FERMAT'S LAST THEOREM

FRANZ LEMMERMEYER

1. Frey Curves

Fermat learned his number theory from the books of Diophantus; it was in the margins of his copy that he wrote down that he had discovered a truly marvelous proof of the fact that $X^n + Y^n = Z^n$ has no solutions in natural numbers for $n > 2$, and that the margin was too small to contain it. The books of Diophantus, including these remarks, were published posthumously by his son. By the early 1800s, all of Fermat's theorems, claims and conjectures had been settled except for this one, which then became known as Fermat's Last Theorem.

The idea that led to the proof by Wiles is due to Frey, who, for any triple A, B, C of integers with $A + B + C = 0$ considered the elliptic curve

$$E_{A,B,C} : y^2 = x(x - A)(x + B).$$

Its discriminant is $\Delta = 16A^2B^2(A + B)^2 = 16A^2B^2C^2$. It seems to be a coninci-dence (?) that the modular curve $X_0(15)$ is the elliptic curve $y^2 = x(x-9)(x+16)$ attached to the Pythagorean triple $(3, 4, 5)$; in particular, this curve is modular.

Frey's idea was that the curves $E_{A,B,C}$ for a Fermat triple $A = a^p$, $B = b^p$, $C = -c^p$ have properties that imply it cannot be modular; since the conjecture of Taniyama-Shimura predicts that such curves do not exist, Fermat's Last Theorem would be a consequence of a proof of the Taniyama-Shimura conjecture.

Actually, Wiles could only prove that semistable elliptic curves are modular. Recall that an elliptic curve E defined over $\mathbb{Q}$ has good reduction at p if the reduction modulo p of E is nonsingular. If E has bad reduction, then there are several cases:

- The reduction modulo p of E has a node; then we say that E has multiplicative reduction. We also distinguish between split (the slopes of the tangents at the node are defined over $\mathbb{F}_p$) and non-split (the slopes of the tangents are defined $\mathbb{F}_{p^2}$ only) reduction.
- The reduction modulo p of E has a cusp. Then we say that E has additive reduction.

The notation here comes from the fact that if E has additive reduction, then $E_{\rm ns}(\mathbb{F}_p)$ is isomorphic to the additive group of $\mathbb{F}_p$. In the case of multiplicative reduction, $E_{\rm ns}(\mathbb{F}_p)$ is isomorphic to the hyperbola or the unit circle over $\mathbb{F}_p$. Now a semistable elliptic curve is one with good or multiplicative reduction.

The prime with bad reduction divide the discriminant $\Delta(E)$ (isomorphisms of elliptic curves change $\Delta(E)$ by 12th powers; we will always work with elliptic curves whose discriminant is minimal among its isomorphism class). The conductor $N(E)$

is a number having the same prime factors as $\Delta(E)$:

$$N(E) = \prod p^{f_p}, \quad f_p = \begin{cases} 0 & \text{if } E \text{ has good reduction at } p, \\ 1 & \text{if } E \text{ has multiplicative reduction at } p, \\ 2 & \text{if } E \text{ has additive reduction at } p \text{ and } p \nmid 6; \end{cases}$$

for the primes $p = 2, 3$, finding f_p is more difficult, but we always have $f_2 \le 8$ and $f_3 \le 5$.

Knowing that semistable curves are modular was sufficient for proving FLT since the Frey curves associated to a solution of the Fermat equation are semistable; more precisely we have

Proposition 1. *For primes $p \ge 5$, let a, b, c be a primitive solution of the Fermat equation $a^p + b^p + c^p = 0$. Assume that $2 \mid b$ and $a \equiv -1 \bmod 4$. Then $E : y^2 = x(x - a^p)(x + b^p)$ is semistable.*

These conditions can always be satisfied since exactly one of the three numbers is even, and since we can always replace a by $-a$ if necessary.

2. Galois Representations

Let $\overline{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$. Its Galois group $G_{\mathbb{Q}} = \mathrm{Gal}\,(\overline{\mathbb{Q}}/\mathbb{Q})$ is the inverse limit of the $\mathrm{Gal}\,(K/\mathbb{Q})$, where $K/\mathbb{Q}$ runs over the finite normal extensions of $\mathbb{Q}$, and therefore carries the profinite topology (called Krull topology when we talk about Galois groups; it was Krull who introduced it in order to save the main theorem of Galois theory for infinite extensions; inverse limits were later invented by Herbrand).

A representation of a group G is a homomorphism into the general linear group of some ring (classically, into the automorphism group of a vector space). The rings we will have to work with are called **coefficient rings**: these are complete Noetherian local rings with finite residue field of characteristic p. Note that local rings A have a unique maximal ideal $\mathfrak{m}$; the field $A/\mathfrak{m}$ is called the residue field of A. The simplest example of such a ring is the ring $A = \mathbb{Z}_p$: its unique maximal ideal is (p), the ring is complete with respect to the induced valuation (all Cauchy sequences in $\mathbb{Z}_p$ converge), it is Noetherian, and the residue field $\mathbb{Z}_p/p\mathbb{Z}_p \simeq \mathbb{F}_p$ is finite of characteristic p.

Now a Galois representation over A is a continuous group homomorphism

$$\rho : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_n(A).$$

Note that $G_{\mathbb{Q}}$ has the Krull topology, and $\mathrm{GL}_n(A)$ the subspace topology induced from A^{n^2}. As a matter of fact, we will be only interested in 2-dimensional Galois representations (those with $n = 2$).

Roots of Unity. Let $\zeta_m = \exp(2\pi i p^{-m})$ be a p^mth root of unity. Raising everything to the p-th power induces a homomorphism $\mu_{m+1} \longrightarrow \mu_m$, and these make

$$1 \longleftarrow \mu_1 \longleftarrow \mu_2 \longleftarrow \mu_3 \longleftarrow \dots$$

into a projective system whose limit we denote by $T_p(\mu) = \varprojlim \mu_m$. Since $\mu_m \simeq \mathbb{Z}/p^m$ as abelian groups, we have $T_p(\mu) \simeq \mathbb{Z}_p$ as abelian groups. In other words, $T_p(\mu)$ is a free $\mathbb{Z}_p$-module of rank 1.

Now the group $G_{\mathbb{Q}}$ acts on the groups μ_m, and this makes $T_p(\mu)$ into a $G_{\mathbb{Q}}$-module. It is the Galois action that makes $T_p(\mu)$ into an interesting object. In

order to understand the Galois action, let us make the isomorphism $T_p(\mu) \simeq \mathbb{Z}_p$ explicit. The element $\zeta = (\zeta_1, \zeta_2, \zeta_3, \ldots)$ is an element of the projective limit $T_p(\mu)$. It is an easy exercise to show that every element of $T_p(\mu)$ can be written uniquely as ζ^a for some $a = (a_1, a_2, a_3, \ldots) \in \mathbb{Z}_p$ (we say that $T_p(\mu)$ is procyclic); here $\zeta^a = (\zeta_1^{a_1}, \zeta_2^{a_2}, \ldots)$. The map $\zeta^a \longmapsto a$ then defines an isomorphism $T_p(\mu) \longrightarrow \mathbb{Z}_p$.

Now let $\sigma \in G_{\mathbb{Q}}$ be an automorphism. Since $T_p(\mu)$ is procyclic, we have $\zeta^\sigma = \zeta^{a(\sigma)}$ for some $a(\sigma) \in \mathbb{Z}_p$. The fact that σ is invertible shows that $a(\sigma) \in \mathbb{Z}_p^\times = \mathrm{GL}_1(\mathbb{Z}_p)$; the map

$$\chi_p : T_p(\mu) \longrightarrow \mathrm{GL}_1(\mathbb{Z}_p) : \sigma \longmapsto a(\sigma)$$

then is a 1-dimensional p-adic representation, or, in our terminology, a 1-dimensional Galois representation over $\mathbb{Z}_p$. For example, if σ denotes complex conjugation, then $a(\sigma) = (-1, -1, -1, \ldots) = -1$ because σ maps every root of unity to its inverse. Another example is given by the automorphisms restricting to the Frobenius automorphism σ_ℓ for primes $\ell \neq p$: it is an easy exercise to show that $\rho_p(\sigma_\ell) = \ell$.

Elliptic Curves. Let E be an elliptic curve defined over $\mathbb{Q}$. The groups $E_m = E[p^m] = \{P \in E(\overline{\mathbb{Q}}) : mP = \mathcal{O}\}$ of p^m-torsion points form a group isomorphic to $\mathbb{Z}/p^m \oplus \mathbb{Z}/p^m$, and multiplication by p induces homomorphisms $E_{m+1} \longrightarrow E_m$. The inverse limit $T_p(E) = \varprojlim E_m$ is called the Tate module of E (with respect to the fixed prime p). Clearly $T_p(E) \simeq \mathbb{Z}_p \oplus \mathbb{Z}_p$, so $T_p(E)$ is a free $\mathbb{Z}_p$-module of rank 2. As above, $T_p(E)$ is also a $G_{\mathbb{Q}}$-module. Fix two generators $P, Q \in T_p(E)$; then every $\sigma \in G_{\mathbb{Q}}$ acts on the generators via $\sigma(P) = aP + bQ$, $\sigma(Q) = cP + dQ$ for $a, b, c, d \in \mathbb{Z}_p$. This defines a map

$$\rho_p : \sigma \longmapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

and the fact that σ is invertible implies that the matrix $\rho_p(\sigma)$ is invertible too. Thus we have a homomorphism

$$\rho_p : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p),$$

and since the map can be shown to be continuous, this defines a 2-dimensional Galois representation over $\mathbb{Z}_p$.

You can think of these representations of lifts of the representations $\overline{\rho}_p : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{Z}/p)$ modulo p defined by the action of the absolute Galois group of $\mathbb{Q}$ on the p-torsion of E.

As an example, consider the elliptic curve $E : y^2 = x(x^2 - d)$ for some nonsquare $d \in \mathbb{Z}$ Here $E[2] = \{\mathcal{O}, (0,0), (\sqrt{d}, 0), (-\sqrt{d}, 0)\}$. Let us choose $P = (\sqrt{d}, 0)$ and $Q = (-\sqrt{d}, 0)$ as the basis of $E[2]$; then $(0,0) = P+Q$. Now let $\sigma \in G_{\mathbb{Q}}$; if σ induces the identity on $K = \mathbb{Q}(\sqrt{d})$, then $P^\sigma = P$ and $Q^\sigma = Q$, hence $\overline{\rho}_2(\sigma) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If the restriction of σ to K is the nontrivial automorphism of $K/\mathbb{Q}$, then $P^\sigma = Q$, $Q^\sigma = P$, and therefore $\overline{\rho}_2(\sigma) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Thus $\overline{\rho}_2(G_{\mathbb{Q}})$ has only two elements.

Now let us look at $E : y^2 = x^3 - 2$. Here $E[2] = \{\mathcal{O}, P, Q, P + Q\}$ for $P = (\zeta \sqrt[3]{2}, 0)$, $Q = (\zeta^2 \sqrt[3]{2}, 0)$, and $P + Q = (\sqrt[3]{2}, 0)$. Here $K = \mathbb{Q}(E[2]) = \mathbb{Q}(\sqrt[3]{2}, \zeta)$ has degree 6 and Galois group $\mathrm{Gal}\,(K/\mathbb{Q}) = S_3$, where $S_3 = \langle \sigma, \tau \rangle$ with

$$\sigma : \sqrt[3]{2} \longmapsto \zeta \sqrt[3]{2}, \zeta \longmapsto \zeta,$$
$$\tau : \sqrt[3]{2} \longmapsto \sqrt[3]{2}, \zeta \longmapsto \zeta^2,$$

and where $\sigma^3 = \tau^2 = \mathrm{id}$ and $\sigma\tau = \tau\sigma^2$.

How does $\mathrm{Gal}\,(K/\mathbb{Q})$ act on $E[2]$? We have $P^\sigma = Q$, $Q^\sigma = P + Q$, hence any lift of σ to $G_\mathbb{Q}$ (which we will still denote by σ) maps to $\overline{\rho}_2(\sigma) = \left(\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right)$. Similarly, $P^\tau = Q$, $Q^\tau = P$, hence $\overline{\rho}_2(\tau) = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$. It is now easy to compute the following table:

id	σ	σ^2	τ	$\sigma\tau$	$\sigma^2\tau$
$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$\left(\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right)$	$\left(\begin{smallmatrix}1&1\\1&0\end{smallmatrix}\right)$	$\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$	$\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)$	$\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$

Here $\overline{\rho}_2(\mathbb{G}_\mathbb{Q}) = \mathrm{GL}_2(\mathbb{Z}/2\mathbb{Z}) \simeq S_3$. By considering the action of $G_\mathbb{Q}$ on $E[4]$ one gets matrices in $\mathrm{GL}_2(\mathbb{Z}/4\mathbb{Z})$; continuing all the way up to $E[2^\infty]$ gives a Galois representation $\rho_2 : G_\mathbb{Q} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_2)$; composing ρ_2 with the natural projection $\mathbb{Z}_2 \longrightarrow \mathbb{Z}/2\mathbb{Z}$ then gives us back the mod 2 representation $\overline{\rho}_2$ computed above.

3. Local Fields

Galois representations $\rho_p : G_\mathbb{Q} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p)$ are used for studying the hardly understood group $G_\mathbb{Q}$ via the images of homomorphisms in well understood groups like $\mathrm{GL}_2(\mathbb{Z}_p)$. We can make the left hand side more manageable by cutting it into little ℓ-adic pieces.

To this end, note that $\mathbb{Q} \subset \mathbb{Q}_v$ for $v = \ell$ (canonically); if K is a finite normal extension of $\mathbb{Q}$ and w a prime ideal in K above v, then $K \hookrightarrow K_w$ (canonically), and we have $\mathrm{Gal}\,(K_w/\mathbb{Q}_v) \hookrightarrow \mathrm{Gal}\,(K/\mathbb{Q})$: in fact, $\mathrm{Gal}\,(K_w/\mathbb{Q}_v)$ can be identified with the decomposition group of w in $\mathrm{Gal}\,(K/\mathbb{Q})$. Choosing a different prime ideal w results in a conjugate subgroup of $\mathrm{Gal}\,(K/\mathbb{Q})$. By taking the inverse limit we find an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ and a corresponding monomorphism $G_{\mathbb{Q}_\ell} = \mathrm{Gal}\,(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \hookrightarrow G_\mathbb{Q}$, which is defined up to conjugacy.

Now $G_{\mathbb{Q}_\ell} = \lim \mathrm{Gal}\,(K_v/\mathbb{Q}_\ell)$ as K_v runs over the finite normal extensions of $\mathbb{Q}_\ell$. For each such extension $K_v/\mathbb{Q}_\ell$, let $\mathcal{O}_v$ denote the ring of integers in K_v and $k_v = \mathcal{O}_v/\mathfrak{m}_v$ the associated residue field (here $\mathfrak{m}_v$ is the maximal ideal in $\mathcal{O}_v$. This is a finite field of characteristic ℓ, and the Hilbert theory of ramification shows that we have an exact sequence

$$1 \longrightarrow I_\ell(K_v) \longrightarrow \mathrm{Gal}\,(K_\ell/\mathbb{Q}_\ell) \longrightarrow \mathrm{Gal}\,(k_v/\mathbb{F}_\ell) \longrightarrow 1,$$

where $I_\ell(K_v)$ is the inertia subgroup of $K_v/\mathbb{Q}_\ell$. Taking inverse limits shows that

$$1 \longrightarrow I_\ell \longrightarrow G_{\mathbb{Q}_\ell} \longrightarrow \mathrm{Gal}\,(\overline{\mathbb{F}}_\ell/\mathbb{F}_\ell) \longrightarrow 1,$$

and here $I_\ell = \varprojlim I_\ell(K_v)$ is called the inertia subgroup of $G_{\mathbb{Q}_\ell}$.

Now to every global Galois representation $\rho : G_\mathbb{Q} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p)$ we can attach a family of local Galois representation $\rho|_\ell$ defined as the composition of $G_{\mathbb{Q}_\ell} \longrightarrow G_\mathbb{Q} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p)$. Note that we also include the case $\ell = \infty$, where $\mathbb{Q}_\infty = \mathbb{R}$, $\overline{\mathbb{Q}}_\infty = \mathbb{C}$, and $G_{\mathbb{Q}_\infty} = \mathrm{Gal}\,(\mathbb{C}/\mathbb{R}) = \langle c\rangle$, where c denotes complex conjugation.

The local representations are simpler objects, and the hope in situations like these is to have some kind of local-global principle that would allow us to construct a global object out of the local data.

Now let ρ be a Galois representation. We say that ρ is **odd** if $\det\rho(c) = -1$, where c is complex conjugation. For example, the 1-dimensional representation on the roots of unity is odd. We say that ρ is **unramified at** ℓ if $I_\ell \subseteq \ker\rho|_\ell$. Finally, ρ is called **flat** at p if for every nonzero ideal $\mathfrak{a} \subseteq \mathbb{Z}_p$ the representation $G_{\mathbb{Q}_p} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p/\mathfrak{a})$ (this is the composition of $\rho|_p$ and reduction modulo $\mathfrak{a}$) extends to a finite flat group scheme over $\mathbb{Z}_p$ (whatever that means).

Theorem 2. *Let $\rho_p : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}(\mathbb{Z}_p)$ be the p-adic Galois representation of an elliptic curve E defined over $\mathbb{Q}$, and let N denote the conductor of E. Then*

(1) $\det \rho_p = \chi_p$;
(2) *ρ_p is unramified outside pN.*

In particular, ρ_p is odd because $\det \rho_p(c) = \chi_p(c) = -1$. If E is semistable with minimal discriminant Δ, then the representation $\overline{\rho}_p : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{F}_p)$ has the following properties:

(1) *$\overline{\rho}_p$ is unramified at primes $\ell \neq p$ if and only if $p \mid v_\ell(\Delta)$ (the exponent of ℓ in the prime factorization of Δ is divisible by p);*
(2) *$\overline{\rho}_p$ is flat at p if and only if $p \mid v_p(\Delta)$.*

4. Modular Forms

Let me recall that modular forms are meromorphic functions on the upper half-plane that behave nicely under the action of subgroups like $\Gamma_0(N)$ of $\mathrm{SL}_2(\mathbb{Z})$ and are holomorphic at the cusps. Cusp forms are modular forms that vanish at the cusps. Let $S(N)$ denote the space of cusp forms of weight 2 with respect to $\Gamma_0(N)$; this is a finite dimensional $\mathbb{C}$-vector space of dimension g, the genus of the curve $X_0(N)$.

Hecke operators are objects T_n that act on almost everything connected with $\Gamma_0(N)$. In particular, the T_n act on $S(N)$, that is, if $f \in S(N)$, then so is $T_n f$; actually it is customary to write $f|T_n$ instead of $T_n f$.

We say that $f \in S(N)$ is an eigenform for T_n if $f|T_n = \lambda f$ for some $\lambda \in \mathbb{C}$. Normalized eigenforms in $S(N)$ are cusp forms $f = \sum_{n\geq 1} a_n q^n$ in $S(N)$ with $a_1 = 1$ and the property that they are simultaneous eigenforms for *all* Hecke operators T_n.

Proposition 3. *If $f \in S(N)$ is an eigenform, then the eigenvalues are just the coefficients of f, i.e., $f|T_n = a_n f$. Moreover, the a_n are algebraic numbers, and $F = \mathbb{Q}(a_1, a_2, a_3, \ldots)$ is a finite extension of $\mathbb{Q}$.*

5. Taniyama-Shimura

Let E be an elliptic curve defined over $\mathbb{Q}$. To E we associate its L-series $L(E, s) = \sum_{n\geq 1} a_n n^{-s}$. The conjecture of Taniyama and Shimura originally said that there E is parametrized by some modular curve, i.e. there is a nonconcstant morphism $X_0(N) \longrightarrow E$ for some integer N. Later it was realized that if there is such an N at all, then it can be taken to be the conductor of E. An equivalent condition is that the Fourier series $f = \sum a_n q^n$ is a newform of weight 2 and level N.

6. Ribet's Theorem

Ribet succeeded in proving Serre's ε-conjecture, thereby reducing FLT to the Taniyama-Shimura conjecture. Note that to every newform f Eichler and Shimura have attached a p-adic Galois representation $G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{Z}_p)$; let $\overline{\rho}_f$ be the corresponding representation $G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{F}_p)$. Then Ribet proved

Theorem 4. *Let N be a positive integer, ℓ a prime dividing N, and $f = \sum a_n q^n$ a newform of weight 2 for $\Gamma_0(N)$. Assume that the action of $G_{\mathbb{Q}}$ on $E[p]$ is absolutely irreducible and that $\overline{\rho}_f$ is unramified at ℓ, or that $\ell = p$ and $\overline{\rho}_f$ is flat at p. Then there is a newform $g = \sum b_n q^n$ of weight 2 for $\Gamma_0(N/\ell)$ such that*

$$a_q \equiv b_q \bmod \mathfrak{p}$$

for almost all primes a and some prime ideal $\mathfrak{p}$ above p in the number field generated by the Fourier coefficients of f.

The following examples are from Schoof's article [2]. Consider $E : y^2 = x^3 - x^2 + 25158x - 775719$. We have $\Delta = -2^4 3^5 7^5 11^7$ and $N = 4 \cdot 3 \cdot 7 \cdot 11$, so E is semistable at 3, 7 and 11. The representations of $G_{\mathbb{Q}}$ on $E[3]$ and $E[7]$ are irreducible. Moreover, the representation ρ_5 is unramified at the primes 3 and 7 because their exponents are divisible by 5.

In fact, the reduction modulo 3 is $\overline{E} : y^2 = x^3 - x^2$, and the same equation describes the reduction modulo 7. Thus the singular point is a node whose tangents have slopes ± 1.

The newform f associated to E thus should be associated to a newform of of level $N/7 = 4 \cdot 3 \cdot 11$. In fact, there is an elliptic curve $E' : y^2 = x^3 - x^2 - 77x + 330$ with $\Delta' = -2^4 3^{10} 11$ and $N' = -4 \cdot 3 \cdot 11 = -132$. There is an associated newform $g = \sum b_n q^n$, and we have the following table:

N	q	2	3	5	7	11	13	17	19
924	a_q	0	−1	−3	−1	−1	1	−4	3
308		0	−1	−1	−1	+1	−4	−6	−2
132	b_q	0	−1	2	2	−1	6	−4	−2
44	c_q	0	1	−3	2	−1	−4	6	8

The curve of conductor 44 is given by $y^2 = x^3 + x^2 + 3x - 1$.

Note that we could also start by reducing the level by 7 first; there is an elliptic curve of conductor $N/7 = 308$, namely $E'' : y^2 = x^3 - x^2 - 21x + 49$ with discriminant $\Delta = -2^8 7^2 11$. The newform attached to E'', however, does not have the desired properties, hence there must be yet another newform of weight 2 and level 308.

Here's the connection with Fermat's Last Theorem: the conductor of the Frey curve E attached to $a^p + b^p + c^p = 0$ is the product of all primes dividing abc. Assume that E is modular and let $f = \sum a_n q^n$ denote the associated newform of weight 2 and level N. By a result of Mazur, the $G_{\mathbb{Q}}$-module $E[p]$ is irreducible for $p \ge 5$. Since the discriminant of E is $\Delta = 2^{-8}(abc)^{2p}$, the conditions of Ribet's theorem are satisfied, and this means we can remove all the odd primes from the level. Thus a solution to the Fermat equation implies the existence of a modular form of weight 2 and level 2; on the other hand, such forms do no exist because $X_0(2)$ has genus 0.

7. Langlands Conjecture

An essential ingredient of Wiles' proof is a little piece of the Langlands conjecture (a vast generalization of class field theory) proved by Langlands and Tunnell.

Theorem 5. *Let $\sigma : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{C})$ be an odd irreducible continuous representation whose image in $\mathrm{PGL}_2(\mathbb{C})$ is solvable. Then there is a normalized eigencuspform $g(z) = \sum_{n \ge 1} b_n q^n$ of weight 2 such that $b_n = \mathrm{Tr}\,(\sigma(\mathrm{Fr}_q))$ for almost all primes q.*

8. Wiles

Wiles proves the following result:

Theorem 6. *Let E be a semistable elliptic curve defined over $\mathbb{Q}$, and consider the associated L-series $L(E,s) = \sum a_n n^{-s}$. Let ℓ be an odd prime such that*

(1) *$E[\ell]$ is an irreducible $G_{\mathbb{Q}}$-module;*
(2) *There is an eigenform $f = \sum b_n q^n$ and a prime ideal $\mathfrak{l}$ above ℓ such that, for almost all primes q,*

$$a_q \equiv b_q \bmod \mathfrak{l}.$$

Then E is modular, i.e., the Fourier series $\sum a_n q^n$ is a modular form of weight 2 for some $\Gamma_0(N)$.

What this result says is that if the Fourier series $f = \sum a_n q^n$ attached to E is congruent to a modular form modulo $\mathfrak{l}$, then it already is modular.

Now start with an elliptic curve E and the Galois representation $\overline{\rho}_3 : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{F}_3)$ on $E[3]$. Now there is an injective group homomorphism $\mathrm{GL}_2(\mathbb{F}_3) \longrightarrow \mathrm{GL}(\mathbb{Z}[\sqrt{-2}])$, and composing ρ_3 with this homomorphism we get a representation $\sigma : G_{\mathbb{Q}} \longrightarrow \mathrm{GL}_2(\mathbb{C})$. The image of σ is contained in a subgroup of $\mathrm{GL}_2(\mathbb{F}_3)$, and its image in $\mathrm{PGL}_2(\mathbb{C})$ is contained in a subgroup isomorphic to S_4, which is solvable. If $\overline{\rho}_3$ is irreducible, then so is σ, and by Langlands-Tunnell, σ is modular. Modifying the corresponding modular form, if necessary, gives us a modular form congruent modulo 3 to $\overline{\rho}_3$. By Wiles' theorem, this means that E must be modular.

What if $E[3]$ is reducible as a $G_{\mathbb{Q}}$-module? In this case, one can prove with methods we discussed (modular curves for $\Gamma_0(N)$ with $N \mid 15$) that for semistable elliptic curves, $G_{\mathbb{Q}}$ on $E[3]$ and $E[5]$ cannot both be reducible $G_{\mathbb{Q}}$-modules. This does not seem to help, because there is no result a la Langlands-Tunnell for $\ell = 5$. What Wiles came up with is his ingenious 3-5 switch: he proved that there is an elliptic curve E' with $E[5] \simeq E'[5]$ and $E'[3]$ irreducible as a $G_{\mathbb{Q}}$-module. Then his proof shows that E' is modular (via $E'[3]$ and Langlands-Tunnell); thus the modular form $\sum b_n q^n$ attached to E' is modular, and $E[5] \simeq E'[5]$ implies that $a_q \equiv b_q \bmod 5$ for almost all primes q. But then the assumptions of Wiles' theorem are satisfied for E with $\ell = 5$, and this shows that E is modular.

References

[1] G. Cornell, J.H. Silverman, G. Stevens (eds.), *Modular Forms and Fermat's Last Theorem*, Springer-Verlag
[2] R. Schoof, *Fermat's Last Theorem*, 1994; available online