Citations
From References: 0
From Reviews: 0

MR3243601 (Review) 11D41 11F11 11F80 11G05 11G18
Saito, Takeshi [Saito, Takeshi^1]

★**Fermat's last theorem.**
The proof.
Translated from the 2009 Japanese original by Masato Kuwata.
Translations of Mathematical Monographs, 245.
Iwanami Series in Modern Mathematics.
American Mathematical Society, Providence, RI, 2014. *xvi*+222 *pp.*
ISBN 978-0-8218-9849-9

The proof of Fermat's last theorem represents one of the highest intellectual accomplishments of the 20th century. The author has attempted to write down an almost complete, self-contained account of this success. It requires an indomitable spirit to launch this daunting task. The author's Japanese book has been translated into a two-volume book in English. The first volume, consisting of seven chapters, studies the three main players in this saga—elliptic curves, modular forms and Galois representations [T. Saito, *Fermat's last theorem*, translated from the Japanese original by Masato Kuwata, Transl. Math. Monogr., 243, Amer. Math. Soc., Providence, RI, 2013; MR3136492]. It gives an outline of the proof of Fermat's last theorem (hereafter to be abbreviated as FLT). This second volume is a technical tour-de-force of four chapters and gives complete proofs of almost all the steps. Even though our task is to review this second volume, it is first necessary quickly to recall the background.

The layout of the FLT proof is well known and we recall it only briefly so as to understand the description of the details of the proof given in this review. In the early 1970's, Hellegouarch related a fictitious solution of FLT to an elliptic curve; his idea was to apply known results on FLT to obtain some consequences for elliptic curves. In 1986, Frey turned it around by looking for consequences in the opposite direction. If l is an odd prime, and if a, b, c form a possible nontrivial triple of integers satisfying $a^l + b^l = c^l$ where (without loss of generality) $a^l \equiv -1 \pmod 4$ and $32 \mid b^l$, Frey looked at the elliptic curve

$$E : y^2 = x(x - a^l)(x + b^l).$$

This curve is semistable, i.e., has square-free conductor. If L is the smallest field over which the points of order l are defined, the action of the absolute Galois group $G_{\mathbb{Q}}$ on $E[l]$ defines a faithful representation

$$\overline{\rho}_{E,l} \colon \mathrm{Gal}(L/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_l)$$

for a choice of basis for $E[l]$. Mazur used the fact that the elliptic curve is semistable to show that $\overline{\rho}_{E,l}$ is irreducible if $l \geq 11$. Combining with Serre's work from 1972, it follows that the Galois group is actually isomorphic to $\mathrm{GL}_2(\mathbb{F}_l)$; therefore, the field L is large in this sense. Moreover, Serre studied two-dimensional Galois representations over finite fields in general in 1987 and conjectured that for Frey's curve, the corresponding Galois representation comes from modular forms modulo l. More precisely, he made the precise conjecture that the corresponding modular form must be of weight 2 and level 2 (which does not exist!). Meanwhile, the Shimura-Taniyama conjecture (originating from a prediction by Taniyama in 1955) produced a direct connection between the representations $\overline{\rho}_{E,l}$ and modular forms of weight 2, but with a rather large level (product of primes dividing abc). In 1990, Ribet came up with the stunning result that if the above representation were indeed to come from a modular form mod l, it would come also from a modular form mod l of weight 2 *and level 2*. In other words, Ribet's theorem implies that the Shimura-Taniyama conjecture (assumed just for semistable elliptic curves) implies FLT. Wiles proved the Shimura-Taniyama conjecture for semistable curves using collaborative work with Taylor to bring the proof into completion. The Shimura-Taniyama conjecture asserts that elliptic curves over $\mathbb{Q}$ are modular; that is, there is a Hecke eigenform f of weight 2 on $\Gamma_0(N)$ for some N so that

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

for almost all primes p, where $f = \sum a_n q^n$. Here, Galois representations enter in the following manner: For an elliptic curve E, the action of $G_{\mathbb{Q}}$ on the subgroup $E[l^n]$ of l^n-division points gives rise to an l-adic representation

$$\rho_{E,l}\colon G_{\mathbb{Q}} \to \mathrm{GL}(\varprojlim E[l^n]) \cong \mathrm{GL}_2(\mathbb{Z}_l).$$

Such an l-adic representation is said to be modular if there is a Hecke eigenform f such that for almost all primes p where $\rho_{E,l}$ is unramified, the trace of the Frobenius at p is the Fourier coefficient a_p of f. Then, E is modular if and only if there is some l for which $\rho_{E,l}$ is modular. Hence, from then on, in the proof of FLT, elliptic curves fade away and only their avatars in the form of Galois representations appear. We mention in passing that semistability of a representation (either over a finite field or over a complete, Noetherian local ring) means it is either good or ordinary. Next, since Serre's conjectures predict the modularity of mod l Galois representations (which was known for primes like 3 and 5), the problem is to lift the modularity property to l-adic representations. Wiles proved the following:

Let l be an odd prime, let K be an l-adic field and $\mathbb{F}$, the residue field of O_K. Suppose $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ is a semistable, irreducible, modular representation. Then, a semistable lifting $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O_K)$, for which $\det \rho$ is the cyclotomic character, is modular of level N_ρ (equal to the product of all primes where ρ is not good).

To do this, Wiles used Mazur's deformation theory. More generally, one considers lifts to representations over complete, Noetherian $\mathbb{Z}_l$-algebras R with residue field $\mathbb{F}$ where Mazur's theory gives a certain universal ring R_Σ. In fact, if Σ is a finite set of primes and one looks for lifts of $\overline{\rho}$ whose restriction to the decomposition groups at l have good properties and whose ramification outside Σ is restricted, then there is a ring R_Σ and a lift $\rho_\Sigma\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R_\Sigma)$ which is universal with respect to all lifts to all such R. Wiles constructed explicitly this universal deformation ring as a certain Hecke algebra. In other words, modular forms are replaced by their avatars: these Hecke algebras. Thus, the rest of the proof is directed towards proving that the natural map from Mazur's deformation ring to the Hecke ring constructed by Wiles is an isomorphism. This requires a lot of commutative algebra at the point where it is needed to prove that the Hecke algebra T corresponding to the empty set Σ is a complete intersection; this was done by Wiles and Taylor.

The main theorems (leading to a proof of the FLT) proved in this book almost completely are the following ones:

(i) the proof of the identification of the deformation ring with a corresponding Hecke algebra (this is proved in Chapter 11 where Selmer groups are studied),

(ii) the modularity theorem for l-adic representations stated above as a consequence of (i), and

(iii) the 'lifting of level' for mod l Galois representations (proved in Chapter 11).

We now attempt to describe the proofs as given in Chapters 8 to 11.

Chapter 8 defines and discusses modular curves over $\mathbb{Z}$. Even though modular curves over $\mathbb{Q}$ were already studied in Chapter 2 of the first book, their arithmetic properties depend also on the behaviour of modular curves over $\mathbb{Z}$ at primes dividing the level (where they may not have good reduction). According to the author, this aspect of having been able to analyze modular curves over $\mathbb{Z}$ is the crucial fact which enabled a proof of the FLT to emerge in the 20th century. Modular curves over $\mathbb{Q}$ are usually defined by means of cyclic subgroups of order N of an elliptic curve. However, a supersingular elliptic curve over a scheme over $\mathbb{F}_p$ has no subgroup of order p in the usual sense. Therefore, a cyclic subgroup scheme is defined via a Drinfeld level structure; viz., if S is a scheme and $N \geq 1$, and G is a finite, flat commutative group scheme of finite presentation type over S of degree N, then G is said to be cyclic of order N if there is a section $P\colon S \to G$ flat locally on S such that $0, P, 2P, \dots, (N-1)P$ is a full set of sections.

Then, for r relatively prime to N, one defines the functor $\mathbb{M}_0(N,r)_{\mathbb{Z}[1/r]}$ over $\mathbb{Z}[1/r]$, which sends any scheme T over the latter to the set of isomorphism classes of triples (E, C, α) where E is an elliptic curve over T, C is a cyclic subgroup scheme of order N, and α is an isomorphism from $\mathbb{Z}/r\mathbb{Z} \times \mathbb{Z}/r\mathbb{Z}$ to $E[r]$.

One similarly defines a functor $\mathbb{M}_1(N,r)_{\mathbb{Z}[1/r]}$ as usual by considering sections of order N instead of cyclic subgroups of order N. The two functors coincide when $N = 1$ and the author first shows that for $r \geq 3$, this functor is representable by a smooth affine connected curve $Y(r)_{\mathbb{Z}[1/r]}$ over $\mathbb{Z}[1/r]$.

For instance, for $r = 3$, it is $\operatorname{Spec} \mathbb{Z}[\frac{1}{3}, \zeta_3, \mu, \frac{1}{\mu^3 - 1}]$.

Further, the author proves that when $(r, N) = 1$ and $r \geq 3$, $\mathbb{M}_1(N,r)_{\mathbb{Z}[1/r]}$ is representable by a finite scheme $\mathbb{Y}_1(N,r)_{\mathbb{Z}[1/r]}$ over $\mathbb{Y}(r)_{\mathbb{Z}[1/r]}$. Looking at the action of $\mathrm{GL}_2(\mathbb{Z}/r\mathbb{Z})$ on $\mathbb{Y}_1(N,r)_{\mathbb{Z}[1/r]}$, the author deduces that the quotient $\mathbb{Y}_1(N)_{\mathbb{Z}[1/r]}$ is a coarse moduli scheme of the restriction of the functor $\mathbb{M}_1(N)$ over $\mathbb{Z}$ to $\mathbb{Z}[1/r]$. Using this, the author can conclude the following theorem, which is one of the two main results of this chapter:

(i) There exists a coarse moduli scheme $\mathbb{Y}_1(N)_{\mathbb{Z}}$ of $\mathbb{M}_1(N)$ over $\mathbb{Z}$. If $N \geq 4$, then $\mathbb{Y}_1(N)_{\mathbb{Z}[1/N]}$ is a fine moduli scheme. Moreover, $\mathbb{Y}_1(N)_{\mathbb{Z}}$ is a normal, connected, affine curve over $\mathbb{Z}$ and is smooth over $\mathbb{Z}[1/N]$. Finally, for a prime p not dividing N, $\mathbb{Y}_1(N)_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{F}_p$ is a coarse moduli scheme of the restriction $\mathbb{M}_1(N)_{\mathbb{F}_p}$.

(ii) Analogous assertions hold for $\mathbb{M}_0(N)$.

The author remarks that even though these theorems imply that the modular curves $\mathbb{Y}_0(N)_{\mathbb{Z}}$ and $\mathbb{Y}_1(N)_{\mathbb{Z}}$ are the integral closures of $\mathbb{Y}_0(N)_{\mathbb{Z}[1/N]}$ and $\mathbb{Y}_0(N)_{\mathbb{Z}[1/N]}$, respectively, it is not sufficient to simply define them in this manner because the study of their detailed structure requires the definition through Drinfeld level structures as given in this chapter.

The second of the two main theorems of this chapter concerns the compactified curves. The compactifications $\mathbb{X}_0(N)_{\mathbb{Z}}$ and $\mathbb{X}_1(N)_{\mathbb{Z}}$ of $\mathbb{Y}_0(N)_{\mathbb{Z}}$ and $\mathbb{Y}_1(N)_{\mathbb{Z}}$ are their integral closures with respect to the j-map to $\mathbb{A}^1_{\mathbb{Z}}$. The main theorem here (from which the fourth assertion is crucially used later) is:

(i) The projective curve $\mathbb{X}_0(N)_{\mathbb{Z}}$ is normal and each of its geometric fibers is connected.

(ii) For any prime p not dividing N, this curve is smooth at p and the fiber $\mathbb{X}_0(N)_{\mathbb{F}_p} = \mathbb{X}_0(N)_{\mathbb{Z}} \otimes \mathbb{F}_p$ is a smooth compactification of $\mathbb{Y}_0(N)_{\mathbb{F}_p}$.

(iii) Analogous assertions of (i), (ii) hold for $\mathbb{X}_1$.

(iv) Let $N = Mp$ with $(p, M) = 1$. Then, $\mathbb{X}_0(N)_{\mathbb{Z}}$ is weakly semistable at p. The closed immersions

$$j_i\colon \mathbb{Y}_i(M)_{\mathbb{F}_p} \to \mathbb{Y}_i(N)_{\mathbb{F}_p}$$

(for $i = 0, 1$) extend to $\mathbb{X}_i(M)_{\mathbb{F}_p}$. The fiber $\mathbb{X}_0(N)_{\mathbb{F}_p}$ is the union of the images $\overline{C}_0$ and $\overline{C}_1$ of j_0, j_1. The intersection of these images is the coarse moduli scheme of $\mathbb{M}_0(M)_{\mathbb{F}_p}$.

Using the above-mentioned theorem from Chapter 8 (especially (iv) above), Chapter 9 proves the following level and ramification mod l result which is important in the deduction of FLT.

Let $l \geq 3$ be a prime, and $\mathbb{F}$ a finite extension of $\mathbb{F}_l$. Let $\rho: G_{\mathbb{Q}} \to GL_2(\mathbb{F})$ be a continuous, irreducible representation. Suppose ρ is modular of level N. If M denotes the prime-to-p part of N, then:

(i) ρ is modular of level $M \iff \rho$ is good at p;

(ii) ρ is modular of level $pM \iff \rho$ is semistable at p.

The main point is to prove that modularity is a consequence of the latter assertions in (i) and (ii) above. The author proves this for (i) when $p \not\equiv 1 \pmod{l}$; as he says, the assertion for $p \equiv 1 \pmod{l}$ requires a lot more preparation involving p-adic uniformization of Shimura curves and Jacquet-Langlands-Shimizu correspondence and he does not give it in the book.

Roughly, the proof of 'good at p implies modular' goes as follows:

Consider the Hecke algebras $\mathbb{T} := T_0(Mp)_{\mathbb{Z}}$ and $\mathbb{T}' := T_0(M)_{\mathbb{Z}}[U_p]/(U_p^2 - T_pU_p + p)$. Here, T_p, U_p are the Hecke operators defined in the usual way. There is a natural surjection from $\mathbb{T}$ to $\mathbb{T}'$ and a sufficient criterion for modularity of a mod l representation can be given in terms of this surjection. Indeed, if ρ is a continuous, absolutely irreducible representation of level Mp and $\phi: \mathbb{T} \to \mathbb{F}$ is a ring homomorphism satisfying

$$\det(1 - \rho(\phi_q)t) = 1 - \phi(T_q)t + qt^2$$

for almost all primes p, and if there is a finite extension $\mathbb{F}'$ of $\mathbb{F}$ and a ring homomorphism $\phi': \mathbb{T} \to \mathbb{F}'$ satisfying an obvious commutativity condition, then ρ is modular of level M.

Let us indicate where the theorem of Chapter 8 quoted above plays a role while using the above criterion.

Let $\mathbf{m}$ be the maximal ideal of $\mathbb{T}$ corresponding to the kernel of a certain homomorphism to $\mathbb{F}$ which arises from the modularity of ρ of level Mp. In fact, if K is a finite extension of $\mathbb{Q}_l$ and f is a primary form over K, there is a ring homomorphism from $T_0(N)_{\mathbb{Q}}$ to its residue field (which can be taken to be our $\mathbb{F}$) which sends each T_p to an element congruent to the eigenvalue $a_p(f)$. Denote by $J_0(Mp)[\mathbf{m}]$, the elements of the Jacobian $J_0(Mp)(\overline{\mathbb{Q}})$ which are killed by $\mathbf{m}$. It can be observed that as a mod l representation, $J_0(Mp)[\mathbf{m}] \otimes_{\mathbb{T}} \mathbb{F}$ is isomorphic to the direct sum of copies of ρ. By the hypothesis, ρ is good at p and so, we have a finite étale group scheme $J_0(Mp)[\mathbf{m}]_{\mathbb{Z}_p}$ over $\mathbb{Z}_p$. By properties of Néron models, the inclusion of $J_0(Mp)[\mathbf{m}]$ in $J_0(Mp)$ extends to a closed immersion from $J_0(Mp)[\mathbf{m}]_{\mathbb{Z}_p}$ to $J_0(Mp)_{\mathbb{Z}_p}$ which can be reduced mod p to obtain a closed immersion from $J_0(Mp)[\mathbf{m}]_{\mathbb{F}_p}$ to $J_0(Mp)_{\mathbb{F}_p}$ over $\mathbb{F}_p$. Of course, here $J_0(Mp)_{\mathbb{F}_p}$ denotes the reduction mod p of a Néron model of $J_0(Mp)$. Then, the $\mathbb{T}$-module scheme $J_0(Mp)_{\mathbb{F}_p}$ over $\mathbb{F}_p$ is a successive extension of three $\mathbb{T}$-module schemes—Φ, $J_0(Mp)_{\mathbb{F}_p}^{\mathrm{ab}}$ and $J_0(Mp)_{\mathbb{F}_p}^{\mathrm{torus}}$—where Φ is the group of connected components, and 'ab' and 'torus' denote, respectively, the abelian and torus parts of $J_0(Mp)_{\mathbb{F}_p}$.

The theorem is then proved by means of the following three steps:

(i) showing that the composite morphism

$$J_0(Mp)[\mathbf{m}]_{\mathbb{F}_p} \to J_0(Mp)_{\mathbb{F}_p} \to \Phi$$

is zero;

(ii) if the morphism $J_0(Mp)[\mathbf{m}]_{\mathbb{F}_p} \to J_0(Mp)_{\mathbb{F}_p}^{\mathrm{ab}}$ induced because of (i) above is not the zero morphism, then showing that ρ is modular of level M; and

(iii) showing that if $J_0(Mp)[\mathbf{m}]_{\mathbb{F}_p} \to J_0(Mp)_{\mathbb{F}_p}^{\mathrm{ab}}$ is the zero morphism, then $p \equiv 1 \pmod{l}$.

As mentioned at the beginning, getting to the modularity of an l-adic representation from the modularity of the reduction mod l uses deformation rings. Chapter 10 studies Hecke modules—these are completions of the homology groups of modular curves. The main theorem of this chapter identifies the deformation ring with a suitable

Hecke algebra. The required properties of the deformation ring which would ultimately complete the proof of the FLT are proved in Chapter 11. In Chapter 11, the author starts with a self-contained introduction to Galois cohomology after which he defines and studies Selmer groups. The main purpose is to relate Selmer groups to deformation rings for this moves the question from deformation rings to Selmer groups. Before describing this relation, we recall that Selmer groups are defined in Galois cohomology via local conditions. More precisely, if S is a finite set of primes, and M is a $G_{\mathbb{Q}}$-module which is unramified outside S, then a family of subgroups $L_p \leq H^1(\mathbb{Q}_p, M)$ for $p \in S$ defines a subgroup of $H^1(G_S, M)$ given as the inverse image of $\bigoplus_{p \in S} L_p$ under the restriction map

$$H^1(G_S, M) \to \bigoplus_{p \in S} H^1(\mathbb{Q}_p, M).$$

Here, G_S is the quotient of $G_{\mathbb{Q}}$ by the normal subgroup generated by all the inertia subgroups at primes in S.

Now, let us describe the local conditions in our situation. Let l be an odd prime, let $\mathbb{F}$ be a finite field of characteristic l, and K be an l-adic field with residue field $\mathbb{F}$. Let $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be a modular, irreducible, semistable mod l representation, f be a corresponding primitive form, and let the associated l-adic representation be $\rho_f\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O_K)$, which is a lift of $\overline{\rho}$ unramified outside $S_{\overline{\rho}} \cup \{l\}$. Let Σ be a finite set of primes not intersecting $S_{\overline{\rho}}$ such that if $l \in \Sigma$, then $\overline{\rho}$ is good and ordinary at l. Taking $S = S_{\overline{\rho}} \cup \Sigma \cup \{l\}$ and considering O_K^2 as a G_S-module, the local conditions defining our Selmer group are as follows: If

$$W = \{f \in \mathrm{End}(O_K^2) : \mathrm{Tr}(f) = 0\}$$

look at the $O_K/\pi O_K$-module $\overline{W} = W/\pi W$. The local subgroups corresponding to the primes in S are defined according to whether ρ is good or ordinary at that prime (which we do not recall here precisely). Denote the corresponding Selmer group by $\mathrm{Sel}_S(\overline{W})$. The relation between the deformation ring and the Selmer group is:

Let R_Σ be the deformation ring and $\pi_\Sigma\colon R_\Sigma \to O_K$ be the homomorphism defined by ρ. Let $\mathbf{m}_\Sigma$ be the maximal ideal of R_σ. Then, there is a natural $\mathbb{F}$-linear isomorphism

$$\mathrm{Hom}_{\mathbb{F}}(\mathbf{m}_\Sigma/(\mathbf{m}_\Sigma^2, \pi), \mathbb{F}) \to \mathrm{Sel}_S(\overline{W}).$$

Further, if $\mathbf{p}_\Sigma$ is the kernel of the ring homomorphism from R_Σ to O_K induced by ρ, there is a natural O_K-module isomorphism

$$\mathrm{Hom}_{\mathbb{O}_{\mathbb{K}}}(\mathbf{p}_\Sigma/\mathbf{p}_\Sigma^2, K/O_K) \to \varinjlim \mathrm{Sel}_S(W/\pi^n W).$$

Following this, the author verifies properties of Selmer groups which—using the above relationship with deformation rings—completes the proof of FLT itself.

The present state of the art is that the full Shimura-Taniyama conjecture as well as Serre's modularity conjecture have been proved.

Here are a few minor comments on the book.

The mathematical writing is lucid and is interspersed with copious comments which are very illuminating. To give an instance, it is explained that in order to obtain arithmetic properties, it is necessary to define the modular curves $Y_0(N)_{\mathbb{Z}}, Y_1(N)_{\mathbb{Z}}$ using Drinfeld level structures and not merely as integral closures of $Y_0(N)_{\mathbb{Z}[1/N]}$ and $Y_1(N)_{\mathbb{Z}[1/N]}$.

In several places, a sentence starts with a mathematical symbol; this could have been avoided as it leads to confusion especially when the previous sentence also finishes with a symbol.

After many lemmata (especially in Chapter 8), examples are provided adding to a quick appreciation of the lemma.

Interesting exercises are given periodically; it is somewhat amusing to find that they appear under the caption ‘question’.

This magnificent book will be used for many years to come. *B. Sury*