Part III Essays Easter 2023

E511C

Attach to Front of Essay

Essay Number: 032

Essay Title:

The Modularity Theorem and the Modular Approach

Contents

Introduction

In 1637 Pierre de Fermat posited that the equation

$$x^n + y^n = z^n \tag{1}$$

has no integer solutions with $xyz \neq 0$ and $n > 2$. This conjecture became known as *Fermat's Last Theorem.* It suffices to consider the cases where n is an odd prime, since if $p \mid n$ and (a, b, c) is a solution to $x^n + y^n = z^n$, then $(a^{n/p}, b^{n/p}, c^{n/p})$ is a solution to $x^p + y^p = z^p$.

Fermat's Last Theorem is remarkable, not only because of how long it took to prove, but also because of the substantial amount of number theory developed with the aim of proving the result.

For instance, Kummer proved that Fermat's Last Theorem was true for regular primes by studying cyclotomic extensions of $\mathbb{Q}$. In looking for a proof, he introduced ideal class groups and the class number; components of the backbone of algebraic number theory.

It would take much more theory to arrive to a proof of Fermat's Last Theorem. In this essay, we focus on the route that proved successful and culminated in Wiles' proof [Wil95] in 1995. This is the so-called modular approach: deriving results using the modularity of elliptic curves (defined in §3).

Before the modular approach, attempts at proving Fermat's Last Theorem typically involved factoring $x^n + y^n$ in cyclotomic extensions. Then in 1984, G.Frey suggested that a proof of the *Taniyama-Shimura conjecture* may prove Fermat's Last Theorem by means of contradiction. The method was to associate to a solution (a, b, c) of (1) with n prime the following elliptic curve

$$E\colon y^2 = x(x - a^n)(x + b^n).$$

This elliptic curve has such strange properties that Frey suspected it would contradict the statement of the Taniyama-Shimura conjecture. Ribet later confirmed Frey's suspicion; by proving the now-called *Ribet's level lowering theorem* in [Rib90], he confirmed that a proof of the Taniyama-Shimura conjecture implies a proof of Fermat's Last Theorem.

Taniyama and Shimura proposed their conjecture in the 1960s. In its most primitive form, it states that to every rational elliptic curve E we can associate a modular form f whose Fourier coefficients encode the reduction properties and arithmetical information of E. This is useful because modular forms are analytical objects whose behaviour is better understood, and also because we can compute dimensions of spaces of types of modular forms.

Since its proof [Wil95] by Wiles, the Taniyama-Shimura conjecture is now referred to as the Modularity Theorem. Similarly, Ribet's level lowering theorem has been renamed since proof. It was originally known as Serre's ε-conjecture, and proposed by Serre in [Ser87].

These results are rooted in the theory of Galois representations. To elliptic curves and certain modular forms known as *newforms*, one can attach two-dimensional representations of the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. These arise by letting the group act on the torsion points of the elliptic curve, or certain abelian varieties. This is worthwhile because much of the information about the original object is then encoded in this representation. The more modern statement of the Modularity Theorem is that one can match to the Galois representation of each rational elliptic curve a Galois representation of a newform.

Thus, the Modularity Theorem is based in three areas of number theory; elliptic curves, modular forms, and Galois representations. This essay explores the interplay between these areas so that one can understand their relevance in the proof of a very simply posed, but formidable problem (1).

Outline of essay

The purpose of this essay is to introduce the statement of the Modularity Theorem in terms of Galois representations and to show how it implies Fermat's Last Theorem.

We provide background material in the first section. Firstly, we recall necessary properties of elliptic curves. We also introduce modular curves and modular forms. We interpret these geometrically so that we can explain later how Galois representations are attached to these objects. We introduce Galois representations without assuming the reader has any prior exposure. We give a thorough introduction to this theory so that the reader can apply their knowledge in other contexts also.

In §2 we describe the two-dimensional Galois representations associated to elliptic curves and newforms. We study the properties of these, and consider the arithmetical data they possess.

The theory of the previous section will enable us to state our main theorems in representation-theoretic terms. In §3 we discuss two substantial theorems: the Modularity Theorem and Ribet's level lowering theorem. With an aim of using these results towards proving Fermat's Last Theorem and solving other Diophantine equations, we consider the implications of these results for Galois representations associated to rational elliptic curves.

At this point we will have discussed all the necessary tools to be able to prove Fermat's Last Theorem, and give a detailed account of how the Modularity Theorem implies Fermat's Last Theorem in §4.

Finally, in §5, we briefly explore some more of the far-reaching consequences of the Modularity Theorem, by using the results of the essay to solve other Diophantine equations.

1 Background material

1.1 Elliptic curves

We assume some familiarity with the topic of elliptic curves. In this section we summarize reduction of an elliptic curve, and introduce the Tate curve associated to an elliptic curve.

1.1.1 Introduction

First we recall the Weierstrass equation of an elliptic curve and quantities associated to it, following [Sil1, Chapter III §1].

Let K be a perfect field with algebraic closure $\overline{K}$.

Definition 1.1 (Elliptic curve over K)**.** An elliptic curve is a smooth projective curve over K of genus 1 with a specified basepoint. Using non-homoegeneous coordinates, any elliptic curve E/K can be written in Weierstrass form:

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{2}$$

with $a_i \in K$, and basepoint denoted $\mathcal{O}_E$.

To an elliptic curve we associate the following standard quantities:

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	$= b_2^3 + 36b_2b_4 - 216b_6$
Δ	$= -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$	j	$= c_4^3/\Delta$

The last two are the *discriminant* Δ and *j-invariant* j respectively. The curve defined by (2) is non-singular if and only if $\Delta \neq 0$. The j-invariant classifies isomorphism classes; two elliptic curves are isomorphic over $\overline{K}$ if and only if they have the same j-invariant (see [Sil1, Chapter 4, Proposition 1.4] for both these results).

We denote the points of an elliptic curve over a field K by $E(K)$. These form a group with addition given by the "chord and tangent process" ([Sil1, Chapter III, §2]).

1.1.2 Reduction of an elliptic curve

Now we summarize reduction of an elliptic curve, following [Sil1, Chapter VII].

Let K be a finite extension of $\mathbb{Q}_p$, with valuation v, valuation ring $\mathcal{O}_K$, uniformizer $\pi \in \mathcal{O}_K$, and residue field $k \coloneqq \mathcal{O}_K/(\pi)$.

Definition 1.2 (Integral and minimal Weierstrass equations, [Sil1, Chapter VII, §1])**.** Let E/K be an elliptic curve with Weierstrass equation (2). Then (2) is integral if $a_i \in \mathcal{O}_K$, and additionally minimal if the valuation $v(\Delta)$ is minimal amongst all integral Weierstrass equations for E.

Remark 1.3 ([Sil1, Chapter VII, §1 Remark 1.1])**.** An integral Weierstrass equation has $v(\Delta) \geq 0$ (by definition of Δ). Since v is discrete, minimal Weierstrass equations thus exist. If an integral Weierstrass equation has $v(\Delta) < 12$, or $v(c_4) < 4$, or $v(c_6) < 6$, then it is minimal.

The reduction of E modulo π is defined as the curve $\tilde{E}/k$ obtained by reducing the coefficients of a minimal Weierstrass equation mod π. The curve $\tilde{E}/k$ is not necessarily non-singular. Its non-singular points form a group denoted by $\tilde{E}_{ns}(\overline{k})$.

If $v(\Delta) = 0$, $\tilde{E}/k$ is an elliptic curve, and we say E has good reduction. Else E has bad reduction.

Table 1: Summary of reduction types ([Sil1, Chapter VII, §5, Proposition 5.1]).

Good reduction:	$\tilde{E}/k$ is non-singular.	$\tilde{E}_{ns}(\overline{k}) = \tilde{E}(\overline{k})$.
Multiplicative split reduction:	$\tilde{E}/k$ has a nodal singularity. Slopes of tangent lines at singular point defined over k.	$\tilde{E}_{ns}(\overline{k}) \cong \mathbb{G}_m(\overline{k}) \cong (\overline{k}^\times, \cdot)$.
Multiplicative non-split reduction:	$\tilde{E}/k$ has a nodal singularity. Slopes of tangent lines at singular point *not* defined over k.	$\tilde{E}_{ns}(\overline{k}) \cong \mathbb{G}_m(\overline{k} \cong (\overline{k}^\times, \cdot)$.
Additive reduction:	$\tilde{E}/k$ has a cuspidal singularity.	$\tilde{E}_{\mathrm{ns}}(\overline{k}) \cong \mathbb{G}_a(\overline{k}) \cong (\overline{k}, +)$.

Now let K be a number field, E/K an elliptic curve. For a prime $\mathfrak{p} \subset \mathcal{O}_K$, let $K_\mathfrak{p}$ denote the $\mathfrak{p}$-adic completion of K with respect to the absolute value on K determined by $\mathfrak{p}$. The reduction type of an elliptic curve E/K at $\mathfrak{p}$ is defined to be the reduction type of $E/K_\mathfrak{p}$. We collect information about the reduction of E at different primes $\mathfrak{p}$ by defining an ideal of $\mathcal{O}_K$ known as the conductor.

Definition 1.4 (Conductor of E/K, [Sil2, Chapter IV, §10])**.** Let K be a number field, E/K an elliptic curve. The conductor $\mathfrak{n} \subset \mathcal{O}_K$ is an ideal divisible by prime ideals of bad reduction only. Explicitly,

$$\mathfrak{n} = \prod_{\substack{\mathfrak{p} \subset \mathcal{O}_K \\ \text{prime}}} \mathfrak{p}^{e_\mathfrak{p}}$$

$$\text{where } e_\mathfrak{p} = \begin{cases} 0 & \text{if } E \text{ has good reduction at } \mathfrak{p}, \\ 1 & \text{if } E \text{ has multiplicative reduction at } \mathfrak{p}, \\ 2 & \text{if } E \text{ has additive reduction at } \mathfrak{p}, \text{ and } \mathfrak{p} \nmid 2,\ \mathfrak{p} \nmid 3, \\ 2 \le e_\mathfrak{p} \le 2 + 6v_\mathfrak{p}(2) + 3v_\mathfrak{p}(3) & \text{if } E \text{ has additive reduction at } \mathfrak{p}, \text{ and } \mathfrak{p} \mid 2 \text{ or } \mathfrak{p} \mid 3. \end{cases}$$

Since $v_\mathfrak{p}(\Delta) \neq 0$ for finitely many primes, this is well-defined. When $\mathfrak{n}$ is squarefree, i.e. E has nowhere additive reduction, E is said to be semistable.

If K has class number 1, then E/K has a global minimum Weierstrass equation, which is a Weierstrass equation that is minimal at each prime $\mathfrak{p}$ of $\mathcal{O}_K$ ([Sil1, Chapter VIII, Corollary 8.3]). This can be computed using Tate's algorithm, and its discriminant is known as the minimal discriminant, sometimes denoted $\Delta_{\min}$.

Definition 1.5. Let $E/\mathbb{Q}$ be an elliptic curve. If E has good reduction at a prime p then the reduction $\tilde{E}/\mathbb{F}_p$ is an elliptic curve and we define

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

We additionally set

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

A theorem of Hasse asserts that $|a_p(E)| \leq 2\sqrt{p}$ ([Sil1, Chapter V, §1, Theorem 1.1].)

1.1.3 The Tate curve

Let $\mathcal{H}$ be the upper-half complex plane. The uniformization theorem for complex elliptic curves ([Sil1, Chapter VI, §5, Proposition 5.2]) shows that for every elliptic curve $E/\mathbb{C}$, there exists $\tau \in \mathcal{H}$ such that

$$E(\mathbb{C}) \cong \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z}) \cong \mathbb{C}^\times/q^{\mathbb{Z}}$$

with $q = e^{2\pi i\tau}$, $q^{\mathbb{Z}} = \{q^n \colon n \in \mathbb{Z}\}$. The second isomorphism is the exponentiation map $\exp(2\pi i \cdot)$. Note that $|q| < 1$.

Additionally, by a change of variables the Weierstrass equation for $E_q \coloneqq E$ can be written as

$$E_q \colon y^2 + xy = x^3 + a_4(q)x + a_6(q)$$

with $a_4(q), a_6(q) \in \mathbb{Z}[[q]]$ convergent power series (see [Sil2, Chapter V, Theorem 1.1] for definitions). The discriminant and j-invariant of E_q are then given by

$$\Delta(q) = q\prod_{n\geq 1}(1-q^n)^{24} \in \mathbb{Z}[[q]], \tag{3}$$

$$j(q) = \frac{1}{q} + 744 + 1968849q + \cdots = \frac{(1+48a_4(q))^3}{\Delta(q)} \in \mathbb{Z}[[q]], \tag{4}$$

where these are both convergent power series. These series are known as the modular discriminant and j-invariant respectively.

There is an analogous story if we consider elliptic curves over a finite extension K of $\mathbb{Q}_p$. The following is due to Tate.

Theorem 1.6 ([Sil2, Chapter V, Theorem 3.1])**.** *Let $K/\mathbb{Q}_p$ be a finite extension, and $q \in K^\times$ with $|q| < 1$. Let $a_4(q)$, $a_6(q)$ be the series defined as in the complex case.*

Then $a_4(q)$ and $a_6(q)$ converge in K. The Tate curve E_q is then an elliptic curve over K, where

$$E_q \colon y^2 + xy = x^3 + a_4(q)x + a_6(q).$$

It has discriminant $\Delta(q)$ as in (3) and j-invariant $j(E_q) = j(q)$ as in (4).

Moreover, there is an isomorphism of G_K-modules

$$\overline{K}^\times/q^{\mathbb{Z}} \xrightarrow{\sim} E_q(\overline{K}).$$

Remark 1.7. The group $G_K \coloneqq \mathrm{Gal}(\overline{K}/K)$ is discussed in §1.3.

Consider a rational elliptic curve $E/\mathbb{Q}$ which has multiplicative reduction at a prime p. There is an unramified extension $K/\mathbb{Q}_p$ with $[K\colon \mathbb{Q}_p] \leq 2$ such that E/K has split multiplicative reduction ($[K\colon \mathbb{Q}_p] = 2$ if $E/\mathbb{Q}_p$ has non-split reduction).

Since $p|\Delta$, we have $p^{-1}|j(E)$, and $|j(E)| > 1$ in $\mathbb{Q}_p$ and in K. Then Tate's p-adic uniformization theorem tells us that E is isomorphic to a Tate curve:

Theorem 1.8 ([Sil2, Chapter V, Theorem 5.3])**.** *Let K be a finite extension of $\mathbb{Q}_p$, and let E/K be an elliptic curve with $|j(E)| > 1$.*

There is a unique $q \in K^\times$ with $|q| < 1$ such that E is isomorphic over $\overline{K}$ to the Tate curve E_q. Additionally, E is isomorphic to E_q over K if and only if E has split multiplicative reduction.

Corollary 1.9. *If $K/\mathbb{Q}_p$ is a finite extension and E/K has split multiplicative reduction then $\exists! q \in K^\times$ with $|q| < 1$ such that*

$$E(\overline{K}) \cong \overline{K}^\times / q^{\mathbb{Z}}$$

as G_K-modules.

Proof. Since $E \simeq E_q$ over K it follows that the isomorphism of G_K-modules in 1.6 is also true for E. □

Then the following description of the l-torsion of E will prove useful later.

Corollary 1.10. *Let $K/\mathbb{Q}_p$ be a finite extension and E/K an elliptic curve with split multiplicative reduction. Then for $l \neq p$ a prime, we have the following isomorphism of G_K-modules:*

$$E[l] \simeq \langle \zeta_l, \sqrt[l]{q} \rangle \subset \overline{K}^\times / q^{\mathbb{Z}}.$$

1.2 Modular forms and modular curves

Firstly, we will introduce moduli spaces and modular curves. We will not discuss the technical definitions of our curves as Riemann surfaces. A good source for this content is [DS05, Chapter 2], and we also follow [DDT95, Chapter 1].

Next, we consider modular forms on congruence subgroups. We restrict to weight two modular forms, being the only weight that arises in this essay.

The statement of the Modularity Theorem involves special modular forms known as new-forms. To define these we introduce Hecke operators, mostly following [DS05, Chapter 5].

1.2.1 Modular curves and moduli spaces

The modular group $\mathrm{SL}_2(\mathbb{Z})$ acts on the upper-half complex plane $\mathcal{H}$ by Möbius transformations. It acts transitively on $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$ via $\gamma \cdot z = \frac{az+b}{cz+d}$ for $z \in \mathbb{Q}$ and $\gamma \cdot \infty = \frac{a}{c}$.

We define the same action of a finite index subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$ and $\mathbb{P}^1(\mathbb{Q})$. The orbit classes of the action on $\mathbb{P}^1(\mathbb{Q})$ are called *cusps*, of which there are finitely many ([DS05, Lemma 2.4.1]). The finite index subgroups we will be most interested in are known as congruence subgroups.

Definition 1.11. Let $N \geq 1$ be an integer.

1. (Principal congruence subgroup) The principal congruence subgroup $\Gamma(N) \subset \mathrm{SL}_2(\mathbb{Z})$ is $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$, the kernel of the map reducing matrix coefficients mod N.

2. (Congruence subgroup) $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup if it contains $\Gamma(N)$ for some $N \geq 1$.

Since $\Gamma(N)$ is a finite index subgroup, it follows that congruence subgroups are finite index subgroups of $\mathrm{SL}_2(\mathbb{Z})$ also. Note that $\mathrm{SL}_2(\mathbb{Z}) = \Gamma(1)$. The following are the most important examples.

Example 1.12 (Hecke subgroups)**.**

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

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

Note that $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ via $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \pmod{N}$.

For $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ a congruence subgroup, denote by Y_Γ the quotient of $\mathcal{H}$ by the action of Γ. By equipping Y_Γ with the analytic structure coming from the projection map $\pi\colon \mathcal{H} \to Y_\Gamma$, one can show that Y_Γ is a Riemann surface. ([DS05, Chapter 2, Corollary 2.1.2]).

The compactification of Y_Γ, denoted by X_Γ, is formed by adding the cusps corresponding to the action of Γ on $\mathbb{P}^1(\mathbb{Q})$. We call Y_Γ and X_Γ *modular curves*. Topologically, X_Γ is a g-holed torus for some $g \geq 1$, an integer known as the *genus* of the surface.

In particular, for $\Gamma = \Gamma_0(N)$, we write $Y_\Gamma = Y_0(N)$, $X_\Gamma = X_0(N)$, and likewise for $\Gamma = \Gamma_1(N)$, $Y_\Gamma = Y_1(N)$, $X_\Gamma = X_1(N)$.

As an incentive, let us state a version of the Modularity Theorem:

Theorem 1.13 (Modularity Theorem in terms of Riemann surfaces)**.** *Let $E/\mathbb{C}$ be an elliptic curve with rational j-invariant $j(q) \in \mathbb{Q}$ as in (4), for some $q \in \mathbb{C}^\times$.*

Then for some positive integer N there exists a surjective holomorphic function of compact Riemann surfaces from the modular curve $X_0(N)$ to the elliptic curve E, known as the modular parametrization of E.

It will become clear in §2 when we introduce Galois representations of elliptic curves why the curve $X_0(N)$ is relevant.

Equivalence classes of $Y_0(1) = Y_1(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ describe the isomorphism classes of complex elliptic curves $E/\mathbb{C}$ (cf. Uniformization theorem for complex elliptic curves). We call $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ a *moduli space* for complex elliptic curves. We can also describe $Y_0(N)$ and $Y_1(N)$ as moduli spaces for arbitrary N.

- For $\Gamma_0(N)$, consider pairs (E, C), with C a cyclic subgroup of order N of $E/\mathbb{C}$. Define an equivalence $(E, C) \sim (E', C')$ if $E \cong E'$, and the isomorphism maps C to C'. Let $[E, C]$ denote an equivalence class, and $S_0(N)$ the set of equivalence classes.
- For $\Gamma_1(N)$, consider pairs (E, Q), with Q a point of exact order N of $E/\mathbb{C}$. Define an equivalence $(E, Q) \sim (E', Q')$ if $E \cong E'$ are isomorphic, and the isomorphism maps Q to Q'. Let $[E, Q]$ denote an equivalence class, and $S_1(N)$ the set of equivalence classes.

Note that

$$S_0(N) = \{[E_\tau, \langle 1/N + \Lambda_\tau \rangle] \mid \tau \in \mathcal{H}\}, \quad S_1(N) = \{[E_\tau, 1/N + \Lambda_\tau] \mid \tau \in \mathcal{H}\}.$$

Theorem 1.14. *[DS05, Theorem 1.5.1]*

1. *$Y_0(N)$ is a moduli space of $S_0(N)$, i.e. equivalence classes $[E_\tau, \langle \frac{1}{N} + \Lambda_\tau \rangle]$ and $[E_{\tau'}, \langle \frac{1}{N} + \Lambda_{\tau'} \rangle]$ are equal if and only if τ and τ' lie in the same $\Gamma_0(N)$-orbit.*
2. *$Y_1(N)$ is a moduli space of $S_1(N)$, i.e. equivalence classes $[E_\tau, \frac{1}{N} + \Lambda_\tau]$ and $[E_{\tau'}, \frac{1}{N} + \Lambda_{\tau'}]$ are equal if and only if τ and τ' lie in the same $\Gamma_1(N)$-orbit.*

Then maps of modular curves become maps of moduli spaces. For example, the map from $Y_1(N)$ to $Y_0(N)$ sending the orbit $\Gamma_1(N)\tau$ to $\Gamma_0(N)\tau$ corresponds to the map from $S_1(N)$ to $S_0(N)$ sending $[E, Q]$ to $[E, \langle Q \rangle]$. We will see another example of this when we define diamond operators shortly.

1.2.2 Weight 2 modular forms

Definition 1.15 (Weak modular forms of weight 2)**.** A holomoprhic function $f\colon \mathcal{H} \to \mathbb{C}$ is a weak modular form of weight 2 for a congruence subgroup Γ if

$$f(\gamma\tau) = (c\tau + d)^2 f(\tau) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

Remark 1.16. If f is a weakly modular of weight 2, then $f(\tau)\, d\tau$ is Γ-invariant.

Definition 1.17 (Modular and cuspidal forms of weight 2)**.** Let f be a weak modular form of weight 2 for the congruence subgroup Γ.

Suppose that for all $\gamma \in \Gamma(1)$ the function $(c\tau + d)^{-2} f(\gamma\tau)$ has a Fourier expansion $\sum_{n=0}^{\infty} b_n e^{2\pi i/h}$ for some $h \in \mathbb{Z}^+$. Then f is a modular form of weight 2.

If the constant term of this Fourier expansion is zero for each $\gamma \in \Gamma$, then we call f a cusp form.

Denote by $M_2(\Gamma)$ and $S_2(\Gamma)$ the space of weight 2 modular and cusp forms corresponding to Γ respectively.

If $f \in S_2(\Gamma)$, then the invariance of $f(\tau)\, d\tau$ under the action of Γ means that it arises as the pullback of a differential on Y_Γ. The vanishing of f at the cusps means that this extends to a holomorphic differential on X_Γ. In fact, we have:

Theorem 1.18 ([DDT95, Lemma 1.12])**.** *The map*

$$f(\tau) \mapsto \omega_f \coloneqq 2\pi i f(\tau) d\tau$$

is an isomorphism between $S_2(\Gamma)$ and the space $\Omega^1(X_\Gamma)$ of holomorphic differentials on the curve X_Γ.

Corollary 1.19. *$S_2(\Gamma)$ is finite-dimensional, with dimension equal to the genus of X_Γ.*

One can compute the genus of g using the Riemann-Hurwitz formula, and knowledge of cusps and ellitpic points. This is surveyed in [DS05, Chapter 3]. We will just provide one example, which we will see to be very important later.

Example 1.20. There are no cusp forms of weight 2 for the congruence subgroup $\Gamma_1(2)$.

Using [DS05, Chapter 1, §1.2, Exercise 1.2.3] one can compute that $[\Gamma(1)\colon \Gamma_1(2)] = 3$. There is one elliptic point $\frac{1}{2}(1 + i)$ of $\Gamma_1(2)$ of order 2 ([DS05, Chapter 2, §2.4, Exercise 2.3.7]). There are two cusps, 0, and ∞. Then the Riemann-Hurwitz formula and [DS05, Chapter 3, Theorem 3.1.1] imply that the genus of $X_1(2)$ is $g = 1 + \frac{3}{12} - \frac{1}{4} - \frac{2}{2} = 0$. Thus $\dim_{\mathbb{C}} S_2(\Gamma_1(2)) = 0$.

1.2.3 Hecke algebra

In defining the Hecke algebra, we focus on the congruence subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$. Since $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right) \in \Gamma_1(N) \subset \Gamma_0(N)$, any modular form corresponding to these groups has a Fourier expansion $f = \sum_{n\geq 0} a_n(f)q^n$ where $q = e^{2\pi i\tau}$, and $a_1(f) = 0$ if f is a cusp form.

Recall that $\Gamma_0(N)/\Gamma_1(N) \simeq (\mathbb{Z}/N\mathbb{Z})^\times$. This quotient acts on $S_1(N)$ via

$$(\gamma + \Gamma_1(N)) \cdot [E_\tau, \frac{1}{N} + \Lambda_\tau] = [E_\tau, \frac{d}{N} + \Lambda_\tau] = [E_{\gamma\tau}, \frac{1}{N} + \Lambda_{\gamma\tau}], \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

The corresponding action on $Y_1(N)$ sends $\Gamma_1(N)\tau$ to $\Gamma_1(N)(\gamma\tau)$. We can extend this to an action on $X_1(N)$. Then the isomorphism of theorem 1.18 leads us to the definition of the diamond operator:

Definition 1.21 (Diamond operator)**.** Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \Gamma_1(N) \in \Gamma_0(N)/\Gamma_1(N)$ correspond to $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, and define the diamond operator $\langle d\rangle \colon S_2(\Gamma_1(N)) \to S_2(\Gamma_1(N))$ by

$$\langle d\rangle f = (c\tau + d)^{-2} f(\gamma\tau)$$

This definition is independent of coset representative. Given an even Dirichlet character $\chi \colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ ([Neu99, Chapter VII, §2]), we let

$$S_2(N, \chi) = \{f \in S_2(\Gamma_1(N)) \mid \langle d\rangle f = \chi(d)f \quad \forall d \in (\mathbb{Z}/N\mathbb{Z})^\times\} \subset S_2(\Gamma_1(N)).$$

We consider even Dirichlet characters, since otherwise the space would be empty, as $\langle -1\rangle f(\tau) = f(\tau)$. We also canonically identify $S_2(N, \mathbb{1})$ with $S_2(\Gamma_0(N))$, where $\mathbb{1}$ is the trivial character.

Then we have the decomposition

$$S_2(\Gamma_1(N)) = \bigoplus_\chi S_2(N, \chi)$$

where the sum ranges over all even Dirichlet characters mod N.

Next, we provide a little motivation for the definition of Hecke operators, following [DDT95, §1.3]. If $p \nmid N$, then there are exactly $p+1$ distinct p-isogenies from $(E_\tau, \frac{1}{N})$ whose images are

$$(E_{\frac{\tau+i}{p}}, \frac{1}{N}) \quad (i = 0, \dots p-1), \quad (E_{p\tau}, \frac{p}{N}).$$

If $p|N$ then the last isogeny is excluded since p/N is not of exact order N. This comes from considering the cyclic sublattices of index p for the lattice $\mathbb{C}/\langle\tau, 1\rangle$.

Definition 1.22 (Hecke operators)**.** Let p be prime. The Hecke operator $T_p \colon S_2(\Gamma_1(N)) \to S_2(\Gamma_1(N))$ is defined for $f \in S_2(\Gamma_1(N))$ as

$$T_p(f) = \begin{cases} \frac{1}{p}\sum_{i=0}^{p-1} f(\frac{\tau+i}{p}) + p\langle p\rangle f(p\tau) & \text{if } p \nmid N, \\ \frac{1}{p}\sum_{i=0}^{p=1} f(\frac{\tau+i}{p}) & \text{if } p \mid N. \end{cases}$$

If $p \nmid N$ then $\omega_{T_p(f)} = \sum \phi_i^*(\omega_f)$ where $\phi_i(\tau) = \frac{\tau+i}{p}$ for $i = 0, \dots p-1$ and $\phi_p(\tau) = \langle p\rangle p\tau$, corresponding to the isogenies listed above (recall definition of ω_f in theorem 1.18).

Proposition 1.23 ([DS05, Proposition 5.2.4]). *Let d, $e \in (\mathbb{Z}/N\mathbb{Z})^\times$ and p, q distinct primes. Then*

$$\langle d\rangle T_p = T_p\langle d\rangle, \qquad \langle d\rangle\langle e\rangle = \langle e\rangle\langle d\rangle = \langle de\rangle, \qquad T_pT_q = T_qT_p.$$

To define T_n for arbitrary n, we set $T_1 = 1$, and inductively define

$$T_{p^r} = T_pT_{p^{r-1}} - p\langle p\rangle T_{p^{r-2}}$$

for $r \geq 2$ and p a prime. Note that $T_{p^r}T_{p^s} = T_{p^s}T_{p^r}$. Then for $n = \prod p_i^{e_i}$, define

$$T_n = \prod T_{p_i^{e_i}}.$$

To define $\langle n\rangle$ for arbitrary n, if $(n, N) = 1$ then let $\langle n\rangle = \langle n \pmod N\rangle$. If $(n, N) > 1$, let $\langle n\rangle = 0$.

Definition 1.24 (Hecke algebra over $\mathbb{Z}$). Let

$$\mathbb{T}_\mathbb{Z} = \mathbb{Z}[\{T_n, \langle n\rangle \colon n \in \mathbb{N}\}]$$

be the subalgebra of $\mathrm{End}(S_2(\Gamma_1(N))$ generated over $\mathbb{Z}$ by T_n and $\langle n\rangle$ for all $n \in \mathbb{N}$.

One can also describe how the Hecke operators act in terms of the Fourier expansions of cusp forms, see [DS05, Proposition 5.3.1].

1.2.4 Newforms

Definition 1.25 (Eigenform and normalized eigenform). A cusp form $f = \sum_{n\geq 1} a_n(f)q^n \in S_2(\Gamma_1(N))$ is an eigenform if it is a simultaneous eigenvector for T_n, $\langle n\rangle$ for all $n \in \mathbb{N}$. Equivalently, there is an algebra homomorphism $\lambda_f \colon \mathbb{T}_\mathbb{Z} \to \mathbb{C}$ such that $Tf = \lambda_f(T)f$ for all $T \in \mathbb{T}_\mathbb{Z}$. If $a_1(f) = 1$, f is a normalized eigenform.

Normalized eigenforms thus have coefficients whose relations satisfy those of the Hecke operators, see [DS05, Proposition 5.8.4].

Up to now, we have considered a fixed level N, but we can move between levels. If $M \mid N$, we can embed $S_2(\Gamma_1(M))$ into $S_2(\Gamma_1(N))$ by the identity map. Another way is to let d be a factor of N/M and consider

$$[d] \colon S_2(\Gamma_1(M)) \to S_2(\Gamma_1(N)), \qquad f(\tau) \mapsto df(d\tau).$$

Thus, we want to distinguish the part of $S_2(\Gamma_1(N))$ coming from lower levels. We do so by defining the space of newforms. This is defined as an orthogonal complement with respect to the Petersson inner product on $S_2(\Gamma_1(N))$ ([DS05, Defintion 5.4.1]).

Definition 1.26 (Space of oldforms and newforms [DS05, Definition 5.6.1]). Let $N \geq 2$ be an integer. For each divisor d of N, let i_d be the map

$$i_d \colon (S_2(\Gamma_1(Nd^{-1})))^2 \to S_2(\Gamma_1(N)), \qquad (f, g) \mapsto f + [d](g)$$

The subspace of oldforms at level N is

$$S_2(\Gamma_1(N))^{\mathrm{old}} = \sum_{p|N \text{ prime}} i_p((S_2(\Gamma_1(N)))^2).$$

The subspace of newforms at level N is the orthogonal complement with respect to the Petersson inner product

$$S_2(\Gamma_1(N))^{\text{new}} = (S_2(\Gamma_1(N)))^{\perp}.$$

Definition 1.27 (Newform)**.** A newform is a normalized eigenform in $S_2(\Gamma_1(N))^{\text{new}}$.

One can show that $S_2(\Gamma_1(N))^{\text{new}}$ is preserved under the action of T_n and $\langle n \rangle$ for all n. (see [DS05, Proposition 5.6.3]). We also have the following important result:

Theorem 1.28 ([DS05, Theorem 5.8.2])**.** *The set of newforms in $S_2(\Gamma_1(N))^{\text{new}}$ forms an orthogonal basis of the space. Each newform is in an eigenspace $S_2(N, \chi)$ for some character χ and satisfies $T_n f = a_n(f) f$ for all $n \in \mathbb{N}$.*

1.3 Galois representations

1.3.1 Topology of absolute Galois groups

Galois representations are representations of absolute Galois groups, which we firstly introduce. We follow the material in [DS05, §9.3]. Another reference is [Neu99, Chapter IV, §1].

Let K be a perfect field, and $\overline{K}$ an algebraic closure of K.

Definition 1.29 (Absolute Galois group)**.** The absolute Galois group of K is $G_K = \text{Gal}(\overline{K}/K)$.

We have

$$\overline{K} = \bigcup_{L/K} L$$

where L runs over all finite Galois extensions of K. For $\sigma \in G_K$, the restriction $\sigma|_L$ is an element of $\text{Gal}(L/K)$ for all L, and $G_K \twoheadrightarrow \text{Gal}(L/K)$. Moreover, the restrictions are compatible: if $K \subset K' \subset K''$ and K'/K, K''/K are finite Galois extensions, then $(\sigma|_{K''})|_{K'} = \sigma|_{K'}$.

Therefore we can form an inverse system

$$G' := \varprojlim_{L} \text{Gal}(L/K) = \{(\sigma_L) \in \prod_{L/K} \text{Gal}(L/K) \ \mid \ (\sigma_L)|_M = \sigma_M \text{ whenever } K \subset M \subset L\}$$

over finite Galois extensions L/K. The absolute Galois group injects $G_K \hookrightarrow G'$, and conversely any compatible sequence in G' gives rise to an automorphism of $\overline{K}$, so that $G_K \simeq G'$.

This description of G_K identifies it as a *profinite group*, to which we attach the *profinite* or *Krull* topology. For L/K a finite Galois extension, equip $\text{Gal}(L/K)$ with the discrete topology. Then we endow G_K with the weakest topology that makes the projection maps $G_K \to \text{Gal}(L/K)$ continuous for all L.

More explicitly, a basis of this topology is given by

$$\text{cosets of } \text{Gal}(\overline{K}/L) \text{ in } G_K \equiv \text{ cosets of finite index normal subgroups of } G_K$$

where on the left hand side, L is a finite Galois extension of K. This equivalence is the fundamental theorem of Galois theory. It is clear from this description that G_K is a topological group. It is also compact ([Neu99, Chapter IV, §1, Proposition 1.1]).

1.3.2 Galois representations

We will define Galois representations as representations of $G_{\mathbb{Q}}$. Let $\mathbb{F}$ be a field that is also a topological group. The topology on $\mathrm{GL}_n(\mathbb{F})$ is the subspace topology given by considering $\mathrm{GL}_n(\mathbb{F}) \subset \mathbb{F}^{n\times n}$, where we equip $\mathbb{F}^{n\times n}$ with the product topology.

Definition 1.30 (Galois representation)**.** A Galois representation is a continuous group homomorphism

$$\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{F}).$$

Define $\dim \rho = n$. If $\rho'\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{F})$ is a Galois representation such that $\exists M \in \mathrm{GL}_n(\mathbb{F})$ with $\rho'(\sigma) = M\rho(\sigma)M^{-1}$ for all $\sigma \in G_{\mathbb{Q}}$, then we say that ρ and ρ' are *equivalent* and write $\rho \sim \rho'$.

Note that since ρ is a homomorphism of topological groups, ρ is continuous if and only if $\rho^{-1}(V)$ is open for each V in the basis of neighbourhoods of the identity matrix in $\mathrm{GL}_n(\mathbb{F})$. We note some consequences of the continuity of Galois representations.

Lemma 1.31 ([DS05, Exercise 9.3.4])**.** *If $\mathbb{F} = \mathbb{C}$ then $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{C})$ has finite image. We call ρ an Artin representation.*

Proof. We admit that there exists an open neighbourhood V of the identity in $\mathrm{GL}_n(\mathbb{C})$ containing no non-trivial subgroups. Let $U = \rho^{-1}(V)$. By continuity of ρ, U is an open neighbourhood of $1 \in G_{\mathbb{Q}}$. Therefore U contains a basis open set $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ for L a finite Galois extension of $\mathbb{Q}$.

Then $\rho(\mathrm{Gal}(\overline{\mathbb{Q}}/L))$ is a subgroup of V, and so must be trivial. Thus ρ factors as $\rho\colon G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(L/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$, and has finite image since $\mathrm{Gal}(L/\mathbb{Q})$ is finite. □

Lemma 1.32. *If $\mathbb{F}$ is a finite extension of $\mathbb{F}_p$, or $\overline{\mathbb{F}}_p$, then $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{F})$ has finite image. We call ρ a mod p Galois representation.*

Proof. $\mathrm{GL}_n(\mathbb{F}_{p^m})$ for all m and $\mathrm{GL}_n(\overline{\mathbb{F}_p})$ are topological groups with respect to the discrete topology. Thus $\ker \rho = \rho^{-1}(\mathrm{Id})$ is an open neighbourhood of $1 \in G_{\mathbb{Q}}$ by continuity of ρ, hence of the form $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ for L a finite Galois extension of $\mathbb{Q}$. Hence ρ has finite image. □

If $\mathbb{F} = \mathbb{Q}_p$ or a finite extension of $\mathbb{Q}_p$, then we call ρ a p-adic Galois representation. The following is a crtierion for continuity in a particular case:

Lemma 1.33. *Consider a p-adic representation $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{Q}_p)$ whose image is contained in $\mathrm{GL}_n(\mathbb{Z}_p) \subset \mathrm{GL}_n(\mathbb{Q}_p)$. Then ρ is continuous if and only if for all m there exists a finite Galois extension $F_m/\mathbb{Q}$ such that $\mathrm{Gal}(\overline{\mathbb{Q}}/F_m) \to \mathrm{Id} \pmod{p^m}$, i.e. $\rho \pmod{p^m}$ factors through a finite extension $F_m/\mathbb{Q}$.*

Proof. The open neighbourhoods of $\mathrm{Id} \in \mathrm{GL}_n(\mathbb{Z}_p)$ are

$$U_1(m) = \{A \in \mathrm{GL}_n(\mathbb{Z}_l) \mid A \equiv \mathrm{Id} \pmod{p^m}\},$$

for all m. If ρ is continuous then $\rho^{-1}(U_1(m))$ is open, hence contains the open set $\mathrm{Gal}(\overline{\mathbb{Q}}/F_m)$ for $F_m/\mathbb{Q}$ some finite Galois extension, and so $\rho(\mathrm{Gal}(\mathbb{Q}/F_m) \equiv \mathrm{Id} \pmod{p^m}$.

Conversely, if every $\rho^{-1}(U_1(m))$ contains an open subgroup $\mathrm{Gal}(\overline{\mathbb{Q}}/F_m)$, then $\rho^{-1}(U_1(m))$ is the union of cosets of $\mathrm{Gal}(\overline{\mathbb{Q}}/F_m)$ in $\rho^{-1}(U_1(m))$, and so is open. □

1.3.3 Galois representations at Frobenius elements

The following definitions can be found in [DS05, Chapter 9, §9.3].

Fix an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$. Let $p \in \mathbb{Z}$ be a prime, and $\mathfrak{p} \subset \overline{\mathbb{Z}}$ a maximal ideal lying over p. Define the decomposition group

$$D_{\mathfrak{p}} = \{\sigma \in G_{\mathbb{Q}} \mid \sigma(\mathfrak{p}) = \mathfrak{p}\}.$$

Then if $\sigma \in D_{\mathfrak{p}}$, σ acts on $\overline{\mathbb{Z}}/\mathfrak{p}$ as $\overline{\sigma}(a + \mathfrak{p}) = \sigma(a) + \mathfrak{p}$. Note that $\overline{\mathbb{Z}}/\mathfrak{p} \cong \overline{\mathbb{F}}_p$. Therefore there is a reduction map

$$D_{\mathfrak{p}} \twoheadrightarrow \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p) = G_{\mathbb{F}_p}.$$

Define the kernel of this map to be the inertia subgroup $I_{\mathfrak{p}}$.

A Frobenius element $\mathrm{Frob}_{\mathfrak{p}}$ is an inverse under this map of the Frobenius element $\sigma_p \in G_{\mathbb{F}_p}$, where $\sigma_p(x) = x^p$ for all $x \in \overline{\mathbb{F}}_p$. Note that this is defined modulo $I_{\mathfrak{p}}$.

Recall that $\mathbb{Q} \subset \mathbb{Q}_p$ is dense. An embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ corresponds to a choice of maximal ideal $\mathfrak{p} \subset \overline{\mathbb{Z}}$ lying over p. To keep track of the embedding we will write $\overline{\mathbb{Q}}_{\mathfrak{p}}$ instead of $\overline{\mathbb{Q}}_p$.

A consequence of Krasner's lemma ([Neu99, Chapter II, §6, Exercise 2]) is that $\overline{\mathbb{Q}} \subset \overline{\mathbb{Q}}_{\mathfrak{p}}$ is dense and $G_{\mathbb{Q}_p}$ embeds into $G_{\mathbb{Q}}$. It can be shown that the image is closed, and isomorphic to $D_{\mathfrak{p}}$ ([Neu99, Chapter II, Proposition 9.6]). Also, if $\mathfrak{p}, \mathfrak{p}' \subset \overline{\mathbb{Z}}$ are different prime ideals lying over p, then $D_{\mathfrak{p}}$ and $D_{\mathfrak{p}'}$ are conjugate subgroups.

Unramified extensions of $\mathbb{Q}_p$ are in one-to-one correspondence with unramified extensions of the residue field $\mathbb{F}_p$. Let $\mathbb{Q}_{\mathfrak{p}}^{\mathrm{un}}$ be the maximal unramified subextension of $\overline{\mathbb{Q}}_{\mathfrak{p}}/\mathbb{Q}_p$. Then $\mathrm{Gal}(\mathbb{Q}_{\mathfrak{p}}^{\mathrm{un}}/\mathbb{Q}_p) \cong G_{\mathbb{F}_p}$. It follows that

$$I_{\mathfrak{p}} \cong \mathrm{Gal}(\overline{\mathbb{Q}}_{\mathfrak{p}}/\mathbb{Q}_{\mathfrak{p}}^{\mathrm{un}}).$$

Now consider a representation $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{F})$.

Definition 1.34 (Unramified Galois representations)**.** A Galois representation $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{F})$ is unramified at a prime $p \in \mathbb{Z}$ if $I_{\mathfrak{p}} \subset \ker\rho$ for any choice of prime ideal $\mathfrak{p} \subset \overline{\mathbb{Z}}$ lying over p.

For this definition to make sense we need that $I_{\mathfrak{p}} \subset \ker\rho$ is independent of the choice of prime $\mathfrak{p}$. But $\ker\rho$ is a normal subgroup of $G_{\mathbb{Q}}$, hence if it contains $I_{\mathfrak{p}}$ for some prime $\mathfrak{p}$, it contains the conjugates of this subgroup, i.e. $I_{\mathfrak{p}'}$ for all other choices of $\mathfrak{p}' \subset \overline{\mathbb{Z}}$ lying over p.

Proposition 1.35. *Suppose that a Galois representation ρ factors as $\rho\colon G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(F/\mathbb{Q}) \to \mathrm{GL}_n(F)$ so that $\ker\rho = \mathrm{Gal}(\overline{\mathbb{Q}}/F)$ for $F/\mathbb{Q}$ some Galois extension. Then ρ is unramified at p if and only if F is unramified at p.*

Proof.

- –> If F is unramified at p, then under the embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_{\mathfrak{p}}$, F is contained in $\mathbb{Q}_{\mathfrak{p}}^{\mathrm{un}}$, for any choice of prime $\mathfrak{p} \subset \overline{\mathbb{Z}}$ lying over p. Therefore $I_{\mathfrak{p}} \subset \ker\rho$, again viewing $\ker\rho$ under the embedding $G_{\mathbb{Q}} \hookrightarrow G_{\mathbb{Q}_p} = D_{\mathfrak{p}}$. Thus ρ is unramified at p.
- <– Conversely, if for $\mathfrak{p} \subset \overline{\mathbb{Z}}$ we have $I_{\mathfrak{p}} \subset \ker\rho$, then by the Galois correspondence $F \subset \mathbb{Q}_{\mathfrak{p}}^{\mathrm{un}}$ where we embed F into $\overline{\mathbb{Q}}_{\mathfrak{p}}$. Therefore $F/\mathbb{Q}$ is unramified at p.

□

The property of ρ being unramified at a prime p is useful because it allows us to consider $\rho(\mathrm{Frob}_\mathfrak{p})$. If we suppress the choice of prime $\mathfrak{p}$ and denote the absolute Frobenius element by Frob_p, then this is defined up to conjugacy, and modulo inertia. Then if $I_\mathfrak{p} \subset \ker \rho$, it follows that $\rho(\mathrm{Frob}_p)$ is well-defined up to conjugacy. However, the trace and determinant of $\rho(\mathrm{Frob}_p)$ will be independent of choice of conjugate elements.

Evaluating a Galois representation at Frobenius elements is important because in certain cases, ρ is determined by its image on these elements. This is a consequence of Chebotarev's density theorem:

Theorem 1.36 (Chebotarev's density theorem, [Neu99, Chapter VII, Theorem 13.4])**.** *Let L/K be a finite Galois extension of number fields with Galois group G. Fix $\sigma \in G$ and let $c = |{}^G\sigma|$ be the size of the orbit of σ under G. Then the set of unramified primes $\mathfrak{p}$ of K with $\left(\frac{L/K}{\mathfrak{p}}\right) \in {}^G\sigma$ has density $c/|G|$.*

Remark 1.37. $\left(\frac{L/K}{\mathfrak{p}}\right)$ is the Artin symbol ([Neu99, Chapter VI, §7]).

Corollary 1.38 ([DDT95, Proposition 2.6])**.** *A semisimple Galois representation $\rho\colon G_\mathbb{Q} \to \mathrm{GL}_2(\mathbb{F})$ where $\mathbb{F}$ is a finite field, $\mathbb{F} = \overline{\mathbb{F}}_p$, or $\mathbb{F} = \mathbb{C}$, is determined by its image on Frobenius elements of unramified primes.*

Proof. Consider two Galois representations $\rho_1, \rho_2\colon G_\mathbb{Q} \to \mathrm{GL}_2(\mathbb{F})$ which are semisimple. These are continuous homomorphisms, so they factor through finite Galois extensions $L_1/\mathbb{Q}$, $L_2/\mathbb{Q}$ for ρ_1, ρ_2 respectively by our results in the previous section. Let L be the compositum of the extensions L_1 and L_2. Let $G = \mathrm{Gal}(L/\mathbb{Q})$.

Suppose these satisfy $\mathrm{Tr}(\rho_1(\mathrm{Frob}_p)) = \mathrm{Tr}(\rho_2(\mathrm{Frob}_p))$ and $\mathrm{Det}(\rho_1(\mathrm{Frob}_p)) = \mathrm{Det}(\rho_2(\mathrm{Frob}_p))$ for all primes p at which both are unramified. This is all but finitely many primes, since $L/\mathbb{Q}$ is ramified at only finitely many primes.

If $\sigma \in G$, then by the Chebotarev's density theorem there exists infinitely many unramified primes p of $\mathbb{Q}$ with $\left(\frac{L/\mathbb{Q}}{p}\right) = \sigma$ modulo conjugation, i.e. $\mathrm{Frob}_p = \sigma$. Since this happens at infinitely many primes, it happens at one where the traces and determinants of ρ_1 and ρ_2 coincide. Therefore ρ_1 and ρ_2 have the same characteristic equation at σ. But our choice of σ was arbitrary, hence they have the same characteristic equation for all $\sigma \in G$ and so for all $\sigma \in G_\mathbb{Q}$. If these are complex representations, then the traces coinciding immediately implies ρ_1 and ρ_2 are isomorphic.

If we are looking at modular representations, we need to quote a result of modular representation theory to infer that ρ_1 and ρ_2 having the same characteristic equation means that they are isomorphic. This is the statement of the Brauer-Nesbitt theorem , so we're done. □

Remark 1.39. Clearly the same holds for one-dimensional Galois representations.

Corollary 1.40. *Let S be a finite set of places of $\mathbb{Q}$. Then the set $\{\mathrm{Frob}_p \mid p \notin S\}$ is dense in $G_\mathbb{Q}$.*

Proof sketch. For each $\sigma \in G_\mathbb{Q}$ our above argument shows that for any finite extension F/K, there are infinitely many primes p outside of S such that $\sigma|_F = \mathrm{Frob}_p|_F$. Density follows by our description of the topology of $G_\mathbb{Q}$. □

Therefore, if an l-adic representation ρ is unramified outside of a finite set of primes, it is determined by its image on Frobenius elements when well-defined. Indeed, $\{\mathrm{Frob}_p \mid \rho \text{ is unramified at } p\}$ is dense in $G_{\mathbb{Q}}$, so the claim follows by continuity of ρ.

1.3.4 Cyclotomic characters

The following examples of one-dimensional Galois representations will be relevant in later sections.

Example 1.41 (mod l cyclotomic character)**.** Let ζ_l denote a primitive l-th root of unity. Let $U_l = \{\zeta_l^i \mid i = 0, \dots l-1\} \simeq \mathbb{F}_l$ be the group of l-th roots of unity. Then $G_{\mathbb{Q}}$ acts on U_l. This action gives rise to a homomorphism $\overline{\chi}_l \colon G_{\mathbb{Q}} \to \mathrm{GL}_1(\mathbb{F}_l) = \mathbb{F}_l^\times$ such that $\sigma \cdot \zeta_l = \zeta_l^{\overline{\chi}_l(\sigma)}$ for all $\sigma \in G_{\mathbb{Q}}$. Thus $\ker \overline{\chi}_l = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_l))$, and $\overline{\chi}_l$ factors as

$$\overline{\chi}_l \colon G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(\mathbb{Q}(\zeta_l)/\mathbb{Q}) \hookrightarrow \mathbb{F}_l^\times.$$

For a prime $p \neq l$, the extension $\mathbb{Q}(\zeta_l)/\mathbb{Q}$ is unramified at p (p splits completely in $\mathbb{Q}(\zeta_l)$). Thus by proposition 1.35, $I_p \subset \ker \overline{\chi}_l$, so $\overline{\chi}_l$ is unramified at p. With respect to the Galois extension $\mathbb{Q}(\zeta_l)/\mathbb{Q}$, the Frobenius element Frob_p satisfies $\mathrm{Frob}_p(x) \equiv x^p \pmod p$. Thus $\mathrm{Frob}_p(\zeta_l) = \zeta_l^p$ and so $\overline{\chi}_l(\mathrm{Frob}_p) = p \pmod l$.

Example 1.42 (l-adic cyclotomic character)**.** For $n \geq 1$, let ζ_{l^n} be a primitive l^n-th root of unity. Then

$$\begin{array}{ccc} \mathrm{Gal}(\mathbb{Q}(\zeta_{l^n})/\mathbb{Q}) & \simeq & (\mathbb{Z}/l^n\,\mathbb{Z})^\times \\ \sigma & \mapsto & k \text{ s.t. } \sigma(\zeta_{l^n}) = \zeta_{l^n}^k \end{array}$$

Then we have a map $G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(\mathbb{Q}(\zeta_{l^n})/\mathbb{Q}) \to (\mathbb{Z}/l^n\mathbb{Z})^\times$ for all $n \geq 1$. Since restrictions of $G_{\mathbb{Q}}$ to finite Galois groups are compatible, for all $n \geq 1$ we have the following commutative diagram:

$$\begin{array}{ccccc} G_{\mathbb{Q}} & \xrightarrow{\text{res}} & \mathrm{Gal}(\mathbb{Q}(\zeta_{l^{n+1}})/\mathbb{Q}) & \xrightarrow{\simeq} & \left(\mathbb{Z}/l^{n+1}\mathbb{Z}\right)^\times \\ \downarrow{\scriptstyle \mathrm{Id}} & & \downarrow & & \downarrow \\ G_{\mathbb{Q}} & \xrightarrow{\text{res}} & \mathrm{Gal}(\mathbb{Q}(\zeta_{l^n})/\mathbb{Q}) & \xrightarrow{\simeq} & (\mathbb{Z}/l^n\mathbb{Z})^\times \end{array}$$

Therefore we get a homomorphism

$$\chi_l \colon G_{\mathbb{Q}} \to \varprojlim_n \mathrm{Gal}(\mathbb{Q}(\zeta_{l^n})/\mathbb{Q}) = \mathrm{Gal}(\mathbb{Q}(\zeta_{l^\infty})/\mathbb{Q}) = \varprojlim_n (\mathbb{Z}/l^n\mathbb{Z})^\times = \mathbb{Z}_l^\times.$$

Note that $\chi_l(\sigma) \pmod{l^n}$ describes how σ acts on ζ_{l^n}. In particular, $\chi_l \pmod l = \overline{\chi}_l$.

Viewing $\mathbb{Z}_l^\times \subset \mathbb{Q}_l^\times = \mathrm{GL}_1(\mathbb{Q}_l)$, we have $\chi_l \colon G_{\mathbb{Q}} \to \mathrm{GL}_1(\mathbb{Q}_l)$. Then $\chi_l \pmod{l^n}$ factors through the finite extension $\mathbb{Q}(\zeta_{l^n})/\mathbb{Q}$ for each $n \geq 1$, and so χ_l is continuous by remark 1.33. Therefore χ_l defines a 1 dimensional l-adic Galois representation, called the l-adic cyclotomic character.

Every embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ corresponds to an order 2 complex conjugation element $c \in G_{\mathbb{Q}}$. This acts on roots of unity by sending ζ to ζ^{-1}, and so $\chi_l(c) = -1$.

Consider a prime p, $p \neq l$. Then p is unramified in $\mathbb{Q}(\zeta_{l^\infty})$ and so $I_p \subset \ker \chi_l$. We can take $\mathrm{Frob}_p \in G_{\mathbb{Q}}$ such that $\mathrm{Frob}_p(\zeta_{l^n}) = \zeta_{l^n}^p$ and so $\chi_l(\mathrm{Frob}_p) = p$.

In particular, the images of the Frobenius elements show that this representation has infinite image.

2 Galois representations associated to elliptic curves and modular forms

In this section we will show examples of two dimensional Galois representations arising from the theory of elliptic curves and modular forms.

In each case, we construct an l-adic Galois representation from the action of $G_{\mathbb{Q}}$ on the *Tate module* of an abelian variety.

2.1 Galois representations from elliptic curves

First, we explain how to associate 2-dimensional Galois representations to elliptic curves. Consider an elliptic curve E defined over $\mathbb{Q}$, with basepoint $\mathcal{O}_E$.

Recall that for $m \geq 2$ an integer, $E[m] = \{P \in E(\overline{\mathbb{Q}}) \mid mP = \mathcal{O}_E\}$ is the kernel of the degree m^2 isogeny $[m]\colon E \to E$ ([Sil1, Chapter III, §4]). Since $\operatorname{char}\mathbb{Q} = 0$, $E[m] \cong (\mathbb{Z}/m\mathbb{Z})^2$ as a $\mathbb{Z}/m\mathbb{Z}$-module. $G_{\mathbb{Q}}$ acts on $E(\overline{\mathbb{Q}})$ by acting on the coordinates of points, and this action preserves addition:

$$(P+Q)^\sigma = P^\sigma + Q^\sigma, \qquad P, Q \in E(\overline{\mathbb{Q}}),\ \sigma \in G_{\mathbb{Q}},$$

since addition is defined over $\mathbb{Q}$. Thus if $P \in E[m]$, then $mP^\sigma = (mP)^\sigma = \mathcal{O}_E$. Therefore $G_{\mathbb{Q}}$ acts on $E[m]$.

Definition 2.1 (The mod l Galois representation associated to E)**.** Let l be a prime. Define the two-dimensional Galois representation $\overline{\rho}_{E,l}\colon G_{\mathbb{Q}} \to \operatorname{Aut}(E[l]) = \mathrm{GL}_2(\mathbb{F}_l)$ via the action of $G_{\mathbb{Q}}$ on $E[l]$.

Let $\mathbb{Q}(E[l])$ denote the smallest Galois field extension of $\mathbb{Q}$ that contains the x and y coordinates of the points of $E[l]$. Then $\ker\overline{\rho}_{E,l} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(E[l]))$ and $\operatorname{Im}\overline{\rho}_{E,l} \cong \mathrm{Gal}(\mathbb{Q}(E[l])/\mathbb{Q})$, and $\overline{\rho}_{E,l}$ is continuous.

Similar to the l-adic cyclotomic character in example 1.42, we make use of the containments

$$E[l] \subset E[l^2] \subset E[l^3] \subset \cdots$$

to construct a two dimensional l-adic representation associated to E. Choose a basis $\{P_n, Q_n\}$ of $E[l^n]$ for all $n \geq 1$ such that

$$lP_n = P_{n-1},\ lQ_n = Q_{n-1}$$

for all n. Since the action of $G_{\mathbb{Q}}$ commutes with addition, it commutes with the multiplication maps $l\colon E[l^n] \to E[l^{n-1}]$ sending $P_n \xrightarrow{l} P_{n-1}$, $Q_n \xrightarrow{l} Q_{n-1}$. We get an inverse system of maps:

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

Definition 2.2 (l-adic Tate module of E)**.** The l-adic Tate module of E is the inverse limit of the above chain:

$$T_l(E) = \varprojlim_n \{E[l^n]\}.$$

We also define $V_l(E) = \mathbb{Q}_l \otimes_{\mathbb{Z}_l} T_l(E)$.

Since $E[l^n] \cong (\mathbb{Z}/l^n\mathbb{Z})^2$ is a $\mathbb{Z}/l^n\mathbb{Z}$-module, it follows that $T_l(E)$ is a $\varprojlim_n \mathbb{Z}/l^n\mathbb{Z} \cong \mathbb{Z}_l$-module (so the definition of the vector space $V_l(E)$ makes sense). We have that $G_\mathbb{Q}$ acts on $T_l(E)$. We get a group homomorphism $\rho_{E,l}\colon G_\mathbb{Q} \to \mathrm{GL}(T_l(E))$ with

$$\mathrm{GL}(T_l(E)) \cong \mathrm{GL}\left(\varprojlim_n E[l^n]\right) \cong \mathrm{GL}\left(\varprojlim_n (\mathbb{Z}/l^n\mathbb{Z})^2\right) \cong \mathrm{GL}(\mathbb{Z}_l^2) \cong \mathrm{GL}_2(\mathbb{Z}_l).$$

View $\mathrm{GL}_2(\mathbb{Z}_l) \subset \mathrm{GL}_2(\mathbb{Q}_l)$.

Definition 2.3 (l-adic Galois representation of E)**.** The two-dimensional l-adic representation of an elliptic curve $E/\mathbb{Q}$ is given by $\rho_{E,l}\colon G_\mathbb{Q} \to \mathrm{GL}_2(\mathbb{Q}_l)$ from above.

Then $\rho_{E,l}(\sigma) \pmod{l^n}$ describes how σ acts on $E[l^n]$ for $\sigma \in G_\mathbb{Q}$ and each $n \geq 1$. If we let $\mathbb{Q}(E[l^n])$ denote the smallest finite Galois extension of $\mathbb{Q}$ containing the x and y coordinates of the points of $E[l^n]$, then clearly $\rho_{E,l} \pmod{l^n}$ factors through the finite Galois group $\mathrm{Gal}(\mathbb{Q}(E[l^n])/\mathbb{Q})$ for all n. Thus $\rho_{E,l}$ is continuous.

Proposition 2.4. *The determinant* $\det \rho_{E,l}$ *is* χ_l*, the* l*-adic cyclotomic character.*

Proof. We show that $\det \rho_{E,l} \equiv \chi_l \pmod{l^n}$ for all $n \geq 1$, thus establishing their equality. Recall that the Weil-pairing ([Sil1, Chapter III, §8])

$$e_{l^n}\colon E[l^n] \times E[l^n] \to (\mathbb{Z}/l^n\mathbb{Z})$$

for each $n \geq 1$ is bilinear, non-degenerate, and Galois-invariant. Let $\sigma \in G_\mathbb{Q}$ with $\rho_{E,l}(\sigma) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}_l)$, and let $\begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \equiv \rho_{E,l}(\sigma) \pmod{l^n}$. Then,

$$\begin{aligned} \sigma \cdot e_{l^n}(P_n, Q_n) &= e_{l^n}(P_n^\sigma, Q_n^\sigma) \\ &\equiv e_{l^n}(a_n P_n + c_n Q_n, b_n P_n + d_n Q_n) \\ &= e_{l^n}(P_n, P_n)^{a_n b_n}\ e_{l^n}(P_n, Q_n)^{a_n d_n}\ e_{l^n}(Q_n, P_n)^{c_n b_n}\ e_{l^n}(Q_n, Q_n)^{c_n d_n} \\ &= e_{l^n}(P_n, Q_n)^{a_n d_n - b_n c_n}. \end{aligned}$$

But $e_{l^n}(P_n, Q_n) = \zeta_{l^n}$ a primitive l^n-th root of unity, and $\sigma \cdot \zeta_{l^n} = \zeta_{l^n}^{\chi_l(\sigma) \pmod{l^n}}$. Thus $\det \rho_{E,l}(\sigma) \equiv \chi_l(\sigma) \pmod{l^n}$ for all $n \geq 1$ and all $\sigma \in G_\mathbb{Q}$. □

Theorem 2.5. *[DS05, Theorem 9.4.1] Let l be prime and let E be an elliptic curve over $\mathbb{Q}$ with conductor N. The Galois representation $\rho_{E,l}$ is unramified at every prime $p \nmid lN$. For any such p let $\mathfrak{p} \subset \overline{\mathbb{Z}}$ be any maximal prime over p. Then the characteristic equation of $\rho_{E,l}(Frob_\mathfrak{p})$ is*

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

where $a_p(E) = p + 1 - \#\tilde{E}(\mathbb{F}_p)$.

Proof. Let p be a prime such that $p \nmid lN$. Then E has good reduction at p. Let $\tilde{E}$ be the reduction of E mod p. Since p and l^n are coprime, the reduction map $E[l^n] \to \tilde{E}(\mathbb{F}_p)$ is injective for all $n \geq 1$ ([Sil1, Chapter VII, Proposition 3.1]).

Consider $\sigma \in I_\mathfrak{p}$. Then $P^\sigma - P \in E[l^n]$ for all $P \in E[l^n]$. By definition of the inertia subgroup, $\overline{\sigma} = \mathrm{Id} \in G_{\mathbb{F}_p}$ and so acts trivially on $\tilde{E}(\mathbb{F}_p)$. Thus under the reduction $E[l^n] \to \tilde{E}(\mathbb{F}_p)$, $P^\sigma - P$ maps to $\mathcal{O}_{\tilde{E}}$. By injectivity, $P^\sigma - P = \mathcal{O}_E$, so σ acts trivially on $E[l^n]$ for

all $n \geq 1$. Therefore $\rho_{E,l}(\sigma) \equiv \mathrm{Id} \pmod{l^n}$ for all $n \geq 1$, and $\sigma \in \ker \rho_{E,l}$. Thus $\rho_{E,l}$ is unramified at all primes $p \nmid lN$. The characteristic equation of $\rho_{E,l}(\mathrm{Frob}_\mathfrak{p})$ is

$$x^2 - \operatorname{Tr} \rho_{E,l}(\mathrm{Frob}_\mathfrak{p})x + \det \rho_{E,l}(\mathrm{Frob}_\mathfrak{p}) = 0.$$

Since $\det \rho_{E,l} = \chi_l$, we have $\det \rho_{E,l}(\mathrm{Frob}_\mathfrak{p}) = p$.

For the trace, we use the calculation $\operatorname{Tr} A = 1 + \det A + \det(\mathrm{Id} - A)$, which holds for any 2×2 matrix A. Let $\sigma_p \colon E \to E$ be the p-power Frobenius automorphism, so that $\#\tilde{E}(\mathbb{F}_p) = \deg(\sigma_p - 1) = \det(\sigma_p - 1)$ ([Sil1, Chapter V, Proposition 2.3]).

Since $\mathrm{Frob}_\mathfrak{p}(x) \equiv x \pmod{\mathfrak{p}}$ for all $x \in \mathbb{Z}$ it follows that $\mathrm{Frob}_\mathfrak{p}$ acts on $E[l^n]$ as σ_p acts on $\tilde{E}[l^n]$. Then $\operatorname{Tr} \sigma_p = 1 + p - \#\tilde{E}(\mathbb{F}_p) = a_p(E)$, and so $\operatorname{Tr} \rho_{E,l}(\mathrm{Frob}_\mathfrak{p}) \equiv a_p(E) \pmod{l^n}$. This holds for all n and so $\operatorname{Tr} \rho_{E,l}(\mathrm{Frob}_\mathfrak{p}) = a_p(E)$. □

Remark 2.6. Since $p \nmid lN$ in the above, we have a commutative diagram for all n

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

As $I_\mathfrak{p} \subset \ker \rho$, as a $G_{\mathbb{Q}_p}$-module $V_l(E)^{I_\mathfrak{p}}$, the submodule of $V_l(E)$ fixed by $I_\mathfrak{p}$, has $V_l(E)^{I_\mathfrak{p}} = V_l(E)$. So we can view $V_l(E)$ as a $G_{\mathbb{F}_p}$-module. The commutativity of the diagram shows that $V_l(E) \cong V_l(\tilde{E})$ as $G_{\mathbb{F}_p}$-modules.

More generally, we have $V_l(E(\overline{\mathbb{Q}}_p))^{I_\mathfrak{p}} \cong V_l(\tilde{E}_{\mathrm{ns}}(\overline{\mathbb{F}}_p))$ as $G_{\mathbb{F}_p}$-modules (see proof of [Sil2, Chapter IV, §10, Theorem 10.2]).

In fact, these are the only primes not equal to l at which $\rho_{E,l}$ is unramified, as stated in the following result:

Theorem 2.7 (Neron-Ogg-Shafarevich, [Sil1, Chapter VII, Theorem 7.1])**.** *Let $E/\mathbb{Q}$ be an elliptic curve, and p a prime not equal to l. Then E has good reduction at p if and only if $\rho_{E,l}$ is unramified.*

The previous results show that the Galois representation $\rho_{E,l}$ contains essential information about our elliptic curve. For $\rho_{E,l}$ to be possibly modular (see §3), it must be irreducible, which is indeed the case:

Theorem 2.8. *The Galois representation $\rho_{E,l}$ is irreducible.*

The proof is beyond the scope of this essay. We will be able to prove that $\overline{\rho}_{E,l}$ is irreducible in certain cases, which will be essential in proving Fermat's Last Theorem.

Now consider a prime p such that $p \neq l$ and $p \| N$ (the notation $p \| N$ means that $p \mid N$ but $p^2 \nmid N$). Then E has multiplicative reduction at p, and so $\rho_{E,l}$ is not unramified at p. However by using the theory of the Tate curve, we can determine when the mod p representation $\overline{\rho}_{E,l}$ is unramified at p.

Proposition 2.9 ([DDT95, Proposition 2.12])**.** *Let $E/\mathbb{Q}$ be an elliptic curve with conductor N and minimal discriminant $\Delta_{\min}$, and let l, p be primes. Suppose that $p \neq l$ and $p \| N$. Then*

$\overline{\rho}_{E,l}$ is unramified at p if and only if $v_p(\Delta_{\min}) \equiv 0 \pmod{l}$. Moreover, if $\overline{\rho}_{E,l}$ is unramified at p, then

$$\mathrm{Tr}(\overline{\rho}_{E,l}(\mathrm{Frob}_p)) \equiv \pm(1+p) \pmod{l}$$
$$\mathrm{Det}(\overline{\rho}_{E,l}(\mathrm{Frob}_p)) \equiv p \pmod{l}$$

Proof (modeled off of that of [Sil2, Chapter V, §6, Proposition 6.1]). We have that $\overline{\rho}_{E,l}$ is unramified at p if and only if p is unramified in $\mathbb{Q}(E[l])$, if and only if the extension $\mathbb{Q}_p(E[l])/\mathbb{Q}_p$ is unramified.

Let K be the unramified extension of $\mathbb{Q}_p$ of degree ≤ 2 such that E/K has split multiplicative reduction. Then by corollary 1.10, $E[l] \cong \langle \zeta_l, \sqrt[l]{q} \rangle$ for $q \in \mathbb{Q}_p^\times$, and so $K(E[l]) \cong K(\zeta_l, \sqrt[l]{q})$. Since $K/\mathbb{Q}_p$ is unramified, $\mathbb{Q}_p(E[l])/\mathbb{Q}_p$ is unramified if and only if $K(\zeta_l, \sqrt[l]{q})/K$ is unramified. We have

$$K \xrightarrow{\text{unramified cyclotomic extension}} K(\zeta_l) \xrightarrow{\text{Kummer extension}} K(\zeta_l, \sqrt[l]{q}) := L$$

with $L/K(\zeta_l)$ an extension of degree dividing l, a prime.

–> Note that $l v_L(\sqrt[l]{q}) = v_L(q) = e_{L/K} v_p(q) = e_{L/K} v_p(\Delta_{\min})$. Thus $e_{L/K} = 1$ implies $v_p(\Delta_{\min}) \equiv 0 \pmod{l}$.

<– Conversely, if $v_p(\Delta_{\min}) \equiv 0 \pmod{l}$ then $v_p(q\ \pi_K^{-xl}) = 0$ for some $x \in \mathbb{Z}$, π_K the uniformizer of K. Let $b = q\ \pi_K^{-xl}$. Then $L = K(\zeta_l)[x]/(x^l - b)$. The discriminant of $x^l - b$ is $\pm l^l b^{l-1}$ which has valuation 0 since $v_p(b) = 0$ and l and p are coprime. Therefore $L/K(\zeta_l)$ is unramified, so L/K is unramified.

Now let $\overline{\rho}_{E,l}$ be unramified at a prime p. It is clear that $\mathrm{Det}(\overline{\rho}_{E,l}(\mathrm{Frob}_p)) \equiv p \pmod{l}$ since $\det \overline{\rho}_{E,l} \equiv \overline{\chi}_l \pmod{l}$, the mod l cyclotomic character (cf. proposition 2.4).

We will prove the formula for the trace in the case of split multiplicative reduction. Then $K = \mathbb{Q}_p$, $G_{\mathbb{Q}_p} = D_p$, and we can describe the action of D_p on $E[l]$ via that of $G_{\mathbb{Q}_p}$ on $\langle \zeta_l, \sqrt[l]{q} \rangle$, since $E[l] \cong \langle \zeta_l, \sqrt[l]{q} \rangle$ as $G_{\mathbb{Q}_p}$-modules.

For $\sigma \in D_p$, $\sigma(\zeta_l) = \zeta_l^{\overline{\chi}_l(\sigma)}$ and $\sigma(\sqrt[l]{q}) = \zeta_l^t \sqrt[l]{q}$ for some t. Thus D_p acts on $E[l]$ as

$$\begin{pmatrix} \chi_l & * \\ 0 & 1 \end{pmatrix},$$

and so the trace formula follows.

In the case of non-split multiplicative reduction we have $\mathrm{Tr}(\overline{\rho}_{E,l}(\mathrm{Frob}_p)) \equiv -(1+p) \pmod{l}$. □

Remark 2.10. Note that $V_l(\mathbb{G}_m(\overline{\mathbb{F}}_l)) \cong \mathbb{Q}_l$ is a $G_{\mathbb{F}_l}$-module with character χ_l, as in example 1.42, and multiplicative reduction at p means $\tilde{E}_{\mathrm{ns}} \cong \mathbb{G}_m$.

2.2 Galois representations from newforms

Given a newform, we describe a two-dimensional Galois representation associated to it. To do so, we need to return to our consideration of modular curves.

2.2.1 Eichler-Shimura theory

Recall that $X_1(N)$ is a complex torus, say of genus g. Let $J_1(N) \coloneqq \mathrm{Jac}(X_1(N))$ be its Jacobian. A consequence of theorem 1.18 and the definition of the Jacobian is the following:

Corollary 2.11 ([DS05, §6.3])**.** *There is a lattice $\Lambda_1(N) \subset S_2(\Gamma_1(N))^\vee$ (the vector space dual of $S_2(\Gamma_1(N))$) such that $J_1(N) \cong S_2(\Gamma_1(N))^\vee/\Lambda_1(N)$.*

$S_2(\Gamma_1(N))$ is a $\mathbb{T}_\mathbb{Z}$-module, and so is $S_2(\Gamma_1(N))^\vee$ by letting $T \in \mathbb{T}_\mathbb{Z}$ act on $\phi \in S_2(\Gamma_1(N))^\vee$ by

$$(T \cdot \phi)(g) = \phi(T(g))$$

for $g \in S_2(\Gamma_1(N))$. One has that $\Lambda_1(N)$ is isomorphic to the lattice of periods $H_1(X_1(N), \mathbb{Z}) \subset \Omega^1_{\mathrm{hol}}(X_1(N))^\vee$. The action of $\mathbb{T}_\mathbb{Z}$ on $S_2(\Gamma_1(N))^\vee$ descends to the quotient. See [DS05, Section 6.3] for details of this. Thus $J_1(N)$ is a $\mathbb{T}_\mathbb{Z}$-module also.

Let $f = \sum_{n\geq 1} a_n(f)q^n \in S_2(N, \chi)$ be a newform. Consider

$$I_f = \ker(\mathbb{T}_\mathbb{Z} \xrightarrow{\lambda_f} \mathbb{C}) = \{T \in \mathbb{T}_\mathbb{Z} \mid Tf = 0\},$$

where λ_f is the ring homomorphism defined in definition 1.25. Then $\mathbb{T}_\mathbb{Z}/I_f \cong \operatorname{Im}\lambda_f \cong \mathbb{Z}[\{a_n(f)\}]$, since $T_n(f) = a_n(f)f$ and one can show that $\chi(d)$ can be written in terms of the $a_n(f)$ ([DS05, Exercise 6.5.2]).

Proposition 2.12. *The ring $\mathbb{Z}[\{a_n(f)\}]$ is finitely generated.*

Proof sketch. The action of $\mathbb{T}_\mathbb{Z}$ on $S_2(\Gamma_1(N))^\vee$ is faithful. Since $\Lambda_1(N) \otimes \mathbb{R} = S_2(\Gamma_1(N))^\vee$, the action of $\mathbb{T}_\mathbb{Z}$ on $S_2(\Gamma_1(N))^\vee$ is determined by that on $\Lambda_1(N)$. Hence we have an injection

$$\mathbb{T}_\mathbb{Z} \to \mathrm{End}(\Lambda_1(N))) \cong M_{2g}(\mathbb{Z}),$$

where $g = \dim S_2(\Gamma_1(N)) = \dim X_1(N)$. Therefore $\operatorname{rk}\mathbb{T}_\mathbb{Z} \leq 4g^2$, and hence is finitely generated. Thus $\mathbb{Z}[\{a_n(f)\}]$ is finitely generated as the quotient of a finitely generated $\mathbb{Z}$-module. □

One can show ([DS05, Theorem 6.5.1]) that the characteristic equation of T_p acting on $S_2(\Gamma_1(N))$ has integer coefficients, and so $a_p(f)$ and $a_n(f)$ for arbitrary n are algebraic integers. Therefore

$$K_f = \mathbb{Q}(\{a_n(f)\})$$

is a number field of dimension $\dim_\mathbb{Q}\mathbb{Q}(\{a_n(f)\}) = \dim_\mathbb{Q}(\mathbb{Z}[\{a_n(f)\}] \otimes \mathbb{Q}) = \operatorname{rk}\mathbb{Z}[\{a_n(f)\}]$. We denote by $\mathcal{O}_f$ the ring of integers of K_f.

Remark 2.13. Note that $\chi(d) \in K_f$ for all d since it can be expressed in terms of $a_n(f)$, and so we can view $\chi\colon (\mathbb{Z}/N\mathbb{Z})^\times \to K_f^\times$.

We can act on the coefficients of f to obtain new eigenforms:

Theorem 2.14 ([DS05, Theorem 6.5.4])**.** *For any embedding $\sigma\colon K_f \to \mathbb{C}$, $f^\sigma \coloneqq \sum_{n\geq 1}\sigma(a_n(f))q^n$ is also a newform, with $f^\sigma \in S_2(N, \chi^\sigma)$, where χ^σ is the character defined as $\chi^\sigma(n) = \sigma(\chi(n))$.*

One associates to f an abelian variety A_f of dimension $[K_f\colon \mathbb{Q}]$. Since $J_1(N)$ is a $\mathbb{T}_{\mathbb{Z}}$-module, we can define

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

The dimension of A_f is computed by showing it is isomorphic to a complex torus of dimension $[K_f\colon \mathbb{Q}]$. Let

$$[f] = \{f^\sigma\colon \sigma \text{ is an automorphism of } \mathbb{C}\}, \quad V_f = \operatorname{span}([f]) \subset S_2(\Gamma_1(N)).$$

Then $\dim V_f = [K_f\colon \mathbb{Q}]$, the number of embeddings of K_f into $\mathbb{C}$. Let $\Lambda_f = \Lambda_1(N)|_{V_f} \subset V_f^\vee$ be the restriction of the functionals $\Lambda_1(N)$ to V_f.

Proposition 2.15 ([DS05, Proposition 6.6.4])**.** *There is an isomorphism $A_f \cong V_f^\vee/\Lambda_f$ given by*

$$[\phi] + I_f J_1(N) \mapsto \phi|_{V_f} + \Lambda_f,$$

for $[\phi] \in J_1(N)$, $\phi \in S_2(\Gamma_1(N))^\vee$. Moreover, the right hand side of this isomorphism is a complex torus of dimension $[K_f\colon \mathbb{Q}]$.

The proof is omitted and can be found in [DS05]. A pivotal characteristic of this construction is the following:

Theorem 2.16. *The abelian variety A_f can be defined over $\mathbb{Q}$. That is, there exists an abelian variety A'_f defined over $\mathbb{Q}$ such that $A'_f \cong A_f$ as complex abelian varieties.*

This result follows by proving the following.

1. $J_1(N)$ can be defined over $\mathbb{Q}$. This is the work of Weil and Chow, who prove this by using the correspondence between function fields and algebraic curves, as well as the Galois theory on function fields.

2. The Hecke operators are defined over $\mathbb{Q}$. The proof of this relies on the moduli interpretation of Hecke operators, that is their action on $S_1(N)$.

These concepts are discussed in Weil's book [Wei48].

2.2.2 Galois representation of the abelian variety associated to a newform

We are now in a position to construct our Galois representation. A reference for this is [DS05, Chapter 9, §9.4].

Let $f \in S_2(\Gamma_1(N))$ be a newform, A_f the corresponding variety of dimension $g = [K_f\colon \mathbb{Q}]$, defined over $\mathbb{Q}$. Since A_f is an abelian variety over $\mathbb{Q}$,

$$A_f[n] = \ker[n] \cong (\mathbb{Z}/n\mathbb{Z})^{2g}.$$

The multiplication by n map is defined over $\mathbb{Q}$, and so $n(P^\sigma) = (nP)^\sigma$ for all $P \in A_f(\overline{\mathbb{Q}})$. Thus $A_f[n]$ is a $G_{\mathbb{Q}}$-module, and we have a continuous homomorphism $G_{\mathbb{Q}} \to \operatorname{Aut}(A_f[n])$.

For l a prime, consider a suitable basis $P_{n,i}$ with $i = 1, \ldots, 2g$ of $A_f[l^n]$ such that for all n

$$lP_{n,i} = P_{n-1,i} \quad \forall i = 1, \ldots, 2g$$

Then the multiplication map $[l]\colon A_f[l^{n+1}] \twoheadrightarrow A_f[l^n]$ is a homomorphism of $G_{\mathbb{Q}}$-modules for all $n \geq 1$. Therefore we get an inverse system of $G_{\mathbb{Q}}$-modules:

$$A_f[l] \leftarrow A_f[l^2] \leftarrow A_f[l^3] \leftarrow \cdots$$

and taking the inverse limit gives the Tate-module

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

The induced action of $G_{\mathbb{Q}}$ yields a continuous homomorphism $\rho_{A_f,l}\colon G_{\mathbb{Q}} \to \mathrm{GL}_{2g}(\mathbb{Z}_l)$, as $\mathrm{Ta}_l(A_f) \cong \mathbb{Z}_l^{2g}$. We can also define the vector space $V_l(A_f) = \mathrm{Ta}_l(A_f) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$, and let $G_{\mathbb{Q}}$ act, yielding $\rho_{A_f,l}\colon G_{\mathbb{Q}} \to \mathrm{GL}_{2g}(\mathbb{Q}_l)$. Note that this construction is entirely analogous to that for elliptic curves.

If $g = 1$ we get a two dimensional Galois representation, but this is not the case for $g > 1$. To remedy this, we consider the action of $\mathbb{T}_{\mathbb{Z}}$ on A_f.

Recall that $\mathbb{T}_{\mathbb{Z}}/I_f \simeq \mathcal{O}_f$. Since $\mathbb{T}_{\mathbb{Z}}$ acts on A_f with I_f acting trivially, there is a faithful action of $\mathcal{O}_f$ on A_f. The action of $\mathcal{O}_f$ on A_f is defined on l-power torsion, so there is an induced action on $\mathrm{Ta}_l(A_f)$. Then $V_l(A_f)$ is a $(\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q}_l)$-module, and

$$\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q}_l \cong (\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_l \simeq K_f \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong \prod_{\lambda|l} K_\lambda,$$

where the product on the right runs over all prime ideals $\lambda \subset \mathcal{O}_f$ dividing l, and K_λ is the λ-adic completion of K_f.

Let $i\colon K_f \otimes_{\mathbb{Q}} \mathbb{Q}_l \xrightarrow{\sim} \prod_{\lambda|l} K_\lambda$ be the isomorphism, and $e_\lambda \in K_f \otimes_{\mathbb{Q}} \mathbb{Q}_l$ such that $i(e_\lambda) = (0, \ldots, 0, 1_{K_\lambda}, 0, \ldots 0)$. Letting $V_\lambda(A_f) \coloneqq e_\lambda(V_l(A_f))$, $V_l(A_f)$ decomposes as

$$V_l(A_f) \cong \prod_{\lambda|l} V_\lambda(A_f),$$

([DS05, Exercise 9.5.2]). The $G_{\mathbb{Q}}$-action on $V_l(A_f)$ restricts to one of $V_\lambda(A_f)$. Additionally, $K_\lambda \subset \prod_{\lambda|l} K_\lambda$ acts on $V_\lambda(A_f)$ and this action commutes with that of $G_{\mathbb{Q}}$. Therefore we have a homomorphism $G_{\mathbb{Q}} \to \mathrm{End}_{K_\lambda}(V_\lambda(A_f))$. The following proposition shows that this yields a two-dimensional Galois representation.

Proposition 2.17 ([DS05, Lemma 9.5.3])**.** *For any $\lambda|l$, $\dim_{K_\lambda} V_\lambda(A_f) = 2$.*

Proof. As a complex abelian variety, $A_f(\mathbb{C}) \cong \mathbb{C}^g/\Lambda$, where Λ is a lattice, and $g = [K_f\colon \mathbb{Q}]$. Thus since A_f is a $\mathcal{O}_f$-module, Λ is one also. Taking the tensor product with $\mathbb{Q}$, $\Lambda_{\mathbb{Q}} = \Lambda \otimes_{\mathbb{Z}} \mathbb{Q}$ is then an $\mathcal{O}_f \otimes_{\mathbb{Z}} \mathbb{Q} = K_f$-vector space.

Now,

$$A_f[l^n] \cong (\mathbb{C}^g/\Lambda)[l^n] \cong \Lambda/l^n\Lambda,$$

and $[l]\colon A_f[l^{n+1}] \to A_f[l^n]$ corresponds to the quotient $\Lambda/l^{n+1}\Lambda \to \Lambda/l^n\Lambda$. Therefore,

$$\mathrm{Ta}_l(A_f) = \varprojlim_n A_f[l^n] = \varprojlim_n \Lambda/l^n\Lambda = \varprojlim_n \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}/l^n\mathbb{Z} \cong \Lambda \otimes_{\mathbb{Z}} \varprojlim_n \mathbb{Z}/l^n\mathbb{Z} = \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}_l,$$

and

$$V_l(A_f) \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Q}_l \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Q} \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong \Lambda_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{Q}_l.$$

Since $\Lambda_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong \Lambda_{\mathbb{Q}} \otimes_{K_f} \prod_{\lambda|l} K_\lambda$, one has

$$V_\lambda(A_f) \cong \Lambda_{\mathbb{Q}} \otimes_{K_f} K_\lambda.$$

Then

$$\dim_{K_\lambda} V_\lambda(A_f) = \dim_{K_\lambda}(\Lambda_{\mathbb{Q}} \otimes_{K_f} K_\lambda) = \dim_{K_f} \Lambda_{\mathbb{Q}} = \frac{1}{[K_f\colon \mathbb{Q}]} \dim_{\mathbb{Q}} \Lambda_{\mathbb{Q}} = \frac{1}{[K_f\colon \mathbb{Q}]} \operatorname{rk} \Lambda = 2,$$

since $\operatorname{rk} \Lambda = 2g = 2[K_f\colon \mathbb{Q}]$.

□

Definition 2.18 (Galois representation associated to a newform)**.** Let $f = \sum_{n\geq 1} a_n(f) q^n \in S_2(N, \chi)$ be a newform with number field K_f. Fix l a prime, and $\lambda \subset \mathcal{O}_f$ a maximal ideal lying over l. Then there is a two-dimensional l-adic Galois representation

$$\rho_{f,\lambda}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(K_\lambda)$$

given by the action of $G_{\mathbb{Q}}$ on the two-dimensional K_λ-vector space $V_\lambda(A_f)$.

We admit that this is continuous, and so defines a Galois representation. As usual, we are interested in the ramification of $\rho_{f,\lambda}$ at primes p, as well as the characteristic equations of the Frobenius elements.

Theorem 2.19 ([DS05, Theorem 9.5.4])**.** *The representation $\rho_{f,\lambda}$ is unramified at every prime $p \nmid lN$. For any such p, and $\mathfrak{p} \subset \overline{\mathbb{Z}}$ a maximal ideal lying over it, the characteristic equation of $\rho_{f,\lambda}(\mathrm{Frob}_{\mathfrak{p}})$ is*

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

Thus if $f \in S_2(\Gamma_0(N))$, the characteristic equation is $x^2 - a_p(f)x + p = 0$.

Remark 2.20.

1. Since $\det(\rho_{f,\lambda}(\mathrm{Frob}_p)) = \chi(p)p$ for infinitely many primes, it follows by Chebotarev's density theorem and the continuity of $\rho_{f,\lambda}$ (cf. corollary 1.40) that $\det \rho_{f,\lambda} = \chi \cdot \chi_l$. Here χ_l is the l-adic cyclotomic character. By remark 2.13, we can view χ as $\chi\colon (\mathbb{Z}/N\mathbb{Z})^\times \to K_f^\times$. We can lift this to a Galois representation $G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q}) \cong (\mathbb{Z}/N\mathbb{Z})^\times \to \mathrm{GL}_1(K_f)$. Then we can consider this landing in $\mathrm{GL}_1(K_\lambda)$ by composing with the map $\mathrm{GL}_1(K_f) \to \mathrm{GL}_1(K_f) \otimes_{\mathbb{Q}} \mathbb{Q}_l \twoheadrightarrow \mathrm{GL}_1(K_\lambda)$.

2. Let $c \in G_{\mathbb{Q}}$ be a complex conjugation automorphism, then $\det \rho_{f,\lambda}(c) = \chi(c) \cdot \chi_l(c) = -1$, since $\chi_l(c) = -1$ and χ is an even Dirichlet character.

3. Suppose that $[K_f\colon \mathbb{Q}] = 1$ so that A_f is an elliptic curve. Then for any prime l, the representations $\rho_{f,l}$ and $\rho_{A_f,l}$ are the same, hence for all $p \nmid N$, $a_p(A_f) = a_p(f)$ by theorem 2.19 and theorem 2.5.

4. The representation $\rho_{f,\lambda}$ is absolutely irreducible ([DDT95, Theorem 3.1]).

The key to showing that $\mathrm{Tr}(\mathrm{Frob}_{\mathfrak{p}}) = a_p(f)$ is the *Eichler-Shimura relation* [DS05, Theorem 8.7.2]. We will not discuss this, but the theory is explained in detail in [DS05, Chapter 8]. The results come from considering the moduli space interpretation of the Hecke operators T_p. In particular these come from p-isogenies of an elliptic curve, so upon reduction we expect to see p-isogenies in characteristic p, i.e. the Frobenius isogeny and its dual.

3 Modularity Theorem and Ribet's level lowering theorem

3.1 Modularity Theorem

The Modularity Theorem for elliptic curves over $\mathbb{Q}$ can be stated in many different ways. An excellent source for these different statements is [DS05]. In particular, much of the content of this section can be found in [DS05, §9.6].

3.1.1 Modularity Theorem and l-adic representations

We will first state the version of the Modularity Theorem that was proven by Wiles in [Wil95] for semistable elliptic curves, and then by Breuil, Conrad, Diamond, and Taylor in [BCDT01] for all elliptic curves.

First, for $E/\mathbb{Q}$ an elliptic curve and l a prime, we define what it means for $\rho_{E,l}$ to be modular.

Definition 3.1 (Modular l-adic representation, [DS05, Definition 9.6.1])**.** An irreducible Galois representation

$$\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_l)$$

such that $\det\rho = \chi_l$ is modular if there exists a newform $f \in S_2(\Gamma_0(M_f))$ with $K_\lambda = \mathbb{Q}_l$ for some maximal ideal $\lambda \subset \mathcal{O}_f$ lying over l and such that $\rho_{f,\lambda} \sim \rho$.

Indeed, $\rho_{E,l}$ is irreducible with determinant χ_l, so is a candidate for being a modular representation. Note that we restrict to newforms corresponding to the congruence subgroup $\Gamma_0(M_f)$ so that the character χ of f is trivial. Then $\det\rho_{f,\lambda} = \chi_l$.

Theorem 3.2 (Modularity Theorem for $\rho_{E,l}$)**.** *Let E be an elliptic curve over $\mathbb{Q}$. There exists a prime l such that $\rho_{E,l}$ is modular.*

The Modularity Theorem can be viewed as a converse to the Eichler-Shimura construction for rational newforms in the previous section. Recall that to a rational newform of level N we can associate an elliptic curve A_f of conductor N such that $a_p(A_f) = a_p(f)$ for all primes $p \nmid N$.

Note that isogenous elliptic curves have good reduction at the same primes ([Sil1, Chapter VII, Corollary 7.2]), and that if E and E' are isogenous, then $a_p(E) = a_p(E')$ for primes p of good reduction ([Sil1, Chapter V, Exercise 5.4]). Thus their arithmetical information can be considered equivalent. The converse is also true, by a theorem of Faltings:

Theorem 3.3 (Faltings' isogeny theorem, [Fal83])**.** *Let E, E' be two elliptic curves over $\mathbb{Q}$ such that*

$$a_p(E) = a_p(E')$$

for all primes p at which E and E' have good reduction. Then E and E' are isogenous.

Therefore we can view the Eichler-Shimura construction for rational newforms as an injective map

{Rational newforms of level N} → {Isogeny classes of rational elliptic curves of conductor N}.

We describe how the Modularity Theorem implies that this map is surjective. Let $E/\mathbb{Q}$ have conductor N and suppose that $\rho_{E,l}$ is modular for l. Then $\rho_{E,l} \sim \rho_{f,\lambda}$ for $f \in S_2(\Gamma_0(M_f))$ a newform of level M_f, and $\lambda \subset \mathcal{O}_f$ a prime ideal lying over l.

For any prime $p \nmid lNM_f$, the traces of $\rho_{E,l}(\mathrm{Frob}_p)$ and $\rho_{f,\lambda}(\mathrm{Frob}_p)$ are equal. Thus

$$a_p(E) = a_p(f) \quad \text{for all but finitely many primes } p.$$

Carayol in [Car86] showed that in fact $a_p(E) = a_p(f)$ for all primes p, and that $M_f = N$. Note that since $a_p(f)$ generate K_f, it follows that $K_f = \mathbb{Q}$ and f is a rational newform. Then by Faltings, E and A_f (an elliptic curve) are isogenous. Hence our map is surjective.

We also define the notion of an elliptic curve E being modular.

Definition 3.4 (Modular elliptic curve)**.** Let $E/\mathbb{Q}$ be an elliptic curve with conductor N. Then E is said to be modular if there exists some rational newform $f = \sum_{n\geq 1} a_n(f)q^n$ with $a_p(E) = a_p(f)$ for all primes p.

Then another formulation of the Modularity Theorem is

Theorem 3.5 (Modularity Theorem for E)**.** *All elliptic curves $E/\mathbb{Q}$ are modular.*

We already discussed how version 3.2 implies version 3.5. The converse is again an application of Chebotarev's theorem. Indeed if $E/\mathbb{Q}$ of conductor N is modular, let f be the rational newform with $a_p(E) = a_p(f)$ for all primes p. Consider $\rho_{A_f,l}$ for an arbitrary prime l. Then for all but finitely many p, the characteristic equations of $\rho_{E,l}$ and $\rho_{A_f,l}$ agree. Thus $\rho_{E,l} \sim \rho_{A_f,l}$ by corollary 1.40 and so $\rho_{E,l}$ is modular for all primes l.

Let us collect these equivalences:

Corollary 3.6 ([DDT95, Proposition 3.20])**.** *Let $E/\mathbb{Q}$ be an elliptic curve and $\rho_{E,l}$ the associated l-adic representation. The following are equivalent:*

1. *E is modular,*
2. *$\rho_{E,l}$ is modular for all primes l,*
3. *$\rho_{E,l}$ is modular for some prime l.*

Thus the usefulness of l-adic representations in the context of the Modularity Theorem is that it can be proven by showing modularity of $\rho_{E,l}$ at just one prime l.

3.1.2 Modularity Theorem and mod l representations

We have a corresponding notion of modular mod l representations. To define this, we first need to define a mod l representation associated to a newform.

Consider an l-adic Galois representation associated to $f \in S_2(\Gamma_1(N))$ a newform, namely

$$\rho_{f,\lambda} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(K_\lambda)$$

for $\lambda \subset \mathcal{O}_f$ a maximal ideal lying over l. Recall that K_λ is the completion of K_f at λ. Let $\mathcal{O}_\lambda$ denote the valuation ring of K_λ, and $\mathfrak{m} \subset \mathcal{O}_f$ the maximal ideal. Denote the residue field by $\mathbb{F}_\lambda = \mathcal{O}_\lambda/\mathfrak{m}$. This is a finite field of characteristic l.

To this we associate a mod l Galois representation $\overline{\rho}_{f,\lambda}$ by the following process:

1. Choose a model of $\rho_{f,\lambda}$ that lands in $\mathrm{GL}_2(\mathcal{O}_\lambda)$. This is always possible up to equivalence; see [DS05, §9.3, Proposition 9.3.5].

2. Reduce the matrix coefficients mod $\mathfrak{m}$ to obtain $\rho'_{f,\lambda}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O}_\lambda) \to \mathrm{GL}_2(\mathbb{F}_\lambda)$.

3. Take the semisimplification of the resulting module to obtain the mod l Galois rep $\overline{\rho}_{f,\lambda}$.

The semisimplification of a module is obtained by replacing it with the direct sum of the simple modules of its Jordan-Hölder decomposition. This does not change the values of the trace or determinant in the matrix representation.

Observe that, regardless of our choice in Step 1, the characteristic equation of $\overline{\rho}_{f,\lambda}(\mathrm{Frob}_p)$ at primes p for which $\rho_{f,\lambda}$ is unramified is the same. Therefore by corollary 1.38, $\overline{\rho}_{f,\lambda}$ is independent of our choice in part 1, and only depends on $\rho_{f,\lambda}$.

We call a representation modular if it arises this way.

Definition 3.7 (Modular mod l representation, [DS05, Definition 9.6.10])**.** Let

$$\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$$

be an irreducible representation, with $\mathbb{F}$ a finite field of characteristic $l \geq 3$. Then $\overline{\rho}$ is modular of level N if there exists a newform $f \in S_2(\Gamma_1(N))$ and $\lambda \subset \mathcal{O}_f$ lying over l, such that $\overline{\rho} \sim \overline{\rho}_{f,\lambda}$.

Remark 3.8. Note here that unlike in definition 3.1, we allow $f \in S_2(\Gamma_1(N))$ to have non-trivial character.

Clearly if $\rho_{E,l}$ is modular, then $\overline{\rho}_{E,l}$ is modular. It's not at all obvious that the converse direction is helpful. However, a remarkable insight of Wiles in the proof of the Modularity Theorem for semistable elliptic curves was in proving the following:

Theorem 3.9 (Modular lifting, [CSS00, Chapter 1, Lemma 7.8])**.** *Let $E/\mathbb{Q}$ be a semistable elliptic curve and suppose that $\overline{\rho}_{E,l}$ is both modular and irreducible for some prime $l \geq 3$. Then E is modular.*

Wiles' argument showed that for $E/\mathbb{Q}$ a semistable elliptic curve, either $\overline{\rho}_{E,3}$ or $\overline{\rho}_{E,5}$ is both modular and irreducible, thus showing that E must be modular. A good overview of Wiles' proof strategy (and one a bit easier to read than Wiles' 100 page paper!) can be found in [CSS00, Chatper 1, §7].

3.1.3 Serre's modularity conjecture

The Modularity Theorem can also be viewed as a consequence of Serre's modularity conjecture, which we will briefly introduce. An excellent resource is Serre's 1987 article [Ser87].

Serre considers mod l Galois representations as those with image contained in $\mathrm{GL}_2(\overline{\mathbb{F}}_l)$. Indeed, any mod l Galois representation which is a $\mathbb{F}$-module can be viewed as such, upon extension of scalars to $\overline{\mathbb{F}}$. Serre's conjecture concerns odd Galois representations.

Definition 3.10 (Odd Galois representation)**.** Let $\mathbb{F}$ be a finite field of characteristic $l \geq 3$. A two-dimensional Galois representation $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ is odd if the matrix $\rho(c)$ has eigenvalues $+1$ and -1, where c is a complex conjugation automorphism in $G_{\mathbb{Q}}$.

For example, $\overline{\rho}_{f,\lambda}$ is odd when $l \geq 3$ since $\det \rho_{f,\lambda}(c) = -1$. Importantly, in the case of odd Galois representations, viewing them as $\mathbb{F}$-modules or $\overline{\mathbb{F}}$ modules does not affect the reducibility.

Lemma 3.11. *Let $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be a mod l odd Galois representation. Then $\overline{\rho}$ is irreducible if and only if it is absolutely irreducible.*

Proof. Absolutely irreducible implies irreducible. Conversely, suppose that $\overline{\rho}$ is not absolutely irreducible, and there is a $G_{\mathbb{Q}}$-invariant one-dimensional subspace

$$\langle w \rangle \subset \overline{\mathbb{F}}^2.$$

The matrix $\overline{\rho}(c)$ has order 2, and eigenvalues ± 1, which are distinct mod l. There are eigenvectors v_1, $v_{-1} \in \mathbb{F}^2$ for the eigenvalues 1, -1 respectively. But then w must be a scalar multiple of either v_1 or v_{-1}, and so we obtain a one dimensional subspace of $\mathbb{F}$ left invariant by $G_{\mathbb{Q}}$, i.e. $\overline{\rho}$ is not irreducible. □

Theorem 3.12 (Serre's modularity conjecture, weak version)**.** *Every irreducible odd representation $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_l)$ is modular.*

The strong version of Serre's conjecture additionally predicts a level $N(\overline{\rho})$ (see [Ser87, §1.2] for definition) and weight $k(\overline{\rho})$ (see [Ser87, §2] for definition) associated to the corresponding newform. Significantly, Serre's modularity conjectures are now theorems, proven by Chandrashekhar Khare and Jean-Pierre Wintenberger in [KW09].

Serre outlines in [Ser87, §4.6] how his modularity conjecture implies the Modularity Theorem. In his paper he also proposed the ε-conjecture, which we now know as Ribet's level lowering theorem. Serre showed how the Taniyama-Shimura conjecture, as well as the ε conjecture, would imply Fermat's Last Theorem, and so proving these became the aim.

3.2 Ribet's level lowering theorem

Ribet's level lowering theorem takes as input a mod l irreducible representation

$$\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}),$$

where $\mathbb{F}$ is a finite field of characteristic $l \geq 3$. Moreover, we assume that $\overline{\rho}$ is modular of level N, as in definition 3.7. Then $\overline{\rho}$ is odd, and hence absolutely irreducible.

Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. The Modularity Theorem implies the existence of a rational newform $f \in S_2(\Gamma_0(N))$ such that $a_p(E) = a_p(f)$ for all p prime.

But definition 3.7 is less stringent than this; $\overline{\rho}_{E,l}$ can be modular of a level different to the conductor of E. Ribet's level lowering theorem lowers this level:

Theorem 3.13 (Ribet's level lowering theorem, [CSS00, Chapter 1, Theorem 4.5])**.** *Let $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be a modular mod l representation of level N. Let p be a prime such that $p \| N$. Suppose that $\overline{\rho}$ is absolutely irreducible and that one of the following is true:*

- *$\overline{\rho}$ is unramified at p, or,*
- *$p = l$ and $\overline{\rho}$ is flat at l.*

Then $\overline{\rho}$ is modular of level N/p.

We have not defined the condition of flatness of a representation at a prime. The definition involves the concept of finite flat group schemes. We will avoid this (though a reference is [CSS00, Chapter V]), and content ourselves with the following:

Theorem 3.14 ([CSS00, Chapter 1, Theorem 2.11])**.** *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N and minimal discriminant $\Delta_{\min}$. Suppose that $l \| N$. Let $\overline{\rho}_{E,l}$ be the mod l Galois representation associated to E. Then $\overline{\rho}_{E,l}$ is flat at l if and only if $l \mid v_l(\Delta_{\min})$.*

Recall from proposition 2.9 that if $p \neq l$ and E has multiplicative reduction at p, i.e. $p \| N$, then $\overline{\rho}_{E,l}$ is unramified at p if and only if $v_p(\Delta_{\min}) \equiv 0 \pmod{l}$. This leads us to the following definition:

Definition 3.15 ([Sik12, §5.2])**.** Let E be an elliptic curve over $\mathbb{Q}$. Let $\Delta_{\min}$ be its minimal discriminant, and N the conductor. Given a prime l, define

$$N_l = N/\Big(\prod_{p \| N,\ l \mid v_p(\Delta_{\min})} p\Big).$$

Notation 3.16. Let l be a prime, $E/\mathbb{Q}$ a rational elliptic curve, and suppose $\overline{\rho}_{E,l}$ is modular of some level N. We write $E \sim_l f$ to mean $\overline{\rho}_{E,l} \sim \overline{\rho}_{f,\lambda}$ for $f \in S_2(\Gamma_1(N_l))$ and $\lambda \subset \mathcal{O}_f$ lying over l.

A direct consequence of theorem 3.13, theorem 3.14, and proposition 2.9 is the following result.

Theorem 3.17. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N, and let l be an odd prime. Suppose that $\overline{\rho}_{E,l}$ is irreducible. Then there exists a newform f of level N_l such that $E \sim_l f$.*

Proof. By the Modularity Theorem, $\overline{\rho}_{E,l}$ is modular of level N. For each prime divisor p of N/N_l, the representation $\overline{\rho}_{E,l}$ is unramified at p (cf. theorems 2.9 and 3.14). Therefore we get that $\overline{\rho}_{E,l}$ is modular of level N/p by Ribet's level lowering theorem. Applying this successively to each prime factor of N/N_l yields the result. □

Corollary 3.18. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N and $f \in S_2(\Gamma_1(N_l))$ a newform. Let l be a prime, and suppose $E \sim_l f$. Then there is some prime ideal $\lambda \subset \mathcal{O}_f$ lying over l such that for all primes $p \nmid lN$, $a_p(E) \equiv a_p(f) \pmod{\lambda}$.*

This gives us a method of finding the newform f of level N_l. The following example runs through this, making use of the LMFDB databases [LMF23], and computations in Magma [BCP97].

Example 3.19 ([Sik12, Exercise 5.2])**.** Consider the elliptic curve

$$E\colon y^2 + xy = x^3 - x^2 - 47808x + 4476064$$

with Cremona label 342f4. This is an elliptic curve of conductor $N = 342 = 2 \cdot 3^2 \cdot 19$. Using Magma, one can compute that E has minimal discriminant $\Delta_{\min} = 2^5 \cdot 3^{18} \cdot 19^4$. Therefore if we consider $l = 5$, we have that $N_5 = N/2 = 171$.

The command `Newforms(CuspForms(171))` in Magma generates a table of the newforms of level 171. There are four rational newforms at this level, and one with $K_f \cong \mathbb{Q}[x]/(x^4 - 9x^2 + 12) \cong \mathbb{Q}(a)$. We just keep track of one of the Galois conjugates of this fifth newform.

Table 2: Fourier coefficients of newforms of level 171 and $a_p(E)$ for primes $p \leq 19$

p	2	3	5	7	11	13	17	19
$a_p(E)$	-1	0	-2	0	4	2	6	-1
$a_p(f_1)$	-1	0	2	0	0	6	6	-1
$a_p(f_2)$	2	0	-1	3	3	-6	-3	-1
$a_p(f_3)$	2	0	3	-5	-1	2	1	-1
$a_p(f_4)$	0	0	-3	-1	-3	-4	3	1
$a_p(f_5)$	a	0	$\frac{1}{2}(-a^3+5a)$	$-a^2+5$	$\frac{1}{2}(a^3-9a)$	2	$\frac{1}{2}(-a^3+5a)$	1

Studying this table allows us to determine which newform f satisfies $E \sim_5 f$.

- Looking at $p = 7 \nmid lN$, corollary 3.18 implies $E \not\sim_5 f_2$ since $a_7(f_2) = 3 \not\equiv 0 = a_7(E) \pmod 5$, and likewise $E \not\sim_5 f_4$.
- Likewise looking at $p = 11$ shows that $E \not\sim_5 f_1$.
- One can check that $x^4 - 9x^2 + 12$ is irreducible mod 5, so (5) is inert in $\mathcal{O}_f$ and is the only prime lying above 5. At $p = 7$, $a_p(f_5) = -a^2 + 5 \notin 5\mathcal{O}_f$ since this implies $a^2, a^4 \in 5\mathcal{O}_f$ and $a^4 - 9a^2 + 12 \in 5\mathcal{O}_f$, so $12 \in 5\mathcal{O}_f \cap \mathbb{Z} = 5\mathbb{Z}$ $\Rightarrow\Leftarrow$. Therefore $a_7(E) \not\equiv a_7(f_5) \pmod{(5)}$ so $E \not\sim_5 f_5$.

By the Modularity Theorem, $\overline{\rho}_{E,l}$ is modular. We admit that $\overline{\rho}_{E,l}$ is irreducible. Thus we can apply Ribet's level lowering theorem, and theorem 3.17 to yield that $E \sim_5 f$ for some newform $f \in S_2(\Gamma_1(171))$. Since f_3 is the only one we haven't ruled out, we must have $E \sim_5 f_3$.

4 Modularity Theorem implies Fermat's Last Theorem

4.1 A certain type of Frey curve

In this subsection we study the properties of a certain Frey curve that we will then use to deduce Fermat's Last Theorem. The following arguments are taken from §4 of Serre's paper [Ser87].

Consider three non-zero, pairwise coprime integers A, B, C such that $A+B+C=0$. To this equation we associate a Frey curve

$$E\colon\ y^2 = x(x-A)(x+B). \tag{5}$$

It is clear that $\Delta = 16A^2B^2(A+B)^2 = 16(ABC)^2$. Also, one calculates $c_4 = 16(A^2+AB+B^2) = 16(-CA+B^2) = 16(A^2-CB)$.

Lemma 4.1. *E has multiplicative reduction at all odd primes q dividing ABC.*

Proof. The odd primes of bad reduction are those dividing Δ, i.e. those dividing ABC. For such a q, it divides exactly one of A, B, C, as these are coprime.

- If $q \mid C$ then $v_q(c_4) = 0$, since if $q|c_4$ this implies $q|B^2$, $q|A^2$, a contradiction. Therefore our Weierstrass equation for E is minimal at q ([Sil1, Chapter VII, §1 Remark 1.1]).

  Then $x(x-A)(x+B) \equiv x(x-A)^2 \pmod{q}$. The reduction mod q has a double root at $(A,0)$, hence reduction of multiplicative type.

- Let $q|AB$, that is $q|A$ or $q|B$. In either case we deduce from our expressions for c_4 that $v_q(c_4) = 0$, and our Weierstrass equation is minimal at q. The reduction mod q has a double root at $(0,0)$, so we have multiplicative reduction also.

□

Since $v_2(c_4) \geq 4$, this equation is not necessarily minimal at 2. In our application of this Frey curve to Fermat's Last Theorem, the integers A, B will additionally satisfy

$$A \equiv -1 \pmod{4}, \quad B \equiv 0 \pmod{32}.$$

In this case, we can describe the reduction at 2, and compute the minimum discriminant.

Proposition 4.2. *Let $E/\mathbb{Q}$ be the elliptic curve with Weierstrass equation* (5)*. Suppose that $A \equiv -1 \pmod 4$ and $B \equiv 0 \pmod{32}$. Then,*

1. *The global minimum discriminant of E is $\Delta_{\min} = 2^{-8}(ABC)^2$,*
2. *E has multiplicative reduction at 2,*
3. *E is semistable, with conductor $N = \mathrm{Rad}(ABC)$.*

Proof.

1. By a change of variables $x = 4X$, $y = 8Y + 4X$, our equation (5) for E takes the form

$$Y^2 + XY = X^3 + cX^2 + dX \tag{6}$$

where $c = (B - 1 - A)/4$, $d = -AB/16$. The conditions on A, B ensure that c, $d \in \mathbb{Z}$. Then $\Delta = 2^{-8}(ABC)^2$ and $c_4 = C^2$. Since B is even, C is odd, so $v_2(c_4) = 0$ and hence (6) is minimal at 2. Noting that $v_q(16(ABC)^2) = v_q(2^{-8}(ABC)^2)$ for all odd primes q, and that the equation (5) is minimal at such a prime, we see that

$$\Delta_{\min} = 2^{-8}(ABC)^2$$

is the global minimum discriminant, and (6) is the global minimum Weierstrass equation for E.

2. Reducing (6) mod 2, we obtain

$$Y^2 + XY = \begin{cases} X^3 & \text{if } A \equiv 7 \pmod 8 \\ X^3 + X^2 & \text{if } A \equiv 3 \pmod 8 \end{cases}$$

and in either case the reduction yields a cubic on $\mathbb{F}_2$ with a double point at $(0, 0)$. Hence E has multiplicative reduction at 2.

3. By lemma 4.1 and part 2, E only has multiplicative bad reduction, hence is semistable. The conductor is $N = \prod_{p|ABC} p = \mathrm{Rad}(ABC)$.

□

By the Modularity Theorem, the Galois representation $\overline{\rho}_{E,l}$ associated to E is modular of level $\mathrm{Rad}(ABC)$. We want to be able to lower this level, and to apply Ribet's level lowering theorem, we need to ensure that $\overline{\rho}_{E,l}$ is absolutely irreducible. We can deduce this from Mazur's theorem on torsion subgroups:

Theorem 4.3 (Mazur, [Sil1, Chapter VIII, Theorem 7.5])**.** *Let $E/\mathbb{Q}$ be an elliptic curve. Then the torsion subgroup $E_{\mathrm{tors}}(\mathbb{Q})$ is isomorphic to one of the following groups:*

- $\mathbb{Z}/M\mathbb{Z}, \quad 1 \leq M \leq 10 \quad or \quad M = 12,$
- $\mathbb{Z} \times \mathbb{Z}/2M\mathbb{Z}, \quad 1 \leq M \leq 4.$

By consequence, as our Frey curve in (5) contains all its 2-torsion in $E(\mathbb{Q})$, we get the following result.

Proposition 4.4 ([Ser87, §4, Proposition 6])**.** *Let $E/\mathbb{Q}$ be the elliptic curve defined in* (5)*. Then for $l \geq 5$, the mod l representation $\overline{\rho}_{E,l}$ is irreducible.*

Proof. Suppose that $\overline{\rho}_{E,l}$ is reducible. Since $E[l] \simeq (\mathbb{Z}/l\mathbb{Z})^2$, this implies $E(\overline{\mathbb{Q}})$ admits a subgroup X of order l that is $G_{\mathbb{Q}}$-invariant. Then there is a rational isogeny

$$\phi \colon E \to E'$$

with $\ker \phi = X$ and E' an elliptic curve over $\mathbb{Q}$ (see proposition 4.12 and remark 4.13.2 in [Sil1, Chapter 3, §4]).

As E is semistable, Serre quotes his results on page 307 of [Ser72] to deduce that either E or E' has a rational point of order l.

But both $E(\mathbb{Q})$ and $E'(\mathbb{Q})$ contain the full 2-torsion subgroup $(\mathbb{Z}/2\mathbb{Z})^2$. Therefore neither can contain the subgroup $\mathbb{Z}/l\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})^2$, as this contradicts Mazur's theorem. The result follows. □

The absolute reducibility of $\overline{\rho}_{E,l}$ follows from lemma 3.11.

4.2 Deducing Fermat's Last Theorem

Now we have all the tools at hand to prove Fermat's Last Theorem. Recall from the introduction that we only need to prove Fermat's Last Theorem for prime exponents.

The case of $p = 3$ can be proven in multiple ways, for instance it was proven by Euler by a method of infinite descent. We're left with proving Fermat's Last Theorem for $p \geq 5$. We show that there are no pairwise coprime solutions. This suffices since any potential solution can be reduced to this case.

Theorem 4.5 (Fermat's Last Theorem for $p \geq 5$)**.** *Let $p \geq 5$ be a prime and consider the Diophantine equation*

$$x^p + y^p = z^p.$$

This has no primitive solutions, i.e. integer solutions $x, y, z \in \mathbb{Z}$ with $xyz \neq 0$ and x, y, z pairwise coprime.

Proof. We prove this by contradiction. Suppose that there does exist a primitive solution (a, b, c). By coprimality one of a, b, c is even, and the other two odd. After possibly rearranging, and changing a, b, c up to sign, we may assume that b is even and $a \equiv -1 \pmod 4$. Since $p \geq 5$, then $b^p \equiv 0 \pmod{32}$.

This solution generates a Frey curve by letting $A = a^p$, $B = b^p$ as in the previous section:

$$E\colon\ Y^2 = X(X - a^p)(X + b^p).$$

Therefore by proposition 4.2

$$N = \mathrm{Rad}(a^p b^p c^p) = \mathrm{Rad}(abc), \quad \Delta_{\min} = 2^{-8}(abc)^{2p}.$$

Note that for all odd primes q with $q \| N$, we have $p \mid v_q(\Delta_{\min})$. Thus it seems profitable to try apply Ribet's level lowering theorem to the representation $\overline{\rho}_{E,p}$. Note that $\overline{\rho}_{E,p}$ is irreducible by proposition 4.4. Therefore, as in definition 3.15, we have $N_p = 2$, and there exists a newform f of level 2 such that $E \sim_p f$ by theorem 3.17.

But as we saw in example 1.20, there are no cusp forms of weight 2 corresponding to the congruence subgroup $\Gamma_1(2)$. Thus we get a contradiction, thereby proving Fermat's Last Theorem. □

5 The modular approach to Diophantine equations

The approach taken in the previous section to prove Fermat's Last Theorem is now known more generally as the modularity approach. It's clear that the methods used should allow us to tackle more general Diophantine equations. In this final section, we explore this approach with a couple of examples. Our main source is Samir Siksek's notes [Sik12].

Here is an example where we do not need to quote the Modularity Theorem or level lowering; it is enough to consider the Frey cure we constructed in §4.1.

Example 5.1 ([Sik16, Exercise 3])**.** We solve the equation

$$2^a = 3^b + 5^c, \qquad a \geq 5$$

for $(a, b, c) \in \mathbb{Z}^3$. Let (a, b, c) be a solution. Let $A = -5^c$, $B = 2^a$, and $C = -3^b$. Then $A + B + C = 0$ and $A \equiv -1 \pmod 4$, $B \equiv 0 \pmod{32}$. Therefore by proposition 4.2 the elliptic curve

$$E\colon y^2 = x(x + 5^c)(x + 2^a)$$

has conductor $N = 30$. We can use the LMFDB database [LMF23] to find rational elliptic curves with conductor 30, and whose rational torsion subgroup contains the subgroup $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

There are two up to isomorphism, with Cremona labels 30a6 and 30a2.

- The elliptic curve $E\colon x(x+5^3)(x+2^7)$ has Cremona label 30a6 and this corresponds to the solution $2^7 = 3 + 5^3$.
- The elliptic curve $E\colon x(x+5)(x+2^5)$ has Cremona label 30a2 and this corresponds to the solution $2^5 = 3^3 + 5$.

Noting that the j-invariant of E depends on the solution, it's clear these are the only solutions to our Diophantine equation.

We consider Frey curves more generally. The following is an ad hoc definition.

Definition 5.2 (Frey curve, [Sik12, §7])**.** For our purposes, a Frey curve is a rational elliptic curve E associated to a solution of a Diophantine equation satisfying the following properties

1. The coefficients of E depend on the solution to the Diophantine equation.
2. The minimal discriminant of E can be written as $\Delta_{\min} = C \cdot D^p$ where D is a term that depends on the solution of the Diophantine equation. The C factor only depends on the equation itself.
3. E has multiplicative reduction at primes dividing D.

The upshot of this definition is that N_p from definition 3.15 will then only depend on C.

An application of the modular approach that has been studied extensively is the consideration of the following type of Diophantine equation

$$Ax^p + By^q = Cz^r \tag{7}$$

with p, q, $r \geq 2$. We look for primitive solutions, namely those with $xyz \neq 0$ and $\gcd(x, y, z) = 1$. We call (p, q, r) the signature of (7). When $A = B = C = 1$, (7) is known as the Generalized Fermat equation.

In [BS04], Bennett and Skinner provide recipes for assigning Frey curves to solutions of equation (7) when the signature is $(p, p, 2)$, for p a prime. They impose that A, B, C are non-zero, and pairwise coprime. We illustrate an example.

Example 5.3 ([Sik16, Exercise 4])**.** We show that

$$u^2 + 5 = v^p, \qquad p \text{ prime } p \geq 7 \tag{8}$$

has no solutions with u, v coprime and non-zero.

A solution (u, v) of (8) yields a solution $(x, y, z) = (v, -1, u)$ of (7) with $A = 1, B = 5, C = 1$, and signature $(p, p, 2)$.

Suppose that (u, v) is a solution. Then the method in [BS04] suggests we look at the Frey curve

$$E\colon y^2 = x^3 + 2ux^2 - 5x. \tag{9}$$

One computes that this has discriminant $\Delta = 2^6 \cdot 5^2 \cdot v^p$, and $c_4 = 2^4(4v^p - 5)$, $c_6 = -2^6 u(8v^p + 5)$. Note that u must be even since u odd implies $u^2 + 1 \equiv 2 \pmod 4$, but we cannot have $v^p \equiv 2 \pmod 4$.

- If $q|v$ is a prime and $q \neq 5$, then $v_q(c_4) = v_q(4v^p - 5) = v_q(-5) = 0$ so (9) is minimal at q, and can be checked to have multiplicative reduction at q, hence $q \| N$.
- Note that $5|u$ implies $5|v$, a contradiction since $\gcd(u, v) = 1$. Therefore $v_5(c_4) = v_5(4u^2 + 15) = v_5(4u^2) = 0$ so (9) is minimal at 5, and again (9) has multiplicative reduction at 5, so $5 \| N$.

To obtain $v_2(N)$ one needs to use Tate's algorithm. The results of [BS04] tell us that

$$\Delta_{\min} = 2^6 \cdot 5^2 \cdot v^p, \qquad N = 2^5 \operatorname{Rad}(5v)$$

therefore, $N_p = 2^5 \operatorname{Rad}(5v)/\operatorname{Rad}(v) = 2^5 \cdot 5 = 160$. We admit that we can apply Ribet's level lowering so that $E \sim_p f$ for some newform $f \in S_2(\Gamma_1(160))$.

There are two rational newforms f_1, f_2 at this level, and one irrational f_3 with $K_{f_3} = \mathbb{Q}(\sqrt{2})$. To derive a contradiction, we will need to use the following addendum to corollary 3.18.

Proposition 5.4 ([Sik12, Proposition 5.1])**.** *Suppose $E/\mathbb{Q}$ has conductor N and $E \sim_l f$. Then there is some prime ideal $\lambda \mid l$ of $\mathcal{O}_f$ such that*

1. *If $p \nmid lNN_l$, then $a_p(E) \equiv a_p(f) \pmod \lambda$,*
2. *If $p \nmid lN_l$ and $p \| N$, then $p + 1 \equiv \pm a_p(f) \pmod \lambda$.*

Example 5.3 continued Consider $q = 3$. By looking at $u^2 + 5 = v^p \pmod 3$ and noting that u is even, we see that $3 \mid v$, so $3 \| N$. But $3 \nmid pN_p$ since $p \geq 7$. Therefore we can apply part 2 of the above proposition. One checks that $a_3(f_1) = -2$, $a_3(f_2) = 2$, $a_3(f_3) = 2\sqrt{2}$. Then $\pm a_3(f_1)$ and $\pm a_3(f_2)$ are not congruent to $3 + 1 = 4 \pmod p$ since $p \geq 7$.

For $a_3(f_3)$, we have $N_{K_{f_3}/\mathbb{Q}}(4 \pm 2\sqrt{2}) = 8$. K_{f_3} is a UFD and $4 \pm 2\sqrt{2}$ cannot be contained in a prime ideal lying over p, i.e part 2 of the above proposition cannot be satisfied.

Therefore $E \not\sim_p f$ for any $f \in S_2(\Gamma_1(N_p))$, a contradiction. Thus our assumption that (u, v) was a solution was incorrect and we are done.

In the spirit of the proof of Fermat's Last Theorem, it is useful to try level-lower to a level at which there are no newforms. A comprehensive list is the following:

Proposition 5.5. *Let $N \in \mathbb{Z}_{>0}$. Then there are no newforms of level N if and only if*

$$N \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 16, 18, 22, 25, 28, 60\}.$$

This is proved by genus computations.

Example 5.6 ([Sik12, Theorem 6, §8])**.** We prove that there are no primitive solutions to the equation

$$x^p + 2^r y^p + z^p = 0, \qquad p \geq 5,\ r \geq 2 \tag{10}$$

This is Fermat's Last Theorem at $r = 0$. We associate to a potential solution (x, y, z) a similar looking Frey curve:

$$E\colon y^2 = X(X - x^p)(X + 2^r y^p).$$

Now $c_4 = 16(z^{2p} + 2^r y^p x^p)$ and $\Delta = 2^{r+4}(xyz)^{2p}$. An application of Tate's algorithm shows that

$$\Delta_{\min} = \begin{cases} 2^{2r+4}(xyz)^{2p} & \text{if } 16 \nmid 2^r y^p, \\ 2^{2r-8}(xyz)^{2p} & \text{if } 16 \mid 2^r y^p, \end{cases}$$

and

$$N = \begin{cases} 2\,\mathrm{Rad}_2(xyz) & \text{if } r \geq 5 \text{ or } y \text{ even}, \\ \mathrm{Rad}_2(xyz) & \text{if } r = 4 \text{ and } y \text{ odd}, \\ 8\,\mathrm{Rad}_2(xyz) & \text{if } r = 2, 3 \text{ and } y \text{ odd}. \end{cases}$$

Here $\mathrm{Rad}_2(N) \coloneqq \prod_{p \mid N, p \neq 2} p$. Applying the definition of N_p, we get

$$N_p = \begin{cases} 2 & \text{if } r \geq 5, \\ 1 & \text{if } r = 4, \\ 2 & \text{if } r = 2, 3 \text{ and } y \text{ even}, \\ 8 & \text{if } r = 2, 3 \text{ and } y \text{ odd}. \end{cases}$$

Note that E contains all its 2-torsion over $\mathbb{Q}$. If E is semistable, the same proof as in proposition 4.4 shows that $\overline{\rho}_{E,p}$ is irreducible. Otherwise, E is not semistable and $v_2(N) = 3$. In this case we can use [Sik12, Theorem 5, §6] to show that $\overline{\rho}_{E,p}$ is irreducible. Since $\overline{\rho}_{E,p}$ is additionally modular, we can apply Ribet's level lowering theorem. But there are no newforms at level N_p in all cases by proposition 5.5. Thus our assumption that (10) has a solution is wrong.

6 Conclusion and outlook

This essay has focused on describing the Modularity Theorem in terms of Galois representations. We have exhibited the power and utility of these objects. With all the necessary theory, the proof of Fermat's Last Theorem 4.5 turns out to be short and elegant. The purpose of this essay was to arrive at this satisfying proof and to give the reader an understanding of the route towards the proof. Of course, the full technical details are vast and encompass numerous bodies of work, but we hope to have provided insight into the area of modularity of elliptic curves.

We conclude with discussing some broader contexts. While Wiles was heavily motivated by proving Fermat's Last Theorem in his endeavours to prove the Modularity Theorem, we should really consider Fermat's Last Theorem as one of the many consequences of his result.

Some of these consequences are results that were already known modulo the Modularity Theorem. For example, an important consequence that we did not explore is that it shows that the L-function $L(E, s)$ of an elliptic curve $E/\mathbb{Q}$ can be analytically continued to the entire complex plane (see [Sil1, Appendix C, §16] for a definition of $L(E, s)$ and discussion of analytic continuation).

Siksek's article [Sik12] gives insight into applications of the modular approach. The Generalized Fermat equation has been studied comprehensively in the case of certain signatures, but this is still an active area of research. Development in this area also relies on more general versions of modularity of elliptic curves over number fields. The article by Bennett, Mihailescu and Siksek [BMS16] surveys this area.

Another consequence of the Modularity Theorem is developments in the area of modularity lifting, inspired by the proofs in Wiles' paper [Wil95]. Breuil, Conrad, Diamond and Taylor first built on Wiles' theory by proving the modularity theorem for all rational elliptic curves [BCDT01]. Approaches at proving modularity theorems over fields other than $\mathbb{Q}$ have all followed the strategy of Wiles' work; using lifting theorems.

Nowadays, the Modularity Theorem is a result in a broader area known as the Langlands Program. This program seeks to attach to automorphic forms (e.g. modular forms) algebraic geometric objects (e.g. elliptic curves). This area is one of the primary focuses of modern-day number theory. The survey article by Calegari [Cal21] discusses the effect of Wiles' and Taylor's results in this area. We have explored one result in this area, and hope that it encourages the reader to delve deeper.

References

[BCDT01] 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(4):843–939, 2001.

[BCP97] Wieb Bosma, John Cannon, and Catherine Playoust. The Magma algebra system. I. The user language. *J. Symbolic Comput.*, 24(3-4):235–265, 1997. Computational algebra and number theory (London, 1993).

[BMS16] Michael Bennett, Preda Mihailescu, and Samir Siksek. *The Generalized Fermat Equation*, pages 173–205. 07 2016.

[BS04] Michael A. Bennett and Chris M. Skinner. Ternary diophantine equations via galois representations and modular forms. *Canadian Journal of Mathematics*, 56(1):23–54, 2004.

[Cal21] Frank Calegari. Reciprocity in the langlands program since fermat's last theorem, 2021.

[Car86] Henri Carayol. Sur les représentations galoisiennes modulo l attachées aux formes modulaires de hilbert. 1986.

[CSS00] Gary Cornell, Joseph H. Silverman, and Glenn Stevens. *Modular Forms and Fermat's Last Theorem*. Springer New York, NY, 2000.

[DDT95] Henri Darmon, Fred Diamond, and Richard Taylor. Fermat's last theorem. 1995.

[DS05] F. Diamond and J. Shurman. *A First Course in Modular Forms*. Springer, 2005.

[Fal83] G. Faltings. Endlichkeitssätze für abelsche varietäten über zahlkörpern. *Inventiones mathematicae*, 73, 1983.

[KW09] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture (I). *Inventiones Mathematicae*, 178(3):485–504, July 2009.

[LMF23] The LMFDB Collaboration. The L-functions and modular forms database. `https://www.lmfdb.org`, 2023.

[Neu99] Jürgen Neukirch. *Algebraic Number Theory*. Springer Berlin, Heidelberg, 1999.

[Rib90] Kevin Ribet. On modular representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms. *Invent. Math. 100, no. 2*, pages 431–47, 1990.

[Ser72] Jean-Pierre Serre. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. *Invent. Math.*, pages 259 – 331, 1972.

[Ser87] Jean-Pierre Serre. Sur les représentations modulaires de degreé 2 de $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. *Duke Mathematical Journal*, 54:179 – 230, 1987.

[Sik12] Samir Siksek. The modular approach to diophantine equations. *Explicit Methods in Number Theory: Rational Points and Diophantine Equations*, 2012.

[Sik16] Samir Siksek. Galois representations and diophantine equations exercises 1. `http://homepages.warwick.ac.uk/staff/S.Siksek/sarajevo/exercises1.pdf`, 2016.

[Sil1] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics, 1986. Springer New York, NY, second edition, 1.

[Sil2] Joseph H. Silverman. *Advanced Topics in the Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics, 1994. Springer New York, NY, 2.

[Wei48] André Weil. *Variétés abéliennes et courbes algébriques*. Actualités scientifiques et industrielles, 1948.

[Wil95] Andrew Wiles. Modular elliptic curves and fermat's last theorem. *Annals of Mathematics*, 141(3):443–551, 1995.