The Modularity Theorem and the Modular Approach

Abstract

The goal of the essay is to show how the modularity theorem implies Fermat's Last Theorem (FLT). We begin the essay by discussing about elliptic curves, modular forms and their Galois representations. Then we state the modularity theorem and Ribet's level lowering theorem. Next we study the properties of the Frey curve and its representations, and apply those theorems to deduce FLT. Finally we make a short introduction on a general version of modularity theorem.

1 Introduction

Fermat's last theorem (FLT) is a famous theorem in number theory, which is stated below.

Theorem 1.1 (**Fermat's Last Theorem**)**.** *For any integer $n \geq 3$, the Fermat's equation $x^n + y^n = z^n$ has no nontrivial solution, i.e., there is no integer triple (x, y, z) with $xyz \neq 0$ satisfying $x^n + y^n = z^n$.*

In history, Euler and Fermat proved the cases for $n = 3, 4$ respectively. Many mathematicians proved it for some special cases, but they failed for the general case. In last century, some mathematicians realized that the modularity theorem implies FLT. And in 1995, Wiles proved the modularity theorem for semistable elliptic curves, which is sufficient for deducing FLT.

The goal of the essay is to explore how the modularity theorem implies FLT. At the beginning, we first introduce some basic knowledge of elliptic curves, modular forms and Galois representations in section 2, which provide the ingredients. Then we explain how to obtain Galois representations from elliptic curves and modular forms respectively in section 3. After that, we state without proof the modularity theorem 4.2 and Ribet's level lowering theorem 4.4 in section 4. The modularity theorem states that a Galois representation associated to an elliptic curve is also associated to some modular form, and the Ribet's level lowering theorem roughly tells us that if a Galois representation associated to a modular form is of some level and satisfies some conditions, then the level can be lowered. Next we define a specific elliptic curve called Frey curve and study its properties carefully in section 5, which is the most important part of the argument. Then we deduce FLT from all of the above in section 6. Finally, we give a general version of modularity theorem 7.3 over a real quadratic field and its applications in section 7. The appendix A includes some background knowledge of the Picard group and the Jacobian.

Let us briefly sketch how to deduce FLT. We adopt the strategy of proof of contradiction. Suppose that we have a nontrivial solution of Fermat's equation. Then we can attach the Frey curve to it. By modularity theorem, the representation associated to the Frey curve comes from a modular form of some level. And by Ribet's level lowering theorem and the properties of representations of the Frey curve, the level can be lowered to precisely 2. But a theory of modular forms tells us that the level 2 is impossible, which leads us to a contradiction.

I have used [4][6] as main sources. Section 2.1 is based on [14, Chapter III]. Section 2.2 is based on [6, Chapter 1,5]. Section 2.3 is mainly based on [4, Chapter I][6, Chapter 9], where the definition of group scheme is taken from [1]. Section 3 is based on [6, Chapter 9]. Section 4 is a composite from both sources [4, Chapter I][6, Chapter 9] where the original papers of modularity theorem and Ribet's theorem can be seen in [15][2][11]. Section 5.1 and 5.2 are based on [9, Exercise 5.6.5] where I provide the solution of the exercise on my own. The statements of main theorems in section 5.3 are taken from [4, Chapter I] whereas some lemmas and the proofs are based on various references [5][10][12][7][13] with my substantial complements and modifications. Section 6 is based on [4, Chapter I]. Section 7 is mainly based on [7] where some background and applications can be seen in [3][8]. Appendix A is based on [6, Chapter 5-7].

2 Elliptic Curves, Modular Forms and Galois Representations

2.1 Introduction of Elliptic Curves

We begin by a brief introduction of elliptic curves from [14].

Definition 2.1. [14, p.59] Let K be a field. An elliptic curve E over K (denoted by E/K) is a smooth projective curve of genus 1 defined over K, with a specified K-rational point O_E.

For an elliptic curve E/K, there is an abelian group $E(K)$ whose underlying set is the K-point of the curve, the identity is exactly O_E, and the addition is defined by tangent process, see [14, p.51]. We call the addition of points the group law of the elliptic curves.

Remark. By saying a point P on the elliptic curve E/K, we actually mean P is a $\overline{K}$-point, i.e., $P \in E(\overline{K})$. It can be shown that an algebraic morphism preserving the identities commutes with the group law [14, p.71]. So an algebraic morphism $E/K \to E'/K'$ sending O_E to $O_{E'}$ can be viewed as a group homomorphism $E(\overline{K}) \to E'(\overline{K'})$.

Lemma 2.2. *[14, p.59] Every elliptic curve E/K is isomorphic over K to a curve in Weierstrass form*

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3,$$

via an isomorphism taking O_E to $(0:1:0)$. Furthermore, if $\operatorname{char}(K) \neq 2, 3$, then it can be put into the short Weierstrass form $y^2 = x^3 + ax + b$ in affine coordinates.

Definition 2.3. [14, p.42] Let K be a field. For a Weierstrass equation with coefficients $a_1, \cdots, a_6 \in K$, we define the following quantities:

$$\begin{aligned}
&b_2 = a_1^2 + 4a_2, && b_4 = 2a_4 + a_1a_3,\\
&b_6 = a_3^2 + 4a_6, && b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2,\\
&c_4 = b_2^2 - 24b_4, && c_6 = -b_2^3 + 36b_2b_4 - 216b_6,\\
&\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \text{ (called discriminant).}
\end{aligned}$$

Remark. For an elliptic curve, there exists at least one Weierstrass equation corresponding to it by lemma 2.2, which is not unique. So the quantities $b_2, b_4, \Delta, c_4, c_6$ depend on the choice of the Weierstrass equations. Conversely, a Weierstrass equation uniquely determines a cubic curve. But the curve may not be an elliptic curve because it is not necessarily smooth. It can be shown that the curve is smooth (i.e., an elliptic curve) if and only if the discriminant $\Delta \neq 0$ [14, p.45].

Definition 2.4. [14, p.186] Let K be a field that is complete with discrete valuation v. A Weierstrass equation for E/K with coefficients $a_1, \cdots, a_6 \in K$ is integral if $a_1, \cdots, a_6 \in \mathcal{O}_K$, and furthermore is minimal if $v(\Delta)$ is minimal among all integral Weierstrass equations for E.

Definition 2.5. [14, p.196] Let K be a number field, $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$, $k := \mathcal{O}_{K_\mathfrak{p}}/\mathfrak{p}\mathcal{O}_{K_\mathfrak{p}}$ be the residue field, and E/K be an elliptic curve given by a Weierstrass equation with coefficients $a_1, \cdots, a_6 \in K$ that is minimal over $K_\mathfrak{p}$. Denote the image of $a_1, \cdots, a_6, c_4$ under the natural projection $\mathcal{O}_{K_\mathfrak{p}} \twoheadrightarrow k$ by $\tilde{a}_1, \cdots, \tilde{a}_6, \tilde{c}_4 \in k$. Then $\tilde{a}_1, \cdots, \tilde{a}_6$ determines a Weierstrass equation and hence a cubic curve $\tilde{E}/k$. We say that E/K has a

- good reduction at $\mathfrak{p}$ if $\tilde{E}/k$ is smooth (i.e., an elliptic curve);
- multiplicative reduction at $\mathfrak{p}$ if $\tilde{E}/k$ is singular and $\tilde{c}_4 \neq 0$;
- additive reduction at $\mathfrak{p}$ if $\tilde{E}/k$ is singular and $\tilde{c}_4 = 0$.

Now we can define the semistability and conductor.

Definition 2.6. [14, p.443] Let K be a number field. An elliptic curve E/K is semistable if it has no additive reduction.

Definition 2.7. [14, p.450] For an elliptic curve $E/\mathbb{Q}$, we define

$$f_p := \begin{cases} 0 & \text{if } E \text{ has a good reduction at } p; \\ 1 & \text{if } E \text{ has a multiplicative reduction at } p; \\ 2 & \text{otherwise, i.e., } E \text{ has an additive reduction at } p. \end{cases}$$

Then we define the conductor of E to be $N := \prod_{p \text{ prime}} p^{f_p}$.

For the last part in this section, let us talk about the torsion points.

Definition 2.8. [14, p.67] Let K be a field and E/K be an elliptic curve. For a positive integer n, we define $[n] : E \to E$ by sending P to $nP = P + \cdots + P$. The kernel of $[n]$ is denoted by $E[n] \subseteq E(\overline{K})$, whose elements are called n-torsion points.

Lemma 2.9. *[14, p.86] If* $\operatorname{char}(K) = 0$*, then* $E[n] \simeq (\mathbb{Z}/n\mathbb{Z})^2$*.*

2.2 Introduction of Modular Forms

Now we give a brief introduction of modular forms from [6]. First we define the modular forms and cuspidal forms. Before doing this, we need a notion of congruence groups and some notations.

Definition 2.10. [6, p.13] A subgroup $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ is called a congruence subgroup if $\Gamma(N) \leq \Gamma$ for some $N \in \mathbb{Z}^+$ where $\Gamma(N)$ is the kernel of the reduction map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$.

In particular, $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. We often denote

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : c \equiv 0 \pmod{N} \right\};$$
$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a \equiv d \equiv 1 \pmod{N} \right\}.$$

Let $\mathbb{H} \subseteq \mathbb{C}$ be the upper half plane. For a function $f : \mathbb{H} \to \mathbb{C}$, $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})^+$, $k \in \mathbb{Z}$ and $\tau \in \mathbb{H}$, we denote

$$g \cdot \tau = \frac{a\tau + b}{c\tau + d},$$
$$f|_k[g](\tau) = f(g \cdot \tau) \det(g)^{k-1}(c\tau + d)^{-k}.$$

Now we can define the modular forms.

Definition 2.11. [6, p.14,17] Let $k \in \mathbb{Z}$ and $\Gamma \leq \Gamma(1)$ be a congruence subgroup. A function $f : \mathbb{H} \to \mathbb{C}$ is a weakly modular form of weight k and level Γ if it is meromorphic in $\mathbb{H}$ such that $f|_k[\gamma] = f$ for all $\gamma \in \Gamma$, and furthermore is a modular form if in addition it is holomorphic in $\mathbb{H}$ and $f|_k[\alpha](\tau)$ is bounded as $\mathrm{Im}(\tau) \to \infty$ for each $\alpha \in \Gamma(1)$.

To define the cuspidal forms, we need the Fourier expansion.

Lemma 2.12. *[6, p.16] Let* $\Gamma \leq \Gamma(1)$ *be a congruence subgroup, then*

$$\left\{ t \in \mathbb{Z}^+ : \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in \Gamma \right\} \neq \emptyset.$$

Denote the minimum by $h \in \mathbb{Z}^+$*, and denote* $q_h(\tau) := e^{2\pi i \tau}/h$ *for* $\tau \in \mathbb{H}$*. Let* $f : \mathbb{H} \to \mathbb{C}$ *be a weakly modular form of weight* k *and level* Γ*. Then* f *has a Fourier expansion*

$$f(\tau) = \sum_{n=0}^{\infty} a_n q_h(\tau)^n \quad \textit{for some } a_n \in \mathbb{C}.$$

Definition 2.13. [6, p.17] A modular form f of weight k and level Γ is called a cuspidal form if the constant term a_0 in the Fourier expansion of $f|_k[\alpha]$ vanishes for all $\alpha \in \Gamma(1)$.

We denote the $\mathbb{C}$-vector space of the modular forms of weight k and level Γ by $M_k(\Gamma)$, and that of the cuspidal forms by $S_k(\Gamma)$.

Next we define the subspace of oldforms and newforms. Before doing it, we introduce the modular curves and Petersson inner product.

Definition 2.14. [6, p.38,58] Let $\Gamma \leq \Gamma(1)$ be a congruence subgroup. Define the modular curves $Y(\Gamma) := \Gamma\backslash\mathbb{H} = \{\Gamma\tau : \tau \in \mathbb{H}\}$ and $X(\Gamma) := \Gamma\backslash\mathbb{H}^* = \{\Gamma\tau : \tau \in \mathbb{H}^*\}$ where $\mathbb{H}^* := \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$. Also denote $X(N) := X(\Gamma(N))$, $X_0(N) := X(\Gamma_0(N))$, and $X_1(N) := X(\Gamma_1(N))$.

Definition 2.15. [6, p.182] Let $\Gamma \leq \Gamma(1)$ be a congruence subgroup. The Petersson inner product $\langle , \rangle_\Gamma : S_k(\Gamma) \times S_k(\Gamma) \to \mathbb{C}$ is defined by

$$\langle f, g\rangle_\Gamma = \frac{1}{[\overline{\Gamma(1)} : \overline{\Gamma}]} \int_{X(\Gamma)} f(x+iy)\overline{g(x+iy)}y^k \frac{dxdy}{y^2}$$

where $\overline{\Gamma} = \Gamma / \{\pm 1\}$ and the integral over $X(\Gamma)$ means over the fundamental domain.

Given $k \in \mathbb{Z}$ and a divisor d of $N \in \mathbb{Z}^+$, denote $\alpha_d = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix}$ and

$$i_{d,k} : S_k(\Gamma_1(Nd^{-1})) \times S_k(\Gamma_1(Nd^{-1})) \to S_k(\Gamma_1(N)), \quad (f, g) \mapsto f + g|_k[\alpha_d].$$

Definition 2.16. [6, p.187-188] Let $N \in \mathbb{Z}^+$. The subspace of oldforms at level N is

$$S_k(\Gamma_1(N))^{\text{old}} := \sum_{p|N, p \text{ prime}} i_{p.k}(S_k(\Gamma_1(Np^{-1})) \times S_k(\Gamma_1(Np^{-1}))).$$

The subspace of newforms at level N is the orthogonal complement of $S_k(\Gamma_1(N))^{\text{old}}$ with respect to the Petersson inner product, i.e., $S_k(\Gamma_1(N))^{\text{new}} = (S_k(\Gamma_1(N))^{\text{old}})^\perp$.

Finally, we define the double coset operator, the diamond operator, the Hecke operator, and the notion of the newform.

Definition 2.17. [6, p.165] Let $k \in \mathbb{Z}$ and $\Gamma_1, \Gamma_2 \leq \Gamma(1)$ be congruence subgroups and $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$. We define the weight-$k$ double coset $\Gamma_1\alpha\Gamma_2$ operator $|_k[\Gamma_1\alpha\Gamma_2] : M_k(\Gamma_1) \to M_k(\Gamma_2)$ by $f|_k[\Gamma_1\alpha\Gamma_2] = \sum_j f|_k[\beta_j]$ where β_j are orbit representations such that $\Gamma_1\alpha\Gamma_2 = \sqcup_j \Gamma_1\beta_j$.

Definition 2.18. [6, p.168,169,178] Let $k \in \mathbb{Z}$ and $N \in \mathbb{Z}^+$. For $\delta \in \mathbb{Z}^+$ coprime to N, define the diamond operator $\langle\delta\rangle = |_k[\Gamma_1(N)\alpha\Gamma_1(N)] : M_k(\Gamma_1(N)) \to M_k(\Gamma_1(N))$ for any $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ with $d \equiv \delta \pmod{N}$. For $\delta \in \mathbb{Z}^+$ that is not coprime to N, define $\langle\delta\rangle = 0$.

Definition 2.19. [6, p.169,178] Let $k \in \mathbb{Z}$ and $N \in \mathbb{Z}^+$. For a prime $p \in \mathbb{Z}$, define the Hecke operator $T_p = |_k[\Gamma_1(N)\alpha\Gamma_1(N)] : M_k(\Gamma_1(N)) \to M_k(\Gamma_1(N))$ where $\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$. Also define

$$\begin{aligned} T_1 &= \mathrm{id}_{M_k(\Gamma_1(N))}; \\ T_{p^r} &= T_p T_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}} \text{ for } r \geq 2. \end{aligned}$$

Then define T_n multiplicatively for general $n \in \mathbb{Z}^+$.

Definition 2.20. [6, p.195] A nonzero modular form $f \in M_k(\Gamma_1(N))$ that is an eigenvector for $\langle n\rangle$ and T_n for all $n \in \mathbb{Z}^+$ is called an eigenform. The eigenform $f(\tau) = \sum_{n=0}^\infty a_n q^n$ is called normalized if $a_1 = 1$. A normalized eigenform in $S_k(\Gamma_1(N))$ is called a newform.

These definitions are well-defined (i.e., independent from the choices of relevant variables), and $S_k(\Gamma_1(N))$ is stable under the Hecke operator and diamond operator.

2.3 Introduction of Galois Representations

For a field K, we define the absolute Galois group of K to be $G_K := \mathrm{Gal}(\overline{K}/K)$, the automorphism group of $\overline{K}$ fixing K. For a prime $p \in \mathbb{Z}$, we write $G_p := \mathrm{Gal}(\overline{\mathbb{Q}_p}/\mathbb{Q}_p)$ and $G_\infty := \mathrm{Gal}(\mathbb{C}/\mathbb{R})$. Now we endow a topology on G_K to make it a topological group.

Definition 2.21. [6, p.377] For each $\sigma \in G_K$, we define $U_\sigma(F) := \sigma \cdot \mathrm{Gal}(\overline{K}/F)$ for a finite Galois extension F/K, and let $\mathcal{N}_\sigma := \{U_\sigma(F) : F/K \text{ is a finite Glois extension}\}$ be an open neighborhood basis of σ. This determines a topology on G_K, called the Krull topology.

By infinite Galois theory, for a finite Galois extension F/K, we have the restriction map $G_K \twoheadrightarrow \mathrm{Gal}(F/K)$ with kernel $\mathrm{Gal}(\overline{K}/F)$. So $U_\sigma(F) = \sigma \cdot \ker(G_K \twoheadrightarrow \mathrm{Gal}(F/K))$, which gives an alternative expression for $U_\sigma(F)$.

With the notations above, now we can define the Galois representation as the following:

Definition 2.22. Let d be a positive integer, K be a field and L be a topological field. A d-dimensional galois representation of G_K on L is a continuous group homomorphism $\rho : G_K \to \mathrm{GL}_d(L)$. Furthermore, the representation is called

- global if K is a global field (we often consider the case $K = \mathbb{Q}$);
- local if K is a local field (we often consider the case $K = \mathbb{Q}_l$ for some prime $l \in \mathbb{Z}$);
- p-adic if L is a finite extension of $\mathbb{Q}_p$ where $p \in \mathbb{Z}$ is a prime.
- mod p if L is a finite extension of $\mathbb{F}_p$ where $p \in \mathbb{Z}$ is a prime.

Next we discuss about the equivalence of representations.

Definition 2.23. [6, p.378-379] Let G be a group, R be a ring. We say that two d-dimensional group representations $\rho : G \to \mathrm{GL}_d(R)$ and $\rho' : G \to \mathrm{GL}_d(R)$ are equivalent (denoted by $\rho \sim \rho'$) if there is $m \in \mathrm{GL}_d(R)$ such that $\rho'(\sigma) = m^{-1}\rho(\sigma)m$ for all $m \in G$. And we say that two d-dimensional Galois representations are equivalent if they are equivalent as group representations.

Lemma 2.24. *[6, p.381] Let $p \in \mathbb{Z}$ be a prime and L be a finite extension of $\mathbb{Q}_p$ with the ring of integers $\mathcal{O}_L$. A p-adic Galois representation $\rho : G_K \to \mathrm{GL}_d(L)$ is equivalent to some representation $\rho' : G_K \to \mathrm{GL}_d(\mathcal{O}_L)$.*

So for a p-adic Galois representation $\rho : G_K \to \mathrm{GL}_d(L)$, we may assume that the image is contained in $\mathrm{GL}_d(\mathcal{O}_L)$. This gives us the definition of the residual representation.

Definition 2.25. [4, p.4] Let d be a positive integer, $p \in \mathbb{Z}$ be a prime and L be a finite extension of $\mathbb{Q}_p$ with the residue field k_L. Let $\mathrm{red} : \mathrm{GL}_d(\mathcal{O}_L) \to \mathrm{GL}_d(k_L)$ be the reduction map. Then the residual representation of the p-adic Galois representation $\rho : G_K \to \mathrm{GL}_d(L)$ with the image contained in $\mathrm{GL}_d(\mathcal{O}_L)$ is the composition $\overline{\rho} : G_K \xrightarrow{\rho} \mathrm{GL}_d(\mathcal{O}_L) \xrightarrow{\mathrm{red}} \mathrm{GL}_d(k_L)$.

For the last part, we define three important concepts for the Galois representation: (absolutely) irreducible representations, unramified representations and flat representations.

Definition 2.26. The Galois representation $\rho : G_K \to \mathrm{GL}_d(L) \simeq \mathrm{GL}_L(V)$ is irreducible if there is no nontrivial proper G_K-invariant subspace of the d-dimensional L-vector space V. And ρ is absolutely irreducible if the corresponding representation $G_K \to GL_L(V \otimes_L \overline{L})$ is irreducible.

Write $\overline{\mathbb{Z}} \subseteq \overline{\mathbb{Q}}$ as the ring of algebraic integers in $\overline{\mathbb{Q}}$. For a prime ideal $\mathfrak{p} \subseteq \overline{\mathbb{Z}}$ over a prime $p \in \mathbb{Z}$, we define the decomposition group of $\mathfrak{p}$ as $D_\mathfrak{p} := \{\sigma \in G_\mathbb{Q} : \sigma(\mathfrak{p}) = \mathfrak{p}\}$. Then each $\sigma \in D_\mathfrak{p}$ acts on $\overline{\mathbb{Z}}/\mathfrak{p}$ as $\sigma(x + \mathfrak{p}) = \sigma(x) + \mathfrak{p}$ for every $x \in \mathbb{Z}$, which is also the action on $\overline{\mathbb{F}_p}$. This induces the reduction map $D_\mathfrak{p} \to G_{\mathbb{F}_p}$ and the kernel is defined to be the inertia group $I_\mathfrak{p}$ at $\mathfrak{p}$.

Definition 2.27. [6, p.380] [4, p.6] Let $p \in \mathbb{Z}$ be a prime and $\rho : G_\mathbb{Q} \to GL_d(L)$ be a global Galois representation. We say that ρ is unramified at p if $I_\mathfrak{p} \subseteq \ker\rho$ for any prime ideal $\mathfrak{p} \subseteq \overline{\mathbb{Z}}$ over p (or equivalently, $I_p \subseteq \ker\rho|_{G_l}$).

To define the flat Galois representation, we need some scheme theory.

Definition 2.28. [1] Let S be a scheme. A group scheme (G, m) over S is a scheme G over S with a morphism of schemes $m : G \times_S G \to G$ such that $G(T) := \mathrm{Hom}_{\mathbf{Sch}/S}(T, G)$ is a group under m for every scheme T over S. Moreover, a group scheme (G, m) is called finite flat if the structure morphism $G \to S$ is finite flat.

Now we define the flatness for a mod p Galois representation.

Definition 2.29. [4, p.6] Let $p, l \in \mathbb{Z}$ be primes and $\rho : G_{\mathbb{Q}} \to GL_d(L)$ be a mod p Galois representation. We say that ρ is flat at l if $\rho|_{G_p} : G_p \to \mathrm{GL}_d(L)$ extends to a finite flat group scheme over $\mathbb{Z}_l$ (means over $\mathrm{Spec}(\mathbb{Z}_l)$).

3 Galois Representations Associated to Elliptic Curves and Modular Forms

In this section, we obtain Galois representations from elliptic curves and modular forms, which is mainly based on [6, Chapter 9].

3.1 Galois Representations Associated to Elliptic Curves

We first introduce the notion of the Tate's module of an abelian group.

Definition 3.1. Let A be an abelian group and $p \in \mathbb{Z}$ be a prime. For $n \in \mathbb{Z}^+$, denote the n-torsion subgroup by $A[n]$. The inverse limit of the chain $A[p] \leftarrow A[p^2] \leftarrow A[p^3] \leftarrow \cdots$ (where the maps are multiplication by p, denoted by $[p]$) is called the p-adic Tate module of A, denoted by $\mathrm{Ta}_p(A) := \varprojlim_n \{A[p^n]\}$.

Lemma 3.2. *[6, p.382] Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve. Then*

1. $\mathrm{Ta}_p(E) := \mathrm{Ta}_p(E(\overline{\mathbb{Q}})) \simeq \mathbb{Z}_p^2$*; and*
2. $\mathrm{Aut}(\mathrm{Ta}_p(E)) \simeq \mathrm{GL}_2(\mathbb{Z}_p)$.

Proof. For 1, we have $E[p^n] \simeq (\mathbb{Z}/p^n\mathbb{Z})^2$ for every $n \in \mathbb{Z}^+$ by lemma 2.9. So the result follows by taking the inverse limit. For 2, note that both sides are isomorphic to $\mathrm{Aut}(\mathbb{Z}_p^2)$. □

Definition 3.3. Let $E/\mathbb{Q}$ be an elliptic curve. We define the action of $G_{\mathbb{Q}}$ on $E = E(\overline{\mathbb{Q}})$ by acting componentwise, i.e., $\sigma(P) = (\sigma(x), \sigma(y))$ for $\sigma \in G_{\mathbb{Q}}$ and $P = (x, y) \in E$.

Since the equation defining E is a polynomial, we have $\sigma(P) \in E$ for any $P \in E$ and $\sigma \in G_{\mathbb{Q}}$. So the action defined above is well-defined.

Lemma 3.4. *[6, p.382] Let p be a prime and $E/\mathbb{Q}$ be an elliptic curve. The action of $G_{\mathbb{Q}}$ on E descends to the action of $G_{\mathbb{Q}}$ on $E[p^n]$ for each $n \in \mathbb{Z}^+$, and hence induces an action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(E)$.*

Proof. Since $[p^n]$ is a rational function of entries for each $n \in \mathbb{Z}^+$, it commutes with σ for every $\sigma \in G_{\mathbb{Q}}$. So the action of $G_{\mathbb{Q}}$ on E preserving p^n-torsion points, and hence descends to the action of $G_{\mathbb{Q}}$ on $E[p^n]$ for each $n \in \mathbb{Z}^+$, which are denoted by $\varphi_n : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p^n])$.

For the second part, one can verify that the following diagram commutes

$$\begin{tikzcd}
& G_{\mathbb{Q}} \arrow[dl, "\varphi_n"'] \arrow[dr, "\varphi_{n+1}"] & \\
\mathrm{Aut}(E[p^n]) & & \mathrm{Aut}(E[p^{n+1}]) \arrow[ll]
\end{tikzcd}$$

where the bottom map is induced by $[p]$ (More precisely, the bottom map is defined by $\psi \mapsto [p] \circ \psi \circ i$ where $i : E[p^n] \hookrightarrow E[p^{n+1}]$ is the inclusion). So by taking the inverse limit, we get the map $G_{\mathbb{Q}} \to \mathrm{Aut}(\mathrm{Ta}_p(E))$, i.e., the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(E)$. □

By choosing a basis (P_n, Q_n) of $E[p^n]$ for each $n \in \mathbb{Z}^+$ with the compatibility $[p]P_{n+1} = P_n$ and $[p]Q_{n+1} = Q_n$, the isomorphism maps in lemma 3.2 can be specified. In summary, we obtain a group homomorphism (call it $\rho_{E,p}$):

$$\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(\mathrm{Ta}_p(E)) \simeq \mathrm{GL}_2(\mathbb{Z}_p) \hookrightarrow \mathrm{GL}_2(\mathbb{Q}_p).$$

Lemma 3.5. *[6, Exercise 9.4.1] Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve. Then $\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_p)$ is continuous, and hence a 2-dimensional global p-adic Galois representation.*

Proof. To prove the continuity, it suffices to show that $\rho_{E,p}^{-1}(1)$ is open in $G_{\mathbb{Q}}$ (since $G_{\mathbb{Q}}$ and $\mathrm{GL}_2(\mathbb{Q}_p)$ are both topological groups). For $\sigma \in G_{\mathbb{Q}}$, $\sigma \in \rho_{E,p}^{-1}(1)$ iff σ acts on $\mathrm{Ta}_p(E)$ trivially, iff σ acts on $E[p^n]$ trivially for every $n \in \mathbb{Z}^+$, iff σ fixes $\mathbb{Q}(E[p^n])$ pointwise for every $n \in \mathbb{Z}^+$ (where $\mathbb{Q}(E[p^n])$ is $\mathbb{Q}$ adjoining entries of points in $E[p^n]$), iff $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(E[p^n]))$ for every $n \in \mathbb{Z}^+$. So

$$\rho_{E,p}^{-1}(1) = \bigcup_{n=1}^{\infty} \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(E[p^n])) = \bigcup_{n=1}^{\infty} U_1(\mathrm{Gal}(\mathbb{Q}(E[p^n])))$$

which is open in $G_{\mathbb{Q}}$. □

Definition 3.6. Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve. $\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_p)$ constructed above is called the p-adic Galois representation associated to the elliptic curve E.

Lemma 3.7. *Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve. The residual representation $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ of $\rho_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_p)$ is exactly the action of $G_{\mathbb{Q}}$ on $E[p]$.*

Proof. It follows from the fact that $(x_1, x_2, \cdots) \in \mathbb{Z}_p \subseteq \prod_{n=1}^{\infty}(\mathbb{Z}/p^n\mathbb{Z})$ mod p is $x_1 \in \mathbb{F}_p$. □

Theorem 3.8. *[6, p.383] Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve with conductor N. If a prime $l \nmid pN$, then both $\rho_{E,p}$ and $\overline{\rho}_{E,p}$ are unramified at l.*

Proof. Let $\mathfrak{l} \subseteq \overline{\mathbb{Z}}$ be any prime ideal lying over $l \in \mathbb{Z}$. Then the following diagram commutes:

$$\begin{tikzcd}
D_{\mathfrak{l}} \arrow[r, "\mathrm{act}_1"] \arrow[d, "\mathrm{red}_1"'] & \mathrm{Aut}(E[p^n]) \arrow[d, "\mathrm{red}_2"] \\
G_{\mathbb{F}_l} \arrow[r, "\mathrm{act}_2"'] & \mathrm{Aut}(\tilde{E}[p^n])
\end{tikzcd}$$

where $\mathrm{act}_1, \mathrm{act}_2$ are actions on $E[p^n]$ and $\tilde{E}[p^n]$ respectively, and $\mathrm{red}_1, \mathrm{red}_2$ are the maps of reduction at l (Since $l \nmid pN$, E has good reduction at l. So $\tilde{E}$ is an elliptic curve and $\tilde{E}[p^n]$ makes sense). Note that red_2 is an isomorphism since the reduction at l preserves p^n-torsion structure. So

$$I_{\mathfrak{l}} = \ker(\mathrm{red}_1) \subseteq \ker(\mathrm{act}_2 \circ \mathrm{red}_1) = \ker(\mathrm{red}_2 \circ \mathrm{act}_1) = \ker(\mathrm{act}_1) \subseteq \ker(G_{\mathbb{Q}} \to \mathrm{Aut}(E[p^n])).$$

It holds for any n. So we conclude that $I_{\mathfrak{l}} \subseteq \ker \rho_{E,p}$, i.e., $\rho_{E,p}$ is unramified at l. In particular, it holds for $n = 1$. So we conclude that $I_{\mathfrak{l}} \subseteq \ker \overline{\rho}_{E,p}$, i.e., $\overline{\rho}_{E,p}$ is unramified at l. □

3.2 Galois Representations Associated to Modular Forms

In this section, we need the notions of the Picard group and the Jacobian. If readers are not familiar with them, see section A.1 in the appendix or [6, Chapter 6] for further details.

3.2.1 Galois Representations Associated to Modular Curves

Let $\Gamma \leq \Gamma(1)$ be a congruence subgroup. In section 2.2, we defined the modular curve $X(\Gamma)$ as a set. Actually, we can endow $X(\Gamma)$ with a topology such that it's Hausdorff, connected and compact. Then by giving appropriate charts, $X(\Gamma)$ becomes a compact Riemann surface (i.e., an algebraic curve over $\mathbb{C}$), denoted by $X(\Gamma)_{\mathbb{C}}$. It also has a model as an algebraic curve over $\mathbb{Q}$, denoted by $X(\Gamma)_{\mathbb{Q}}$. The details can be seen in the reference [6, Chapter 2,7].

Lemma 3.9. *[6, p.386-387] Let p be a prime, $N \in \mathbb{Z}^+$, and g be the genus of $X_1(N)_{\mathbb{C}}$. Then*

1. $\mathrm{Ta}_p(\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})) \simeq \mathbb{Z}_p^{2g}$*; and*

2. $\mathrm{Aut}(\mathrm{Ta}_p(\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}}))) \simeq \mathrm{GL}_{2g}(\mathbb{Z}_p)$.

Proof. For 1, we have $\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})[p^n] \simeq \mathrm{Pic}^0(X_1(N)_{\mathbb{C}})[p^n]$ for each $n \in \mathbb{Z}^+$ [6, p.386]. By Abel's theorem A.4 and lemma A.3,

$$\mathrm{Pic}^0(X_1(N)_{\mathbb{C}})[p^n] \simeq \mathrm{Jac}(X_1(N)_{\mathbb{C}})[p^n] \simeq (\mathbb{C}^g/\Lambda_g)[p^n] \simeq (\mathbb{Z}/p^n\mathbb{Z})^{2g}$$

for some lattice Λ_g in $\mathbb{C}^g$. So the result follows by taking the inverse limit.

For 2, both sides are isomorphic to $\mathrm{Aut}(\mathbb{Z}_p^{2g})$. □

Definition 3.10. [6, p.387] The action of $G_{\mathbb{Q}}$ on $X_1(N)_{\mathbb{Q}}$ is defined componentwise, and the action of $G_{\mathbb{Q}}$ on $\mathrm{Div}^0(X_1(N)_{\mathbb{Q}})$ is defined by $\sigma(\sum n_P(P)) = \sum n_P(\sigma(P))$.

Lemma 3.11. *[6, p.387] The action of $G_{\mathbb{Q}}$ on $\mathrm{Div}^0(X_1(N)_{\mathbb{Q}})$ defined above descends to the action of $G_{\mathbb{Q}}$ on $\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})[p^n]$ for every $n \in \mathbb{Z}^+$, and hence induces to the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$.*

Proof. Note that div commutes with σ for any $\sigma \in G_{\mathbb{Q}}$. So the action descends to the action of $G_{\mathbb{Q}}$ on $\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})$. Also note that $[n]$ commutes with σ for every $n \in \mathbb{Z}^+$ and $\sigma \in G_{\mathbb{Q}}$ by the formula $\sigma(\sum n_P(P)) = \sum n_P(\sigma(P))$. So the action descends to the action of $G_{\mathbb{Q}}$ on $\mathrm{Pic}(X_1(N)_{\mathbb{Q}})[p^n]$ for every $n \in \mathbb{Z}^+$.

For the second part, one can verify the diagram (analogous to the one in the proof of lemma 3.4) commutes, which induces the action on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$ by taking the inverse limit. □

Choosing a basis of $\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})[p^n]$ for each $n \in \mathbb{Z}^+$ specifies the isomorphism maps in lemma 3.9. Thus, we obtain a group homomorphism (call it $\rho_{X_1(N),p}$):

$$\rho_{X_1(N),p} : G_{\mathbb{Q}} \to \mathrm{Aut}(\mathrm{Ta}_p(\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}}))) \simeq \mathrm{GL}_{2g}(\mathbb{Z}_p) \hookrightarrow \mathrm{GL}_{2g}(\mathbb{Q}_p).$$

The proof of continuity of $\rho_{X_1(N),p}$ is omitted here.

Definition 3.12. Let N be a positive integer and $p \in \mathbb{Z}$ be a prime. $\rho_{X_1(N),p}$ constructed above is called the p-adic Galois representation associated to the modular curve $X_1(N)$.

3.2.2 Galois Representations Associated to the Abelian Variety of Newforms

Definition 3.13. [6, p.387] The Hecke algebra over $\mathbb{Z}$ is $\mathbb{T}_{\mathbb{Z}} := \mathbb{Z}[\{T_n, \langle n\rangle : n \in \mathbb{Z}^+\}]$ which is the sub-algebra of $\mathrm{End}(S_2(\Gamma_1(N)))$ where T_n and $\langle n\rangle$ are Hecke operator and diamond operator respectively. For a newform $f \in S_2(\Gamma_1(N))$, define the ideal $I_f := \{T \in \mathbb{T}_{\mathbb{Z}} : Tf = 0\}$.

There are natural actions of the Hecke algebra $\mathbb{T}_{\mathbb{Z}}$ on Picard groups and the Jacobian. For the definition of these actions, see section A.2, A.3 in the appendix or [6, Chapter 5-7] for further details. We comment that the action of $\mathbb{T}_{\mathbb{Z}}$ on $\mathrm{Pic}^0(X_1(N)_{\mathbb{Q}}$ induces the action of $\mathbb{T}_{\mathbb{Z}}$ on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$, which commutes with the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$ [6, p.387].

Definition 3.14. [6, p.388] Denote $J_1(N) := \mathrm{Jac}(X_1(N)_{\mathbb{C}})$. The abelian variety of a newform $f \in S_2(\Gamma_1(N))$ is $A_f := J_1(N)/I_f J_1(N)$ where $I_f J_1(N)$ is the image of the action of $I_f \subseteq \mathbb{T}_{\mathbb{Z}}$ on $J_1(N)$.

Actually A_f is a complex torus, but we won't define this structure here.

Definition 3.15. [6, p.389] For a modular form $f : \mathbb{H} \to \mathbb{C}$ with Fourier coefficients $\{a_n(f)\}_{n\geq 0}$, define $\mathcal{O}_f := \mathbb{Z}[\{a_n(f) : n \in \mathbb{Z}^+\}]$ and K_f as the fractional field of $\mathcal{O}_f$.

Lemma 3.16. *[6, p.389] The dimension of A_f as a complex torus is $[K_f : \mathbb{Q}]$.*

Lemma 3.17. *[6, p.389] Let p be a prime, $f \in S_2(\Gamma_1(N))$ be a newform, and $d = [K_f : \mathbb{Q}]$.*

1. $\mathrm{Ta}_p(A_f) \simeq \mathbb{Z}_p^{2d}$*; and*
2. $\mathrm{Aut}(\mathrm{Ta}_p(A_f)) \simeq \mathrm{GL}_{2d}(\mathbb{Z}_p)$.

Proof. For 1, we have $A_f[p^n] \simeq (\mathbb{Z}/p^n\mathbb{Z})^{2d}$ for every $n \in \mathbb{Z}^+$ by lemma 3.16, and the result follows by taking the inverse limit. For 2, both sides are isomorphic to $\mathrm{Aut}(\mathbb{Z}_p^{2d})$. □

Lemma 3.18. *[6, p.389] The map $\psi : \mathrm{Pic}^0(X_1(N)_{\mathbb{Q}})[p^n] \simeq \mathrm{Pic}^0(X_1(N)_{\mathbb{C}})[p^n] \simeq J_1(N)[p^n] \to A_f[p^n]$ is surjective with kernel stable under $G_{\mathbb{Q}}$.*

Definition 3.19. The action of $G_{\mathbb{Q}}$ on $A_f[p^n]$ for each $n \in \mathbb{Z}^+$ is defined in the following way. For $\sigma \in G_{\mathbb{Q}}$ and $x \in A_f[p^n]$, $\sigma \cdot x$ is given by $\psi(\sigma \cdot x')$ where x' is a preimage of x under ψ.

$\sigma \cdot x'$ makes sense since we defined the action of $G_{\mathbb{Q}}$ on $\mathrm{Pic}^0(X_1(N))[p^n]$ previously. The lemma 3.18 guarantees that the action defined in definition 3.19 is well-defined. Similar to lemma 3.4 and 3.11, it induces the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(A_f)$.

Choosing a basis of $A_f[p^n]$ for each $n \in \mathbb{Z}^+$ specifies the isomorphism maps in lemma 3.17. Thus, we obtain a group homomorphism (call it $\rho_{A_f,p}$):

$$\rho_{A_f,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(\mathrm{Ta}_p(A_f)) \simeq \mathrm{GL}_{2d}(\mathbb{Z}_p) \hookrightarrow \mathrm{GL}_{2d}(\mathbb{Q}_p).$$

The continuity of $\rho_{A_f,p}$ follows from the continuity of $\rho_{X_1(N),p}$.

Definition 3.20. Let $p \in \mathbb{Z}$ be a prime and $f \in S_2(\Gamma_1(N))$ be a newform. $\rho_{A_f,p} : G_{\mathbb{Q}} \to \mathrm{GL}_{2d}(\mathbb{Q}_p)$ constructed above is the Galois representation associated to the abelian variety A_f.

We can also define the action of $\mathcal{O}_f$ on A_f by the following lemma.

Lemma 3.21. *[6, p.389] Let $N \in \mathbb{Z}^+$ and $f \in S_2(\Gamma_1(N))$ be a newform. Then $\mathbb{T}_{\mathbb{Z}}/I_f \simeq \mathcal{O}_f$.*

Under the isomorphism of the above lemma, let $a_p(f)$ act on A_f as $T_p + I_f$ act on A_f, which determines the action of $\mathcal{O}_f$ on A_f. This action restricts to $A_f[p^n]$ and thus induces the action of $\mathcal{O}_f$ on $\mathrm{Ta}_p(A_f)$, which commutes with the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(A_f)$ because the action of $\mathbb{T}_{\mathbb{Z}}$ on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$ commutes with the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(\mathrm{Pic}(X_1(N)_{\mathbb{Q}}))$.

3.2.3 Galois Representations Associated to Newforms

Define the module $V_p(A_f) := \mathrm{Ta}_p(A_f) \otimes \mathbb{Q}$ over $\mathcal{O}_f \otimes \mathbb{Q}_p = K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p$.

Lemma 3.22. *[6, p.390] Let $N \in \mathbb{Z}^+$ and $f \in S_2(\Gamma_1(N))$ be a newform.*

- *$V_p(A_f) \simeq (K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p)^2$ as an isomorphism of $(K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p)$-module.*
- *$K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p \simeq \prod_{\lambda | p} K_{f,\lambda}$.*

By the above lemma, choosing a basis of $V_p(A_f)$ specifies the isomorphism maps

$$\mathrm{Aut}(V_p(A_f)) \simeq \mathrm{Aut}((K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p)^2) \simeq \mathrm{GL}_2(K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p) \simeq \mathrm{GL}_2(\prod_{\lambda|p} K_{f,\lambda}).$$

Since $G_{\mathbb{Q}}$ acts on $\mathrm{Ta}_p(A_f)$ as we defined previously, $G_{\mathbb{Q}}$ acts on $V_p(A_f)$. Also the action of $G_{\mathbb{Q}}$ on $V_p(A_f)$ is $K_f \otimes_{\mathbb{Q}} \mathbb{Q}_p$-linearly, since the action of $G_{\mathbb{Q}}$ on $\mathrm{Ta}_p(A_f)$ and the action of $\mathcal{O}_f$ on $\mathrm{Ta}_p(A_f)$ commutes. Thus, we obtain a group homomorphism (call it $\rho_{f,\lambda}$):

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to \mathrm{Aut}(V_p(A_f)) \simeq \mathrm{GL}_2(\prod_{\lambda|p} K_{f,\lambda}) \twoheadrightarrow \mathrm{GL}_2(K_{f,\lambda}).$$

The continuity of $\rho_{f,\lambda}$ follows from the continuity of $\rho_{A_f,p}$.

Definition 3.23. Let $N \in \mathbb{Z}^+$, $f \in S_2(\Gamma_1(N))$ be a newform, and $\lambda \subseteq \mathcal{O}_{K_f}$ be a prime ideal lying over a prime $p \in \mathbb{Z}$. $\rho_{f,\lambda} : G_{\mathbb{Q}} \to \mathrm{GL}_2(K_{f,\lambda})$ constructed above is called the Galois representation associated to the newform f.

4 Modularity Theorem and Ribet's Level Lowering Theorem

4.1 Modularity Theorem

Now we define the notion of the modularity for p-adic Galois representations.

Definition 4.1. Let $p \in \mathbb{Z}$ be a prime. A p-adic global Galois representation ρ is modular if there exists $N \in \mathbb{Z}^+$ and a newform $f \in S_2(\Gamma_0(N))$ with $K_{f,\lambda} = \mathbb{Q}_p$ for some prime ideal $\lambda \subseteq \mathcal{O}_{K_f}$ such that $\rho \sim \rho_{f,\lambda}$.

Wiles proved the modularity theorem for semistable elliptic curves, which roughly states that the p-adic Galois representation $\rho_{E,p}$ associated to a semistable Elliptic curve $E/\mathbb{Q}$ is modular for every prime $p \in \mathbb{Z}$. We can furthermore require that the conductor of the elliptic curve and the newform coincide. The precise statement of the modularity theorem (for semistable Elliptic curves) is quoted here (called modularity theorem strong version R in the reference [6, p.392]):

Theorem 4.2 (**Modularity Theorem for Semistable Elliptic Curves**). *[6, p.392] Let $E/\mathbb{Q}$ be a semistable elliptic curve with conductor N. Then for any prime $p \in \mathbb{Z}$, we have $\rho_{E,p} \sim \rho_{f,p}$ for some newform $f \in S_2(\Gamma_0(N))$ with $K_f = \mathbb{Q}$.*

The proof of modularity theorem for semistable elliptic curves can be seen in Wiles's paper [15]. Later, C. Breuil, B. Conrad, F. Diamond, and R. Taylor proved the modularity theorem for the general elliptic curves (see the proof in [2]). But we only need the semistable case in this essay because it is strong enough to deduce the Fermat's last theorem.

4.2 Ribet's Level Lowering Theorem

Next we turn to the notion of modularity for mod p Galois representations. And we define the level of the representation to be the conductor of the newform, as shown in the following.

Definition 4.3. Let $p \in \mathbb{Z}$ be a prime and $N \in \mathbb{Z}^+$. A mod p global Galois representation $\overline{\rho}$ is modular of level N if there exists a newform $f \in S_2(\Gamma_0(N))$ with $K_{f,\lambda} = \mathbb{Q}_p$ for some prime ideal $\lambda \subseteq \mathcal{O}_{K_f}$ such that $\overline{\rho} \sim \overline{\rho}_{f,\lambda}$.

A question is how to find the minimal level of such representation. The method is given by Ribet's level lowering theorem, which is quoted as the following:

Theorem 4.4 (**Ribet's Level Lowering Theorem**). *[4, p.9] Let N be a positive integer, p be a prime, $l \nmid N$ be a prime, and $f \in S_2(\Gamma_0(Nl))$ be a newform. Suppose that $\overline{\rho}_{f,p}$ is absolutely irreducible and one of the followings is true:*

- *$\overline{\rho}_{f,p}$ is unramified at l; or*
- *$\overline{\rho}_{f,p}$ is flat at p.*

Then $\overline{\rho}_{f,p} \sim \overline{\rho}_{g,p}$ for some newform $f \in S_2(\Gamma_0(N))$.

The proof of Ribet's level lowering theorem can be seen in his original article [11].

5 Frey Curves

5.1 Frey Curves Attached to a Nontrivial Solution of the Fermat's Equation

Now we discuss the Fermat's last theorem 1.1. Since the cases for $n = 3, 4$ were proven, we assume $n \geq 5$ now. If n is the power of 2, then $4|n$, and hence a nontrivial solution (a, b, c) for $x^n + y^n = z^n$ gives a nontrivial solution $(a^{n/4}, b^{n/4}, c^{n/4})$ for $x^4 + y^4 = z^4$. But the case $n = 4$ was proven. So we assume that there is an odd prime p dividing n now, say $n = pk$. But then a nontrivial solution (a, b, c) for $x^n + y^n = z^n$ gives a nontrivial solution (a^k, b^k, c^k) for $x^p + y^p = z^p$. This observation tells us that we only need to prove FLT for the case that $n = p$ is an odd prime. Since the case $n = 3$ was proven, we assume that $n = p \geq 5$ is prime now.

Our strategy is the proof of contradiction. From now on, we assume that there is a nontrivial solution (a, b, c) of the Fermat's equation $a^p + b^p = c^p$ for prime $p \geq 5$ and want to deduce the contradiction. We rewrite it as $a^p + b^p + c^p = 0$ by replacing c with $-c$. We can further require that a, b, c be coprime. By analyzing the parity, we find that at least one of a, b, c must be even, say b is even. Since a, b, c are coprime, both a and c are odd. By mod 4 analysis, we find that at least one of a and c must be congruent to -1 mod 4, say $a \equiv -1 \pmod 4$.

Therefore, we reduce FLT to the following easier version:

Theorem 5.1 (**Reduced Fermat's Last Theorem**). *There is no $a, b, c, p \in \mathbb{Z}$ satisfying the condition ★: $p \geq 5$ is prime, $abc \neq 0$, $a^p + b^p + c^p = 0$, a, b, c are coprime, b is even, and $a \equiv -1$* (mod 4).

Now we can define the Frey curve:

Definition 5.2. [4, p.2] Under ★, the Frey curve attached to them is the elliptic curve over $\mathbb{Q}$ defined by $y^2 = x(x-a^p)(x+b^p)$, denoted by E_{a^p,b^p,c^p}.

By the simple computation of partial derivatives, it is easy to see that the Frey curve is nonsingular, i.e., it is indeed an Elliptic curve.

5.2 Semistability, Conductor and Minimal Discriminant

We first introduce some theorems in the theory of elliptic curves to determine whether a Weierstrass equation is minimal and whether a reduction is good or multiplicative.

Theorem 5.3. *[14, p.186] Let $l \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation with coefficients $a_1, \cdots, a_6 \in \mathbb{Z}$. If $v_l(\Delta) < 12$ or $v_l(c_4) < 4$, then the equation is minimal at l.*

Theorem 5.4. *[14, p.196] Let $l \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation that is minimal at l, with coefficients $a_1, \cdots, a_6 \in \mathbb{Z}$. Then*

- *E has a good reduction at l iff $v_l(\Delta) = 0$.*
- *E has a multiplicative reduction at l iff $v_l(\Delta) > 0$ and $v_l(c_4) = 0$.*

By the above theorems, we have the following critically important lemma by following the routes in the reference[9, p.137, Exercise 5.6.5] with some modifications, which finds a minimal model for Frey curve with some key properties.

Lemma 5.5. *Write*

$$E^{\min}_{a^p,b^p,c^p}/\mathbb{Q} : y'^2 - 3x'y' = x'^3 + \frac{b^p - a^p - 9}{4}x'^2 - \frac{(ab)^p}{16}x'.$$

Under ★, we have:

1. *E_{a^p,b^p,c^p} is isomorphic to $E^{\min}_{a^p,b^p,c^p}$ by a linear change of coordinates.*
2. *All of the coefficients $a_1 = -3$, $a_2 = \frac{b^p-a^p-9}{4}$, $a_3 = 0$, $a_4 = -\frac{(ab)^p}{16}$, $a_6 = 0$ of the equation for $E^{\min}_{a^p,b^p,c^p}$ are integers.*
3. *For $E^{\min}_{a^p,b^p,c^p}$, the discriminant is $\Delta = 2^{-8}(abc)^{2p}$ and $c_4 = c^{2p} - (ab)^p$.*
4. *If a prime $l \nmid abc$, then $v_l(\Delta) = 0$.*
5. *If a prime $l | abc$, then $v_l(c_4) = 0$ and $v_l(\Delta) > 0$.*
6. *The equation for $E^{\min}_{a^p,b^p,c^p}$ is minimal (i.e., minimal at every prime).*
7. *$E^{\min}_{a^p,b^p,c^p}$ has good reduction at all primes $l \nmid abc$, whereas has multiplicative reduction at all primes $l|abc$ (and so does E_{a^p,b^p,c^p}, since E_{a^p,b^p,c^p} and $E^{\min}_{a^p,b^p,c^p}$ are isomorphic).*

Proof.

1. Under the linear change of coordinates

$$\begin{cases} x = 4x' \\ y = 8y' - 12x' \end{cases},$$

$E_{a^p,b^p,c^p} : y^2 = x(x-a^p)(x+b^p)$ turns to $E^{\min}_{a^p,b^p,c^p} : y'^2 - 3x'y' = x'^3 + \frac{b^p-a^p-9}{4}x'^2 - \frac{(ab)^p}{16}x'$.

2. $a_2 \in \mathbb{Z}$ because b is even, $p \geq 5$ odd, and $a \equiv -1 \pmod 4$; $a_4 \in \mathbb{Z}$ because b is even and $p \geq 5$.
3. By the formulae of Δ and c_4, we directly compute $\Delta = 2^{-8}(ab)^{2p}(a^p + b^p)^2 = 2^{-8}(abc)^{2p}$ and $c_4 = a^{2p} + (ab)^p + b^{2p} = c^{2p} - (ab)^p$.
4. If a prime $l \nmid abc$, then note $l \neq 2$ and hence $l \nmid \Delta$, i.e., $v_l(\Delta) = 0$.

5. If a prime $l|abc$, then l divides at least one of a, b, c. But note that each of a, b, c cannot divide $c_4 = c^{2p} - (ab)^p$ since a, b, c are coprime. Thus, $l \nmid c_4$ and hence $v_l(c_4) = 0$. For $v_l(\Delta) > 0$, we split it into two cases. If $l \neq 2$, from $\Delta = 2^{-8}(abc)^{2p}$ we know $v_l(\Delta) \geq 2p \geq 10 > 0$; If $l = 2$, by the fact that b is even, we know $v_l(\Delta) \geq 2p - 8 \geq 2 > 0$.
6. For all primes l, either $v_l(\Delta) = 0 < 12$ (when $l \nmid abc$) or $v_l(c_4) = 0 < 4$ (when $l|abc$). Also in part 2 we proved the coefficients $a_1, \cdots, a_6 \in \mathbb{Z}$. So the minimality follows from theorem 5.3.
7. It follows from part 4,5 and theorem 5.4.

□

Remark. The equation $y^2 = x(x-a^p)(x+b^p)$ for the Frey curve E_{a^p,b^p,c^p} is minimal at all primes except $l = 2$. So the purpose of introducing $E^{\min}_{a^p,b^p,c^p}$ is dealing with the tricky case $l = 2$.

The following theorem is the summary of the above lemma.

Theorem 5.6. *[4, p.3] Under ★, the Frey curve E_{a^p,b^p,c^p} satisfies:*

1. *E_{a^p,b^p,c^p} is semistable; and*
2. *The conductor of E_{a^p,b^p,c^p} is $N = \prod_{l \text{ prime}, l|abc} l$; and*
3. *The minimal discriminant of E_{a^p,b^p,c^p} is $\Delta = 2^{-8}(abc)^{2p}$.*

Proof. The first two parts follow from part 7 of lemma 5.5, and the last part follows from part 1,3,6 of lemma 5.5. □

5.3 Residual Galois Representations Associated to Frey Curves

We write $\rho_{a^p,b^p,c^p} := \rho_{E,p}$ as the p-adic Galois representation associated to the Frey curve $E = E_{a^p,b^p,c^p}$, and $\overline{\rho}_{a^p,b^p,c^p} := \overline{\rho}_{E,p}$ as the residual representation mod p.

5.3.1 Absolutely Irreducible Representations

In this part we will show that $\overline{\rho}_{a^p,b^p,c^p}$ is absolutely irreducible by Mazur's theory.

Mazur proved the following theorem of the structure of $E(\mathbb{Q}_{\text{tors}})$:

Theorem 5.7. *[10, Theorem 2] Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})_{\text{tors}}$ has only 15 possibilities up to isomorphism: $\mathbb{Z}/n\mathbb{Z}$ ($1 \leq n \leq 10$, $n = 12$), $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2n\mathbb{Z}$ ($1 \leq n \leq 4$).*

It can be shown that if $E/\mathbb{Q}$ has a rational subgroup (i.e., $G_{\mathbb{Q}}$-invariant) of order p, then it admits an p-isogeny defined over $\mathbb{Q}$ [14, p.74]. If we further require that E is semistable, then we obtain more as in the following lemma.

Lemma 5.8. *[5, p.56] Let $p \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be a semistable elliptic curve. If E has a rational subgroup of order p, then E is p-isogenous to a curve with a rational point of order p.*

The condition of Ribet's level lowering theorem requires that the representation is absolutely irreducible. The following lemma shows that we do not need to worry about the absoluteness.

Lemma 5.9. *[12, Lemma 5] Let $p \geq 3$ be a prime and E/K be an elliptic curve where the field K embeds into $\mathbb{R}$. If $\overline{\rho}_{E,p}$ is irreducible, then $\overline{\rho}_{E,p}$ is absolutely irreducible.*

Now we can establish a sufficient condition for $\overline{\rho}_{E,p}$ to be absolutely irreducible.

Theorem 5.10. *[5, p.56] Let $E/\mathbb{Q}$ be a semistable elliptic curve. If $|E(\mathbb{Q})[2]| = 4$, then $\overline{\rho}_{E,p}$ is absolutely irreducible for any prime $p \geq 5$.*

Proof. Since $\mathbb{Q}$ embeds into $\mathbb{R}$, by lemma 5.9, it suffices to show that $\overline{\rho}_{E,p}$ is irreducible.

Suppose that the two-dimensional representation $\overline{\rho}_{E,p}$ is reducible. Then there is a one-dimensional $G_{\mathbb{Q}}$-invariant subspace of $E[p]$ over $\mathbb{F}_p$, i.e., a $G_{\mathbb{Q}}$-invariant subgroup $\Phi \subseteq E[p]$ of order p. By lemma 5.8, there is an p-isogeny $E \to E'$ where E' has an p-torsion rational point. So $E'(\mathbb{Q})_{\text{tors}}$ has a subgroup isomorphic to $\mathbb{Z}/p\mathbb{Z}$. Since $p \geq 5$, $E'(\mathbb{Q})_{\text{tors}}$ cannot be of the form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2n\mathbb{Z}$ ($1 \leq n \leq 4$). Then by theorem 5.7, $E'(\mathbb{Q})_{\text{tors}}$ is of the form $\mathbb{Z}/n\mathbb{Z}$ ($1 \leq n \leq 10, n = 12$). Since the isogeny has degree $p \neq 2$, $E(\mathbb{Q})[2]$ injects into $E'(\mathbb{Q})[2]$ which makes $|E'(\mathbb{Q})[2]| = 4$ as well. But $\mathbb{Z}/n\mathbb{Z}$ has only two 2-torsion points, which is a contradiction. □

Theorem 5.11. *[4, p.7] Under ★, $\overline{\rho}_{a^p,b^p,c^p}$ is absolutely irreducible.*

Proof. Note $E_{a^p,b^p,c^p}(\mathbb{Q})[2] = \{(0,0),(0,a^p),(0,-b^p),O\}$ which has 4 points. Also we proved that E_{a^p,b^p,c^p} is semistable. So $\overline{\rho}_{a^p,b^p,c^p}$ is absolutely irreducible by theorem 5.10 and $p \geq 5$. □

5.3.2 Unramified Representations

In this part we will show that $\overline{\rho}_{a^p,b^p,c^p}$ is unramified outside $2p$ by Tate's theory.

For an elliptic curve $E/\mathbb{Q}$ with multiplicative reduction at a prime $l \in \mathbb{Z}$, there is a quantity $q \in \mathbb{Q}_l^\times$ such that $E/\mathbb{Q}_l$ is isomorphic to the Tate curve $E_q/\mathbb{Q}_l$ (which we will not define here, see [13, Chapter V] for details). We state without the proof the following lemma:

Lemma 5.12. *[13, p.423] Let $l \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ with multiplicative reduction at l and minimal discriminant Δ. Then $\Delta = q\prod_{n\geq 1}(1-q^n)^{24}$. In particular, $v_l(\Delta) = v_l(q)$.*

Also we need the following lemma (without proof) in algebraic number theory:

Lemma 5.13. *Let $l \neq p$ be two primes. Then*

- *the extension $\mathbb{Q}_l(\zeta_p)/\mathbb{Q}_l$ is unramified; and*
- *the extension $\mathbb{Q}_l(q^{1/p})/\mathbb{Q}_l$ is unramified iff $p|v_l(q)$.*

Later we will see that the unramified extension is closely related to the unramified representation, and $p|v_l(q) = v_l(\Delta)$ is exactly the criterion of whether the representation is unramified.

To state the Tate's theorem, we define the character $\delta_{E,l}$ as the following.

Definition 5.14. [13, p.440] [5, p.20] Let l be a prime and $E/\mathbb{Q}$ be an elliptic curve with multiplicative reduction at l. Let $\chi_{E,l} : \mathrm{Gal}(\mathbb{Q}_l(\sqrt{-c_4/c_6})/\mathbb{Q}_l) \to \{\pm 1\}$ be the unique isomorphism and $\mathrm{res}_{E,l} : G_l \twoheadrightarrow G_{\mathbb{Q}_l(\sqrt{-c_4/c_6})/\mathbb{Q}_l}$. Define the character $\delta_{E,l} = \chi_{E,l} \circ \mathrm{res}_{E,l} : G_l \to \{\pm 1\}$.

Then we can formulate the Tate's theorem.

Theorem 5.15 (**Tate**)**.** *[5, p.20] Let $l \in \mathbb{Z}$ be a prime and $E/\mathbb{Q}$ be an elliptic curve with multiplicative reduction at l. Then there is an isomorphism $\Phi : \overline{\mathbb{Q}}_l^\times/q^{\mathbb{Z}} \simeq E(\overline{\mathbb{Q}}_l)$ with the property*

$$\forall \sigma \in G_l, \forall x \in \overline{\mathbb{Q}}_l^\times/q^{\mathbb{Z}}, \sigma(\Phi(x)) = \Phi(\sigma(x)^{\delta_{E,l}(\sigma)})$$

where $q^{\mathbb{Z}} := \{q^n : n \in \mathbb{Z}\}$.

For the proof of Tate's theorem, see [13, p.423, Theorem 3.1 + p.439, Lemma 5.2 + p.442, Theorem 5.3]. The following lemma is a direct corollary of the Tate's theorem.

Lemma 5.16. *Let l, p be two primes and $E/\mathbb{Q}$ be an elliptic curve with multiplicative reduction at l. Then there is an isomorphism $\phi : \langle \zeta_p \rangle \times \langle q^{\frac{1}{p}} \rangle \simeq E[p]$ with the property*

$$\forall \sigma \in G_l, \sigma(\phi(\zeta_p)) = \phi(\sigma(\zeta_p)^{\delta_{E,l}(\sigma)}) \text{ and } \sigma(\phi(q^{1/p})) = \phi(\sigma(q^{1/p})^{\delta_{E,l}(\sigma)})$$

where $q^{1/p}$ here means $q^{1/p}$ mod $q^{\mathbb{Z}}$.

Now we can establish a criterion of whether $\overline{\rho}_{E,p}$ is unramified at l.

Theorem 5.17. *[4, p.6-7] Let $l \neq p$ be two primes and $E/\mathbb{Q}$ be a semistable elliptic curve with minimal discriminant Δ. Then $\overline{\rho}_{E,p}$ is unramified at l iff $p|v_l(\Delta)$.*

Proof. If E has a good reduction at l, then $l \nmid N$ and hence $l \nmid pN$ since $l \neq p$. By theorem 3.8, $\overline{\rho}_{E,p}$ is unramified. Also by theorem 5.4, $v_l(\Delta) = 0$ and hence $p|v_l(\Delta)$. Therefore, we assume that E has a bad reduction at l. By semistability of E, the reduction at l is multiplicative.

Now let us prove the lemma. $\overline{\rho}_{E,p}$ is unramified at l means $\forall \sigma \in I_l, \sigma$ acts on $E[p]$ trivially by definition 2.27. Since $E[p]$ is generated by $\phi(\zeta_p)$ and $\phi(q^{1/p})$ by lemma 5.16, we only need the actions on $\phi(\zeta_p)$ and $\phi(q^{1/p})$ to be trivial for $\overline{\rho}_{E,p}$ to be unramified, i.e., the condition is $\forall \sigma \in I_l$, $\sigma(\phi(\zeta_p)) = \phi(\zeta_p)$ and $\sigma(\phi(q^{1/p})) = \phi(q^{1/p})$. By lemma 5.16 and $\forall \sigma \in I_l, \delta_{E,l}(\sigma) = 1$, we compute

$$\sigma(\phi(\zeta_p)) = \phi(\sigma(\zeta_p)^{\delta_{E,l}(\sigma)}) = \phi(\sigma(\zeta_p)),$$
$$\sigma(\phi(q^{1/p})) = \phi(\sigma(q^{1/p})^{\delta_{E,l}(\sigma)}) = \phi(\sigma(q^{1/p})).$$

So the condition now simplifies as $\forall \sigma \in I_l, \sigma(\zeta_p) = \zeta_p$ and $\sigma(q^{1/p}) = q^{1/p}$. It just means the inertia subgroups $I_{\mathbb{Q}_l(\zeta_p)/\mathbb{Q}_l}$ and $I_{\mathbb{Q}_l(q^{1/p})/\mathbb{Q}_l}$ are both trivial, i.e., the extensions $\mathbb{Q}_l(\zeta_p)/\mathbb{Q}_l$ and $\mathbb{Q}_l(q^{1/p})/\mathbb{Q}_l$ are both unramified. By lemma 5.13, it is equivalent to $p|v_l(q)$, and thus equivalent to $p|v_l(\Delta)$ by lemma 5.12. □

Theorem 5.18. *[4, p.7] Under ★, $\overline{\rho}_{a^p,b^p,c^p}$ is unramified at all primes $l \nmid 2p$.*

Proof. By theorem 5.6, the minimal discriminant of E_{a^p,b^p,c^p} is $\Delta = 2^{-8}(abc)^{2p}$. For prime $l \nmid 2p$, if $l \nmid abc$, then $l \nmid \Delta$ and hence $v_l(\Delta) = 0$; if $l|abc$, say $abc = l^k r$ where $l \nmid r$, then $v_l(\Delta) = 2kp$. Thus for either case, we always have $p|v_l(\Delta)$. Then the result follows from theorem 5.17. □

5.3.3 Flat Representations

In this part, we just state the criterion for $\overline{\rho}_{E,p}$ to be flat at p without proof.

Theorem 5.19. *[4, p.6-7] Let $E/\mathbb{Q}$ be a semistable elliptic curve with minimal discriminant Δ. Then $\overline{\rho}_{E,p}$ is flat at p iff $p|v_p(\Delta)$.*

Theorem 5.20. *[4, p.7] Under ★, $\overline{\rho}_{a^p,b^p,c^p}$ is flat at p.*

Proof. By theorem 5.6, the minimal discriminant of E_{a^p,b^p,c^p} is $\Delta = 2^{-8}(abc)^{2p}$. Note $p \neq 2$. If $p \nmid abc$, then $p \nmid \Delta$ and hence $v_p(\Delta) = 0$; if $p|abc$, say $abc = p^k r$ where $p \nmid r$, then $v_p(\Delta) = 2kp$. Thus for either case, we always have $p|v_p(\Delta)$. Then the result follows from theorem 5.19. □

6 Deduction of Fermat's Last Theorem

First we give a lemma on the dimension of the space $S_2(\Gamma_0(2))$.

Lemma 6.1. $\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = 0$.

Proof. By dimension formula [6, Theorem 3.5.1], $\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = g$ where g is the genus of $X(\Gamma_0(2))$. We note that $\Gamma_0(2) = \Gamma_1(2)$. By [6, Figure 3.4], the genus of $X(\Gamma_1(2))$ is 0. So $\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = g = 0$. □

The Fermat's Last Theorem is stated in theorem 1.1. By the discussion in section 5.1, it suffices to prove the reduced theorem 5.1 which states that there are no a, b, c, p satisfying ★. The following shows the proof of theorem 5.1 which is taken from [4, p.10].

Proof. Suppose that there is some $a, b, c, p \in \mathbb{Z}$ that satisfy ★. By theorem 5.6, the Frey curve E_{a^p,b^p,c^p} is semistable with conductor $N = \prod_{l \text{ prime}, l|abc} l$. Then by modularity theorem 4.2, $\rho_{a^p,b^p,c^p} \sim \rho_{f,p}$ for some newform $f \in S_2(\Gamma_0(N))$. So after mod p, we have $\overline{\rho}_{a^p,b^p,c^p} \sim \overline{\rho}_{f,p}$, which is absolutely irreducible by theorem 5.11.

Write $N = N_1 N_2$ where

$$N_1 = \prod_{l \text{ prime}, l|N, l|2p} l \quad , \qquad N_2 = \prod_{l \text{ prime}, l|N, l\nmid 2p} l \quad .$$

By theorem 5.18, $\overline{\rho}_{f,p} \sim \overline{\rho}_{a^p,b^p,c^p}$ is unramified outside $2p$, i.e., unramified inside N_2. So by Ribet's level lowering theorem 4.4, $\overline{\rho}_{f,p} \sim \overline{\rho}_{g,p}$ for some newform $g \in S_2(\Gamma_0(N_1))$. In summary,

$$\overline{\rho}_{a^p,b^p,c^p} \sim \overline{\rho}_{f,p} \sim \overline{\rho}_{g,p}.$$

Now we define a newform $h \in S_2(\Gamma_0(2))$ in the following way.

- If $p \nmid N$, then $N_1 = 2$ and we simply let $h = g$;
- Otherwise, $N_1 = 2p$. By theorem 5.20, $\overline{\rho}_{g,p} \sim \overline{\rho}_{a^p,b^p,c^p}$ is flat at p. So by Ribet's level lowering theorem 4.4, $\overline{\rho}_{g,p} \sim \overline{\rho}_{h,p}$ for some newform $h \in S_2(\Gamma_0(2))$.

In this way, we found a newform $h \in S_2(\Gamma_0(2))$, which implies $\dim_{\mathbb{C}} S_2(\Gamma_0(2)) \geq 1$ since a newform is nonzero. But by lemma 6.1, $\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = 0$, which leads us to a contradiction. □

7 More General Version of Modularity Theorem

The modularity theorem 4.2 involves elliptic curves over $\mathbb{Q}$. Now we give a general version of the modularity theorem with the base field a real quadratic field. In this case, we need to consider a general version of modular forms in higher dimensions, which is the Hilbert modular form.

Definition 7.1. [3] Let F be real quadratic field, $\mathrm{Gal}(F/\mathbb{Q}) = \{1, \sigma\}$, $k \in \mathbb{Z}$, and $\Gamma \leq \mathrm{SL}_2(F)$ be a subgroup such that $\Gamma \cap \Gamma_F$ has finite index in both Γ and Γ_F where $\Gamma_F := \mathrm{SL}_2(\mathcal{O}_F)$. A Hilbert modular form f over F of parallel weight k and level Γ is a holomorphic function $f : \mathbb{H}^2 \to \mathbb{C}$ such that for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $z_1, z_2 \in \mathbb{C}$,

$$(cz_1 + d)^{-k}(\sigma(c)z_2 + \sigma(d))^{-k} f\left(\frac{az_1 + b}{cz_1 + d}, \frac{\sigma(a)z_2 + \sigma(b)}{\sigma(c)z_2 + \sigma(d)}\right) = f(z_1, z_2).$$

For a Hilbert modular form, we have analogous notions of cuspidal forms, Hecke operators and eigenforms. Also, we can similarly define the Galois representation $\rho_{E,p} : G_F \to \mathrm{GL}_2(\mathbb{Q}_p)$ associated to an elliptic curve E over a real quadratic field F and $\rho_{f,p}$ associated to a Hilbert modular form f over F as section 3. Now we can state the definition of a modular representation and the modularity theorem over a real quadratic field.

Definition 7.2. Let p be a prime, F be a real quadratic field, and E/F be an elliptic curve. $\rho_{E,p}$ is modular if $\rho_{E,p} \sim \rho_{f,p}$ for some Hilbert cuspidal eigenform f over F of parallel weight 2. Similarly, we can define the notion of modularity for a residual representation $\overline{\rho}_{E,p}$.

Theorem 7.3. *[7, Theorem 1] Let p be a prime, F be a real quadratic field, and E/F be an elliptic curve. Then $\rho_{E,p}$ is modular.*

The paper [7] adopts the following steps to prove it, taken from the introduction of [7].

- Step 1: If $\overline{\rho}_{E,p}$ is modular and the image of $\overline{\rho}_{E,p}$ is "sufficiently large", then $\rho_{E,p}$ is modular.
- Step 2: If $\overline{\rho}_{E,3}$ is irreducible then it is modular.
- Step 3: If $\overline{\rho}_{E,5}$ is irreducible then it is modular.
- Step 4: By the previous steps, $\rho_{E,p}$ is modular except possibly if the image of $\overline{\rho}_{E,p}$ is "small" for $p = 3$ and 5 simultaneously. This step deals with the exceptions by considering rational points on certain modular curves.

The general version of modularity theorem 7.3 has some applications. The first application is to generalize the modularity theorem by the same method even further to elliptic curves over a particular totally real field satisfying some technical conditions, see [7, Theorem 7]. The second application is dealing with Fermat's equation over a real quadratic field. To be more precise, if a square-free integer $d > 6$ satisfies $d \equiv 3, 6, 10, 11 \pmod{16}$, then there is an effectively computable constant B_d such that for all primes $p > B_d$, the Fermat's equation $x^p + y^p = z^p$ has no nontrivial solution (x, y, z) over $\mathbb{Q}(\sqrt{d})$. See [7, section 6.2] and [8, Theorem 1].

Appendix A The Picard Group and the Jacobian

A.1 Introduction of the Picard Group and the Jacobian

First we introduce the notion of Picard groups.

Definition A.1. [6, p.214] Let X be an algebraic curve. The Picard group of X is defined by $\mathrm{Pic}^0(X) := \mathrm{Div}^0(X)/\mathrm{Prin}(X)$, the divisors of degree 0 quotient by the principal divisors.

Next we introduce the notion of the Jacobian.

Definition A.2. [6, p.214] Let X be a compact Riemann surface. The Jacobian of X is given by $\mathrm{Jac}(X) := \Omega^1_{\mathrm{hol}}(X)^*/H_1(X; \mathbb{Z})$, where $\Omega^1_{\mathrm{hol}}(X)^*$ is the dual space of holomorphic differentials of degree 1, generated by some $\mathbb{R}$-combination of integrals over some loops, and $H_1(X; \mathbb{Z})$ is the first homology group of X with integer coefficients, generated by a $\mathbb{Z}$-combination of such.

Lemma A.3. *[6, p.214] Let g be the genus of a compact Riemann surface X. Then* $\mathrm{Jac}(X) \simeq \mathbb{C}^g/\Lambda_g$ *for some lattice* Λ_g *in* $\mathbb{C}^g$*, which makes* $\mathrm{Jac}(X)$ *a g-dimensional complex torus.*

Proof. It follows from $\Omega^1_{\mathrm{hol}}(X)^* \simeq \mathbb{C}^g$ and $H_1(X;\mathbb{Z})$ is isomorphic to a lattice Λ_g in $\mathbb{C}^g$. □

For $X = X(\Gamma)_\mathbb{C}$, it can be shown that there is an isomorphism $\omega : S_2(\Gamma_\mathbb{C}) \simeq \Omega^1_{\mathrm{hol}}(X(\Gamma)_\mathbb{C})$ [6, p.82], and hence $\mathrm{Jac}(X(\Gamma)_\mathbb{C}) \simeq S_2(X(\Gamma)_\mathbb{C})^*/\omega^*(H_1(X(\Gamma)_\mathbb{C};\mathbb{Z}))$ where ω^* is the dual map of ω.

The Picard group and Jacobian are related by Abel's theorem.

Theorem A.4 (**Abel**). *[6, p.214] For a compact Riemann surface X,* $\mathrm{Pic}^0(X) \simeq \mathrm{Jac}(X)$.

A.2 The Action of the Hecke Algebra on the Picard Group

First we describe how a double coset operator induces a map between divisors.

Definition A.5. [6, p.167] For a weight-2 double coset $\Gamma_1\alpha\Gamma_2$ operator $|_2[\Gamma_1\alpha\Gamma_2] : S_2(\Gamma_1) \to S_2(\Gamma_2)$ with β_j defined in definition 2.17, it induces a map

$$X(\Gamma_2) \to \mathrm{Div}(X(\Gamma_1)), \quad \Gamma_2\tau \mapsto \sum_j \Gamma_1\beta_j(\tau).$$

By extending it $\mathbb{Z}$-linearly, the operator induces a map $\mathrm{Div}(X(\Gamma_2)) \to \mathrm{Div}(X(\Gamma_1))$.

In particular, for $T = T_p$ or $\langle\delta\rangle$ (where p is prime and δ is coprime to N), it induces a map $\mathrm{Div}(X_1(N)) \to \mathrm{Div}(X_1(N))$ by definition A.5, i.e., T acts on $\mathrm{Div}(X_1(N))$.

Now we do the following three steps.

- Step 1: The action of T on $\mathrm{Div}(X_1(N))$ passes to the action of T on $\mathrm{Div}(X_1(N)_\mathbb{Q})$.
- Step 2: The action of T on $\mathrm{Div}(X_1(N)_\mathbb{Q})$ descends to the action of T on $\mathrm{Div}^0(X_1(N)_\mathbb{Q})$.
- Step 3: The action of T on $\mathrm{Div}^0(X_1(N)_\mathbb{Q})$ descends to the action of T on $\mathrm{Pic}^0(X_1(N)_\mathbb{Q})$.

The details of Step 1 can be seen in the reference [6, Chapter 7]. Step 2 follows from the fact that the action on $\mathrm{Div}(X_1(N)_\mathbb{Q})$ preserves the degree 0. Step 3 follows from the fact that the action on $\mathrm{Div}^0(X_1(N)_\mathbb{Q})$ preserves the principal divisors.

These three steps determine the action of $T = T_p$ or $\langle\delta\rangle$ on $\mathrm{Pic}^0(X_1(N)_\mathbb{Q})$, and hence gives us the action of the Hecke algebra $\mathbb{T}_\mathbb{Z}$ on $\mathrm{Pic}^0(X_1(N)_\mathbb{Q})$.

A.3 The Action of the Hecke Algebra on the Jacobian

For a weight-2 double coset $\Gamma_1\alpha\Gamma_2$ operator $|_2[\Gamma_1\alpha\Gamma_2] : S_2(\Gamma_1) \to S_2(\Gamma_2)$, it has the pullback

$$|_2[\Gamma_1\alpha\Gamma_2]^* : S_2(\Gamma_2)^* \to S_2(\Gamma_1)^*.$$

The pullback descends to a well-defined map [6, p.228]

$$|_2[\Gamma_1\alpha\Gamma_2]^* : \mathrm{Jac}(X(\Gamma_2)_\mathbb{C}) \to \mathrm{Jac}(X(\Gamma_1)_\mathbb{C}).$$

By the definition of the pull-back,

$$\forall \psi \in S_2(\Gamma_2)^*, \quad |_2[\Gamma_1\alpha\Gamma_2]^*([\psi]) = [\psi \circ (|_2[\Gamma_1\alpha\Gamma_2])]$$

where $[\psi] \in \mathrm{Jac}(X(\Gamma_2)_\mathbb{C})$ is the image of ψ under the quotient map $S_2(\Gamma_2)^* \twoheadrightarrow \mathrm{Jac}(X(\Gamma_2)_\mathbb{C})$.

In particular, for $T = T_p$ or $\langle\delta\rangle$ (where p is a prime and δ coprime to N), the above gives

$$T^* : J_1(N) \to J_1(N), \quad T^*([\psi]) = [\psi \circ T] \text{ for } \psi \in S_2(\Gamma_1(N))^*$$

(Recall $J_1(N) = \mathrm{Jac}(X_1(N)_\mathbb{C})$).

This determines the action of $T = T_p$ or $\langle\delta\rangle$ on $J_1(N)$ via T^*, and hence gives us the action of the Hecke algebra $\mathbb{T}_\mathbb{Z}$ on $J_1(N)$.

References

[1] Section 39.4 (022r): Group schemes—the stacks project. URL: https://stacks.math.columbia.edu/tag/022R.

[2] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor. On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises. *Journal of the American Mathematical Society*, 14:843–939, 10 2001. URL: https://www.ams.org/journals/jams/2001-14-04/S0894-0347-01-00370-8/, doi:10.1090/S0894-0347-01-00370-8.

[3] Jan Hendrik Bruinier. Hilbert modular forms and their applications. *arXiv (Cornell University)*, pages 105–179, 01 2008. doi:10.1007/978-3-540-74119-0_2.

[4] Gary Cornell, Joseph H Silverman, and Glenn Stevens. *Modular forms and Fermat's last theorem*. Springer, 1997.

[5] Henri Darmon, Fred Diamond, and Richard Taylor. Fermat's last theorem, 2007. URL: https://people.math.wisc.edu/~nboston/ddt.pdf.

[6] Fred Diamond and Jerry Shurman. *A First Course in Modular Forms*. Springer, 2005.

[7] Nuno Freitas, Bao V Le, and Samir Siksek. Elliptic curves over real quadratic fields are modular. *Inventiones Mathematicae*, 201:159–206, 11 2015. doi:10.1007/s00222-014-0550-z.

[8] Nuno Freitas and Samir Siksek. The asymptotic fermat's last theorem for five-sixths of real quadratic fields. *Compositio Mathematica*, 151:1395–1415, 08 2015. doi:10.1112/s0010437x14007957.

[9] Álvaro Lozano-Robledo. *Elliptic curves, modular forms, and their L-functions*. American Mathematical Society ; Princeton, N.J, 2011.

[10] Barry Mazur. Rational isogenies of prime degree. *Inventiones Mathematicae*, 44:129–162, 06 1978. doi:10.1007/bf01390348.

[11] K. A. Ribet. On modular representations of $(\bar{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms. *Inventiones Mathematicae*, 100:431–476, 12 1990. doi:10.1007/bf01231195.

[12] Karl Rubin. Modularity of mod 5 representations. *Springer eBooks*, pages 463–474, 01 1997. doi:10.1007/978-1-4612-1974-3_16.

[13] Joseph H Silverman. *Advanced topics in arithmetic of elliptic curves*. Springer-Verlag, 1994.

[14] Joseph H Silverman. *The Arithmetic of elliptic curves*. Springer-Verlag, 2009.

[15] Andrew Wiles. Modular elliptic curves and fermat's last theorem. *The Annals of Mathematics*, 141:443, 05 1995. doi:10.2307/2118559.