak{p}$-integral Weierstrass equations as in Definition 2.8. A $\mathfrak{p}$-integral Weierstrass equation with good or multiplicative reduction is called **$\mathfrak{p}$-minimal**.

Proposition 6.2. *Any elliptic curve over $\overline{\mathbb{Q}}$ has a $\mathfrak{p}$-minimal Weierstrass equation. Ordinary, supersingular, and multiplicative reduction are well defined on equivalence classes of $\mathfrak{p}$-minimal Weierstrass equations. If E and E' are equivalent $\mathfrak{p}$-minimal Weierstrass equations with good reduction at $\mathfrak{p}$, then their reductions define isomorphic elliptic curves over $\overline{\mathbb{F}}_p$.*

Proof. See Section 8.4 of [4]. □

Proposition 6.3. *Let E be an elliptic curve over $\overline{\mathbb{Q}}$. Then*

1. *If E has good reduction at $\mathfrak{p}$, then $E[N] \to \widetilde{E}[N]$ is surjective for all N.*
2. *For any isogenous elliptic curve E', E has good reduction at $\mathfrak{p}$ iff E' does.*

Proof. (1) can be shown using VII.2.1 and VII.3.1 while (2) is VII.7.2 of [3]. □

6.2. **Reduction of modular curves.** Modular curves are algebraic curves. We first define generally the reduction of algebraic curves mod p.

Definition 6.4.

(1) Let $C = \mathrm{Spec}(\mathbb{Q}[x_1, \dots, x_n]/(\varphi_1, \dots, \varphi_m))$ be a non-singular affine algebraic curve over $\mathbb{Q}$. Then C has **good reduction at** p if
   (a) $\mathrm{Spec}(\mathbb{Z}_{(p)}[x_1, \dots, x_n]/(\varphi_1, \dots, \varphi_m))$ is a variety (i.e., $(\varphi_1, \dots, \varphi_m)$ is prime in $\mathbb{Z}_{(p)}[x_1, \dots, x_n]$).
   (b) Let $\widetilde{\varphi}_1, \dots \widetilde{\varphi}_m \in \mathbb{F}_p[x_1, \dots, x_n]$ be the polynomials obtained by reducing the coefficients mod p. Then $\widetilde{C} = \mathrm{Spec}(\mathbb{F}_p[x_1, \dots, x_n]/(\widetilde{\varphi}_1, \dots, \widetilde{\varphi}_m))$ defines a non-singular curve over $\mathbb{F}_p$.

   In this case $\widetilde{C}$ is the **reduction of** C **at** p.

(2) Let C be a non-singular projective curve over $\mathbb{Q}$. Then C has **good reduction at** p if every affine piece C_i either has good reduction at p or has empty reduction at p. The curve $\widetilde{C}$ is defined to be the projectivization of the reduction of any affine piece C_i that has good reduction at p.

Theorem 6.5. *Let E and E' be elliptic curves over $\overline{\mathbb{Q}}$ with good reduction at $\mathfrak{p}$, and let $\widetilde{E}$ and $\widetilde{E'}$ denote their respective reductions. Let $\varphi : E \to E'$ be an isogeny over $\overline{\mathbb{Q}}$ of elliptic curves over $\overline{\mathbb{Q}}$. Then there is an isogeny*

$$\widetilde{\varphi} : \widetilde{E} \to \widetilde{E'}$$

such that

(1) *If $\psi : E' \to E''$ is also an isogeny, then $\widetilde{\psi \circ \varphi} = \widetilde{\psi} \circ \widetilde{\varphi}$.*
(2) *The following diagram commutes*

$$\begin{array}{ccc} E & \xrightarrow{\varphi} & E' \\ \downarrow & & \downarrow \\ \widetilde{E} & \xrightarrow[\widetilde{\varphi}]{} & \widetilde{E'} \end{array}$$

(3) $\deg(\widetilde{\varphi}) = \deg(\varphi)$.

Proof. See Section 8.5 of [4]. □

Proposition 6.6. *Let E be an elliptic curve over $\overline{\mathbb{Q}}$. If E has ordinary (resp. supersingular) reduction at $\mathfrak{p}$, then so does E/C.*

Proof. Let $\varphi : E \to E'$ be an isogeny. By Theorem 6.5, φ reduces to an isogeny $\widetilde{\varphi} : \widetilde{E} \to \widetilde{E'}$. Heading towards a contradiction, assume without loss of generality that E has ordinary reduction and E' has supersingular reduction at $\mathfrak{p}$ (for the other way around, simply use the dual isogeny of φ). By definition, $\widetilde{E}[p^e] \cong (\mathbb{Z}/p\mathbb{Z})^e$, and $\widetilde{E'}[p^e] = 0$. Then $\widetilde{\varphi}(\widetilde{E}[p^e]) = 0$ for all $e \in \mathbb{Z}_+$, which contradicts that $\ker \widetilde{\varphi}$ is a finite set. □

Next, consider the special case of $X_1(N)$. Define

$$S_1(N)'_{\mathrm{gd}} = \{[E, Q] \in S_1(N) : E \text{ has good reduction at } \mathfrak{p}, \widetilde{j(E)} \notin \{0, 1728\}\}.$$

$$\widetilde{S}_1(N) = \{[E, Q] : E \text{ is an elliptic curve over } \overline{\mathbb{F}}_p, Q \in E \text{ is a point of order } N\}$$

$$\widetilde{S}_1(N)' = \{[E, Q] \in \widetilde{S}_1(N) : j(E) \notin \{0, 1728\}\}$$

Theorem 6.7. ***(Igusa)*** *Let N be a positive integer and let p be a prime with $p \nmid N$. The modular curve $X_1(N)$ has good reduction at p. Moreover, reducing the modular curve is compatible with reducing the moduli space in the sense that the following diagram commutes*

$$\begin{array}{ccc} S_1(N)'_{gd} & \xrightarrow{\psi_1} & X_1(N) \\ \downarrow & & \downarrow \\ \widetilde{S}_1(N)' & \xrightarrow[\widetilde{\psi}_1]{} & \widetilde{X}_1(N) \end{array}$$

Proof. See [18]. □

6.3. **The Eichler–Shimura Relation.** The Eichler–Shimura relation is a neat description of Hecke operators on the reduction of modular curve at p using the Frobenius morphism. In this subsection, we will provide some intuition on why such a description is true. We will assume that there is a natural reduction of T_p acting as $T_p : \mathrm{Pic}^0(\widetilde{X}_1(N)) \to \mathrm{Pic}^0(\widetilde{X}_1(N))$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{Pic}^0(X_1(N)) & \xrightarrow{T_p} & \mathrm{Pic}^0(X_1(N)) \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(\widetilde{X}_1(N)) & \xrightarrow[\widetilde{T}_p]{} & \mathrm{Pic}^0(\widetilde{X}_1(N)) \end{array}$$

(cf. Theorem 9.5.1 of [19].) Then we will show a version of the Eichler–Shimura relation in the context of moduli spaces. The rest of the proof of the Eichler–Shimura relation is just diagram chasing (using the above assumption) and passing between the various equivalent notions of Hecke operators. For this part of the proof, we refer the readers to [4, Section 8.7].

Let E be an elliptic curve over $\overline{\mathbb{Q}}$ with ordinary reduction at $\mathfrak{p}$. Let $Q \in E$ be a point of order N. Let C_0 be the kernel of the reduction map $E[p] \to \widetilde{E}[p]$. By the structure theorem of torsion points of elliptic curves, $E[p] \cong (\mathbb{Z}/p\mathbb{Z})^2$, and by definition $|\widetilde{E}[p]| = p$. From Proposition 6.3(1), it follows that C_0 is an order p subgroup.

Lemma 6.8. *Under the hypotheses of the previous paragraph, for any order p subgroup C of E,*

$$[\widetilde{E/C}, \widetilde{Q+C}] = \begin{cases} [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}] & \text{if } C = C_0, \\ [\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}] & \text{if } C \neq C_0. \end{cases}$$

Proof sketch. (For full details, see Lemma 8.7.1 in [4].) Let $E' = E/C$, $\varphi : E \to E'$ be the quotient isogeny, and $\psi : E' \to E$ be the dual isogeny so that $\psi \circ \varphi = [p]$ and $\varphi \circ \psi = [p]$. Let $Q' = \varphi(Q)$.

Case 1: $C = C_0$**.** A diagram chase on the following diagram

$$\begin{array}{ccccc} E'[p] & \xrightarrow{\psi} & E[p] & \xrightarrow{\varphi} & E'[p] \\ {\scriptstyle \pi'}\downarrow & & {\scriptstyle \pi}\downarrow & & \\ \widetilde{E}'[p] & \xrightarrow[\widetilde{\psi}]{} & \widetilde{E}[p] & & \end{array}$$

shows that $\ker([p]_{\widetilde{E'}}) = \ker(\widetilde{\psi}) = \widetilde{E'}[p]$. Since $\widetilde{\psi}$ is an isogeny,

$$\deg_{\text{sep}}([p]_{\widetilde{E'}}) = |\ker[p]_{\widetilde{E'}}| = p \implies \deg_{\text{ins}}([p]_{\widetilde{E'}}) = \deg([p]_{\widetilde{E'}})/\deg_{\text{sep}}([p]_{\widetilde{E'}}) = p,$$

and since $[p]_{\widetilde{E'}} = \widetilde{\psi} \circ \widetilde{\varphi}$, it follows that

$$\deg_{\text{sep}}(\widetilde{\psi}) = p, \qquad \deg_{\text{ins}}(\widetilde{\psi}) = 1, \qquad \deg_{\text{sep}}(\widetilde{\varphi}) = 1, \qquad \deg_{\text{ins}}(\widetilde{\varphi}) = p.$$

Thus, it follows from subsection 2.4 that $\widetilde{\varphi} = i \circ \sigma_p$, where $i : \widetilde{E}^{\sigma_p} \to \widetilde{E'}$ is an isomorphism taking $\widetilde{Q}^{\sigma_p}$ to $\widetilde{Q'}$. Thus,$[\widetilde{E'}, \widetilde{Q'}] = [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}]$.

Case 2: $C \neq C_0$**.** Let $C' = \ker\psi$ and $C'_0 = \ker\pi'$. In this case, a diagram chase on

$$\begin{array}{ccccc} E[p] & \xrightarrow{\varphi} & E'[p] & \xrightarrow{\psi} & E[p] \\ {\scriptstyle\pi}\downarrow & & {\scriptstyle\pi'}\downarrow & & \\ \widetilde{E}[p] & \xrightarrow[\widetilde{\varphi}]{} & \widetilde{E'}[p] & & \end{array}$$

allows us to show that $C' = C'_0$. Then, applying the argument in case 1, replacing E, Q, φ by E', Q', ψ, respectively, one obtains $\widetilde{\psi} = i \circ \sigma_p$, where

$$i : \widetilde{E'}^{\sigma_p} \xrightarrow{\sim} \widetilde{E} \qquad \widetilde{Q'}^{\sigma_p} \to \widetilde{\psi(Q')} = [p]\widetilde{Q}.$$

Applying σ_p^{-1} to the coefficients of i gives

$$i^{\sigma_p^{-1}} : \widetilde{E'} \to \widetilde{E}^{\sigma_p^{-1}} : \qquad \widetilde{Q'} \mapsto [p]\widetilde{Q}^{\sigma_p^{-1}}.$$

Thus, $[\widetilde{E'}, \widetilde{Q'}] = [\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}]$, which completes the proof. □

Now define the Diamond operators on $\widetilde{S}_1(N)$ in the usual way:

$$\widetilde{\langle d\rangle} : \widetilde{S}_1(N) \to \widetilde{S}_1(N) \qquad [E, Q] = [E, d \cdot Q] \qquad (d, N) = 1.$$

For an elliptic curve E over $\overline{\mathbb{Q}}$ with ordinary reduction at $\mathfrak{p}$, take the sum over all order p subgroups $C \subset E$:

$$\sum_C [\widetilde{E/C}, \widetilde{Q + C}] = (\sigma_p + p\widetilde{\langle p\rangle}\sigma_p^{-1})[\widetilde{E}, \widetilde{Q}].$$

Lemma 6.8 also applies to the case of supersingular reduction (the proof is similar). Thus, the above formula holds for any elliptic curve over $\overline{\mathbb{Q}}$ with good reduction at $\mathfrak{p}$.

This is arguably the meat of the Eichler–Shimura relation. One can transfer this formula on the moduli space to the divisor group of the modular curves, then to the Pic^0 of modular curves, replacing σ_p, $\widetilde{\langle p\rangle}$ and $p\sigma_p^{-1}$ with the suitable induced maps on Picard groups (cf. subsection 2.1 and subsection 2.4). The end product of these translations is:

Theorem 6.9. ***(Eichler–Shimura relation)*** *Let $p \nmid N$. The following diagram commutes:*

$$\begin{array}{ccc} Pic^0(X_1(N)) & \xrightarrow{T_p} & Pic^0(X_1(N)) \\ \downarrow & & \downarrow \\ Pic^0(\widetilde{X}_1(N)) & \xrightarrow[\sigma_{p,*} + \widetilde{\langle p\rangle}_* \sigma_p^*]{} & Pic^0(\widetilde{X}_1(N)) \end{array}$$

In particular, since $\widetilde{\langle p \rangle}$ *acts trivially on* $\widetilde{X}_0(N)$*, the following diagram commutes as well.*

$$\begin{array}{ccc} Pic^0(X_0(N)) & \xrightarrow{T_p} & Pic^0(X_0(N)) \\ \downarrow & & \downarrow \\ Pic^0(\widetilde{X}_0(N)) & \xrightarrow[\sigma_{p,*}+\sigma_p^*]{} & Pic^0(\widetilde{X}_0(N)) \end{array}$$

6.4. **Characteristic polynomial of Frobenius; Modularity Theorem.**

Theorem 6.10. ***(Modularity Theorem, Version*** a_p***)*** *Let* E *be an elliptic curve over* $\mathbb{Q}$ *with conductor* N_E *(cf. Definition 2.9). Then for some newform* $f \in \mathcal{S}_2(\Gamma_0(N_E))$*,*

$$a_p(f) = a_p(E) \qquad \text{for all primes } p.$$

This version of Modularity Theorem tells us that the information about the number of points on the reduction of elliptic curves mod p is always encoded in the coefficient of a modular form. In the rest of the subsection, we will demonstrate the relationship between this and the (strong) version R of the Modularity Theorem, defined in the previous section.

Recall that $J_1(N) = \mathrm{Jac}(X_1(N))$ is an abelian variety of dimension g, where g is the genus of $X_1(N)$. Theorem 5.4 tells us that $J_1(N)[l^n] \cong (\mathbb{Z}/l^n\mathbb{Z})^{2g}$. Thus, we can use the same construction to obtain

$$V_l(X_1(N)) = \varprojlim_n (J_1(N)[l^n]) \otimes \mathbb{Q}_l \cong \mathbb{Q}_l^{2g}.$$

and the Galois action, which takes a point $Q \in J_1(N)(\overline{\mathbb{Q}})$ to Q^σ, defines a representation

$$\rho_{X_1(N),l} : G_{\mathbb{Q}} \to \mathrm{GL}(V_l(X_1(N))).$$

Theorem 6.11. *Let* l *be a prime and* N *be a positive integer. The Galois representation* $\rho_{X_1(N),l}$ *is unramified at every prime* $p \nmid lN$*. For such* p*, let* $\mathfrak{p} \subset \overline{\mathbb{Z}}$ *be a prime over* p*. Then* $\rho_{X_1(N),l}(Frob_{\mathfrak{p}})$ *satisfies the polynomial equation*

$$x^2 - T_p x + \langle p \rangle p = 0.$$

Proof. Let $\mathfrak{p}$ lie over p and $p \nmid lN$. For each $n \in \mathbb{Z}_+$ the following diagram commutes:

$$\begin{array}{ccc} D_{\mathfrak{p}} & \longrightarrow & \mathrm{Aut}(J_1(N)[l^n]) \\ \downarrow & & \downarrow \pi \\ G_{\mathbb{F}_p} & \longrightarrow & \mathrm{Aut}(\mathrm{Jac}(\widetilde{X}_1(N))[l^n]) \end{array}$$

The map π is induced from the reduction map $X_1(N) \to \widetilde{X}_1(N)$. We will take as a fact that the map $\mathrm{Pic}^0(X_1(N))[l^n] \to \mathrm{Pic}^0(\widetilde{X}_1(N))[l^n]$ is an injection (for details, see the discussion in [20]), which also shows that this map is an isomorphism, since by Theorem 5.4 both sides are isomorphic to $(\mathbb{Z}/l^n\mathbb{Z})^{2g}$. Thus, the map π is also an isomorphism. By diagram chasing, the kernel of the left vertical map is $I_{\mathfrak{p}} \subset \ker \rho_{X_1(N),l}$. Thus, $\rho_{X_1(N),l}$ is unramified at p.

To prove the polynomial equation, use the Eichler–Shimura relation

$$\begin{array}{ccc} \mathrm{Pic}^0(X_1(N)) & \xrightarrow{T_p} & \mathrm{Pic}^0(X_1(N)) \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(\widetilde{X}_1(N)) & \xrightarrow[\sigma_{p,*}+\widetilde{\langle p\rangle}_*\sigma_p^*]{} & \mathrm{Pic}^0(\widetilde{X}_1(N)) \end{array}$$

(Recall that $J_1(N)$ is funtorially isomorphic to $\mathrm{Pic}^0(X_1(N))$, so we use the two notations interchangeably.) The same diagram but with T_p replaced by $\mathrm{Frob}_\mathfrak{p} + p\langle p\rangle\mathrm{Frob}_\mathfrak{p}^{-1}$ also commutes. Since the vertical arrows are isomorphisms, $T_p = \mathrm{Frob}_\mathfrak{p} + p\langle p\rangle\mathrm{Frob}_\mathfrak{p}^{-1}$ on $\mathrm{Pic}^0(X_1(N))[l^n]$. Multiplying by $\mathrm{Frob}_\mathfrak{p}$ on both sides gives the desired polynomial relation. Since n is arbitrary, this extends to $V_l(X_1(N))$. □

Next, we will show that the representation $\rho_{X_1(N),l}$ is compatible with $\rho_{A_f,l}$.

Lemma 6.12. *The map $J_1(N)[l^n] \to A_f[l^n]$ is a surjection.*

Proof. Use the model of $J_1(N)$ and A_f over $\mathbb{C}$. Let $y \in A_f[l^n]$. Then $y = x + I_f J_1(N)$ for some $x \in J_1(N)$ such that $l^n x \in I_f J_1(N)$. Thus, $l^n x = l^n x'$ for some $x' \in I_f J_1(N)$. Then $x - x' \in J_1(N)[l^n]$ maps to y. □

Since the morphism $J_1(N) \to A_f$ is defined over $\mathbb{Q}$, and hence locally defined by rational functions over $\mathbb{Q}$, the Galois action commutes with the morphism, i.e., for $\sigma \in G_\mathbb{Q}$, the following diagram commutes:

$$\begin{array}{ccc} J_1(N) & \xrightarrow{\sigma} & J_1(N) \\ \downarrow & & \downarrow \\ A_f & \xrightarrow[\sigma]{} & A_f \end{array}$$

Since T_p acts as $a_p(f)$ and $\langle p\rangle$ acts as $\chi(p)$, it follows that $\rho_{A_f,l}(\mathrm{Frob}_\mathfrak{p})$ also satisfies the polynomial equation

$$x^2 - a_p(f)x + \chi(p)p = 0.$$

Now the modular Galois representation $\rho_{f,\lambda}$ is $\rho_{A_f,l}$ followed by a projection onto $\mathrm{GL}_2(K_{f,\lambda})$; thus, $\rho_{f,\lambda}(\mathrm{Frob}_\mathfrak{p})$ also satisfies the above polynomial equation. This proves:

Theorem 6.13. *Let $f \in \mathcal{S}_2(N,\chi)$ be a newform with number field K_f. Let l be a prime. For each prime λ of $\mathcal{O}_{K_f}$ lying over l, there is a Galois representation $\rho_{f,\lambda} : G_\mathbb{Q} \to GL_2(K_{f,\lambda})$ which is unramified at every prime $p \nmid lN$ such that for any $\mathfrak{p}$ lying over such p, $\rho_{f,\lambda}(Frob_\mathfrak{p})$ satisfies the polynomial equation*

$$x^2 - a_p(f)x + \chi(p)p = 0.,$$

In particular, if $f \in \mathcal{S}_2(\Gamma_0(N))$, then the relation is $x^2 - a_p(f)x + p = 0$.

Theorem 6.14. *The three versions of Modularity Theorems (Version R, Strong Version R, and Version a_p) are all equivalent.*

Proof. Assume Version R. Let E be an elliptic curve over $\mathbb{Q}$ with conductor N. Then by Version R, there is a newform $f \in \mathcal{S}_2(\Gamma_0(M_f))$ such that $\rho_{f,\lambda} \sim \rho_{E,l}$ for some maximal ideal λ of $\mathcal{O}_{K_f}$ lying over l. Thus, $\rho_{E,l}(\mathrm{Frob}_\mathfrak{p})$ satisfies the polynomial $x^2 - a_p(f)x + p$ for any $\mathrm{Frob}_\mathfrak{p}$ where $\mathfrak{p}$ lies over $p \nmid lM_f$. But we also know from

Theorem 2.12 that $\rho_{E,l}(\text{Frob}_\mathfrak{p})$ has characteristic polynomial $x^2 - a_p(E) + p$. Thus, $a_p(f) = a_p(E)$ for almost all p. By [21], this implies that $a_p(f) = a_p(E)$ for all prime p.

Conversely, suppose Version a_p of Modularity is true. Let E be an elliptic curve over $\mathbb{Q}$ with conductor N. There is a newform $f \in \mathcal{S}_2(\Gamma_1(N))$ such that $a_p(f) = a_p(E)$ for all p. Since $a_p(f) \in \mathbb{Z}$, it follows that $K_f = \mathbb{Q}$ and A_f is an elliptic curve. The respective characteristic polynomials for $\rho_{f,l}(\text{Frob}_\mathfrak{p})$ and $\rho_{E,l}(\text{Frob}_\mathfrak{p})$ are $x^2 - a_p(f)x + p$ and $x^2 - a_p(E)x + p$ for all but finitely many p. But then this means that the characteristic polynomials are equal on a dense subset of $G_\mathbb{Q}$, and since trace and determinant are continuous, the characteristic polynomials are always equal. Consequently, the representations are equivalent (cf. Exercise 9.6.1 of [4]). This implies strong Version R of Modularity, which clearly implies Version R of Modularity.

□

The proof above also shows the following remarkable fact:

Proposition 6.15. *Let E be an elliptic curve over $\mathbb{Q}$. If $\rho_{E,l}$ is modular for some l, then $\rho_{E,l}$ is modular for all l.*

Acknowledgments

It is a pleasure to thank my mentor, Raghuram Sundararajan, for his invaluable guidance in helping me understand many of the crucial ideas and details in the material presented in this paper.

References

[1] A. J. Wiles. *Modular Elliptic Curves and Fermat's Last Theorem.* Ann. of Math. **141** (1995), no. 3, 443-551, DOI 10.2307/2118559.
[2] T. Saito. *Fermat's Last Theorem: Basic Tools.* American Mathematical Society, 2013.
[3] J. H. Silverman. *The Arithmetic of Elliptic Curves.* Springer, 2009.
[4] F. Diamond and J. Shurman. *A First Course in Modular Forms.* Springer, 2005.
[5] J. P. Serre. *Abelian l-adic representations and elliptic curves.* Addison-Wesley, 1989.
[6] Y. Qiu and Mathmo123. *Irreduciblity of l-adic representation attach to the elliptic curve over $\mathbb{Q}$ with complex multiplication.* Math StackExchange. Available at https://math.stackexchange.com/questions/3112343/irreduciblity-of-ell-adic-representation-attach-to-the-elliptic-curve-over/3112609#3112609
[7] T. Weston. *The Modular Curves $X_0(11)$ and $X_1(11)$.*
[8] W. Fulton. *Algebraic Curves.* Addison-Wesley, 2008.
[9] The Stacks Project Authors. *Stacks Project.* Available at http://stacks.math.columbia.edu
[10] O. Forster. *Lectures on Riemann Surfaces.* Springer, 1981.
[11] J. P. Serre. *A Course in Arithmetic.* Springer, 1973.
[12] S. Ramanujan. *On Certain Arithmetical Functions.* Trans. Camb. Philos. Soc. **22** (1916), 159–184 (English).
[13] C. Perret-Gentil. *Associating abelian varieties to weight-2 modular forms: the Eichler-Shimura construction.* Available at https://c.pgdm.ch/publications/eichler-shimura.pdf.
[14] A. Polishchuk. *Abelian Varieties, Theta Functions and the Fourier Transform.* Cambridge University Press, 2003.
[15] B. Cais, Brian Conrad, and Francesco Polizzi. *Quotient of abelian variety by an abelian subvariety.* MathOverflow. Available at http://mathoverflow.net/q/37536 (version: 2010-09-02).
[16] K. A. Ribet and W. A. Stein. *Lectures on Modular Forms and Hecke Operators.* Available at https://wstein.org/books/ribet-stein/main.pdf.
[17] J. S. Milne. *Abelian Varieties (v2.00).* 2008. Available at www.jmilne.org/math/.

[18] J. Igusa. *Kroneckerian Model of Fields of Elliptic Modular Functions*. Amer. J. Math., **81** (1959), 561-577, DOI 10.2307/2372914.
[19] S. Bosch, W. Lütkebohmert, and M. Raynaud. *Néron Models*. Springer-Verlag, 1990.
[20] J. Apple and A. Youcis. *Seeking References for Facts from Diamond & Shurman*. Math StackExchange. Available at https://math.stackexchange.com/questions/4295108/seeking-references-for-facts-from-diamond-shurman
[21] H. Carayol. *Sur les Représentations l-adiques Associées aux Formes Modulaires de Hilbert*. Ann. Sci. E. N. S., **19** (1986), 409-468, DOI 10.24033/asens.1512.