THE MODULARITY THEOREM

PETER ZHOU

Abstract. This paper aims to provide an elementary introduction to different versions of the modularity theorems, a remarkable theorem in number theory. Starting by constructing the modular curves, this paper introduces all necessary concepts to state the modularity theorems, including modular curves, oldforms and newforms, Hasse L-function of elliptic curves, and Galois representations. Finally, the last section proves some of the equivalences between the modularity theorems with certain results in algebraic geometry taken as granted. The paper assumes some familiarity in complex analysis, elliptic curves, and algebraic curves.

Contents

Introduction 1
1. Modular Curve 2
1.1. Complex Elliptic Curve and Lattice 2
1.2. Construction of Modular Curve 4
1.3. Modular Curve as Algebraic Curve 7
2. Modular Form and Modularity Theorem 8
2.1. Modular forms 8
2.2. Hecke Operators and Eigenform 9
2.3. Oldform and Newform 11
2.4. Hasse L-function and Modularity Theorem 13
3. Galois Representation and Modularity Theorem 16
3.1. Basics of Galois representation 16
3.2. Galois Representations by Tate Module 17
4. Eichler Shimura Relation 19
4.1. Hecke Operators on Modular Curves 20
4.2. Shifting from complex analytic curves to algebraic curves over $\mathbb{Q}$ 20
4.3. Reduction of Modular Curve and Hecke Operators 22
4.4. Eichler-Shimura Relation 24
5. Equivalences between Modularity Theorems 27
Acknowledgments 29
References 29

Introduction

This paper aims to provide an introduction to different versions of the Modularity Theorem. We start by introducing the modular objects ,i.e. modular curves and

Date: December 14th, 2023.

modular forms, which pin down the concept of modularity. We then associate them with L-function and Galois representation, through which their relations to elliptic curve become explicit. In Section 1, we motivate the concept of modular curves through the moduli problem of elliptic curves over $\mathbb{C}$ and then state one version of modularity that uses modular curves. Section 2 aims to associate L-functions with a special kind of modular form and with elliptic curves over $\mathbb{Q}$, enabling the statement of the modularity conjecture using L-functions. Relevant background on modular forms, Hecke operators, and elliptic curves will be introduced. Similar to Section 2, Section 3 associates Galois representations with both elliptic curves and modular forms and introduces the version of the modularity conjecture in Galois representations, which is the version that Andre Wiles partially proved in 1996. In Section 4, we sketch the proof of equivalences between different versions of modularity conjectures introduced in the previous sections.

1. Modular Curve

Modular curves naturally arises when considering the following question: can we find a parametization of elliptic curves, say over $\mathbb{C}$? If we want to parameterize the simpler curves such as conics instead, it suffices to provide the six coefficients $a_1, a_2, a_3, a_4, a_5, a_6$ so that they determine the conic $a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6 = 0$ up to a scalar. Such tuple $[a_1, ..., a_6]$ are precisely the points in space $\mathbb{P}^5$, the projective space. Thus, the class of cubics can be parameterized by points in $\mathbb{P}^5$. We call $\mathbb{P}^5$ a moduli space of the conics.

Question: Can we find a moduli space *Ell* of the elliptic curves?

Unfortunately, elliptic curves over a general field K are abstract and are hard to parameterize (and is not parameterizable by varieties!). Elliptic curves over $\mathbb{C}$, on the other hand, can be identified with the complex analytic objects: compact Riemann surfaces of genus 1, which correspond to lattices in the complex plane. Thus, we start our parameterization on the class of lattices. (Terminology reminder: All moduli spaces in this paper refer to coarse moduli space.)

1.1. Complex Elliptic Curve and Lattice.

In this subsection, we establish an equivalence between the category of elliptic curves over $\mathbb{C}$ and the category of lattices in $\mathbb{C}$ to transform the moduli problem of elliptic curve into a simpler moduli problem of lattices.

Definition 1.1. A *lattice in* $\mathbb{C}$ is a set $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ with $\frac{\omega_1}{\omega_2} \in \mathbb{C} - \mathbb{R}$. A complex torus is the quotient $\mathbb{C}/\Lambda$. The parallelogram D with vertices $O, \omega_1, \omega_2, \omega_1 + \omega_2$ is called the fundamental parallelogram of the torus.

The fundamental parallelogram contains exactly one representative in each equivalence class of the quotient $\mathbb{C}/\Lambda$, except on the edge where the opposite edges are identified, making the quotient $\mathbb{C}/\Lambda$ topologically a torus.

Proposition 1.2. *Suppose $\phi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ is a holomorphic map between complex tori which perserves the addition structure. Then there exists $m \in \mathbb{C}$ such that $m\Lambda_1 \subset \Lambda_2$ and that $\phi(z_1 + \Lambda_1) = mz_1 + \Lambda_2$. ϕ is invertible if and only if $m\Lambda_1 = \Lambda_2$. We call such morphism an isogeny between complex tori.*

A surprising fact is that a complex torus is actually a complex elliptic curve! In particular, we may find an embedding $i : \mathbb{C}/\Lambda \hookrightarrow \mathbb{CP}^2$ such that $i(\mathbb{C}/\Lambda)$ is a complex

elliptic curve. The key to find such embedding lies in the field of meromorphic functions over $\mathbb{C}/\Lambda$.

A meromorphic function $f : \mathbb{C}/\Lambda \to \mathbb{C}$ can be naturally identified with a meromorphic function $f : \mathbb{C} \to \mathbb{C}$ such that $f(z+\omega) = f(z)$ for $\omega \in \Lambda$, which we call Λ-periodic. It would be useful to construct some Λ-periodic meromorphic functions explicitly to understand the function field of $\mathbb{C}/\Lambda$. The most natural way to achieve this is to write a sum of terms over the whole lattice, e.g. $f(z) = \sum_{\omega\in\Lambda}(z-\omega)^{-2}$. Since it sums over all the lattice, $f(z+\omega) = f(z)$ always holds. However, the series does not converge absolutely, so we need to modify it slightly.

Definition 1.3. The *Weierstrass $\wp$ function* is defined to be

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega\in\Lambda,\omega\neq 0}\left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right)$$

It is straightforward from the construction that $\wp$ and its derivative $\wp'$ are Λ-periodic.

Proposition 1.4. *The field of meromorphic function on $\mathbb{C}/\Lambda$ is $\mathbb{C}(\wp, \wp')$.*

Proof. One may construct a rational function of $\wp$ that has the same number of zeros and poles away from O of any given Λ-periodic even function. By Liouville's theorem and the fact that the number of zeros and poles are equal for meromorphic function on compact Riemann surface, it follows that the two functions differ by a scalar multiple. The general case follows since $f(z) = \frac{1}{2}(f(z)+f(-z)) + \frac{1}{2\wp'(z)}(f(z)-f(-z))\wp'(z)$, where both $f(z)+f(-z)$ and $(f(z)-f(-z))\wp'(z)$ are Λ-periodic even functions. □

Since $\mathbb{C}(\wp, \wp')$ as the meromorphic function field over $\mathbb{C}/\Lambda$ has transcendence degree 1, we should be able to find a polynomial $F(x, y)$ such that $F(\wp, \wp') = 0$. We can find that $F(x, y)$ is a cubic polynomial by computing Laurent series of $\wp, \wp'$ and match the negative terms, so it defines an elliptic curve over $\mathbb{C}$.

Theorem 1.5. *The functions $\wp$ and $\wp'$ satisfy the relation*

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

where g_2, g_3 are constants determined by Λ.

Therefore, complex tori are complex elliptic curves. More explicitly, there is an embedding $i : \mathbb{C}/\Lambda \hookrightarrow \mathbb{CP}^2$ given by

$$z + \Lambda \to [\wp(z), \wp'(z), 1]$$

whose image is the zero set of the homogenization of $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$.

The mapping taking complex torus $\mathbb{C}/\Lambda_1$ to the complex elliptic curve $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$ turns out to be a bijection between the set of complex tori and set of complex elliptic curve. The surjectivity is demonstrated through a construction involving theta function (See [2]). The isomorphism can be further extended to an identification between the category of complex torus and the category of complex elliptic curve, where the morphisms are isogenies.

The preceding discussion, coupled with Proposition 1.2, leads to the following theorem that establishes the correspondence between the categories of elliptic curves over $\mathbb{C}$ and lattices. Further details of the proof can be found in [3]

Theorem 1.6. *The following categories are equivalent:*

(1) *Objects: Elliptic curves over $\mathbb{C}$ up to isomorphism.*
Morphism: Isogenies.
(2) *Objects: Complex tori over $\mathbb{C}$ up to isomorphism.*
Morphism: Isogenies.
(3) *Objects: Lattices $\Lambda \subset \mathbb{C}$, up to homothety.*
Morphism: $Mor(\Lambda_1, \Lambda_2) = \{a \in \mathbb{C} : a\Lambda_1 \subset \Lambda_2\}$

Where two lattices are called homothetic if $\Lambda_1 = a\Lambda_2$ for some $a \in \mathbb{C}$.

1.2. **Construction of Modular Curve.** By Theorem 1.6, we may transform the problem of finding a moduli space of elliptic curves over $\mathbb{C}$ (up to isomorphism) into finding a moduli space of lattices in $\mathbb{C}$ (up to homothety). This is a much simpler problem as the following proposition shows.

Proposition 1.7. *Two lattices $\Lambda_1 = \mathbb{Z}\tau \oplus \mathbb{Z}$ and $\Lambda_2 = \mathbb{Z}\tau' \oplus \mathbb{Z}$ are the same up to homothety if and only if $\tau' = \frac{a\tau+b}{c\tau+d}$ where a, b, c, d are entries of the matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$. Define the group action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$ by $Az = \frac{az+b}{cz+d}$. Each orbit of the group action of $SL_2(\mathbb{Z})$ on the upperhalf plane $\mathcal{H}$ determines a unique lattice up to homothety. Thus the set of orbits $SL_2(\mathbb{Z})/\mathcal{H}$ is a moduli space of complex elliptic curves. We denote it by $Y_0(1)$.*

Proof. Let $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$, $\Lambda' = \omega_1'\mathbb{Z} \oplus \omega_2'\mathbb{Z}$ be two lattices. Without loss of generality, we may assume that $\frac{\omega_1}{\omega_2}, \frac{\omega_1'}{\omega_2'} \in \mathcal{H}$. Suppose $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$, $\Lambda' = \mathbb{Z}\omega_1' \oplus \mathbb{Z}\omega_2'$. By solving linear equations we have that $\Lambda = \Lambda'$ if and only if

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

for some matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in SL_2(\mathbb{Z})$. Thus $\mathbb{Z}\tau \oplus \mathbb{Z}$ is homothetic to $\mathbb{Z}\tau' \oplus \mathbb{Z} \iff \mathbb{Z}\tau \oplus \mathbb{Z} = \mathbb{Z}\alpha\tau' \oplus \mathbb{Z}\alpha \iff \alpha\tau' = a\tau + b, \alpha = c\tau + d \iff \tau' = \frac{a\tau+b}{c\tau+d}$. □

We now extend the definition above to a more general class of modular curves, the significance of which we will soon see.

Definition 1.8. The group $\Gamma(N)$ is defined by

$$\Gamma(N) : \{A \in SL_2(\mathbb{Z}), A \equiv I \bmod N\}$$

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

Example 1.9. The subgroup $\Gamma(N)$ itself is a congruence subgroup. Also, the subgroup

$$\Gamma_1(N) : \{A \in SL_2(\mathbb{Z}), A \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) (mod\, N)\}$$

and

$$\Gamma_0(N) : \{A \in SL_2(\mathbb{Z}), A \equiv \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) (mod\, N)\}$$

both contains $\Gamma(N)$ and are thus congruent as well.

We define the modular curve $Y(\Gamma)$ to be the quotient $\Gamma\backslash\mathcal{H}$. We will use the following notation throughout the rest of this paper: $Y(\Gamma_0(N)) = Y_0(N)$, $Y(\Gamma_1(N)) = Y_1(N)$, $Y(\Gamma(N)) = Y(N)$.

Let Γ be a congruence subgroup. Since $\Gamma \subset SL_2(\mathbb{Z})$, there is a natural surjection $\pi : Y(\Gamma) \to Y(1)$ defined by $\Gamma\tau \to SL_2(\mathbb{Z})\tau$. Consider the collection of cosets $\Gamma\backslash SL_2(\mathbb{Z}) = \{\Gamma\beta_i\}$. Recall that the points in $Y(1)$ corresponds to complex elliptic

curves up to isomorphism. Will the points in $Y(\Gamma)$ corresponds to some equivalence classes of elliptic curves as well? Note that the points in $Y(\Gamma)$ are essentially a further division of points in $Y(1)$, so it should break the equivalence class of isomorphic elliptic curves further. The next theorem shows that this is indeed the case.

Theorem 1.10. *The modular curve $Y_1(N)$ is the moduli space of*

$$\{[E, Q], E \text{ is an elliptic curve, } Q \text{ a } N\text{-torsion point on } E\}$$

where two pairs $[E, Q], [E', Q']$ are considered isomorphic when the two elliptic curves E, E' are isomorphic and the N-torsion point Q, Q' are mapped to each other by the isomorphism maps.The explicit correspondence is given by

$$\Gamma_1(N)\tau \to [\mathbb{C}/\Lambda, \frac{1}{N} + \Lambda]$$

where $\Lambda = \mathbb{Z}\tau \oplus \mathbb{Z}$.

Similarly, the modular curve $Y_0(N)$ is the moduli space of

$$\{[E, \langle Q\rangle], E \text{ is an elliptic curve, } \langle Q\rangle \text{ a cyclic subgroup of order } n \text{ on } E\}$$

where two pairs are considered isomorphic if the isomorphism between the elliptic curves take the order N cyclic subgroups to each other.

Proof. (Sketch) For an arbitrary pair $[E, Q]$, we may find some lattice Λ such that E is identified with $\mathbb{C}/\Lambda$ by Theorem 1.6. One may then choose an appropriate basis of the lattice by Proposition 1.7 such that the point Q corresponds to $\frac{1}{N} + \Lambda$. The correspondence can be checked by simply examining the condition of $SL_2(\mathbb{Z})$ action preserving the N-torsion point, which shows that the matrix belongs to $\Gamma_1(N)$. The argument for $Y_0(N)$ is exactly the same, except that $SL_2(\mathbb{Z})$ action only needs to map a N-torsion point to some of power of the given generating N-torsion points, which shows that the matrix is in $\Gamma_0(N)$. □

The modular curves $Y(\Gamma)$ can be endowed with the structure of Riemann surfaces.

Lemma 1.11. *Take $\tau_1, \tau_2 \in \mathcal{H}$ (not necessarily distinct), then there exists U_1, U_2 open in $\mathcal{H}$, such that $\tau_1 \in U_1$, $\tau_2 \in U_2$, and that for any $\gamma \in SL_2(\mathbb{Z})$,*

$$\gamma(U_1) \cap U_2 \neq \emptyset \Longrightarrow \gamma(\tau_1) = \tau_2$$

The lemma is essentially saying that the action of Γ on $\mathcal{H}$ is *discrete*.

Corollary 1.12. *$Y(\Gamma)$ is Hausdorff.*

Proof. Take distinct points $\Gamma\tau_1, \Gamma\tau_2 \in Y(\Gamma)$. By Lemma 1.11, there exists U_1, U_2 such that $\tau_1 \in U_1, \tau_2 \in U_2$, and that $\gamma(U_1) \cap U_2 = \emptyset$ for every $\gamma \in SL_2(\mathbb{Z})$ since $\gamma(\tau_1) \neq \tau_2$. On the other hand, $\pi(U_1) \cap \pi(U_2) \neq \emptyset$ implies that there exists $a \in U_1$, $b \in U_2$ and $SL_2(\mathbb{Z})a = SL_2(\mathbb{Z})b$, i.e. $b = \gamma a$ for some $\gamma \in SL_2(\mathbb{Z})$. But this means $b \in \gamma(U_1) \cap U_2$, which is a contradiction. $Y(\Gamma)$ is Hausdorff. □

Another useful corollary of Lemma 1.11 greatly reduces the work of laying down atlas on $Y(\Gamma)$: It tells us the projection map $\pi : \mathcal{H} \to Y(\Gamma)$ is locally injective at most points.

Corollary 1.13. *Define the fixing group* $\Gamma_\tau = \{\gamma \in \Gamma, \gamma(\tau) = \tau\}$. *If* Γ_τ *contains matrices other than possibly* $\pm I$, *then we call such* τ *an elliptic point.* $\pi : \mathcal{H} \to Y(\Gamma)$ *is locally injective at non-elliptic points.*

Proof. Take $\tau_1 = \tau_2 = \tau$ in Lemma 1.11 where τ is not an elliptic point, we get open set U such that $\tau \in U$ satisfying that if $\gamma(U) \cap U \neq \emptyset$ then $\gamma(\tau) = \tau$, i.e. $\gamma \in \Gamma_\tau$, so $\gamma = \pm I$ and acts trivially. Now U cannot contains $\Gamma-$ equivalent points: suppose that $a, b \in U$ such that $\gamma(a) = b$, then $b \in \gamma(U) \cap U$ implies $\gamma = \pm I$ and $a = b$. This proves no two points are Γ equivalent in U and thus the mapping $U \to \pi(U)$ is bijective, i.e. $\pi : \mathcal{H} \to Y(\Gamma)$ is locally injective at non-elliptic point. □

We have proved that $Y(\Gamma)$ is Hausdorff and put coordinate charts at non-elliptic points. It remains to put coordinate chart at the (finitely many) elliptic points. We omit the proof here. See [3] Chapter 2.

Now, we move on to study the fundamental domain of modular curves, which allows us to see the modular curves visually.

Definition 1.14. The fundamental domain $D \subset \mathcal{H}$ of the modular form $Y(\Gamma)$ is a region that contains exactly one point from each orbit of Γ action except some possible duplication on the edge.

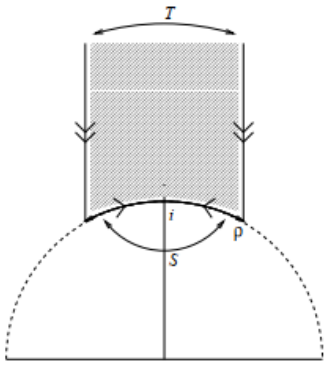


FIGURE 1. Fundamental Region of $X_0(1)$

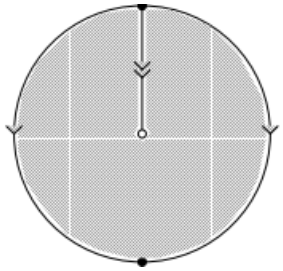


FIGURE 2. Edges Identified

The fundamental region D of $Y(1)$ is given by $D : \{z \in \mathcal{H}, |z| > 1, |Re(z)| < \frac{1}{2}\}$. The $\mathbb{Z}$ periodicity comes from the acction of $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ sending z to $z + 1$. The other matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ send z to $-\frac{1}{z}$. Checking that the two matrices generate $SL_2(\mathbb{Z})$ (which essentially follows from Euclidean algorithm), with some calculation one may show D is the fundamental domain satisfying Definition 1.14. Its two straight edges are identified, and the circular part is symmetrically identified. After identifying the edges, the fundamental region of $X_0(1)$ is almost a sphere with the center point missing, which is the point at infinity.(See Figure 1 and 2 below. These are taken from [1]).

Thus we may compactify modular curve $Y(1)$ by adding the point at infinity to it. More generally, to compactify $Y(\Gamma)$, note that matrices in $SL_2(\mathbb{Z})$ may send ∞

to any rational number $\frac{a}{b}$ by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. This proves that $\Gamma\infty = \mathbb{Q} \cup \{\infty\}$. The points lying above $\Gamma\infty$ (i.e. the Γ equivalence points in $\mathbb{Q} \cup \{\infty\}$) are missing in $Y(\Gamma)$ and we must add all these points to the curve $Y(\Gamma)$ to compactify it.

Definition 1.15. The compactified modular curve $X(\Gamma)$ is given by the set of orbits $\Gamma\backslash\mathcal{H}^*$, where $\mathcal{H}^* = \mathcal{H} \cup \mathbb{Q} \cup \{\infty\}$, which can be made into a compact Riemann surface by choosing appropriate topology and charts. It is the modular curve $Y(\Gamma)$ adjoining the Γ equivalence points in $\mathbb{Q} \cup \{\infty\}$, which we call the *cusps* of $X(\Gamma)$.

Lemma 1.16. *$X(1)$ has one cusp. More generally, $X(\Gamma)$ has finitely many cusps.*

Proof. Again uses $SL_2(\mathbb{Z}) = \sqcup_{j=1}^{d} \Gamma\gamma_j$ to conclude there is at most d cusps. □

One important motivation of compactifying modular curves is that compact Riemann surfaces, as we have seen in the special case of tori, are algebraic.

We have now set the stage to present the initial version of the Modularity Theorem

Theorem 1.17 (Modularity Theorem, Modular Curve Version)**.** *Let $E/\mathbb{C}$ be a complex elliptic curve with $j(E) \in \mathbb{Q}$. Then there exists some N such that a surjection exists between the modular curve $X_0(N)$ and E as compact Riemann surfaces.*

However, this version of modularity theorem is naive in the sense that complex analytic objects reveal little arithmetic information. To get meaningful arithmetic information, we shall move from Riemann surfaces to algebraic varieties over $\mathbb{Q}$.

1.3. **Modular Curve as Algebraic Curve.** The first version of the modularity theorem is complex analytic. However, given the general fact that every compact Riemann surface is projective curve, one may expect an algebraic version of the theorem exists. The first step is to realize the modular curves as complex projective curves. To do this, we use the same strategy in Theorem 1.5.

Theorem 1.18. *The meromorphic function field on the modular curve $X_0(N)$ is $\mathbb{C}(j, f_0)$. Similarly, the meromorphic function field on the modular curve $X_1(N)$ is given by $\mathbb{C}(j, f_1)$, , where*

$$f_1 = \frac{g_2(\tau)}{g_3(\tau)} \wp(\frac{1}{N})$$

$$f_0 = \frac{g_2(\tau)}{g_3(\tau)} \sum_{d=1}^{N-1} \wp(\frac{d}{N})$$

The polynomials relating the generators all have coefficients in $\mathbb{Q}$.

Proof. See [3], Chapter 7. □

By embedding the modular curves into projective space $\mathbb{CP}^2$, we get the complex algebraic version of the modularity theorem: A surjection exists between $X_0(N)$ and E while this time both objects are complex algebraic varieties and the map is morphism between varieties.

However, what we really want is to change both side to varieties defined over $\mathbb{Q}$. Any elliptic curve $E/\mathbb{C}$ with $j(E) \in \mathbb{Q}$ is isomorphic to the universal elliptic curve E_j over $\mathbb{C}$, and E_j is defined by the Weierstrass equation with coefficient in $\mathbb{Q}$. See [2]. Realizing modular curves as algebraic curves over $\mathbb{Q}$, on the other hand, needs more work. Unlike the case of elliptic curve which we can construct the $\mathbb{Q}$ model explicitly by the universal elliptic curve, we give a function field over $\mathbb{Q}$. By the well

known equivalence between the category of nonsingular curves over k and function field of transcendence degree 1 defined over k, This is equivalent to provide a curve over $\mathbb{Q}$.

Definition 1.19. The algebraic modular curve $X_0(N)^{alg}$ is the nonsingular projective curve over $\mathbb{Q}$ given by the function field $\mathbb{Q}(j, f_0)$. Similarly $X_1(N)^{alg}$ is also a nonsingular projective curve over $\mathbb{Q}$ given by the function field $\mathbb{Q}(j, f_1)$.

Now we can state the algebraic version of the modularity theorem.

Theorem 1.20 (Modularity Theorem, Modular Curve Version $\mathbb{Q}$)**.** *Let $E/\mathbb{Q}$ be an elliptic curve over $\mathbb{Q}$. Then there exists some N such that a surjection exists between the modular curve $X_0(N)^{alg}$ and E as algebraic variety over $\mathbb{Q}$. The smallest such N is called the conductor of E.*

It turns out that the two modularity theorems are equivalent. One direction is easy: a morphism defined over $\mathbb{Q}$ is automatically a morphism defined over $\mathbb{C}$. The other direction is quite complicated. We refer readers to the appendix of [6] for a proof.

2. Modular Form and Modularity Theorem

In this section, we turn to the second version of the modularity theorem. We introduces modular forms, Hecke operators, and the theory of newforms in order to associate a L-function to a newform. We also define the Hasse L-functions associated to elliptic curves. These would provide enough background to state the L-function version of the modularity theorem.

2.1. **Modular forms.** What would be a holomorphic function from $X(\Gamma)$ to $\mathbb{C}$? Since $X(\Gamma) = \Gamma\backslash\mathcal{H}^*$, such function should be invariant when its input are two Γ-equivalent points. Furthermore, it should be holomorphic on the cusps we added to $Y(\Gamma)$. How should we define the latter? Since cusps are all transformed from ∞, we just need to find a definition for f to be holomorphic at ∞: We define it for $\frac{1}{f}$ to be holomorphic near 0, and define holomorphic condition at other cusps by saying after performing the same transformation taking the cusp to ∞ on f, the resulting function is holomorphic at infinity. We will make this idea precise in this section.

Definition 2.1. Let Γ be a congruence subgroup. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k with respect to* Γ if $f(\gamma\tau) = (c\tau + d)^k f(\tau)$ for all $\tau \in \mathcal{H}$ and $\gamma \in \Gamma$, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

One may notice that we need specific entries of γ to define modular condition, which is based on coordinate. Here is a coordinate free version of it that also simplifies the notation.

Definition 2.2. Let $\gamma \in SL_2(\mathbb{Z})$. We denote $(c\tau + d)^{-1}$ by $j(\gamma, \tau)$, and define the operator $[\gamma]_k$ that sends f to $f[\gamma]_k = j(\gamma, \tau)^k f$. The weak modularity condition can be rephrased into $f[\gamma]_k = f$ for all $\gamma \in \Gamma$.

The holomorphic condition on cusps are reflected in the following definition.

Definition 2.3. A weakly modular function is *modular* with respect to Γ if it further satisfies

(1) f is holomorphic on $\mathcal{H}$
(2) $f[\gamma]_k$ is defined and holomorphic at ∞

Now we turn to some examples of modular forms:

Example 2.4. The function $G_k(\tau)$ defined by $\tau \to \sum_{(m,n)\in\mathbb{Z}^2-\{(0,0)\}}(m\tau+n)^{-k}$ is a weight k modular form corresponds to $SL_2(\mathbb{Z})$.

$$G_k(\gamma\tau) = \sum_{(m,n)\neq(0,0)} (m\frac{a\tau+b}{c\tau+d}+n)^{-k}$$

$$= (c\tau+d)^{-k} \sum_{(m,n)\neq(0,0)} (ma+nc)\tau+mb+nd)^{-k}$$

When (m,n) varies through $\mathbb{Z}^2-(0,0)$, $(ma+nc, mb+nd)$ also varies through $\mathbb{Z}^2-(0,0)$, so the equation reduces to

$$f(\gamma(\tau)) = (c\tau+d)^{-k} f(\tau)$$

It is holomorphic at infinity by checking its q-expansion, which we define below.

Next, we introduce q-expansion of modular form.

The mapping $\tau \to e^{2\pi i\tau}$ takes $\mathcal{H}$ to $D^\times$, the unit disk without center on $\mathbb{C}$. Given a modular form f with respect to Γ, since $\Gamma(N) \subset \Gamma$ for some N The function $g(z)$ defined by $g(e^{\frac{2\pi i\tau}{N}}) = f(\tau)$ is well defined and holomorphic as composition of *log* function with f, and f holomorphic at cusps correpsonds to the fact that g can be extended holomorphically at 0, so we get a holomorphic function $g : D \to \mathbb{C}$, for which we may compute its power series expansion $g(z) = \sum_{i=0}^{\infty} a_i z^i$. Substituting back to f gives $f(\tau) = \sum_{n=0}^{\infty} a_n q^n$ where $q = e^{\frac{2\pi i\tau}{N}}$.

Definition 2.5. A modular form f is a cusp form if $a_0 = 0$ in its q-expansion.

2.2. **Hecke Operators and Eigenform.** The motivation to define Hecke operators is as follows. We may define the *Petersson Inner Product* on $S_k(\Gamma)$, under which this special class of operators in $End(V)$ would be normal. The Spectral theorem tells us that the vectorspace $S_k(\Gamma)$ has a basis of modular forms that are eigenvectors of all Hecke operators $\{T_n, \langle n\rangle | gcd(n,N) = 1\}$. Such modular forms are called eigenforms. A special kind of eigenforms has nice properties, that they are eigenvectors of all Hecke operators, and that their coefficient $a_n(f)$ in q-expansions are given by the eigenvalue of T_n. This allows us to construct a L-function that look like the Hasse-Weil L-function of elliptic curve.

To motivate the specific construction of Hecke operator, consider the following general problem. Let Γ_1, Γ_2 be two congruence subgroup. How can we define operators mapping $M_k(\Gamma_1)$ to $M_k(\Gamma_2)$? Once we know this, taking $\Gamma_2 = \Gamma_1$ gives us a way to construct endomorphisms on $S_k(\Gamma_1)$. There is the double coset operator that does the trick and can construct lots of such operators: Take any $\alpha \in GL_2(\mathbb{Q})$ where $det(\alpha) > 0$. Then consider the set $\Gamma_1\alpha\Gamma_2 = \{\gamma_1\alpha\gamma_2, \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2\}$. Γ_1 acts on this set by left multiplication. Consider the orbits $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2 = \sqcup_j \Gamma_1\beta_j$. The orbits are actually finite by the following lemma:

Lemma 2.6. *Let Γ_1, Γ_2 be congruence subgroups, and $\alpha \in GL_2(\mathbb{Q})$, $det(\alpha) > 0$. Then $\Gamma_3 = \alpha\Gamma_1\alpha^{-1} \cap SL_2(\mathbb{Z})$ is again a congruence subgroup, and there exists a bijection between $\Gamma_3\backslash\Gamma_2$ and $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2$, sending $\Gamma_3\gamma$ to $\Gamma_1\alpha\gamma$.*

Also, any two congruence subgroups are commensurable, i.e. $[G_1 : G_1 \cap G_2], [G_2 : G_1 \cap G_2] < \infty$. Hence the number of orbits in $\Gamma_1 \backslash \Gamma_1 \alpha \Gamma_2$ is finite, and we may define the double coset operator as

Definition 2.7. The double coset operator $[\Gamma_1 \alpha \Gamma_2]_k : M_k(\Gamma_1) \to M_k(\Gamma_2)$ is defined as $f[\Gamma_1 \alpha \Gamma_2]_k = \sum f[\beta_j]_k$

In particular, take $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$, we get $[\Gamma_1 \alpha \Gamma_2]_k \in End(M_k(\Gamma_1(N)))$. Among these operators, there are two kinds that stand out to be crucial in the theory.

Definition 2.8. The Diamond operators $\langle d \rangle \in End(M_k(\Gamma_1(N)))$ are the double coset operators $[\Gamma_1(N) \alpha \Gamma_1(N)]$ where $\alpha \equiv \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) (\text{mod } N)$. The Hecke operators T_p are the double coset operators $[\Gamma_1(N) \alpha \Gamma_1(N)]$ where $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ and p any prime number.

As we mentioned at the start of the section, the q-expansion coefficient of eigenforms behave nicely under the Hecke operators T_p. In order to know this, we need to calculate the q-expansion coefficient of $T_p(f)$ from q-expansion of f, which requires us to find coset representatives of $\Gamma_1(N)/\Gamma_1(N)\alpha\Gamma_1(N)$.

Proposition 2.9. *When $p|N$, the coset representatives of $\Gamma_1(N)/\Gamma_1(N)\alpha\Gamma_1(N)$ are given by*

$$\gamma_j = \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}$$

If $p \nmid N$, then there is an additional representative

$$\begin{pmatrix} m & n \\ N & p \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$$

Therefore,

$$T_p(f) = (\sum_{j=0}^{p-1} f[\left(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix}\right)]_k) + 1_N(p) f[\left(\begin{smallmatrix} m & n \\ N & p \end{smallmatrix}\right)]_k [\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)]_k$$

where $1_N(p) = 1$ if $p \nmid N$ and 0 otherwise, m, n are any integer that satisfy $mp - nN = 1$. and

$$\langle d \rangle f = f[\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)]_k$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$.

We state the way to calculate the q expansion of $T_p(f)$ and $\langle d \rangle f$ explicitly here, which will be useful when constructing L-functions associated to modular forms.

Proposition 2.10. *Suppose $f \in M_k(\Gamma_1(N))$ have q expansion $f = \sum_{n=0}^{\infty} a_n(f) q^n$. (Here we make the convention that $a_n(f)$ denotes the n-th q-coefficient of f. Also $a_{\frac{n}{p}}(f) = 0$ if $p \nmid n$)*

$$a_n(T_p(f)) = a_{np}(f) + 1_N(p) p^{k-1} a_{\frac{n}{p}}(\langle P \rangle f) \tag{2.11}$$

In particular, let $\chi : (\mathbb{Z}/n\mathbb{Z})^* \to \mathbb{C}$ be a character. If $f \in M_k(N, \chi)$ then $T_p(f) \in M_k(N, \chi)$ as well, and

$$a_n(T_p(f)) = a_{np}(f) + \chi(p) p^{k-1} a_{\frac{n}{p}}(f) \tag{2.12}$$

We have defined Hecke operators $\langle d \rangle$ for $gcd(d, N) = 1$ and T_p for p prime. Now we define Hecke operator $\langle n \rangle$ and T_n for arbitrary n in an inductive manner.

Definition 2.13. The Hecke operators T_n, $\langle n\rangle$ are defined as follows:$T_1 = Id$, $T_{p^n} = T_p T_{p^{n-1}} - p^{k-1}\langle P\rangle T_{p^{n-2}}$ for $n > 1$, and finally $T_{ab} = T_a T_b$ if $gcd(a,b) = 1$. Define $\langle n\rangle = 0$ if $gcd(n, N) > 1$.

We have a nice-looking formula to calculate the coefficient of $T_n f$.

Proposition 2.14.

$$a_m(T_n(f)) = \sum_{d|gcd(m,n)} d^{k-1} a_{\frac{mn}{d^2}}(\langle d\rangle f) \tag{2.15}$$

To close this subsection, we state without proof the definition of Petersson Inner product on $S_k(\Gamma_1(N))$ and the fact that Hecke operators are normal with respect to the inner product. The proof comes from calculating adjoints of double coset operator in general and the commutativeness of Hecke operators. See [3] Chapter 5.

Definition 2.16. The Petersson Inner Product on $S_k(\Gamma_1(N))$ is defined by

$$\langle f, g\rangle = \frac{1}{V_\Gamma}\int_{X_1(N))} f(\tau)\overline{g(\tau)} Im(\tau)^k \, d\mu$$

Where $d\mu$ is $\frac{dxdy}{y^2}$, the hyperbolic measure on the upper-half plane $\mathcal{H}$, $v_\Gamma = \int_{X_{\Gamma_1(N)}} d\mu$.

Proposition 2.17. *The Hecke operators T_n and $\langle n\rangle$ in $End(S_k(\Gamma_1(N)))$ are normal for all n satisfies that $gcd(n, N) = 1$*

Corollary 2.18. *The vectorspace $S_k(\Gamma_1(N))$ has a basis consists of simultaneous eigenvectors for all $T_n, \langle n\rangle$, $gcd(n, N) = 1$. We call the simultaneous vectors eigenforms.*

2.3. **Oldform and Newform.** This section aims to define a special kind of eigenforms. They have the magical property that while apriori we only know they are eigenvectors of Hecke operators coprime to the level, they are in fact eigenvectors of all Hecke operators, and the eigenvalues of T_n gives the n-th q-expansion coefficient of f. Such eigenforms are the so called new forms.

We begin with an observation: Since $\Gamma_1(N) \subset \Gamma_1(M)$ when $M|N$, a modular form at level M is also a modular form at level N. Furthermore, take d satisfies $dM|N$, and $\gamma = \left[\begin{smallmatrix} a_0 & b_0 \\ c_0 & d_0 \end{smallmatrix}\right] \in \Gamma_1(N)$, the modular form $f(d\tau)$ satisfy that $f(d\tau)\left[\begin{smallmatrix} a_0 & b_0 \\ c_0 & d_0 \end{smallmatrix}\right]_k = f\left[\begin{smallmatrix} da_0 & b_0 \\ dc_0 & d_0 \end{smallmatrix}\right]_k$. Since $da_0 \equiv d \bmod N$ hence $da_0 \equiv 1 \bmod M$ and $f \in M_k(\Gamma_1(M))$, $f\left[\begin{smallmatrix} da_0 & b_0 \\ dc_0 & d_0 \end{smallmatrix}\right]_k = f$, hence $f(d\tau) \in M_k(\Gamma_1(N))$.

In other words, some forms at level N are essentially forms coming from lower levels. Since the vectorspace $S_k(N)$ is closed under addition, these cusp forms coming from lower levels span out a subspace of $S_k(\Gamma_1(N))$. We summarize the idea in the following definition:

Definition 2.19. Let $Im_d(S_k(\Gamma_1(\frac{N}{d}))$ denotes the image of $S_k(\Gamma_1(\frac{N}{p}))$ in $S_k(\Gamma_1(N))$ under the embeddings $i_d : f \to f[\gamma]_k$, where $\gamma = d \in \Gamma_0(N)/\Gamma_1(N)$. The subspace of old forms at level N is

$$S_k(\Gamma_1(N))^{old} = \sum_{p|N}\sum_{d} Im_d(S_k(\Gamma_1(\frac{N}{p})))$$

The subspace of new forms at level N is

$$S_k(\Gamma_1(N))^{new} = S_k(\Gamma_1(N))^{old\perp}$$

By computation, we may prove that the subspace of old forms and new forms are stable under Hecke operators. (See [3] Chapter 5.) Therefore, $S_k(\Gamma_1(N))^{old}$ and $S_k(\Gamma_1(N))^{new}$ each has an orthogonal basis consisting of simultaneous eigenforms of Hecke operators $T_n, \langle n\rangle$, where $gcd(n, N) = 1$. We proceed to show that an eigenform that is also a new form will be the eigenvector of all Hecke operators, including those that are not coprime to the level, and that the eigenvalue of T_n is $a_n(f)$, the n-th coefficient in the q-expansion of f.

Theorem 2.20. *Let $f \in S_k(\Gamma_1(N))^{new}$ be an eigenform with respect to Hecke operators $T_n, \langle n\rangle$, where $gcd(n, N) = 1$. Then f is in fact a newform. It is unique in the sense that if f' is another eigenform with respect to Hecke operators $T_n, \langle n\rangle | gcd(n, N) = 1$ and have same eigenvalues for all Hecke operators, then they differ by a scalar multiple. Furthermore, the eigenvalues for all T_n are given by $a_n(f)$, the n-th coefficient in the q expansion of f, respectively.*

We need a nontrivial lemma to prove the theorem.

Lemma 2.21. *Suppose a cusp form $f \in S_k(\Gamma_1(N))$ has q-expansion $f = \sum_n a_n q^n$ that satisfies $a_n(f) = 0$ for all $gcd(n, N) = 1$. Then $f = \sum_{p|N} i_p f_p(p\tau)$ with each $f_p \in S_k(\Gamma_1(\frac{N}{p}))$.*

Proof. The *only if* direction is easy. Observe that the q-coefficients of forms $g \in S_k(\Gamma_1(N))$ obtained by $g(\tau) = \sum_d f_d(d\tau)$ for $f \in S_k(\Gamma_1(\frac{N}{d}))$ for $d > 1, d|N$ satisfy that $a_n(f) = 0$ if $gcd(n, N) = 1$. Indeed, suppose $f(\tau) = \sum_n a_n q^n$. Then the q expansion is given by $f(d\tau) = \sum a_n q^{nd}$, so the coefficient of q^n will always be zero if $n \neq dx$ for some $d|N$, i.e. $gcd(n, N) = 1$. The proof of the *if* direction involves rather technical calculation and we omit the details here. See [3] section 5.7. □

Now we may prove Theorem 2.20.

Proof. Let $f \in S_k(\Gamma_1(N))^{new}$ be an eigenform for the Hecke operators T_n and $\langle n\rangle$ for all $gcd(n, N) = 1$. Suppose $T_n f = c_n f$ and $\langle n\rangle f = d_n f$. The map $n \to d_n$ defines a Dirichlet character $\chi : (\mathbb{Z}/n\mathbb{Z})^* \to \mathbb{C}^*$, so $f \in S_k(N, \chi)$. Proposition 2.14 shows that $a_1(T_n(f)) = a_n(f)$ for all $n \in \mathbb{Z}^+$. Since f is eigenform away from the level, $a_1(T_n(f)) = c_n a_1(f)$ when $gcd(n, N) = 1$. Thus if $a_1(f) = 0$, then $a_n(f) = 0$ whenever $gcd(n, N) = 1$, and f is an old form by Lemma 2.21, thus $f = 0$.

Now suppose $f \neq 0$ a new form. Then $a_1(f) \neq 0$, and we may normalize it so that $a_1(f) = 1$. The form $T_n f - a_n(f)f$ is a new form since the space of new forms is stable under Hecke operators. Its first coefficient is zero, so the discussion above shows that $T_n f - a_n(f)f = 0$. □

We say a nonzero modular form $f \in S_k(\Gamma_1(N))$ is a Hecke eigenform if it is the eigenvector for all $T_n, \langle n\rangle, n \in \mathbb{Z}^+$. The Hecke eigenform $f = \sum a_n q^n$ is said to be normalized when $a_1(f) = 1$. A Hecke eigenform in $S_k(\Gamma_1(N))^{new}$ is called a newform.

We have seen that newforms give a basis of the subspace of new forms. In fact, the newforms and the forms grow out of it provides a basis the whole space $S_k(\Gamma_1(N))$. This will become useful when we construct Galois representation of modular forms in the next section.

Theorem 2.22. *The set*

$$B_k(N) = \{f(n\tau) : f \text{ is a newform of level } M \text{ and } nM|N\}$$

span $S_k(\Gamma_1(N))$. In fact, it is a basis of $S_k(\Gamma_1(N))$.

Proof. (partial) To prove linear independence, we need a result called Strong Multiplicity One. So we only prove that the set span all of $S_k(\Gamma_1(N))$ here. We prove by induction. When $N = 1$ there are no old forms, so $S_k(\Gamma_1(N)) = S_k(\Gamma_1(N))^{new}$ which has a basis of newforms. Suppose $S_k(\Gamma_1(M))$ has basis in the form of $B_k(M)$ for all $M < N$.

Consider the decomposition

$$\begin{aligned} S_k(\Gamma_1(N)) &= S_k(\Gamma_1(N))^{new} \oplus S_k(\Gamma_1(N))^{old} \\ &= S_k(\Gamma_1(N))^{new} \oplus \sum_{p|N} Im(S_k(\Gamma_1(\frac{N}{p}))) \end{aligned}$$

The first subspace is spanned by newforms as proved in . By induction hypothesis each $Im(S_k(\Gamma_1(\frac{N}{p})))$ is spanned by $B_k(\frac{N}{p})$. Hence $B_k(N)$ spans all of $S_k(\Gamma_1(N))$. □

From Theorem 2.20, we know that the coefficients in q expansion are exactly eigenvalues for the newforms. Recall the recursion definition of Hecke operators T_n, we may check whether a modular form is normalized newform by looking at its coefficients in q expansion:

Proposition 2.23. *Let $f \in S_k(N, \chi)$. Then f is a normalized newform if and only if its q expansion coefficients satisfy the conditions*

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

L function of modular forms. The definition Let f be a modular form with q expansion $f = \sum a_n q^n$. Then the L function associated to f is given by $L(s, f) = \sum a_n n^{-s}$.

Theorem 2.24. *Let $f \in S_k(N, \chi)$. Then f is a normalized newform if and only if the L function $L(f, s)$ has the Euler product form*

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

Remark 2.25. The L-function defined above can always be extended meromorphically to the whole complex plane by Mellin transformations. For details, see [3]. The existence of meromorphic extensions of various L-functions can be really hard in general. For example, the Artin's conjecture just asks about whether every Artin L-functions have such an extension. The Hasse L-functions associated to elliptic curves, which we will define in the next section, is also very hard to extend. In contrast, the L-functions associated to eigenforms are fairly easy to extend. Therefore, the modularity theorem becomes significant as it relates hard L-functions to much easier ones! In particular, it solves the problem of extending Hasse L-functions by showing that they are always L-functions of some eigenform.

2.4. **Hasse L-function and Modularity Theorem.** In the last section, we defined the L-functions associated to normalized Hecke eigenforms. In this section, we define the Hasse L-functions associated to elliptic curves $E/\mathbb{Q}$. The L-function version of modularity theorem tells us that any Hasse L-function coincides with a

L-function of some normalized Weight 2 Hecke eigenform. We will soon recognize some observations that convince us the validity of the theorem.

The reduction of an elliptic curve E is defined by the following process. Consider a general Weierstrass equation E defined over $\mathbb{Q}$,

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, a_i \in \mathbb{Q}$$

The linear change of variable $(x, y) = (u^2x', u^3y')$ gives a Weierstrass equation E' with $a_i' = a_iu^{-i}$. Any two Weierstrass equation that can be transformed by such map is said to be equivalent. Therefore, any Weierstrass equation is equivalent to one Weierstrass equation with coefficients $a_i \in \mathbb{Z}$

For any prime p and Weierstrass equation E, let $v_p(E)$ be the largest integer N such that $p^N|\triangle(E)$. Also define $v_p(0) = +\infty$. Define

$$v_p(E)_{min} = min\{v_p(E'), E' \text{ equivalent to } E, \triangle(E') \in \mathbb{Z}\}$$

It can be shown that for Weierstrass equation E defined over $\mathbb{Q}$, there exists an equivalent Weierstrass equation E' that satisfies $v_p(E') = v_p(E)_{min}$ for all p simutaneously. We call E' the global minimal Weierstrass equation. The reduction $\tilde{E}/\mathbb{F}_p$ is an algebraic curve over $\mathbb{F}_p$ defined by the equation

$$\tilde{E} : y^2 + \tilde{a_1}xy + \tilde{a_3}y = x^3 + \tilde{a_2}x^2 + \tilde{a_4}x + \tilde{a_6}$$

where $\tilde{a_i}$ is the image of coefficient a_i from the global minimal Weierstrass equation E' under mod p reduction.

The reduced curve $\tilde{E}/\mathbb{F}_p$, while is defined by a Weierstrass form with coefficient $\tilde{a_i} \in \mathbb{F}_p$, may not be an elliptic curve as it will be singular if $v_p(E)_{min} > 0$, in which case $\triangle(\tilde{E}) = 0$ in $\mathbb{F}_p$. In particular, we have the following categorization of different reduction phenomena.

Definition 2.26. An elliptic curve $E/\mathbb{Q}$

(1) has *good reduction (stable) at* p if $\tilde{E}$ is singular, or equivalently if $p \not| \Delta_{min}(E)$. Such reduction is *ordinary* if $\tilde{E}[p] = \mathbb{Z}/p\mathbb{Z}$, or it is *supersingular* if $\tilde{E}[p] = 0$.
(2) has *multiplicative reduction* at p if $\tilde{E}$ has a node, i.e. if $p|\Delta_{min}(E)$ and the two tangent lines have different slopes. It is said to *split* if the two tangent lines at the node have slope in F_p and *nonsplit* otherwise.
(3) have additive reduction at p if $\tilde{E}$ has a cusp, i.e. $p|\Delta_{min}(E)$ and the two tangent lines have the same slope.

The reduction behavior of elliptic curve $E/\mathbb{Q}$ at different primes combine to indicate global properties of E.

Proposition 2.27. *Define* $a_p(E) = p+1-\tilde{(}E)(\mathbb{F}_p)$*, and* σ_p *the Frobenius map on* E*. Let* E *be an elliptic curve over* $\mathbb{Q}$ *and let* p *be a prime such that* E *has good reduction modulo* p*. Let* $\sigma_{p,*}$ *and* σ_p^* *be the forward and reverse maps on* $\mathrm{Pic}^0(\widetilde{E})$ *induced by* σ_p*. (Here* $Pic^0(E)$ *denote Picard group of degree* 0*. Then*

$$a_p(E) = \sigma_{p,*} + \sigma_p^* \text{ as endomorphisms of } \mathrm{Pic}^0(\widetilde{E}).$$

(Here the left side means multiplication by $a_p(E)$*.)*

Proof. An element $x \in \overline{\mathbb{F}}_p$ satisfies $x^p = x$ if and only if $x \in \mathbb{F}_p$. Thus

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

and so since $\sigma_p - 1$ is separable,

$$\left|\widetilde{E}\left(\mathbb{F}_p\right)\right| = \left|\ker\left(\sigma_p - 1\right)\right| = \deg\left(\sigma_p - 1\right).$$

Now the result follows from the fact that $[deg(\phi)] = \phi_*\phi^*$. □

Definition 2.28. The Zeta-counting function is defined by

$$Z_p(E, X) = \exp(\sum_{e=1}^{\infty} \#\tilde{E}(\mathbb{F}_{p^r})\frac{X^e}{e})$$

Finally, the Hasse L-function associated to E is defined by $L(E, s) = \prod_p Z_p(E, p^{-s})$.

The Hasse L-function can be thought as some way to pack up information of all reductions.

Lemma 2.29. *Define the solution counting coefficients by $t_{p^r} = p^r + 1 - \tilde{E}(\mathbb{F}_{p^r})$. Then they satisfy the recursive relation*

$$t_{p^r}(E) = t_p(E)t_{p^{r-1}}(E) - 1_E(p)pt_{p^{r-2}}(E)$$

Proof. By direct computation. See [2]. □

Theorem 2.30.

$$Z_p(E, p^{-s}) = (1 - a_p(E)p^{-s} + 1_E(p)p^{1-2s})^{-1}$$

where $a_p(E)$ are the same as defined in Proposition 2.27.

Thus, the Hasse L-function has the following product form

$$L(E, s) = \prod_p (1 - a_p(E)p^{-s} + 1_E(p)p^{1-2s})^{-1}$$

Proof. It suffices to prove

$$\sum_{e=1}^{\infty} \frac{t_{p^e}(E)}{e} X^e = -\log(1 - a_p(E) + 1_E(p)pX^2)$$

When $X = 0$, both sides are zero, so it suffices to show their derivatives are equal, i.e.

$$\sum_{e=1}^{\infty} t_{p^e}(E)X^{e-1} = \frac{a_p(E) - 1_E(p)2pX}{1 - a_p(E)X + 1_E(p)pX^2}$$

which follows from the recursive relation stated in Lemma 2.29. □

Compare Theorem 2.30 and Theorem 2.24, we notice that the Hasse L-functions has similar Euler product form to the L-function associated to weight 2 Hecke eigenforms. This is not merely a coincidence. In fact,

Theorem 2.31 (Modularity Theorem, L function version)**.** *Let $E/\mathbb{Q}$ be an elliptic curve over $\mathbb{Q}$ with conductor N. Then there exists $f \in S_2(\Gamma_0(N))$ that satisfies*

$$L(f, s) = L(E, s)$$

This version of modularity theorem, despite looks vastly different from the modularity theorem we stated at Theorem 1.20, is in fact equivalent to the modular curve version by the Eichler-Shimura relation, which we will come to in section 4.

3. Galois Representation and Modularity Theorem

The last version of the modularity theorem is perhaps the least intuitive version at the first glance. The idea is that the arithmetic information of both elliptic curves and cusp forms at level N are encoded in some l-adic representations of the absolute Galois group $G_{\mathbb{Q}}$, and, just as other versions of modularity, the representations coming from elliptic curves always correspond to the representations coming from cusp forms. This version of modularity theorem is the one Andre Wiles proved directly. By Taylor-Wiles method one may study the deformation of Galois representations to study modularity. For an introduction on Galois deformation, see [8].

3.1. **Basics of Galois representation.** In this section we will show how to extract Galois representations from elliptic curve and modular forms and state our last version of modularity theorem.

The algebraic closure $\overline{\mathbb{Q}}$ is the union of all algebraic elements in $\mathbb{C}$ over $\mathbb{Q}$. Define the algebraic integer $\overline{\mathbb{Z}}$ to be the collection of all algebraic integers. This is a ring in $\overline{\mathbb{Q}}$ and we have that $Frac(\overline{\mathbb{Z}}) = \overline{\mathbb{Q}}$. Its prime ideals include prime ideals β in $\mathcal{O}_K$, the ring of integers of in any number fields.

Recall that the Galois group of the infinite dimensional extension $\overline{\mathbb{Q}}$ over $\mathbb{Q}$ consists of $g \in Aut(\overline{\mathbb{Q}})$ fixing elements in $\mathbb{Q}$. From classical Galois theory we know every $\rho \in Gal(K/\mathbb{Q})$ can be extended into some $\rho' : Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ and $\rho(K) = K$ for any $\rho \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ and $K/\mathbb{Q}$ Galois extension. These fact combine to give the alternative definition of the Galois group $G_{\mathbb{Q}} = Gal(\overline{\mathbb{Q}}/\mathbb{Q})$:

$$Gal(\overline{\mathbb{Q}}/\mathbb{Q}) = \varprojlim Gal(K/\mathbb{Q})$$

To recognize some interesting elements in $G_{\mathbb{Q}}$, we investigate the concept of decomposition group.

Let $D_\beta = \{\sigma \in G_{\mathbb{Q}}, \sigma(\beta) = \beta\}$. Then any automorphism σ on $\overline{\mathbb{Q}}$ first restricts to $\overline{\mathbb{Z}}$ and then to $\overline{\mathbb{Z}}/\beta$, and we have

$$\pi : D_\beta \to Gal((\overline{\mathbb{Z}}/\beta)/(\mathbb{Q}/(\beta \cap \mathbb{Q}))) = Gal(\overline{\mathbb{F}_p}/\mathbb{F}_p)$$

This map is surjective with kernel $I_\beta = \{\sigma \in G_{\mathbb{Q}}, \sigma(x) \equiv x (\text{mod } \beta \text{ for } x \in \overline{\mathbb{Z}})\}$, which we call inertia group. The Frobenius element $Frob_\beta$ is defined to be the preimage of $frob_p : x \to x^p \in Gal(\overline{\mathbb{F}_p}/F_p)$. Thus $Frob_\beta$ is defined up to conjugacy class of the inertia subgroup and satisfies that $\sigma(x) \equiv x^p \text{mod } \beta$, for all $x \in \overline{\mathbb{Z}}$.

We also note that $Frob_{\sigma(\beta)} = \sigma Frob_\beta \sigma^{-1}$, so if β is inside an abelian extension of $\mathbb{Q}$, then $D_\beta = D_{\sigma(\beta)}$ for any $\sigma \in G_Q$. In this case we may denote it by $Frob_p$. (Recall that prime ideals in $\mathcal{K}$ lies over a unique prime $p = \beta \cap \mathbb{Z} \subset \mathbb{Z}$, and suppose $p\mathcal{K} = \prod \beta_i^{e_i}$ and $K/\mathbb{Q}$ Galois, then the Galois group acts transitively on β_i and $e_i = e, f_i = f$ for all i.)

Definition 3.1. A l-adic Galois representation with respect to $\mathbb{Q}$ is a continuous homomorphism $\rho : Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_d(K)$ where the Galois group is endowed with Krull topology and K is finite dimensional over $\mathbb{Q}_l$. The representation is said to be unramified at prime $p \in \mathbb{Z}$ if $I_\beta \subset ker(\rho)$ for some prime ideal $\beta \subset \overline{\mathbb{Z}}$ lying over p.

When ρ is unramified at p, the image $\rho(Frob_\beta)$ is well defined for any β lies over p since $I_{\beta'} = \sigma I_\beta \sigma^{-1} \subset \sigma ker(\rho)\sigma^{-1} = ker(\rho)$. In our context, we will see that ρ is unramified at all but finitely many primes.

Example 3.2. The l-adic character $\chi_l : G_{\mathbb{Q}} \to \mathbb{Q}_l^*$ is given by the composition $G_{\mathbb{Q}} \to Gal(\mathbb{Q}(\mu_{l^\infty}/\mathbb{Q})) \cong \mathbb{Q}_l^*$. It is continuous, unramified at $p \neq l$, and $\chi_l(Frob_p) = p$ for $p \neq l$.

We also need a lemma about the Krull topology of G_Q.

Lemma 3.3. *The sets $U_\sigma(F) = \{\sigma\sigma', \sigma'_{|F} = id\}$ are open and form a basis for $\sigma \in G_Q$ and F Galois over $\mathbb{Q}$. Furthermore, the elements which takes the form $Frob_\beta$ are dense.*

A subset $U \subset G_Q$ is open normal subgroup if and only if it is given by $U_1(F)$.

3.2. **Galois Representations by Tate Module.** Both the elliptic curves and modular forms have their corresponding Galois representation constructed by Tate modules, which we now define.

Given elliptic curve $E/\mathbb{Q}$, the absolute Galois group act on points of E, fixing the n-torsion points $E[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. (The former holds because the map $[n] : E \to E$ is essentially a rational map, thus $[n](\sigma(P)) = \sigma(nP) = 0$. To see the latter holds, we consider the case when the elliptic curve is defined over $\mathbb{C}$. In that case we may view it as a complex torus $\mathbb{C}/\Lambda$, whose n-torsion points is given by $\{\frac{a+b\tau}{N} + \Lambda_\tau, a, b = 0, 1, .., n-1\}$, which is isomorphic to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.

Consider the set of l^n-torsion points $E[l^n]$ for $n = 1, 2, ...$ There is a natural inclusion $i : E[l^{n-1}] \to E[l^n]$ as a subset, hence compatible with the action by $G_{\mathbb{Q}}$. Thus $G_{\mathbb{Q}}$ acts on the inverse limit

$$\lim_{\leftarrow} E[l^n] = \lim_{\leftarrow} \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_l \times \mathbb{Z}_l$$

Definition 3.4. The l-adic Tate Module $T_l(E)$ is defined to be the inverse limit of the l^n-torsion points.

$$T_l(E) := \lim_{\leftarrow} E[l^n]$$

which is isomorphic to $\mathbb{Z}_l \times \mathbb{Z}_l$ as a group. Thus, we have associated elliptic curve E with a representation $\rho_E : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}_l)$.

Associating a Galois representation to a weight 2 normalized eigenform with respect to level N is much more complicated. The idea is to find some geometry object for each Hecke eigenform, and then defines the Tate module of that object similar to the case of elliptic curve.

Definition 3.5. Let Γ be a congruence subgroup. The Jacobian of the corresponding modular curve $X(\Gamma)$ is

$$Jac(X(\Gamma)) = S_2(\Gamma)^* \backslash H_1(X, \mathbb{Z})$$

where $S_2(\Gamma)^*$ is the dual space of weight 2 cusp forms with respect to Γ.

The dimension of $S_2(\Gamma)$ as well as the the number of generators of $H_1(X_1(N), \mathbb{Z})$ are both given by the genus g of $X_1(\Gamma)$. Thus the Jacobian is a $2g$ dimensional torus.

Definition 3.6. The Hecke algebra over $\mathbb{Z}$ is the subalgebra of the endomorphisms of $S_2(\Gamma_1(N))$ generated by the Hecke operators,

$$T_{\mathbb{Z}} = \mathbb{Z}[\{T_n, \langle n \rangle\}]$$

The Hecke algebra acts on f by composing the actions of Hecke operators. The kernel of the eigenvalue map

$$I_f = \{T \in T_{\mathbb{Z}} : Tf = 0\}$$

Theorem 3.7. *There exists an isomorphism*

$$T_{\mathbb{Z}}/I_f \cong \mathcal{O}_f \quad \textit{where } \mathcal{O}_f = \mathbb{Z}[\{a_n(f)\}]$$

Proof. Consider the homomorphism $\lambda_f : T_{\mathbb{Z}} \to \mathbb{C}$ which satisfy that

$$Tf = \lambda_f(T)f$$

Since $T_{\mathbb{Z}}$ is generated by $T_n, \langle n \rangle$, its image $im(\lambda_f)$ is generated by $\lambda_f(T_n) = a_n(f)$ and $\chi(n)$ over $\mathbb{Z}$ (where $f \in S_2(N, \chi)$. The $\chi(n)$ terms are actually redundant. □

The Hecke operator also acts on the Jacobian by composition.

Proposition 3.8. *The number field K_f associated to normalized eigenform f given by $\mathbb{Q}[\{a_n(f)\}]$ is finite dimensional.*

Proof. The Hecke operators T_p acts on $S_2(\Gamma_1(N))^*$ and also descends to $J_1(N) = Jac(X_1(N))$, so it restricts to an endomorphism on $H_1(X_1(N), \mathbb{Z})$, which is a free abelian group. Hence, the eigenvalues of T_p should be algebraic integers as the roots of the characteristic polynomial corresponds to T_p restricted to $H_1(X_1(N), \mathbb{Z})$. Moreover, viewing $T_{\mathbb{Z}}$ has the ring of endomorphism of $H_1(X_1(N), \mathbb{Z})$ shows that $T_{\mathbb{Z}}$ is finitely generated as a $\mathbb{Z}$-module as well. □

Definition 3.9. The (analytic) abelian variety A_f associated to f is

$$A_f := J_1(N)/I_f J_1(N)$$

Proposition 3.10. *The abelian variety A_f is a $2d$ dimensional torus, where $d = [K_f, \mathbb{Q}]$.*

Now, similar to the construction Tate module for elliptic curve, the l^n-torsion points on the abelian variety $A_f[l^n] \cong \mathbb{Z}_{l^n}^{2d}$ forms the l-adic Tate module $T_l(A_f) \cong \mathbb{Z}_l^d$ by taking inverse limit. Meanwhile, the Hecke algebra $T_{\mathbb{Z}}$ also acts on A_f, or really the quotient $T_{\mathbb{Z}}/I_f \cong \mathcal{O}_f$, since I_f action is inert on A_f. This makes $T_l(A_f)$ a $\mathcal{O}_\{$ module.

Proposition 3.11. *$V_l(A_f) = T_l(A_f) \otimes \mathbb{Q}$ is a free module of rank 2 over $K_f \otimes_{\mathbb{Q}} \mathbb{Q}_l$.*

Proof. Since $\mathrm{Ta}_\ell(A_f)$ is the inverse limit of the torsion groups $A_f[\ell^n]$, we need to describe $A_f[\ell^n]$ in a fashion that will help establish the freeness.

As above, let $\mathcal{S}_2 = \mathcal{S}_2(\Gamma_1(N))$ and let $H_1 = \mathrm{H}_1(X_1(N)_{\mathbb{C}}, \mathbb{Z}) \subset \mathcal{S}_2^\wedge$. Consider the quotients $\overline{\mathcal{S}_2^\wedge} = \mathcal{S}_2^\wedge / I_f \mathcal{S}_2^\wedge$ and $\bar{H}_1 = (H_1 + I_f \mathcal{S}_2^\wedge)/I_f \mathcal{S}_2^\wedge$, both $\mathcal{O}_f$ modules. Compute that

$$\begin{aligned} A_f = \mathrm{J}_1(N)/I_f\ \mathrm{J}_1(N) &= (\mathcal{S}_2^\wedge/H_1) \,/\, ((I_f\mathcal{S}_2^\wedge + H_1)/H_1) \\ &\cong \mathcal{S}_2^\wedge / (I_f\mathcal{S}_2^\wedge + H_1) \\ &\cong (\mathcal{S}_2^\wedge/I_f\mathcal{S}_2^\wedge) \,/\, ((H_1 + I_f\mathcal{S}_2^\wedge)/I_f\mathcal{S}_2^\wedge) = \overline{\mathcal{S}_2^\wedge}/\bar{H}_1. \end{aligned}$$

Thus $A_f[\ell^n] \cong \ell^{-n}\bar{H}_1/\bar{H}_1$ for any $n \in \mathbb{Z}^+$. The $\mathcal{O}_f$-linear isomorphisms $\ell^{-n}\bar{H}_1/\bar{H}_1 \longrightarrow \bar{H}_1/\ell^n\bar{H}_1$ induced by multiplication by ℓ^n on $\ell^{-n}\bar{H}_1$ assemble to give an isomorphism of $\mathcal{O}_f \otimes \mathbb{Z}_\ell$-modules,

$$\mathrm{Ta}_\ell(A_f) = \lim_n \{A_f[\ell^n]\} = \lim_n \left\{\ell^{-n}\bar{H}_1/\bar{H}_1\right\} \cong \lim_n \left\{\bar{H}_1/\ell^n\bar{H}_1\right\} \cong \bar{H}_1 \otimes \mathbb{Z}_\ell,$$

where the transition maps in the last inverse limit are the natural projection maps.

The fact that A_f is a complex torus of dimension d and the calculation a moment ago that $A_f \cong \overline{\mathcal{S}_2^\wedge}/\bar{H}_1$ combine to show that the $\mathcal{O}_f$-module $\bar{H}_1 \cong H_1/(H_1 \cap I_f\mathcal{S}_2^\wedge)$

has $\mathbb{Z}$-rank $2d$. Since $\mathbb{K}_f$ is a field, $\bar{H}_1 \otimes \mathbb{Q}$ is a free $\mathbb{K}_{f-}$ module whose $\mathbb{Q}$-rank is $2d$ and whose $\mathbb{K}_f$-rank is therefore 2 . Consequently, $\bar{H}_1 \otimes \mathbb{Q}_\ell = \bar{H}_1 \otimes \mathbb{Q} \otimes \mathbb{Q}\mathbb{Q}_\ell$ is free of rank 2 over $\mathbb{K}_f \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$. So finally,

$$V_\ell\left(A_f\right) = \mathrm{Ta}_\ell\left(A_f\right) \otimes \mathbb{Q} \cong \bar{H}_1 \otimes \mathbb{Z}_\ell \otimes \mathbb{Q} \cong \bar{H}_1 \otimes \mathbb{Q}_\ell$$

□

Lemma 3.12. *Let K_f be a number field. Then*

$$K_f \otimes \mathbb{Q}_l = \prod_{\lambda|l} K_{f,\lambda}$$

where λ are prime ideals in K_f lying over l and $K_{f,\lambda}$ is the local field obtained from completing K_f at λ.

Proof.

$$K_f \otimes \mathbb{Q}_l = K_f \otimes \mathbb{Z}_l = K_f \otimes (\varprojlim \mathbb{Z}/l^n\mathbb{Z}) = \varprojlim (K_f \otimes \mathbb{Z}/l^n\mathbb{Z})$$

$$= \varprojlim (K_f/l^n K_f) = \varprojlim (\prod_{\lambda|l} (K_f/\lambda^n K_f) = \prod \varprojlim K_f/\lambda^n K_f = \prod K_{f,\lambda}$$

. □

Proposition 3.11 and Lemma 3.12 combine to give the desired representation associated to the weight 2 eigenform: The fact that $V_l(A_f)$ is a $G_{\mathbb{Q}}$ module and is free over $K_f \otimes \mathbb{Q}_l$ of rank two gives a Galois representation $\rho_f : G_{\mathbb{Q}} \to GL_2(K_f \otimes \mathbb{Q}_l)$ and the factorization by Lemma 3.12 gives projections $\pi : K_f \otimes \mathbb{Q}_l \to K_{f,\lambda}$. Composing then gives a representation

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to GL_2(K_{f,\lambda})$$

Theorem 3.13 (Modularity Theorem, Galois Representation Version)**.** *Let E be an elliptic curve defined over $\mathbb{Q}$. Then there exists some prime l and a weight two newform f, such that $K_f = \mathbb{Q}$, and that*

$$\rho_{E,l} \sim \rho_{f,l}$$

4. Eichler Shimura Relation

So far, we have settled all the necessary ground to state different versions of the modularity theorems. What we have not proved is the equivalences between them, so that they can be regarded as one single theorem. This would be partially achieved in this section by Eichler-Shimura relation that connects Hecke operators over modular curves in positive characteristics with Frobenius maps on Elliptic curve, which builds bridges between the modularity conjectures. We will define the Hecke operators on modular curves over $\mathbb{C}$ in the first section, following by a brief discussion shifting them to modular curves over $\mathbb{Q}$. This allows us to explore reduction of modular curves into positive characters and to define Hecke operators on the reduced curves.

We have to skipped over a few proofs since they require heavy machineries from advanced algebraic geometry.

4.1. **Hecke Operators on Modular Curves.** The section begins with an observation: the double coset operators defined in Definition 2.7 carry modular forms on Γ_1 to those on Γ_2. Can we imitate the process to obtain a map from $X(\Gamma_1)$ to $X(\Gamma_2)$? Such map should take $\Gamma_1\tau$, a point in $X(\Gamma_1)$, to $\sum\Gamma_1\beta_j\tau$, where β_j are again the coset representatives of $\Gamma_1/\Gamma_1\alpha\Gamma_2$. However, unlike the case of modular forms, addition of points on $X(\Gamma_2)$ does not make sense, so what we actually get is a map from $Div(X(\Gamma_1))$ to $Div(X(\Gamma_2)$.

Explicitly, the Hecke operators on divisor group of modular curve is given by

$$T_p : Div(X_1(N)) \to Div(X_1(N)), \quad \Gamma_1(N)\tau \to \sum_j \Gamma_1(N)\beta_j(\tau)$$

where again, the βs are the coset representative given by $\beta_j = \left[\begin{smallmatrix}1 & j\\ 0 & 1\end{smallmatrix}\right]$ for $0 \le j < p$ and if $p \nmid N$, also includes the term $\beta_\infty = \left[\begin{smallmatrix}m & n\\ N & p\end{smallmatrix}\right]\left[\begin{smallmatrix}1 & p\\ 0 & 1\end{smallmatrix}\right]$.

Furthermore, the isomorphism between the modular curve $Y_1(N)$ and the moduli space of enhanced elliptic curve $S_1(N)$ motivates the following definition of Hecke operators acting on $Div(S_1(N))$:

Definition 4.1. The Hecke operators $T_p : Div(S_1(N)) \to Div(S_1(N))$ sends $[\mathbb{C}/\Lambda_\tau, \frac{1}{N} + \Lambda_\tau]$ to $\sum[\mathbb{C}/\Lambda_{\beta_j\tau}, \frac{1}{N} + \Lambda_{\beta_j\tau}]$. Moreover, we may define the Hecke operators $T_p : Div(S_1(N)) \to Div(S_1(N)$, such that the following diagram commute:

$$\begin{array}{ccc} Div(S_1(N)) & \xrightarrow{T_p} & Div(S_1(N)) \\ \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} \\ Div(X_1(N)) & \xrightarrow{T_p} & Div(X_1(N)) \end{array} \tag{4.2}$$

Remark 4.3. With the perspective that modular forms can be regarded as sections of line bundles over the modular curves, this formulation of Hecke operators is in fact more intrinsic. We will not come to it in this paper.

The hecke operators T_p action on an arbitrary class of complex elliptic curve is given in the following way (which can be proved by combining Definition 4.1 and Theorem 1.10)

$$T_p : [E, P] \to \sum_C [E/C, P + C] \tag{4.4}$$

where C are all order p subgroups of the elliptic curve E such that $C \cap \langle P\rangle = O$.

To summarize, we have three compactible versions of Hecke operator T_p, starting from the double coset $\Gamma_1(N)\left[\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right]\Gamma_1$, we get T_p as a linear operator in $End(M_k(\Gamma_1(N))$. On the other hand, the Hecke operator is defined on divisor group of modular curve $X_1(N)$, which is identified with the mapping on divisor group of the moduli space $S_1(N)$.

Similarily, the diamond operators $\langle d\rangle$ also acts on $Div(X_1(N))$ and $Div(S_1(N))$, given by $\langle d\rangle : Div(X_1(N)) \to Div(X_1(N))$, $\langle d\rangle(\Gamma_1(N)\tau) = \Gamma_1(N)\beta\tau$, $\langle d\rangle : S_1(N) \to S_1(N)$, $\langle d\rangle[\mathbb{C}/\Lambda_\tau] = \mathbb{C}/\Lambda_{\beta\tau}$, where $\beta = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$. The following diagram also commutes by definition.

4.2. **Shifting from complex analytic curves to algebraic curves over $\mathbb{Q}$.** After providing a $\mathbb{Q}$ model of modular curves $X_0(N), X_1(N)$, we shall expect the moduli spaces $S_0(N) \cong Y_0(N), S_1(N) \cong Y_1(N)$ has a corresponding $\mathbb{Q}$ model as

$$\begin{array}{ccc} Div(S_1(N)) & \xrightarrow{\langle d \rangle} & Div(S_1(N)) \\ \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} \\ Div(X_1(N)) & \xrightarrow{\langle d \rangle} & Div(X_1(N)) \end{array}$$

well. Then, we hope to find (4.2) still works in such context after some appropriate modifications.

Definition 4.5. The pairs (E, Q) consists of complex algebraic elliptic curve with Q an N-torsion point. Two such pairs $(E, Q), (E', Q)$ are equivalent if there exists an isomorphism $\phi : E \to E'$ such that $\phi(Q) = Q'$. The *complex algebraic moduli space for* $\Gamma_1(N)$ is the set of equivalence classes

$$S_1(N)_{alg,\mathbb{C}} = \{\text{Enhanced complex algebraic elliptic curves}/\sim\}$$

Similarly, the *algebraic moduli space for* $\Gamma_1(N)$ is the set of equivalence classes

$$S_1(N)_{alg} = \{\text{Enhanced algebraic elliptic curves over } \overline{\mathbb{Q}}/\sim\}$$

Lemma 4.6. *The intersection of the set of equivalence class* $[E, Q]$ *in* $S_1(N)_{alg,\mathbb{C}}$ *and* $S_1(N)_{alg}$ *is an equivalence class in* $S_1(N)$

By Lemma 4.6, $S_1(N)_{alg}$ can be seen as subset of $S_1(N)_{alg,\mathbb{C}}$, which can be identified with $S_1(N)$, the moduli space of complex analytic elliptic curve, via the map

$$[\mathbb{C}/\Lambda_\tau, \frac{1}{N} + \Lambda_{tau}] \to [E_\tau, (\wp(\frac{1}{N}), \wp'(\frac{1}{N}]$$

Also, the N-torsion group is retained from $\mathbb{C} \to \overline{Q}$ (See Theorem 7.1.3). Thus, Hecke operator T_p on the top row of diagram (4.2) restricts to $Div(S_1(N)_{alg})$. The bottom row of diagram (4.2) also restricts to $Div(X_1(N)_{alg})$ once we prove T_p as an endomorphism on $Div(X_1(N))$ is defined over $\mathbb{Q}$. (See P306,[3]. The sides of diagram (4.2) extends the map given in Theorem 1.10,

$$i : S_1(N) \to X_1(N) \;\; [\mathbb{C}/\Lambda_\tau, \frac{1}{N} + \Lambda_\tau] \to \Gamma_1(N)\tau$$

Since we have identified $S_1(N)$ with $S_1(N)_{alg,\mathbb{C}}$ and $X_1(N)$ with $X_1(N)_{alg,\mathbb{C}}$, we obtain a map $i_{alg,\mathbb{C}} : S_1(N)_{alg,\mathbb{C}} \to X_1(N)_{alg,\mathbb{C}}$. This map also restricts to an algebraic map $i_{alg} : S_1(N)_{alg} \to X_1(N)_{alg}$. To see this, consider the following diagram with mapping given by Since the element $[E, Q] \in S_1(N)_{alg,\mathbb{C}}$ belongs to

$$\begin{array}{ccc} S_1(N)_{alg,\mathbb{C}} & \longrightarrow & S_1(N)_{alg,\mathbb{C}} \\ \downarrow{\scriptstyle i_{alg,\mathbb{C}}} & & \downarrow \\ X_1(N)_{alg,\mathbb{C}} & \longrightarrow & X_1(1)_{alg,\mathbb{C}} \end{array}$$

$$\begin{array}{ccc} [E, Q] & \longrightarrow & [E] \\ \downarrow & & \downarrow \\ P & \longrightarrow & j(E) \end{array}$$

$S_1(N)$ when E is defined over $\overline{\mathbb{Q}}$, which means $j(E) \in \overline{\mathbb{Q}}$, i.e. $P \in \overline{\mathbb{Q}}$.

From the discussion above, we obtain the algebraic version of (4.2):

$$\begin{array}{ccc} Div(S_1(N)_{alg}) & \xrightarrow{T_p} & Div(S_1(N)_{alg}) \\ {\scriptstyle i_{alg}}\downarrow & & \downarrow{\scriptstyle i_{alg}} \\ Div(X_1(N)_{alg}) & \xrightarrow[T_p]{} & Div(X_1(N)_{alg}) \end{array} \tag{4.7}$$

The diagram restricts to the degree 0 divisors groups. Since T_p takes f to $\sum f[\beta_j]_k$, it takes principal divisors to principal divisors, which is also true as morphism defined over $\overline{\mathbb{Q}}$, the bottom line induces mapping between the Picard groups. That is, we get

$$\begin{array}{ccc} Div_0(S_1(N)_{alg}) & \xrightarrow{T_p} & Div_0(S_1(N)_{alg}) \\ {\scriptstyle i_{alg}}\downarrow & & \downarrow{\scriptstyle i_{alg}} \\ Pic(X_1(N)_{alg}) & \xrightarrow[T_p]{} & Pic(X_1(N)_{alg}) \end{array} \tag{4.8}$$

Similarly, we have another commutative diagram for Diamond operators $\langle d \rangle$:

$$\begin{array}{ccc} Div_0(S_1(N)_{alg}) & \xrightarrow{\langle d \rangle} & Div_0(S_1(N)_{alg}) \\ {\scriptstyle i_{alg}}\downarrow & & \downarrow{\scriptstyle i_{alg}} \\ Pic(X_1(N)_{alg}) & \xrightarrow[\langle d \rangle]{} & Pic(X_1(N)_{alg}) \end{array} \tag{4.9}$$

4.3. **Reduction of Modular Curve and Hecke Operators.** Starting from this section, we will drop the index alg and denote the modular curve over $\mathbb{Q}$ by $X_1(N)$.

First, we reduce the moduli space $S_1(N)$ at prime p. Let $\mathfrak{p}$ be an maximal ideal over p. Only elliptic curves over $\overline{Q}$ with good reduction at $\mathfrak{p}$ reduce to elliptic curve over $\overline{\mathbb{F}_p}$, so we need to restrict the moduli space a little bit. Also, for purpose will be clear very soon, we further exclude the elliptic curves with j invariant $\overline{j(E)} = 0, 1728$. We define

$$S_1(N)' = \{[E, Q] : E/\overline{Q} \text{ has good reduction at } \mathfrak{p},\, \tilde{j(E)} \neq 0, 1728$$

and similarly

$$\tilde{S}_1(N)' = \{[E, Q], E/\overline{F_p}, j(E) \neq 0, 1728\}$$

We may then define

$$S_1(N)' \to \tilde{S}_1(N)' \quad , [E, Q] \to [\tilde{E}, \tilde{Q}]$$

The reduction map surjects since any Weierstrass equation with coefficients in $\overline{F_p}$ naturally lifts to a Weierstrass equation over $\overline{\mathbb{Z}}$ with non-zero discriminant. The N-torsion point $\tilde{Q}$ also has a lift since the torsion group of elliptic curves over surjects in reduction. (See [3] Proposition 8.4.4)

We now construct a planar curve C that will be useful to the reduction of modular curve. Take any $[E, Q] \in \tilde{S}_1(N)$, and denote $j(E) = j$. Consider the universal elliptic curve $\tilde{E}_j$ defined over $\mathbb{F}_p(j)$,

$$\tilde{E}_j : y^2 + xy = x^3 - (\frac{36}{j - 1728})x - \frac{1}{j - 1728}$$

This curve has discriminant $\frac{j^2}{(j-1728)^3}$ and j-invariant j. Since $j \neq 0, 1728$, it is an elliptic curve over $\overline{\mathbb{F}_p}$ and is isomorphic to E over $\overline{\mathbb{F}_p}$.

Take a point $Q \in \tilde{E}_j$ of order N, and let $\phi_{1,N} \in \mathbb{F}_p(j)[x]$ be the minimal polynomial of the x-coordinate of point Q. Define the field

$$K_1(N) = \mathbb{F}_p(j)[x]/(\phi_{1,N})$$

It is true that $K_1(N) \cap \overline{\mathbb{F}}_p = \mathbb{F}_p$, so $K_1(N)$ is a function field over $\mathbb{F}_p$. Viewing j as a variable, the polynomial $\phi_{1,N} \in \mathbb{F}_p(j)[x]$ defines a planar curve whose points are $(j, x(Q))$.

Reducing the modular curve as an algebraic curve is hard. The way we constructed modular curve as algebraic curve over $\mathbb{Q}$ is implicit, and it is almost impossible to find a general polynomial equation that defines $X_1(N)$ for all N. Thus, we confine ourselves to only give a definition of reduction of curves in general. Then we state the Igusa theorem, which says that our planar curve C is birationally equivalent to $\tilde{X}_1(N)$.

Definition 4.10. Let C be a nonsingular affine algebraic curve over $\mathbb{Q}$, defined by polynomials $\phi_1, ..., \phi_m \in \mathbb{Z}_{(p)}[x_1, ..., x_n]$. Then C has good reduction at p if

(1) The ideal $I = \langle \phi_1, ..., \phi_m \rangle$ of $\mathbb{Z}_{(p)}[x_1, ..., x_n]$ is prime.
(2) The reduced polynomials $\tilde{\phi}_i \in \mathbb{F}_p[x_1, ..., x_n]$ defines a nonsingular affine algebraic curve $\tilde{C}$ over $\mathbb{F}_p$.

Definition 4.11. Let C be a nonsingular projective curve over $\mathbb{Q}$ defined by the homogenization $I \in \mathbb{Z}_{(p)}[x_0, ..., x_n]$ of a prime ideal $I_{(0)} \in \mathbb{Z}_{(p)}[x_1, ..., x_n]$. Then C has good reduction at p if for $i = 1, 2, ..., n$, either the affine curve C_i defined by dehomogenizing I at x_i has good reduction at p or I reduces to all of $\mathbb{F}_p[x_0, ..., x_{i-1}, x_{i+1}, ..., x_n]$.

Theorem 4.12 (Igusa)**.** *The modular curve $X_1(N)$ has good reduction at p if $p \nmid N$. There is an isomorphism of function fields*

$$\mathbb{F}_p(\tilde{X}_1(N)) \to K_1(N)$$

Moreover, reducing the modular curve is compatible with reducing the moduli space in the sense that the following diagram commutes:

$$\begin{array}{ccc} S_1(N)' & \longrightarrow & X_1(N) \\ \downarrow & & \downarrow \\ \tilde{S}_1(N)' & \longrightarrow & \tilde{X}_1(N) \end{array}$$

The top and bottom map is given by $[E, Q] \to (j, x(Q))$ followed by the birational equivalence from the planar model C to $X_1(N)$.

Proof. See [4] □

Remark 4.13. The modular curve $X_0(N)$ may be reduced in the same way by replacing $\phi_{1,N}$ with $\phi_{0,N}$, the minimal polynomial of $\sum_Q x([d]Q)$ (i.e. sum over x-coordinate in the cyclic group $\langle Q \rangle$ attached to elliptic curve E) over $\mathbb{F}_p(j)[x]$. The theorem still holds by replacing $S_1(N)$ by $S_0(N)$ and $X_1(N)$ by $X_0(N)$.

4.4. **Eichler-Shimura Relation.** We start this section by stating a theorem of the reduction of algebraic curves that we need to reduce the Diamond operators $\langle d \rangle$.

Theorem 4.14. *Let C and C' be nonsingular projective algebraic curves over $\mathbb{Q}$ with good reduction at p, and let C' has positive genus. For any morphism $h : C \to C'$, there exists a unique morphism $\tilde{h} : \tilde{C} \to \tilde{C}'$ that commutes with the reduction maps. It satisfies that $deg(\tilde{h}) = deg(h)$. Furthermore, let $h_* : Pic^0(C) \to Pic^0(C')$ and $\tilde{h}_* : Pic^0(\tilde{C}) \to Pic^0(\tilde{C}')$ be the induced pushforward maps, the following diagram commutes:*

$$\begin{array}{ccc} Pic^0(C) & \xrightarrow{h_*} & Pic^0(C') \\ \downarrow & & \downarrow \\ Pic^0(\tilde{C}) & \xrightarrow{\tilde{h}_*} & Pic^0(\tilde{C}') \end{array}$$

Proof. See Theorem 9.5.1 in [7]. □

Now, recall that the Diamond operator $\langle d \rangle$ gives a morphism $\langle d \rangle : X_1(N) \to X_1(N), \quad \Gamma_1(N)\tau \to \Gamma_1(N)\tau \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

Theorem 4.15. *The Hecke operator $\langle d \rangle$ on $X_1(N)$ reduces modulo p and passes to the Picard groups to give a commutative diagram*

$$\begin{array}{ccc} Pic^0(X_1(N)) & \xrightarrow{\langle d \rangle_*} & Pic^0(X_1(N)) \\ \downarrow & & \downarrow \\ Pic^0(\tilde{X}_1(N)) & \xrightarrow{\langle \tilde{d} \rangle_*} & Pic^0(\tilde{X}_1(N)) \end{array}$$

Proof. When the genus of $X_0(N)$ is zero, the Picard group is trivial and there is nothing to prove. Otherwise, apply Theorem 4.14 to the morphism $\langle d \rangle$. □

The reduction of the Hecke operators, on the other hand, cannot be obtained in the same manner, since T_p is not a morphism from $X_1(N)$ to itself. Nevertheless, it can be seen as an endomorphism of $Pic^0(X_1(N))$, which can be identified as an abelian variety.

Theorem 4.16. *There exists an unique operator $\tilde{T}_p$ such that the diagram commutes:*

$$\begin{array}{ccc} Pic^0(X_0(N)) & \xrightarrow{T_p} & Pic^0(X_0(N)) \\ \downarrow & & \downarrow \\ Pic^0(\tilde{X}_0(N)) & \xrightarrow{\widetilde{T_p}} & Pic^0(\tilde{X}_0(N)) \end{array}$$

Proof. See Theorem 9.5.1 in [7]. □

The brilliant idea is that we may compute the reduction by Frobenius morphism. To prove this relation, we study the Hecke operators on modular curves while identifying them as moduli spaces.

We may define a map $f : Pic^0(\tilde{S}_1(N)') \to Pic^0(\tilde{S}_1(N)')$ by

$$f([\widetilde{E}, \widetilde{Q}]) = \sum_{\beta_i} [\widetilde{E/C}, \widetilde{Q + C}]$$

By equation (4.4), the following diagram commutes:

$$\begin{array}{ccc} Div^0(S_1(N)') & \xrightarrow{T_p} & Div^0(S_1(N)') \\ \downarrow & & \downarrow \\ Div^0(\tilde{S}_1(N)') & \xrightarrow{f} & Div^0(\tilde{S}_1(N)') \end{array}$$

The relation between Hecke operators and Frobenius maps finally becomes explicit by the following proposition that express f in Frobenius maps.

Proposition 4.17. *Let E be an elliptic curve over $\overline{\mathbb{Q}}$ with ordinary reduction at $\mathfrak{p}$ and let $Q \in E$ be a point of order N. Let C_0 be the kernel of the reduction map $E[p] \to \tilde{E}[p]$. Then C_0 is an order p subgroup of E. For any order p subgroup C of E,*

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

If E is an elliptic curve over $\mathbb{Q}$ with supersingular reduction at $\mathfrak{p}$, then $[\tilde{E}^{\sigma_p}, \tilde{Q}^{\sigma_p}] = [\tilde{E}^{\sigma_p^{-1}}, [p]\tilde{Q}^{\sigma_p^{-1}}]$ and the proposition still holds.

Proof. First suppose $C = C_0$. Let $E' = E/C$, and $\psi : E' \to E$ the dual of the quotient isogeny. Consider the commutative diagram

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

Since E is over field of characteristic 0, all extensions are separable, and we have $deg_s(\psi) = deg(\psi) = deg(\phi) = p$. The kernel of ψ should contain p points. The kernel is contained in $E'[p]$ since $[p]Q = \phi \circ \psi Q = 0$. Thus $\psi(E'[p])$ contains $\#(E'[p])/p = p$ points. We claim that $\psi(E'[p])$ coincide with $C_0 = ker(\phi)$. Since $\phi \circ \psi(E'[p]) = [p]\psi(E'[p]) = 0$, $\psi(E'[p]) \subset ker(\phi)$, and they both contain p points.

Therefore $\tilde{\phi} \circ r = r \circ \phi = 0$, and since the reduction map on torsion points is surjective, $\tilde{\phi} = 0$. This means kernel of $\tilde{\phi}$ contains all of $\tilde{E}[p]$, which has p points since E has ordinary reduction. We get $p \leq deg_s(\tilde{\psi}) \leq deg(\tilde{\psi})$, which by Theorem 4.14 equals to $deg(\psi) = p$. Since $deg_s(\tilde{\psi}) \cdot deg_s(\tilde{\phi}) = deg_s(\psi \circ \phi) = deg_s([\tilde{p}]) = \#ker([\tilde{p}]) = p$, we conclude that $deg_s(\tilde{\phi}) = 1$, i.e. $\tilde{\phi} : \tilde{E} \to \widetilde{E'}$ is a purely inseparable extension of degree p. Thus, there exists a canonical isomorphism between $\widetilde{E'}$ and $\tilde{E}^{\sigma_p}$. The morphism sends Q to Q^{σ_p}.

If $C \neq C_0$, consider the diagram

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

The image $\phi(C_0)$ is order p since C_0 and $C = ker(\phi)$ intersect only at O. It is thus all of $\phi(E[p])$. Composition of $r \circ \phi = \tilde{\phi} \circ r$ is 0 when restricted to C_0 since $C_0 := ker(r)$. Therefore $r \circ \phi$ is 0 on all of $E[p]$ since $\phi(E[p]) = \phi(C_0)$. Again by surjectivity of reduction, we get $\tilde{\phi} = 0$, which similarly implies that $\tilde{\psi}$ is purely

inseparable extension of degree p. This gives a canonical isomorphism between $\tilde{E}$ and $\tilde{E}'^{\sigma_p}$. Thus $\tilde{E}' = \tilde{E}^{\sigma_p^{-1}}$. Since ψ takes $Q' = \phi(Q)$ to $\psi \circ \phi(Q) = [p]Q$, $\tilde{\psi} = i \circ \sigma_p$ takes $\widetilde{Q'}$ to $[p]\tilde{Q}$.

We omit the case of supersingular reduction here. See [3]. □

Proposition 4.18. *The following diagram commutes:*

$$\begin{array}{ccc} Div^0(\tilde{S}_1(N)') & \xrightarrow{\sigma_p + p\widetilde{\langle p\rangle}\sigma_p^{-1}} & Div^0(\tilde{S}_1(N)') \\ \downarrow & & \downarrow \\ Pic^0(\tilde{X}_1(N)) & \xrightarrow[\sigma_{p,*} + \widetilde{\langle p\rangle}\sigma_p^*]{} & Pic^0(\tilde{X}_1(N)) \end{array}$$

Proof. It suffices to verify commutativeness term by term. See [3] Chapter 8. □

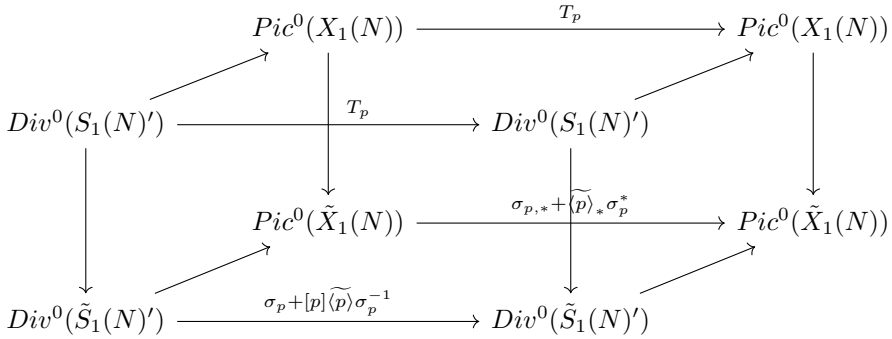
Theorem 4.19 (Eichler-Shimura Relation)**.** *Let $p \nmid N$. The following diagram commutes:*

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

In particular, $\widetilde{\langle p\rangle}_*$ acts trivially on $X_0(N)$, thus

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

Proof. Consider this cubic diagram that commutes except possibly for the back of the cube.



Consider

$$Div^0(S_1(N)') \longrightarrow Pic^0(X_1(N)) \longrightarrow Pic^0(\tilde{X}_1(N)) \xrightarrow{\sigma_{p,*} + \widetilde{\langle p\rangle}_*\sigma_p^*} Pic^0(\tilde{X}_1(N))$$

By applying the commutativeness of the face on the left, bottom, front, right, top faces respectively, this map equals to

$$Div^0(S_1(N)') \longrightarrow Pic^0(X_1(N)) \xrightarrow{T_p} Pic^0(X_1(N)) \longrightarrow Pic^0(\tilde{X}_1(N))$$

However, this does not show the commutativeness of the back square as the mapping $Div^0(S_1(N)') \to Pic^0 X_1(N)$ is not surjective. We further require that there exists a unique map $\tilde{T}_p : Pic^0(\tilde{X}_1(N)) \to Pic^0(\tilde{X}_1(N))$ such that the diagram commutes. This is where we need Theorem 4.16. Now

$$Div^0(S_1(N)') \longrightarrow Pic^0(X_1(N)) \longrightarrow Pic^0(\tilde{X}_1(N)) \xrightarrow{\tilde{T}_p} Pic^0(\tilde{X}_1(N))$$

will agree with

$$Div^0(S_1(N)') \longrightarrow Pic^0(X_1(N)) \xrightarrow{T_p} Pic^0(X_1(N)) \longrightarrow Pic^0(\tilde{X}_1(N))$$

back the commutativeness of the back square, which equals to

$$Div^0(S_1(N)') \longrightarrow Pic^0(X_1(N)) \longrightarrow Pic^0(\tilde{X}_1(N)) \xrightarrow{\sigma_{p,*}+\widetilde{\langle p\rangle}_*\sigma_p^*} Pic^0(\tilde{X}_1(N))$$

The composition $Div^0(S_1(N)') \to Pic^0(X_1(N)) \to Pic^0(\tilde{X}_1(N))$ is surjective. This is true since we may skip finitely many points and still get surjectivity in reduction of algebraic curves. See Theorem 7.3.1 in [3] for a complete discussion. The surjectivity together with the uniqueness of reduction map by Theorem 4.16 implies that $\tilde{T}_p = \sigma_{p,*} + \widetilde{\langle p\rangle}_*\sigma_p^*$. □

5. Equivalences between Modularity Theorems

Theorem 5.1. *Let E be an elliptic curve over $\mathbb{Q}$. The following statements are equivalent:*

1. *There exists some N' and a nonconstant morphism from $X_0(N')$ and E as complex Riemann surface.*
2. *There exists some N and a nonconstant morphism from $X_0(N)$ and E as algebraic variety over $\mathbb{Q}$.*
3. *There exists a weight two newform of level N such that $L(f,s) = L(E,s)$.*
4. *There exists a weight two newform of level N such that there exists a nonconstant morphism from the abelian variety A_f to E.*
5. *There exists a weight two newform of level N and some prime l such that $\rho_{E,l} \sim \rho_{f,l}$.*
6. *There exists a weight two newform of level N such that $\rho_{E,l} \sim \rho_{f,l}$ for all prime l.*

If any of these conditions holds, we say the elliptic curve E is modular.

Proof. (Partial, Sketch) The sketch below skips some technical details without addressing them. (1) $\iff$ (2): See [6].

(2) $\to$ (3): Consider the diagram

$$\begin{array}{ccccc}
Pic^0(X_0(N)) & \xrightarrow{T_p - a_p(E)} & Pic^0(X_0(N)) & \xrightarrow{\alpha_*} & Pic^0(E) \\
\downarrow{\scriptstyle r} & & \downarrow{\scriptstyle r} & & \downarrow{\scriptstyle r} \\
Pic^0(\tilde{X}_0(N)) & \xrightarrow{\sigma_p^* + \sigma_{p,*} - a_p(E)} & Pic^0(\tilde{X}_0(N)) & \xrightarrow{a_*} & Pic^0(\tilde{E}) \\
\downarrow{\scriptstyle 1} & & & & \downarrow \\
Pic^0(\tilde{X}_0(N)) & \xrightarrow{\tilde{\alpha}_*} & Pic^0(\tilde{E}) & \xrightarrow{\sigma_p^* + \sigma_{p,*} - a_p(E)} & Pic^0(\tilde{E})
\end{array}$$

Which is commutative by Eichler-Shimura relation Theorem 4.19, and that the frobenius morphism $\sigma_p : E \to E$ commutes with $\alpha : X_0(N) \to E$. The bottom row is 0, thus the top row is also zero. A morphism from $X_0(N)$ to E give rise to a morphism from $J_0(N)$ to E. Now $J_0(N)$ is a direct sum of A_f (see [3] Chapter 6), and since morphism between abelian variety is either surjective or constant, this gives rise to a morphism $A_f \to E$. The operator $T_p - a_p(E) = 0$ evaluates as $a_p(f) - a_p(E)$ on A_f, which is proves $a_p(E) = a_p(f)$. Proving the levels agree and the other directions here require much more machinary. See [9] for a complete and rigorous proof.

To prove that (5), (6) $\iff$ (3), we need a result on the characteristic polynomial of the Galois representations. □

Theorem 5.2. *Let l be a prime, $N \in \mathbb{Z}^+$. The Galois representation $\rho_{X_1(N),l}$ is unramified at $p \not| lN$. The linear map $\rho_{X_1(N),l}(Frob_{\mathfrak{p}})$ satisfies the polynomial equation*

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

Similarly, the Galois representation $\rho_{A_f,\lambda}$ is unramfied and the linear map $\rho_{A_f,\lambda}(Frob_{\mathfrak{p}})$ satisfies the polynomial equation

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

Proof. (Sketch) To check the Galois representation of Tate module, we first check on each l^n torsion. Since the Frobenius morphism is purely inseparable, $\sigma_{p,*}$ send $[P]$ to $[P^{\sigma_p}]$ and σ_p^* send $[P]$ to $p[P^{\sigma_p^{-1}}]$. Thus $\sigma_{p,*} = \rho_{X_1(N),l}(Frob_{\mathfrak{p}})$, $\sigma_p^* = p\rho_{X_1(N),l}(Frob_{\mathfrak{p}})^{-1}$ as linear maps in $Aut(Pic^0(\tilde{X}_1(N))[l^n])$. The Eichler Shimura relation from Theorem 4.19 restricted to l^n torsion points gives the commutative diagram

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

The vertical rows are in fact isomorphisms. (See[3] Chapter 9). This shows $T_p = \rho_{X_1(N),l}(Frob_{\mathfrak{p}}) + p\langle p\rangle\rho_{X_1(N),l}(Frob_{\mathfrak{p}})^{-1}$. Since this is true for all n and is compatible with the natural inclusions between l^n torsion points, the result passes to Tate module, and the characteristic polynomial follows. To show the A_f version, since T_p act as $a_p(f)$ on $A_f := J_1(N)/I_f J_1(N)$, and the mapping $Pic^0(X_1(N))[l^n] \to A_f[l^n]$ is stable under the Galois group action.

□

continued. Similar to the proof in Theorem 5.2, Proposition 2.27 implies that the Galois representation associated to elliptic curve E satisfies that

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

where $x = \rho_{E,l}(Frob_p)$. Now, if the two Galois representations are similar, then the two characteristic polynomial agrees, which implies $a_p(E) = a_p(f)$. On the other hand, given that there exists newform f such that $a_p(f) = a_p(E)$, consider the Galois representation $\rho_{f,l}$ and $\rho_{E,l}$ associated to the abelian variety A_f and elliptic curve E respectively, for any prime l. Theorem 5.2 shows that the characteristic polynomial of $\rho_{E,l}(Frob_p), \rho_{A_f,l}(Frob_p)$ agrees at all unramified primes p, which is

the case at all but finitely many p. Such Frobenius element is dense in $G_{\mathbb{Q}}$ by a weak form of Chebotarev density theorem. Since both trace and determinant are continuous function, this implies that the two Galois representations are equivalent. Thus (5) $\to$ (3) $\to$ (6) $\to$ (5), and we are done. □

In the end, we are able to state the modularity theorem in a concise and elegant way.

Theorem 5.3 (The Modularity Theorem)**.** *All elliptic curves over* $\mathbb{Q}$ *are modular.*

Acknowledgments

It is my pleasure to thank my mentor, Wei Yao, for suggesting and guiding me through this interesting topic. I am also indebted to Professor Matthew Emerton for inspiring me with invaluable concrete examples. Special thanks to my friend Michael Barz, who has always encouraged me and provided valuable guidance when I am writing the paper. Lastly, I extend my thanks to Professor Peter May for his dedication on organizing such an amazing REU program.

References

[1] T. Weston. The Modular Curves $X_0(11)$ And $X_1(11)$.
[2] J.H. Silverman. The Arithmetic of Elliptic Curve. Springer.
[3] Fred Diamond, Jerry Shurman. A First Course in Modular Forms. Springer
[4] J. Igusa. Kroneckerian model of fields of elliptic modular functions. Amer. J. Math., 81:561–577, 1959
[5] Weston. The Modular Curves $X_0(11)$ And $X_1(11)$.
[6] B. Mazur, Number Theory as Gadfly, MAA
[7] Siegfried Bosch, Werner Lutkebohmert, and Michel Raynaud. N´eron Models, volume 21 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3). Springer-Verlag, 1990
[8] Cornell, G., Stevens, G. and Silverman, J. (2000) Modular forms and Fermat's last theorem. New York: Springer.
[9] H. Carayol. Sur les repr´esentations -adiques associ´ees aux formes modulaires de Hilbert. Ann. Sci. E. N. S., 19:409–468, 1986