25 Fermat's Last Theorem

In this final lecture we give an overview of the proof of Fermat's Last Theorem. Our goal is to explain what Andrew Wiles [21], with the assistance of Richard Taylor [19], proved, and why it implies Fermat's Last Theorem. This implication is a consequence of earlier work by several mathematicians, including Richard Frey, Jean-Pierre Serre, and Ken Ribet. We will say very little about the details of Wiles' proof, which are beyond the scope of this course, but we will provide references for those who wish to learn more.

25.1 Fermat's Last Theorem

In 1637, Pierre de Fermat famously wrote in the margin of a copy of Diophantus' *Arithmetica* that the equation

$$x^n + y^n = z^n$$

has no integer solutions with $xyz \neq 0$ and $n > 2$, and claimed to have a remarkable proof of this fact. As with most of Fermat's work, he never published this claim (mathematics was a hobby for Fermat, he was a lawyer by trade). Fermat's marginal comment was apparently discovered only after his death, when his son Samuel was preparing to publish Fermat's mathematical correspondence, but it soon became well known and is included as commentary in later printings of *Arithmetica*.

Fermat did prove the case $n = 4$, using a descent argument. It then suffices to consider only cases where n is an odd prime, since if $p|n$ and (x_0, y_0, z_0) is a solution to $x^n + y^n = z^n$, then $(x_0^{n/p}, y_0^{n/p}, z_0^{n/p})$ is a solution to $x^p + y^p = z^p$.

A brief chronology of the progress made toward proving Fermat's Last Theorem prior to Wiles' work is listed below below.

1637	Fermat makes his conjecture and proves it for $n = 4$.
1753	Euler proves FLT for $n = 3$ (his proof has a fixable error).
1800s	Sophie Germain proves FLT for $n \nmid xyz$ for all $n < 100$.
1825	Dirichlet and Legendre complete the proof for $n = 5$.
1839	Lamé addresses $n = 7$.
1847	Kummer proves FLT for all primes $n \nmid h(\mathbb{Q}(\zeta_n))$, called *regular* primes. This leaves 37, 59, and 67 as the only open cases for $n < 100$.
1857	Kummer addresses 37, 59, and 67, but his proof has gaps.
1926	Vandiver fills the gaps and addresses all irregular primes $n < 157$.
1937	Vandiver and assistants handle all irregular primes $n < 607$.
1954	Lehmer, Lehmer, and Vandiver introduce techniques better suited to mechanical computation and use a computer to address all $n < 2521$.
1954-1993	Computers verify FLT for all $n < 4,000,000$.

All of the results above are based on work in algebraic number theory, none of it uses elliptic curves.[1] The first person to suggest a connection between elliptic curves and Fermat's Last Theorem was Yves Hellegouarch. In his 1972 doctoral thesis [7], Hellegouarch associates

[1] Work in this direction continued even after FLT was proved. We now know that the Kummer-Vandiver conjecture $p \nmid h(\mathbb{Q}(\zeta_p)^+)$ holds for $p \leq 2^{31}$ [6]. This conjecture is a key ingredient to approaches to proving FLT using algebraic number theory (in particular, the theory of cyclotomic fields); see [20, Ch. 9] for details. We still do not know if the Kummer-Vandiver conjecture is true or not (but we do know FLT is true).

to any non-trivial solution (a,b,c) of $x^p + y^p = z^p$, with p an odd prime, the elliptic curve

$$E_{a,b,c}: \qquad y^2 = x(x-a^p)(x+b^p).$$

Without loss of generality we assume $\gcd(a,b,c) = 1$, which implies that a, b, c must be pairwise relatively prime, and that $a \equiv 3 \bmod 4$ and $b \equiv 0 \bmod 2$ (we can always swap a and b and/or multiply both sides by -1 in order to achieve this). Proving Fermat's Last Theorem then amounts to showing that no such elliptic curve $E_{a,b,c}$ can exist.

Hellegouarch did not make much progress with this, but in 1984 Gerhard Frey suggested that the elliptic curve $E_{a,b,c}$, if it existed, could not possibly be modular [5]. Shortly thereafter, Jean-Pierre Serre [15] reduced Frey's conjecture to a much more precise statement about modular forms and Galois representations, known as the *epsilon conjecture*, which was proved by Ken Ribet a few years later [13]. With Ribet's result in hand, it was then known that the modularity conjecture, which states that every elliptic curve over $\mathbb{Q}$ is modular, implies Fermat's Last Theorem: it guarantees that $E_{a,b,c}$, and therefore the solution (a,b,c) to $x^p + y^p = z^p$, cannot exist. At that time no one expected the modularity conjecture to be proved any time soon; indeed, the fact that it implies Fermat's Last Theorem was taken as evidence of how difficult it would be to prove the modularity conjecture.

25.2 A strange elliptic curve

To get a sense of what makes the elliptic curve $E_{a,b,c}$ so strange that one might question its very existence, let us compute its discriminant:

$$\Delta(E_{a.b,c}) = -16(0-a^p)^2(0+b^p)^2(a^p+b^p)^2 = -16(abc)^{2p}.$$

As explained in the last lecture, the definition of the L-series of an elliptic curve E requires us to determine the minimal discriminant of E its reduction type at each prime dividing the minimal discriminant (additive, split multiplicative, or non-split multiplicative) at each prime which divide it. It turns out that the discriminant Δ is not quite minimal, the minimal discriminant is

$$\Delta_{\min}(E_{a,b,c}) = 2^{-8}(abc)^{2p},$$

(assuming $p > 3$, which we know must be the case), which differs from Δ only at 2.

On the other hand, the conductor of $E_{a,b,c}$ is much smaller than its minimal discriminant. Recall from the previous lecture that for odd primes ℓ an elliptic curve $E : y^2 = f(x)$ can have additive reduction at ℓ only if the cubic $f \in \mathbb{Z}[x]$ has a triple root modulo ℓ. This is clearly not the case for the curve $E_{a,b,c}: y^2 = f(x) = x(x-a^p)(x+b^p)$, since 0 is always a root modulo ℓ, but a and b are relatively prime and cannot both be divisible by ℓ, so 0 is not a triple root. One can also show that $E_{a,b,c}$ does not have additive reduction at 2. This implies that $E_{a,b,c}$ is semistable, so Its conductor is the squarefree integer

$$N_{E_{a,b,c}} = \prod_{\ell \mid abc} \ell,$$

which we note is divisible by 2 (since b is).

For the elliptic curve $E_{a,b,c}$ the ratio $\Delta_{a,b,c}/N_{a,b,c}$ grows exponentially with p. But it is very unusual (conjecturally impossible) for the minimal discriminant of an elliptic curve to

be so much larger than its conductor. Szpiro's conjecture [17], which is closely related to the ABC conjecture,[2] states that we for every $\epsilon > 0$ there is a constant c_ϵ such that

$$\Delta_{\min}(E) \le c_\epsilon N_E^{6+\epsilon}$$

for every elliptic curve $E/\mathbb{Q}$. This cannot possibly be true for $E_{a,b,c}$ if p is sufficiently large. This does not imply that $E_{a,b,c}$ cannot be modular, but it suggests that there is something very strange about this elliptic curve (so strange that one might expect it cannot exist).

25.3 Galois representations

Let E be an elliptic curve over $\mathbb{Q}$, let ℓ be a prime, and let $K := \mathbb{Q}(E[\ell])$ be its *ℓ-torsion field*, the extension of $\mathbb{Q}$ obtained by adjoining the coordinates of all the points in $E[\ell]$ to $\mathbb{Q}$. The field K is a Galois extension of $\mathbb{Q}$ (it is either the splitting field of the ℓth division polynomial, or a quadratic extension of it), and its Galois group acts on the ℓ-torsion subgroup $E[\ell]$ via its action on the coordinates of each point. This yields a group representation

$$\rho\colon \mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{Aut}(E[\ell]) \simeq \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z}),$$

that maps each $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$ to the automorphism of $E[\ell] \simeq \mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}$ given by applying σ to the coordinates of each ℓ-torsion point (all of which lie in $K = \mathbb{Q}(E[\ell])$, by definition). We consider two representations $\rho, \rho'\colon \mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ to be isomorphic if there exists $A \in \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ such that $\rho'(\sigma) = A\rho(\sigma)A^{-1}$ for all $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$, in which case we write $\rho \simeq \rho'$.

Let S be the finite set of primes consisting of ℓ and the primes of bad reduction for E. Every prime $p \notin S$ is unramified in K. As explained in Lecture 20, this means that the $\mathcal{O}_K$-ideal generated by p factors into a product of distinct prime ideals:

$$p\mathcal{O}_K = \mathfrak{p}_1 \cdots \mathfrak{p}_r.$$

The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ acts transitively on the set $\{\mathfrak{p}|p\} := \{\mathfrak{p}_1, \dots, \mathfrak{p}_r\}$, and for each prime ideal $\mathfrak{p}|p$ we have a corresponding *decomposition group*

$$D_\mathfrak{p} := \{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) : \sigma(\mathfrak{p}) = \mathfrak{p}\}$$

equipped with an isomorphism

$$\begin{aligned} \varphi\colon D_\mathfrak{p} &\xrightarrow{\sim} \mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p) \\ \sigma &\mapsto \bar\sigma \end{aligned}$$

where $\mathbb{F}_\mathfrak{p} := \mathcal{O}_K/\mathfrak{p}$ is the residue field at $\mathfrak{p}$ and the automorphism $\bar\sigma$ is defined by $\bar\sigma(\bar x) = \overline{\sigma(x)}$, where $\bar x$ denotes the image of $x \in \mathcal{O}_K$ in the quotient $\mathcal{O}_K/\mathfrak{p} = \mathbb{F}_\mathfrak{p}$. The Galois group $\mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p)$ is cyclic, generated by the p-power Frobenius automorphism $\pi_p\colon x \mapsto x^p$, and we define the *Frobenius element*

$$\mathrm{Frob}_\mathfrak{p} := \varphi^{-1}(\pi_p) \in D_\mathfrak{p} \subseteq \mathrm{Gal}(K/\mathbb{Q}).$$

[2]The ABC conjecture states that for all $\epsilon > 0$ there is a constant c_ϵ such that only finitely many integer solutions to $a + b = c$ satisfy $\mathrm{rad}(abc)^{1+\epsilon} < c_\epsilon$, where $\mathrm{rad}(abc)$ denotes the squarefree part of abc. This is equivalent to a modified version of Szpiro's conjecture in which one replaces $\Delta_{\min}(E)$ with $\max(|A|^3, B^2)$, where A and B are the coefficients in a short Weierstrass equation for $E\colon y^2 = x^3 + Ax + B$. Mochizuki announced a proof of the ABC conjecture in 2012 that was finally published in 2021, but as of this writing, most number theorists do not consider the ABC conjecture to have been proved.

Different choices of $\mathfrak{p}|p$ yield conjugate $\mathrm{Frob}_\mathfrak{p}$ (and every conjugate of $\mathrm{Frob}_\mathfrak{p}$ arises for $\mathfrak{p}|p$), and we let Frob_p denote this conjugacy class; as an abuse of terminology we may speak of the Frobenius element Frob_p as an element of $\mathrm{Gal}(K/\mathbb{Q})$ representing this conjugacy class, with the understanding that Frob_p is determined only up to conjugacy.

Thus for each prime $p \notin S$ we get a Frobenius element $\mathrm{Frob}_p \in \mathrm{Gal}(K/\mathbb{Q})$, and may consider its image $A_p := \rho(\mathrm{Frob}_p) \in \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ under the Galois representation ρ. The characteristic polynomial of A_p (which depends only on the conjugacy class of Frob_p) is

$$\det(\lambda I - A_p) = \lambda^2 - (\operatorname{tr} A_p)\lambda + \det A_p,$$

with

$$\operatorname{tr} A_p \equiv a_p \bmod \ell \qquad \text{and} \qquad \det A_p \equiv p \bmod \ell.$$

Here $a_p := p + 1 - \#E_p(\mathbb{F}_p)$ is the trace of the Frobenius endomorphism of the reduction $E_p/\mathbb{F}_p$ of E modulo p, equivalently, the pth coefficient in the Dirichlet series of the L-function $L_E(s) = \sum_{n\geq 1} a_n n^{-s}$ of the elliptic curve E.

For any positive integer n we can similarly consider the Galois representation

$$\rho\colon \mathrm{Gal}(\mathbb{Q}(E[\ell^n])/\mathbb{Q}) \to \mathrm{Aut}(E[\ell^n]) \simeq \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z}).$$

For primes $p \notin S$ with $4\sqrt{p} \leq \ell^n$, the value of the integer $a_p \equiv \operatorname{tr}\rho(\mathrm{Frob}_p) \bmod \ell^n$ is uniquely determined. Note that this holds no matter which auxiliary prime ℓ we pick.

The discussion above applies not only to $\mathbb{Q}(E[\ell^n])$, but to any Galois extension K of $\mathbb{Q}$ containing $\mathbb{Q}(E[\ell^n])$. Even if the extension $K/\mathbb{Q}$ is ramified at primes outside of S, the image of $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$ under ρ depends only on the restriction of the automorphism σ to $\mathbb{Q}(E[\ell^n])$, so given a Galois representation $\rho(\mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{Aut}(E[\ell^n]) \simeq \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$ we can determine $\rho(\mathrm{Frob}_p) \in \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$ up to conjugacy. Here we use $\mathrm{Frob}_p \in \mathrm{Gal}(K/\mathbb{Q})$ to denote any element whose restriction to $\mathrm{Gal}(\mathbb{Q}(E[\ell^n])/\mathbb{Q})$ lies in the conjugacy class represented by the Frobenius element $\mathrm{Frob}_p \in \mathrm{Gal}(\mathbb{Q}(E[\ell^n])/\mathbb{Q})$. The conjugacy class of $\rho(\mathrm{Frob}_p)$ in $\mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$, and in particular its trace, is independent of this choice.

We now define the *ℓ-adic Tate module*

$$T_\ell(E) \coloneqq \varprojlim_n E[\ell^n]$$

as the projective limit of the inverse system

$$E[\ell] \xleftarrow{[\ell]} E[\ell^2] \xleftarrow{[\ell]} \cdots \xleftarrow{[\ell]} E[\ell^n \xleftarrow{[\ell]} E[\ell^{n+1}] \xleftarrow{[\ell]} \cdots,$$

whose the connecting homomorphisms are multiplication-by-ℓ maps. Elements of $T_\ell(E)$ are infinite sequences of points $(P_1, P_2, P_3, \ldots)$ with $P_n \in E[\ell^n]$ such that $\ell P_{n+1} = P_n$.

We now let $G_\mathbb{Q} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and define the *ℓ-adic Galois representation*

$$\rho_{E,\ell}\colon G_\mathbb{Q} \to \mathrm{Aut}(T_\ell(E)) \simeq \mathrm{GL}_2(\mathbb{Z}_\ell),$$

where $\mathbb{Z}_\ell = \varprojlim \mathbb{Z}/\ell^n\mathbb{Z}$ is the *ring of ℓ-adic integers*, which contains $\mathbb{Z}$ as a subring.[3] Each $\sigma \in G_\mathbb{Q}$ acts on $(P_1, P_2, P_3, \ldots) \in T_\ell$ via its action on the coordinates of each $P_n \in E[\ell^n]$.

[3]You can view elements of $\mathbb{Z}_\ell$ as infinite sequences of integers $(a_1, a_2, a_3, \ldots)$ with $a_n \equiv a_{n+1} \bmod \ell^n$, and ring operations defined coordinate-wise. We embed $\mathbb{Z}$ in $\mathbb{Z}_\ell$ via the map $a \mapsto (a, a, a, \ldots)$. Note that $\mathbb{Z}_\ell$ has characteristic 0 but comes equipped with reduction maps to the positive characteristic rings $\mathbb{Z}/\ell^n\mathbb{Z}$.

For primes $p \notin S$ we now use $\mathrm{Frob}_p \in G_{\mathbb{Q}}$ to denote an element whose restriction to $\mathrm{Gal}(\mathbb{Q}([\ell^n])/\mathbb{Q})$ is conjugate to $\mathrm{Frob}_p \in \mathrm{Gal}(\mathbb{Q}(E[\ell^n])/\mathbb{Q})$ for each $n \geq 1$; this amounts to choosing a compatible sequence of Frobenius elements $\mathrm{Frob}_{p,n} \in \mathrm{Gal}(\mathbb{Q}(E[\ell^n])/\mathbb{Q})$ such that $\mathrm{Frob}_{p,n}$ is the restriction of $\mathrm{Frob}_{p,n+1}$ to $\mathbb{Q}(E[\ell^n])$. The conjugacy class of $\rho(\mathrm{Frob}_p)$ in $\mathrm{GL}_2(\mathbb{Z}_\ell)$ is independent of these choices; in particular its trace in $\mathbb{Z}_\ell$ is well defined.

We then have $\operatorname{tr}\rho_{E,\ell}(\mathrm{Frob}_p) = a_p$, as elements of $\mathbb{Z} \subseteq \mathbb{Z}_\ell$. The representation $\rho_{E,\ell}$ thus determines the coefficients a_p of the L-series $L_E(s)$ at all primes $p \notin S$. By the Tate-Faltings Theorem (see Theorem 24.38), this determines E up to isogeny, and therefore determines the entire L-function $L_E(s)$, including the values of a_p for $p \in S$.

We also have the *mod-ℓ Galois representation*

$$\overline{\rho}_{E,\ell} \colon G_{\mathbb{Q}} \to \mathrm{Aut}(E[\ell]) \simeq \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z}),$$

which is equivalent to composing $\rho_{E,\ell}$ with the map from $\mathrm{GL}_2(\mathbb{Z}_\ell)$ to $\mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ that reduces each matrix coefficient modulo ℓ.

25.4 Serre's modularity conjecture

Let us forget about elliptic curves for a moment and consider an arbitrary[4] ℓ-adic Galois representation $\rho \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}_\ell)$ with $\ell > 3$ prime. We say that ρ is *modular* (of weight k and level N), if there is a modular form $f_\rho = \sum a_n q^n$ in $S_k(\Gamma_1(N))$ with $a_n \in \mathbb{Z}$ such that[5]

$$\operatorname{tr}\rho(\mathrm{Frob}_p) = a_p$$

for all primes $p \nmid \ell N$ (if $\rho = \rho_{E,\ell}$ and $N = N_E$ this excludes the same finite set of primes S as the previous section). Similarly, if we have a mod-ℓ representation $\overline{\rho} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$, we say that $\overline{\rho}$ is modular if

$$\operatorname{tr}\overline{\rho}(\mathrm{Frob}_p) \equiv a_p \bmod \ell$$

for all primes $p \nmid \ell N$.

Let $c \in G_{\mathbb{Q}}$ be the automorphism of $\overline{\mathbb{Q}} \subseteq \mathbb{C}$ corresponding to complex conjugation. The automorphism c has order 2, so $\det\rho(c) = \pm 1$. We say that a Galois representation ρ is *odd* when $\det\rho(c) = -1$. This is necessarily the case if $\rho = \rho_{E,\ell}$ is a Galois representation associated to an elliptic curve. One way to see this is to base change E to $\mathbb{C}$ and view $E_{\mathbb{C}}$ as isomorphic to a torus $\mathbb{C}/L$ for some lattice $L = [1, \tau]$. For a suitable choice of basis (P, Q) for the ℓ^n-torsion subgroup of $\mathbb{C}/L$ in which P has real coordinates, complex conjugation fixes P and sends Q to $-Q$ (this is easy to see when $\operatorname{re}\tau = 0$ and holds in general). Since we already know that every $f = \sum a_n q^n$ in $S_2^{\mathrm{new}}(\Gamma_0(N))$ with $a_n \in \mathbb{Z}$ gives rise to an elliptic curve (see Theorem 24.37), this constraint necessarily applies to Galois representations associated to modular forms of weight 2 with integral q-series.

We want to impose a further constraint on the Galois representations we shall consider that is not always satisfied by the representation $\overline{\rho}_{E,\ell}$ associated to an elliptic curve $E/\mathbb{Q}$, but usually is (always for $\ell > 163$). We call a Galois representation $\rho \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ *irreducible* if its image does not fix any of one-dimensional subspaces of $(\mathbb{Z}/\ell\mathbb{Z})^2$; equivalently,

[4] As profinite groups, both $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and $\mathrm{GL}_2(\mathbb{Z}_\ell)$ are topological groups and we always require ℓ-adic Galois representation to be continuous with respect to this topology; this is automatically true for the representations $\rho_{E,\ell}$ of interest to us.

[5] In the previous lecture we focused on $S_k(\Gamma_0(N))$, which suffices for everything we need in the sections that follow (and we only need $k = 2$), but in order to state Serre's conjecture we temporarily work in greater generality; note that $\Gamma_1(N) \subseteq \Gamma_0(N)$ implies $S_k(\Gamma_0(N)) \subseteq S_k(\Gamma_1(N))$.

its image is not conjugate to a group of upper triangular matrices in $\mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$. For an elliptic curve $E/\mathbb{Q}$, the mod-ℓ Galois representation $\bar{\rho}_{E,\ell}$ is irreducible if and only if E does not admit a rational ℓ-isogeny. Mazur's isogeny theorem [11] implies that this necessarily holds for $\ell \notin \{2, 3, 5, 7, 11, 13, 17, 19, 37, 43, 67, 163\}$ (the cases $19, 43, 67, 163$ can arise only when E has complex multiplication).

In 1975 Serre made the following remarkable conjecture, which he refined in [15]. This conjecture is now a theorem, proved in 2008 by Khare and Wintenberger [8, 9], but this work came long after the proof of Fermat's Last Theorem (and built on the modularity lifting techniques used to prove it).

Conjecture 25.1 (Serre's modularity conjecture)**.** *Every odd irreducible Galois representation $\bar{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ is modular.*[6]

Serre gave an explicit recipe for what the optimal weight $k(\bar{\rho})$ and level $N(\bar{\rho})$ of the corresponding modular form should be. Given a newform $f \in S_2^{\mathrm{new}}(\Gamma_0(N))$ with Fourier coefficients $a_n \in \mathbb{Z}$, the Eichler-Shimura Theorem (see Theorem 24.37) gives us a corresponding elliptic curve $E/\mathbb{Q}$ whose mod-ℓ Galois representation $\rho_{E,\ell}$ is modular of weight 2 and level $N = N_E$, and $\bar{\rho}_{E,\ell}$ will typically also be irreducible. The weight 2 agrees with the optimal weight $k(\bar{\rho}_{E,\ell})$ conjectured by Serre (at least when $\ell \nmid N_E$), but the optimal level $N(\bar{\rho}_{E,\ell})$ may properly divide N_E. In certain (rare) circumstances, distinct newforms of weight 2 with different levels may have Fourier coefficients a_n that are congruent modulo ℓ.

The mod-ℓ Galois representation associated to the "strange" elliptic curve $E_{a,b,c}$ arising from a Fermat solution $a^\ell + b^\ell = c^\ell$ gives rise to one of these rare circumstances. For an irreducible mod-ℓ Galois representations $\bar{\rho}_{E,\ell}$ arising from a semistable elliptic curve $E/\mathbb{Q}$, Serre's optimal level $N(\bar{\rho}_{E,\ell})$ is a product of primes p for which $v_p(\Delta_{\min}(E)) \not\equiv 0 \bmod \ell$, where $v_p(\cdot)$ denotes the p-adic valuation.

For the elliptic curve $E_{a,b,c}$ we have

$$N_{E_{a,b,c}} = \prod_{p|abc} p, \qquad\qquad \Delta_{\min}(E_{a,b,c}) = 2^{-8}(abc)^{2\ell},$$

which means that for every odd prime $p|N_E$ we have $v_p(\Delta_{\min}(E_{a,b,c})) \equiv 0 \bmod \ell$, in which case Serre's optimal level is $N(\bar{\rho}_{E_{a,b,c},\ell}) = 2$. But there are no (nonzero) modular forms of weight 2 and level 2, because $\dim S_2^{\mathrm{new}}(\Gamma_1(2)) = \dim S_2^{\mathrm{new}}(\Gamma_0(2)) = g(X_0(2)) = 0$. We must have $\ell > 163$, since Fermat's Last Theorem has long been known for $\ell \le 163$, so $E_{a,b,c}$ cannot admit a rational ℓ-isogeny, by Mazur's isogeny theorem, which means that $\bar{\rho}_{E_{a,b,c},\ell}$ must be irreducible. Thus if $E_{a,b,c}$ is modular, then $\bar{\rho}_{E_{a,b,c},\ell}$ represents a counterexample to Serre's conjecture. Serre's *epsilon-conjecture*, proved by Ribet in 1986, implies that this cannot happen. Below is a form of Ribet's theorem [13] that suffices to prove this.

Theorem 25.2 (Ribet)**.** *Let ℓ be prime, let E be an elliptic curve of conductor $N = mN'$, where m is the product of all primes $p|N$ such that $v_p(N) = 1$ and $v_p(\Delta_{\min}(E)) \equiv 0 \bmod \ell$. If E is modular and $\bar{\rho}_{E,\ell}$ is irreducible, then $\bar{\rho}_{E,\ell}$ is modular of weight 2 and level N'.*

Corollary 25.3. *The elliptic curve $E_{a,b,c}$ is not modular.*

[6] In fact Serre made his conjecture for all odd irreducible representations $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_{\ell^n})$, which includes the special case considered here with $\mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z}) \simeq \mathrm{GL}_2(\mathbb{F}_\ell)$.

25.5 The modularity lifting theorem

The final and by far the most difficult step to proving Fermat's Last Theorem is to show that if the elliptic curve $E_{a,b,c}$ exists, then it *is* modular. Andrew Wiles, with the assistance of Richard Taylor,[7] proved the stronger statement that every semistable elliptic curve over $\mathbb{Q}$ is modular (recall that $E_{a,b,c}$ is semistable).

A key element of Wiles' proof is a technique now known as *modularity lifting*. Let E be an elliptic curve over $\mathbb{Q}$ and let ℓ be a prime. Wiles uses modularity lifting to show that if the mod-ℓ Galois representation $\overline{\rho}_{E,\ell}$ of semistable elliptic curve $E/\mathbb{Q}$ is modular, then the ℓ-adic representation $\rho_{E,\ell}$ is also modular, which in turn implies that E is modular.

Given a representation $\rho_0 \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$, a representation $\rho_1 \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}_\ell)$ whose reduction modulo ℓ is equal to ρ_0 is called a *lift* of ρ_0. More generally, if R is a suitable ring[8] with a reduction map to $\mathbb{Z}/\ell\mathbb{Z}$, and $\rho_1 \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R)$ is a representation whose reduction is equal to ρ_0, then we say that ρ_1 is a lift of ρ_0 (to R). Two lifts of ρ_0 are said to be *equivalent* if they are conjugate via an element in the kernel of the reduction map from $\mathrm{GL}_2(R)$ to $\mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$. A *deformation* of ρ_0 is an equivalence class of lifts of ρ_0 to the ring R, which is sometimes called the *deformation ring*.

Building on work by Mazur, Hida, and others that established the existence of certain *universal deformations* $\rho_{\mathbb{T}} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T})$, where $\mathbb{T}$ is a certain Hecke algebra, Taylor and Wiles were able to show that if ρ_0 is modular, then *every* lift of ρ_0 satisfying a specified list of properties is modular (this result and generalizations of it are now known as "$R = \mathbb{T}$" theorems), and Wiles was able to show that this list of properties is satisfied by the representation $\rho_{E,\ell}$ associated to a semistable elliptic curve $E/\mathbb{Q}$.

We are intentionally glossing over a massive amount of detail that is beyond the scope of this course. We refer the interested reader to [3], which contains not only a detailed overview of the proof, but many chapters devoted to the background necessary to understand these details, and also the lecture notes from 2009-2010 Modularity lifting seminar held at Stanford [2] which covers refinements of the Taylor-Wiles method and subsequent results.

Theorem 25.4 (Taylor-Wiles)**.** *Let $E/\mathbb{Q}$ be a semistable elliptic curve. If $\overline{\rho}_{E,\ell}$ is modular, then $\rho_{E,\ell}$ is also modular (and therefore E is modular).*

25.6 Proof of Fermat's Last Theorem

It remains only to find a modular representation $\rho_0 \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/\ell\mathbb{Z})$ that we can lift to $\rho_{E,\ell}$. The obvious candidate is $\overline{\rho}_{E,\ell}$, for some suitable choice of ℓ. It is not clear that proving the modularity of $\overline{\rho}_{E,\ell}$ modular is necessarily any easier than proving the modularity of $\rho_{E,\ell}$, but thanks to work of Langlands and Tunnel on a special case of Langlands' Reciprocity Conjecture [3, Ch. 6], we have the following result for $\ell = 3$.

Theorem 25.5 (Langlands-Tunnel)**.** *Let E be an elliptic curve over $\mathbb{Q}$. If $\overline{\rho}_{E,3}$ is irreducible, then it is modular.*

The one remaining difficulty is that $\overline{\rho}_{E,3}$ is need not be irreducible; indeed there are infinitely many semistable elliptic curves $E/\mathbb{Q}$ that admit a rational 3-isogeny, and for these curves $\bar{\rho}_{E,3}$ is not irreducible. However, if E is semistable and $\overline{\rho}_{E,3}$ is reducible then $\overline{\rho}_{E,5}$ must be irreducible. This follows from the fact that if neither $\bar{\rho}_{E,3}$ nor $\bar{\rho}_{E,5}$ is irreducible

[7] Wiles' retracted his initial proof due to a gap that was found. Richard Taylor helped Wiles to circumvent this gap, which was the last critical step required to obtain a complete proof; see [4] for an accessible account.

[8] A complete local Noetherian ring with residue field $\mathbb{F}_\ell$.

then E admits both a rational 3-isogeny and a rational 5-isogeny; the cyclic group of order 15 generated by their kernels is then the kernel of a rational 15-isogeny, but this cannot be the case if E is semistable.

Theorem 25.6. *No semistable elliptic curve $E/\mathbb{Q}$ admits a rational 15-isogeny.*

Proof. Let $E/\mathbb{Q}$ be an elliptic curve that admits a rational 15-isogeny. Let $\langle P\rangle \subseteq E(\overline{\mathbb{Q}})$ be the kernel of this isogeny, which we note is necessarily cyclic. The pair $(E, \langle P\rangle)$ corresponds to a non-cuspidal $\mathbb{Q}$-rational point on $X_0(15)$, the modular curve that parameterizes $\overline{\mathbb{Q}}$-isomorphism classes of 15-isogenies. The modular curve $X_0(15)$ is a smooth projective curve of genus 1, and it has a rational point (take the cusp at infinity, for example), so it can be viewed as an elliptic curve. A minimal Weierstrass model for $X_0(15)$ is given by

$$X_0(15)\colon y^2 + xy + y = x^3 + x^2 - 10x - 10.$$

Additional information about this curve can be found on its home page in the LMFDB [10]. This information includes the fact that $X_0(15)$ has rank 0 and a torsion subgroup of order 8. Its 8 rational points include 4 cusps and 4 non-cuspidal points that represent $\overline{\mathbb{Q}}$-isomorphism classes $(E, \langle P\rangle)$ of elliptic curves $E/\mathbb{Q}$ that admit a rational 15-isogeny with kernel $\langle P\rangle$. None of these elliptic curves E has j-invariant 0 or 1728, so each isomorphism class is a family of quadratic twists. Any family of quadratic twists of elliptic curves over $\mathbb{Q}$ contains a minimal representative whose conductor divides the conductor of all others; for the 4 non-cuspidal points on $X_0(15)$ these minimal quadratic twists all have conductor $50 = 2\cdot 5^2$ (you can find a list of them and the 15-isogenies they admit here). None of these curves is semistable, since 50 is not squarefree, nor are any of their quadratic twists. The theorem follows. □

There is unfortunately no analog of the Langlands-Tunnel theorem for $\ell = 5$. Indeed, the case $\ell = 3$ is quite special: the group $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ is solvable, which is not true for any prime $\ell > 3$ (and $\ell = 2$ has other problems). So we would seem to be stuck. But Wiles cleverly proved the following result, which is now known as the *three-five trick*.

Theorem 25.7 (Wiles)**.** *Let $E/\mathbb{Q}$ be a semistable elliptic curve for which $\overline{\rho}_{E,5}$ is irreducible. There exists a semistable elliptic curve $E'/\mathbb{Q}$ such that*

- *$\overline{\rho}_{E',3}$ is irreducible,*
- *$\overline{\rho}_{E',5} \simeq \overline{\rho}_{E,5}$.*

Now we are in business.

Theorem 25.8 (Wiles)**.** *Let $E/\mathbb{Q}$ be a semistable elliptic curve. Then E is modular.*

Proof. There are two cases. If $\overline{\rho}_{E,3}$ is irreducible then:

- $\overline{\rho}_{E,3}$ is modular, by the Langlands-Tunnel theorem,
- $\rho_{E,3}$ is modular, by the modularity lifting theorem,
- E is modular, since $f_E = f_{\rho_{E,3}}$.

On the other hand, if $\overline{\rho}_{E,3}$ is reducible, then:

- $\overline{\rho}_{E,5}$ is irreducible, because no semistable $E/\mathbb{Q}$ admits a rational 15-isogeny,
- there exists a semistable $E'/\mathbb{Q}$ with $\overline{\rho}_{E',3}$ irreducible and $\overline{\rho}_{E'5} \simeq \overline{\rho}_{E,5}$, by the 3-5 trick,

- $\overline{\rho}_{E',3}$ is modular, by the Langlands-Tunnel theorem,
- $\rho_{E',3}$ is modular, by the modularity lifting theorem,
- E' is modular, since $f_{E'} = f_{\rho_{E',3}}$,
- $\rho_{E',5}$ and therefore $\overline{\rho}_{E',5}$ is modular, since $f_{\rho_{E',5}} = f_{E'}$,
- $\overline{\rho}_{E,5} \simeq \overline{\rho}_{E',5}$ is modular,
- $\rho_{E,5}$ is modular, by the modularity lifting theorem,
- E is modular, since $f_E = f_{\rho_{E,5}}$.

Q.E.D. □

Corollary 25.9. *$x^n + y^n = z^n$ has no integer solutions with $xyz \neq 0$ for $n > 2$.*

References

[1] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises*, Journal of the American Mathematical Society **14** (2001), 843–939.

[2] B. Baran, R. Bellovin, B. Conrad, S. Dasgupta, B. Levin, S. Lichtenstein, M. Lipnowski, A. Paulin, N. Mok, A. Snowden, D. Trotabas, A. Venkatesh, and M. Weissman, *Modularity lifting seminar*, lecture notes from Stanford seminar, 2009–2010.

[3] G. Cornell, J.H. Silverman, G. Stevens, *Modular forms and Fermat's Last Theorem*, Springer, 1998.

[4] G. Faltings, *The proof of Fermat's last theorem by R. Taylor and A. Wiles*, Notices of the American Mathematical Society **42** (1995), 743–746.

[5] G. Frey, *Links between stable elliptic curves and certain diophantine equations*, Annales Universitatis Saraviensis. Series Mathematicae **1** (1986), 1–40.

[6] W. Hart, D. Harvey, and W. Ong, *Irregular primes to two billion*, Math. Comp. **86** (2017), 3031–3049.

[7] Y. Hellegouarch, *Courbes elliptiques et équation de Fermat*. Thèse, Besançon, (1972).

[8] C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture (I)*, Inventiones Mathematicae **178** (2009), 485–586.

[9] C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture (II)*, Inventiones Mathematicae **178** (2009), 485–586.

[10] The LMFDB collaboration, *The L-functions and Modular Forms Database*, published electronically at `http://www.lmfdb.org`, accessed May 19, 2021.

[11] B. Mazur, *Rational isogenies of prime degree*, Invent. Math. **44** (1978), 129–162.

[12] J.S. Milne, *Elliptic curves*, BookSurge Publishers, 2006.

[13] K. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms*, Inventiones Mathematicae **100** (1990), 431–476.

[14] K. Ribet, *Galois representations and modular forms*, Bulletin of the AMS **32** (1995), 375–402.

[15] J.-P. Serre, *Sur les représentationes modulaires de degré 2 de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Duke Mathematics Journal **54** (1987), 179–230.

[16] J.H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Springer, 1994.

[17] L. Szpiro, *Discriminant et conducteur des courbes elliptiques*, in *Séminaire sur les Pinceaux de Courbe Elliptiques (Paris, 1988)*, Astérique **183** (1990), 7–18.

[18] A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Mathematische Annalen **168** (1967), 149–156.

[19] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Annals of Mathematics **141** (1995), 553–572.

[20] L. Washington, *Cyclotomic fields*, Springer, 1997.

[21] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Annals of Mathematics **141** (1995), 443-551.