MF Overview Notes

Niven Achenjang

Fall 2023

These are notes on an overview of the proof of Fermat's Last Theorem, written for the Modularity/Fermat Seminar.. They reflect my understanding (or lack thereof) of the material, so are far from perfect. They are likely to contain some typos and/or mistakes, but ideally none serious enough to distract from the mathematics. With that said, enjoy and happy mathing.

These are much more detailed than my talk is gonna be. In general, speaker's notes *don't* have to be this long. Also, if you decide to read these, keep in mind that much of what's written can be safely skipped (for example, most of the remarks are unnecessary).

Introduction

Theorem A (FLT)**.** *Fix an integer $n > 2$. Then, every triple of integers (a, b, c) satisfying*

$$a^n + b^n = c^n$$

also satisfy $abc = 0$.

I'll skip the history of this equation. If you're interested, there's a short description at the beginning of [CSS97] and a longer one at the beginning of [DDT07]. As you likely already know, the ultimate proof of Theorem A begins by reducing it to modularity of elliptic curves (really, only modularity of Frey curves). This reduction is what I'd first like to describe.

Remark 0.1. FLT says that $a^n + b^n = c^n$ has no *non-trivial* solutions. It does have plenty of solutions (e.g. $(a, b, c) = (0, 1, 1)$), but we want to say that these obvious/stupid/trivial ones are all of them. Probably FLT would be easier to prove if there were literally no solutions whatsoever. ○

In the below discussion, I'm assuming the reader knows what it means for an elliptic curve (or a Galois representation) to be 'modular'. I'm also imagining the reader is already vaguely aware that FLT is proven by attaching some elliptic curve to solutions to $a^n + b^n = c^n$, showing that said curve can't be modular, and then showing that all (semistable) rational elliptic curves are in fact modular.

1 Frey Curves and reduction to Modularity

Solving $a^n + b^n = c^n$ is hard, so let's start by looking at an easier equation: $a + b = c$.

Aside on $a + b = c$. If you're reading these notes, feel free to skip this part. It's just a long-winded motivation for the definition of Frey curves, and it's not gonna be in my talk. However, if you choose to not skip this part, then indulge me as I spend a nontrivial amount of time looking at the geometry of $a + b = c$.

Imagine you're interested in *non-trivial* solutions to $a + b = c$. First note that $a + b = c$ is homogeneous, and then consider the hyperplane $X' := V(a + b - c) \subset \mathbb{P}^2$. We don't care about all points on X', only those which are non-trivial – i.e. where none of a, b, c are zero – so the really curve to consider is $X := X' \setminus (V(a) \cup V(b) \cup V(c))$. What scheme is this? Well, $X' \xrightarrow{\sim} \mathbb{P}^1$ via $[a : b : c] \mapsto [a : -b]$ (the minus sign is to get 1 instead of -1 appearing in the next sentence) and this isomorphism sends X to $\mathbb{P}^1 \setminus \{[0 : 1], [1 : 0], [1 : 1]\}$. If you set $\infty := [1 : 0]$ so $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$, then we conclude that

$$X \simeq \mathbb{P}^1 \setminus \{0, \infty, 1\} \simeq \mathbb{A}^1 \setminus \{0, 1\}.$$

What was the point of all that? The point is that $\mathbb{A}^1 \setminus \{0, 1\}$ is famous for being the base of the Legendre family of elliptic curves[1]

$$E_\lambda : y^2 = x(x - 1)(x - \lambda).$$

This curve is smooth (and so elliptic) iff $\lambda \neq 0, 1, \infty$ (and 2 is invertible), i.e. iff $\lambda \in \mathbb{A}^1 \setminus \{0, 1\}$. So it defines a family of elliptic curves over $\mathbb{A}^1 \setminus \{0, 1\}$.

Remark 1.1 (extremely unimportant). Let $\mathcal{Y}(2)$ denote the (fine) moduli space of elliptic curves equipped with a (naive) full level 2 structure.[2] The Legendre family has a level 2 structure obtained by declaring $(0, 0), (1, 0)$ to be a basis for the 2-torsion, and so it defines a map $\mathbb{A}^1 \setminus \{0, 1\} \to \mathcal{Y}(2)$. This map turns out to be an étale cover, and in fact a $\underline{\mathbb{Z}/2\mathbb{Z}}$-torsor (with trivial $\mathbb{Z}/2\mathbb{Z}$ action),[3] so $\mathcal{Y}(2)$ is (isomorphic to) the classifying stack $B\underline{\mathbb{Z}/2\mathbb{Z}}_{\mathbb{A}^1 \setminus \{0,1\}}$ of $\underline{\mathbb{Z}/2\mathbb{Z}}$-torsors over $\mathbb{A}^1 \setminus \{0, 1\}$. This is a long-winded way of saying that it's not a coincidence that elliptic curves are connected to (non-trivial) solutions to $a + b = c$. $\circ$

Anyways, the isomorphism $X \simeq \mathbb{A}^1 \setminus \{0, 1\}$ lets us realize the Legendre family as a family of elliptic curves of X. In particular, it pulls back to the following family of elliptic curves over X:

$$E'_{a,b,c} : y^2 = x(x - 1)\left(x + \frac{a}{b}\right).$$

[1] One should really think of this family as an elliptic scheme $\mathcal{E} \longrightarrow \mathbb{A}^1_{\mathbb{Z}[1/2]} \setminus \{0, 1\} = \operatorname{Spec} \mathbb{Z}[1/2][\lambda, \lambda^{-1}, (1 - \lambda)^{-1}]$ whose fiber over some $\lambda \in \mathbb{A}^1 \setminus \{0, 1\}$ is E_λ.

[2] So this is a stack over $\mathbb{Z}[1/2]$

[3] This being a cover means that given any family of elliptic curves w/ full level 2 structure, you can (étale locally) put it in Legendre form. It's also a $\underline{\mathbb{Z}/2\mathbb{Z}}$-torsor w/ trivial $\mathbb{Z}/2\mathbb{Z}$-action. The action being trivial is essentially the statement that if a family can be put in Legendre form, then it can do so in a unique way (you can't change the Legendre parameter λ w/o changing the level 2 structure). Given this, being a $\underline{\mathbb{Z}/2\mathbb{Z}}$-torsor comes from the fact that the only automorphisms (which preserve the level 2 structure) of such a family are ± 1.

If you've seen Frey curves before, this is probably not the equation you expected me to end up at. We will remedy this in two steps. First, we can form the (quadratic) twist of $E'_{a,b,c}$ by Galois cover $X[\sqrt{b}] \to X$ (this is étale since both 2 and b are invertible on X) in order to obtain the family

$$E''_{a,b,c} : y^2 = x(x-b)(x+a)$$

(make the substitution $(x,y) \rightsquigarrow (x/b, y/b^{3/2})$). Next, we relabel our family by setting $E_{a,b,c} := E''_{b,a,c}$ in order to arrive at the usual form of the Frey curves

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

Back to Fermat The upshot of the above discussion is that, over the scheme $X = \{a+b-c : abc \neq 0\}$, there is a natural family

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

of elliptic curves. So, given any non-trivial solution to $a^n + b^n = c^n$ (i.e. $abc \neq 0$ and $\gcd(a,b) = 1$), we can form the corresponding Frey curve E_{a^n,b^n,c^n} and hope that its geometry tells us something about our imagined solution (ideally, it tells us that it can't exist).

Let's take a minute to investigate some basic properties of the the curves $E = E_{a,b,c} : y^2 = x(x-a)(x+b)$ before specializing to those coming from a Fermat triple.

- By definition, its discriminant is

$$\Delta = \Delta_{a,b,c} = 16(abc)^2.$$

  If ℓ is an odd prime, one can use Tate's algorithm [Sil94, Section IV.9] (only needing to go as far as Step 2) to show that $y^2 = x(x-a)(x+b)$ is minimal at ℓ (where it has good or multiplicative reduction). Let $N = N(E)$ denote its conductor. We always have

$$\Delta^{\min} = 2^{4-12s}(abc)^2 \text{ and } N = 2^t \prod_{\text{odd } \ell \mid abc} \ell$$

  for some integers s, t. In particular, the conductor of E is always a multiple of the radical $\operatorname{rad}(abc)$ of abc, i.e. the product of prime dividing abc (and $N/\operatorname{rad}(abc)$ is always a power of 2).

- What happens at $\ell = 2$? Step 4 of Tate's algorithm shows that E has additive reduction at 2 if $4 \nmid b$. Step 6 shows that E has additive reduction at 2 if $16 \nmid b$. [DDT07, Page 58,"The Frey curve"] suggests that you also get additive reduction at 2 if $A \not\equiv -1 \pmod 4$, but I was too lazy to check this.

- If $a \equiv -1 \pmod 4$ and $b \equiv 0 \pmod{16}$, then E has semistable reduction at 2 (like everywhere else). In this case, $E = E_{a,b,c}$ has minimal Weierstrass equation

$$y^2 + xy = x^3 + \frac{b-a-1}{4}x^2 - \frac{ab}{16}x,$$

and has

$$\Delta^{\min} = 2^{-8}(abc)^2 \text{ and } N = \left(\prod_{\text{odd } \ell \mid abc} \ell\right) \cdot \begin{cases} 2 & \text{if } 32 \mid b \\ 1 & \text{otherwise.} \end{cases} = \operatorname{rad}\left(\frac{abc}{16}\right).$$

(note the minimal discriminant $\Delta^{\min}$ above is an integer since $2^4 \mid b$).

Now we specialize to the case of a Fermat triple.

Remark 1.2. Fermat proves FLT in the case of exponent 4 and Euler proved it in the case of exponent 3. Thus, it suffices to prove it in the case of a prime exponent $p \geq 5$. Note that $(-c)^p = -c^p$, so Fermat's equation is equivalent for exponent p is equivalent to the equation

$$a^p + b^p + c^p = 0,$$

with a, b, c playing more symmetric roles. Mod 4 considerations show that for any non-trivial solution to this equation, one must have $\{a \bmod 4, b \bmod 4, c \bmod 4\} = \{0, 1, -1\}$. Thus, without loss of generality, we may suppose that $a \equiv -1 \bmod 4$ and $2 \mid b$. ○

Definition 1.3. A Fermat triple for prime exponent $p \geq 5$ is a triple of coprime integers (a, b, c) such that $a^p + b^p = c^p$, $a \equiv -1 \bmod 4$, and $2 \mid b$. ◇

By our previous discussion, for any Fermat triple (a, b, c) for exponent $p \geq 5$, the elliptic curve $E = E_{a^p, b^p, c^p}$ has everywhere semistable reduction, and its minimal discriminant and conductor are given by

$$\Delta^{\min} = 2^{-8}(abc)^{2p} \text{ and } N = \prod_{\ell \mid abc} \ell.$$

Start talk here, stating definition/properties of Frey curve w/o justification

Remark 1.4. Szpiro's conjecture (which is equivalent to ABC) says that for any $\varepsilon > 0$, there is a constant $C_\varepsilon > 0$ such that the minimal discriminant $\Delta(E)$ and conductor $N(E)$ of any elliptic curve $E/\mathbb{Q}$ satisfy

$$|\Delta(E)| < C \cdot N(E)^{6+\varepsilon}.$$

Thus, (an effective form of) Szpiro's conjecture would imply FLT (for large exponents). ○

For now, Szpiro's conjecture remains a conjecture, so a different route is needed to rule out the existence of the curve E.

Theorem 1.5 (Frey, Serre). *Let $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \operatorname{Aut}(E[p]) \simeq \operatorname{GL}_2(\mathbb{F}_p)$ be the Galois representation on $E[p]$. Then,*

(a) *$\overline{\rho}_{E,p}$ is absolutely irreducible;*

(b) *$\overline{\rho}_{E,p}$ is odd;*

(c) *$\overline{\rho}_{E,p}$ is unramified outside $2p$ and is flat at p.*

(More about this in Section 3)

Remark 1.6. You can find a proof of this in [Ser87, §4 Applications]. You can find more information about Frey curves and applications to other Fermat-like equations in that section. This is also the paper where Serre made Conjecture 1.8, though (if I'm not mistaken) I think the formulation of his conjecture in that paper is technically incorrect in some cases, so [CSS97, Chapter VII] might be a better place to start looking for more information on it. ○

Definition 1.7. Above $\overline{\rho}_{E,p}$ being "`flat at` p" means that the restriction $G_{\mathbb{Q}_p} \xrightarrow{\overline{\rho}_{E,p}} \mathrm{GL}_2(\mathbb{F}_p)$ is the representation coming from (the action of $G_{\mathbb{Q}_p}$ on the $\overline{\mathbb{Q}}_p$-points of) some finite, flat $\mathbb{Z}_p$-group scheme, if I'm understanding [CSS97, Chapter I, Definition (2.9)] correctly. ◇

Theorem 1.5 shows that the representation $\overline{\rho}_{E,p}$ is "too good to be true" in the sense that it contradicts (a stronger version of) the following conjecture of Serre.

Conjecture 1.8 (Serre)**.** *Every odd, absolutely irreducible Galois representation $\rho_0 : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_q)$ is modular.*

Remark 1.9. Apparently, this is now a theorem due to Chandrashekhar Khare and Jean-Pierre Wintenberger (including the stronger statement where you specify the weight and level of the form giving rise to ρ_0), see [KW09a, KW09b]. It's probably worth remarking that the full version of Serre's conjecture not only proves FLT, but also proves modularity as well; this is explained, for example, in Theorem 5.4 of these notes. ○

Because $\overline{\rho}_{E,p}$ is unramified outside $2p$ and flat at p, a more precise form a Serre's conjecture (described e.g. in [CSS97, Chapter VII, Edixhoven's article]) would predict that this representation comes from a modular form of level 2. This should be worrying, because $S_2(\Gamma_0(2)) = 0$ (the corresponding curve $X_0(2)$ has genus 0.[4]).

Remark 1.10. It's maybe worth emphasizing that the reason Serre's conjecture predicts the modular form will be of level 2 (instead of level 2^n for some $n > 1$) is that E has (at worst) multiplicative reduction at 2. Indeed, the level predicted by Serre is the conductor N of $\overline{\rho}_{E,p}$.[5] For any prime $\ell \neq p$, one has $v_\ell(N) \geq 2$ only if ℓ is a prime of additive reduction (see [Sil94, Proof of Theorem IV.10.2]); I am not 100% sure if this is an 'iff' (it would be if we were talking about $\rho_{E,p}$ instead of $\overline{\rho}_{E,p}$). ○

Serre's full conjecture was still open when FLT was proven, so the following "level-lowering" result (which would follow from the full form of Serre's conjecture) of Ribet was used instead.

[4] Remark 1.1 shows that $Y(2)$ is $\mathbb{A}^1 \setminus \{0,1\}$, so $X(2)$ has genus 0. $X_0(2)$ is a quotient of $X(2)$, so it better be of genus 0 as well.

[5] Except for some funny business related to the exponent of p in the level. What's important for us, though, is that p does not appear in the level if $\overline{\rho}_{E,p}$ is flat at p.

Theorem 1.11 (Ribet, Serre's epsilon conjecture)**.** *Let f be a weight two newform of conductor $N\ell$, where $\ell \nmid N$ is prime. Suppose $\overline{\rho}_f$ is absolutely irreducible and that either*

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

Then, there is a weight two newform g of conductor N such that $\overline{\rho}_f \cong \overline{\rho}_g$.

Proof of FLT (Theorem A), assuming modularity. Assume that (a, b, c) is a triple of coprime integers satisfying $a^p + b^p = c^p$ (for some prime $p \geq 5$), $2 \mid b$, and $a \equiv -1 \pmod 4$. By the previous discussion, FLT will hold if we can show that no such triple exists. By assumption, the Frey curve $E = E_{a^p,b^p,c^p}$ is modular, so its mod p representation $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ is modular. Combining Theorem 1.5 with (several applications of) Theorem 1.11, we conclude that there exists some weight two newform g of conductor 2 such that $\overline{\rho}_{E,p} \cong \overline{\rho}_g$. However, as remarked earlier, $S_2(\Gamma_0(2)) = 0$, so there is no weight two newform g of conductor 2, a contradiction. ■

2 Proof Strategy for Modularity

We end by saying a few words of how Wiles, Taylor-Wiles prove modularity of semistable elliptic curves. More words can be found in [CSS97, Chapter 1, Section 7]. Even more words can be found in the rest of that book and also in [DDT07].

2.1 The case when $\overline{\rho}_{E,3}$ is irreducible

Setup 2.1. Let $E/\mathbb{Q}$ be a semistable elliptic curve. Furthermore, suppose that $\overline{\rho}_{E,3}$ is irreducible.

The strategy here is to show that $\overline{\rho}_{E,3}$ is modular, and then to "lift" this modularity from $\overline{\rho}_{E,3}$ to $\rho_{E,3}$ by studying deformations of this representation.

Theorem 2.2. *$\overline{\rho}_{E,3}$ is modular.*

Proof Sketch, stated roughly. One uses a few coincidences to deduce this from a result of Langlands-Tunnel about modularity of *complex* Galois representations.

- There is an injective homomorphism $\psi : \mathrm{GL}_2(\mathbb{F}_3) \hookrightarrow \mathrm{GL}_2(\mathbb{Z}[\sqrt{-2}]) \subset \mathrm{GL}_2(\mathbb{C})$ which is a splitting of the reduction mod $(1+\sqrt{-2})$ map $\mathrm{GL}_2(\mathbb{Z}[\sqrt{-2}]) \twoheadrightarrow \mathrm{GL}_2(\mathbb{F}_3)$.

  This let's one extend $\overline{\rho}_{E,3}$ to the complex representation $G_{\mathbb{Q}} \xrightarrow{\overline{\rho}_{E_3}} \mathrm{GL}_2(\mathbb{F}_3) \xrightarrow{\psi} \mathrm{GL}_2(\mathbb{C})$.

- Using that $\overline{\rho}_{E,3}$ is odd and that $\mathrm{GL}_2(\mathbb{F}_3)$ is solvable, Langlands-Tunnel [CSS97, Chapter VI, Theorem 1.3] show that $\psi \circ \overline{\rho}_{E,3}$ is "modular of weight 1" in the sense that there exists a normalized eigenform $g(\tau) = \sum_{n \geq 1} b_n q^n \in S_1(\Gamma_0(N), \psi)$ (for some level N and character ψ) such that $b_\ell = \mathrm{Tr}(\psi \circ \overline{\rho}_{E,3}(\mathrm{Frob}_\ell))$ for almost all primes ℓ.

  Remark 2.3. g above is an eigenvector for *all* Hecke operators, *not* just those T_n with $\gcd(n, N) = 1$. ○

Somehow (to my naive eyes), it seems this Langlands-Tunnel result is the source of modularity underlying to whole argument.

- There exists some Eisenstein series E of weight 1 such that $E \equiv 1 \pmod 3$. Thus, Eg is a weight 2 cusp form whose coefficients are congruent to $b_\ell \equiv \operatorname{Tr}(\overline{\rho}_{E,3}(\mathrm{Frob}_\ell)) \pmod 3$.

I'm being imprecise. These coefficients lie in some number field, so really you should look mod some prime above 3, but I'll continue to ignore this subtlety

- One can find a weight 2 *eigenform* whose coefficients are congruent mod 3 to those of Eg. This eigenform is now a witness to $\overline{\rho}_{E,3}$'s modularity.

  See Lemma 3.7 for more details on this. ■

Theorem 2.4. *$\rho_{E,3}$ (and hence E) is modular. In fact, Wiles proves that for any prime $p \geq 3$, if $\overline{\rho}_{E,p}$ is modular and irreducible, then E is modular.*

This is proved using some deformation theory argument I understand nothing about, so I won't say any more than that.

2.2 The case when $\overline{\rho}_{E,3}$ is reducible

Setup 2.5. Let $E/\mathbb{Q}$ be a semistable elliptic curve. Furthermore, suppose that $\overline{\rho}_{E,3}$ is reducible.

In this case, one can use a trick to connect modularity of E to modularity of some other elliptic curve E'.

Theorem 2.6 (3-5 Trick)**.** *Suppose $\overline{\rho}_{E,5}$ is irreducible. Then, there is another semistable elliptic curve $E'/\mathbb{Q}$ such that $\overline{\rho}_{E',3}$ is irreducible and $\overline{\rho}_{E',5} \cong \overline{\rho}_{E,5}$.*

Proof Sketch. Consider the (fine) moduli space $Y(E[5])$ parameterizes pairs $(E', E'[5] \xrightarrow{\sim} E[5])$ of elliptic curves E' equipped with Galois-equivariant, symplectic (i.e. preserves the Weil pairing) isomorphism $E'[5] \xrightarrow{\sim} E[5]$. This moduli problem is a twist of $\mathcal{Y}(5) = Y(5)$ and so is (representable by) a smooth, affine curve of genus 0 (Indeed, $Y(E[5]) \simeq Y(5)$ over the field $\mathbb{Q}(E[5])$). Thus, $Y(E[5])(\mathbb{Q})$ is huge, so there are plenty of elliptic curves E' with $\overline{\rho}_{E',5} \cong \overline{\rho}_{E,5}$. At least one of these will be semistable and also have $\overline{\rho}_{E',3}$ irreducible. ■

Given such an E', Theorem 2.2 shows that E' is modular (using modularity of $\overline{\rho}_{E',3}$). Hence, $\overline{\rho}_{E',5} \cong \overline{\rho}_{E,5}$ is modular, so Theorem 2.4 now shows that E is modular. Finally,

Lemma 2.7. *At least one of the representations $\overline{\rho}_{E,3}$ or $\overline{\rho}_{E,5}$ is irreducible.*

Proof. If not, then $E[15]$ would contain Galois invariant subgroup of order 15, so E would admit a (cyclic) 15-isogeny. However, the modular curve $X_0(15)$ has 4 non-cuspidal $\mathbb{Q}$-points and one can check that none of the corresponding elliptic curves are semistable (at 5). ■

3 Miscellaneity

Here, I want to collect proofs of some lemmas/facts that would have taken up too much space if I included them above.

3.1 Proof of Theorem 1.5

The main thing needed to prove Theorem 1.5 is an understanding of ramification in the mod p representation of a semi-stable elliptic curve. The below two propositions are essential parts (c),(d) of [DDT07, Proposition 2.12].

Proposition 3.1. *Let $E/\mathbb{Q}$ be a semistable elliptic curve, and consider*

$$\overline{\rho}_{E,p} : G_{\mathbb{Q}} \longrightarrow \operatorname{Aut}(E[p]) \cong \mathrm{GL}_2(\mathbb{F}_p).$$

Let Δ denote the minimal discriminant of E. For any prime $\ell \neq p$, $\overline{\rho}_{E,p}$ is unramified at ℓ if and only if $p \mid v_\ell(\Delta)$.

Proof. If ℓ is a prime of good reduction, then $v_\ell(\Delta) = 0$ is divisible by p and $\overline{\rho}_{E,p}$ is unramified at ℓ (e.g. since the reduction map $E[p](\mathbb{Q}) \hookrightarrow E[p](\mathbb{Q}_\ell) = E[p](\mathbb{Z}_\ell) \hookrightarrow E[p](\mathbb{F}_\ell)$ is injective). Say ℓ is a prime of bad reduction. Because E is semistable, ℓ necessarily is of multiplicative reduction. Let $K/\mathbb{Q}_\ell$ be an unramified extension (possibly of degree 1) at which E obtains *split* multiplicative reduction. Then, the theory of Tate curves [DDT07, Proposition 1.5] shows that $E(\overline{\mathbb{Q}}_\ell) \cong \mathbb{G}_m/q^{\mathbb{Z}}$ as G_K-modules, where $q \in K$ is some number satisfying $v_\ell(q) = v_\ell(\Delta)$. Since $K/\mathbb{Q}_\ell$ is unramified, $\overline{\rho}_{E,p}$ will be unramified at ℓ if and only if its restriction to G_K is unramified.

Note that $E[p](\overline{\mathbb{Q}}_\ell) = \mu_p \times (p\text{th roots of } q)$, so $\overline{\rho}_{E,p}|_{G_K}$ is unramified if and only if $K(\zeta_p, q^{1/p})/K$ is an unramified extension. This is the case if and only if $p \mid v_\ell(q)$ (in which case, q already has a pth root in K). Since $v_\ell(q) = v_\ell(\Delta)$, the claim follows. ■

Recall 3.2 (Definition 1.7)**.** With notation as above, $\overline{\rho}_{E,p}$ is *flat at p* if there exists some finite, flat $\mathbb{Z}_p$-group scheme G such that $E[p](\overline{\mathbb{Q}}_p) \cong G(\overline{\mathbb{Q}}_p)$ Galois-equivariantly. ⊙

Proposition 3.3. *Let $E/\mathbb{Q}$ be a semistable elliptic curve, and consider*

$$\overline{\rho}_{E,p} : G_{\mathbb{Q}} \longrightarrow \operatorname{Aut}(E[p]) \cong \mathrm{GL}_2(\mathbb{F}_p).$$

Let Δ denote the minimal discriminant of E. Then, $\overline{\rho}_{E,p}$ is flat at p if and only if $p \mid v_p(\Delta)$.

Remark 3.4. I am going to give two arguments for this. The first one only proves one direction (though it is the direction we need). The second one is essentially the argument given in [Ser87, Propositions 4 and 5]; in principle, it proves both directions, but I am kinda iffy on some of the details so you can decide what exactly is proven by what I've written down. ○

Proof 1. We will simply show that p-torsion in E's Néron model has constant fiber degree p^2 and so is finite.[6]

Suppose that $p \mid v_p(\Delta)$. Let $\mathscr{E}/\mathbb{Z}_p$ denote E's Néron model, and let $\mathscr{E}_0/\mathbb{F}_p$ denote its special fiber. If $v_p(\Delta) = 0$, then $\mathscr{E}[p]$ is a finite, flat $\mathbb{Z}_p$-scheme extending $E[p]$, so $\overline{\rho}_{E,p}$ is flat at p. Suppose that $v_p(\Delta) > 0$. By assumption, E has multiplicative reduction in this case, so $\mathscr{E}_0^0[p]$ (p-torsion in the identity component) is a finite $\mathbb{F}_p$-group scheme of order p (it's a twist of μ_p) and the group $\Phi := \mathscr{E}_0/\mathscr{E}_0^0$ of components is a cyclic finite, étale $\mathbb{F}_p$-group scheme of order $n := v_p(\Delta)$ [Sil94, Table 4.1]. Since $\mathbb{F}_p$ is perfect, we have $\mathscr{E}_0 \simeq \Phi \times \mathscr{E}_0^0$, so $\mathscr{E}_0[p] \simeq \Phi[p] \times \mathscr{E}_0^0[p]$ is a finite $\mathbb{F}_p$-group scheme of order p^2 (note $\#\Phi[p] = p$ since $p \mid v_p(\Delta) = \#\Phi$). Hence, $\mathscr{E}[p]$ is a quasi-finite, flat (flatness e.g. by [Ces15, Lemma B.4]) $\mathbb{Z}_p$-group scheme whose fibers all have order p^2. As a consequence of the structure theorem for quasi-finite, separated schemes over a local henselian base [Con, Theorem 4.10] this means that $\mathscr{E}[p]$ is actually finite over $\mathbb{Z}_p$, and thus $\overline{\rho}_{E,p}$ is flat at p. ■

Proof 2. By the theory of Tate curves there exists an unramfied extension $K/\mathbb{Q}_p$ (of degree ≤ 2) over which we obtain an exact sequence

$$0 \longrightarrow \mu_p \longrightarrow E[p] \longrightarrow \mathbb{Z}/p\mathbb{Z} \longrightarrow 0 \tag{3.1}$$

of G_K-modules (equivalently, étale K-group schemes). We will show that this extension comes from an extension of $\mathbb{Z}/p\mathbb{Z}$ by μ_p *over* $\mathscr{O}_K$ if and only if $p \mid v_p(\Delta)$.[7] Serre [Ser87, Proof of Proposition 4] seems to suggest that having such an extension over $\mathscr{O}_K$ suffices to obtain one over $\mathbb{Z}_p$, but it's not clear to me why.

The trick to understanding extensions of $\mathbb{Z}/p\mathbb{Z}$ by μ_p is applying $\operatorname{Hom}(-, \mu_p)$ to the short exact sequence $0 \to \mathbb{Z} \xrightarrow{p} \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \to 0$. Doing this (and using that $K/\mathbb{Q}_p$ is unramified, so $\mu_p(K) = 1$) gives the first isomorphism in the sequence below

$$\operatorname{Ext}^1_K(\mathbb{Z}/p\mathbb{Z}, \mu_p) \xrightarrow{\sim} \operatorname{Ext}^1_K(\mathbb{Z}, \mu_p)[p] \simeq \mathrm{H}^1(K, \mu_p)[p] = \mathrm{H}^1(K, \mu_p) \simeq K^\times/(K^\times)^p,$$

The same argument shows that $\operatorname{Ext}^1_{\mathscr{O}_K}(\mathbb{Z}/p\mathbb{Z}, \mu_p) \simeq \mathrm{H}^1(\mathscr{O}_K, \mu_p) \simeq \mathscr{O}_K^\times/(\mathscr{O}_K^\times)^p$ (fppf cohomology). The explicit construction of the exact sequence (3.1) in terms of Tate curves shows that its corresponding element of $K^\times/(K^\times)^p$ is the Tate period $q \in \mathbb{Q}_p^\times \subset K^\times$. This will be in the image of the map $\mathscr{O}_K^\times/(\mathscr{O}_K^\times)^p \to K^\times/(K^\times)^p$ if and only if q is of the form $u\alpha^p$ for some $u \in \mathscr{O}_K^\times$ and $\alpha \in K^\times$ if and only if $p \mid v_p(q)$. Since $v_p(q) = v_p(\Delta)$, we win. ■

Now, we can prove Theorem 1.5.

Proof of Theorem 1.5. Let $E = E_{a^p, b^p, c^p}$ be the Frey curve attached to some Fermat triple (a, b, c).

[6] By the proof, this is true if and only if $p \mid v_p(\Delta)$. However, I don't know if $\overline{\rho}_{E,p}$ being flat at p is equivalent to the Néron model having finite, flat p-torsion. I mean, this must be true by Proposition 3.3, but I don't know if this is obviously true w/o proving Proposition 3.3.

[7] I think it might be possible to show that any finite, flat $\mathscr{O}_K$-scheme extending $E[p]$ must also be an extension of $\mathbb{Z}/p\mathbb{Z}$ by μ_p by using Raynaud's theorem [CSS97, Chapter V, Theorem 4.5.1]. I haven't thought too carefully about this though.

Recall that it has minimal discriminant

$$\Delta = 2^{-8}(abc)^{2p}$$

which is a p-power away from 2. Consider the mod p representation $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p]) \cong \mathrm{GL}_2(\mathbb{F}_p)$.

(a) The argument (which I stole from [Ser87, Proposition 6]) uses the following fact that I'm still not sure how to prove.

Fact (See Proposition 3.8 for more info)**.** For $E/\mathbb{Q}$ a semistable elliptic curve, if the mod p represetnation $\overline{\rho}_{E,p}$ is not surjective, then E either contains a point of order p or E is p-isogenous to a curve w/ a point of order p.

If you accept this, it's easy to show that $\overline{\rho}_{E,p}$ is irreducible (even surjective). If not, E (or a p-isogenous curve E') contains a point of order p; but E has full 2-torsion, this would force $\#E_{\text{tors}} \geq 4p \geq 20$ (or force $\#E'_{\text{tors}} \geq 20$), contradicting Mazur.

Exercise. Show that if $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ is irreducible, then it's absolutely irreducible. (Hint[8]).

(b) The (perfect, alternating, Galois-equivariant) Weil pairing $E[p] \times E[p] \to \mathbb{G}_m$ shows that $\det \overline{\rho}_{E,p}$ is the mod p cyclotomic character, so $\det \overline{\rho}_{E,p}(\text{cmplx conj}) = -1$. Hence, $\overline{\rho}_{E,p}$ is odd.

(c) The facts that E has everywhere semistable reduction and that Δ is a p-power away from 2 show (by Propositions 3.1 and 3.3) that $\overline{\rho}_{E,p}$ is unramified away from $2p$ and is flat at p. ■

3.2 Deligne-Serre Lifting Lemma

We want to add more details to the last bullet point of Theorem 2.2. In particular, we prove a special case of a lemma due to Deligne and Serre; in brief, mod p eigenforms of weight k lift to actual eigenforms of weight k. More information can be found e.g. in these notes (especially in their Lemma 1.2).

Notation 3.5. Let $\mathbb{T}_k(N) \subset \mathrm{End}(S_k(\Gamma_0(N)))$ denote weight Hecke algebra acting on weight k cusp forms of level $\Gamma_0(N)$. This is generated by the Hecke operators $\{T_n : n \in \mathbb{N}\}$ (maybe sometimes people write e.g. U_p instead of T_p if $p \mid N$).

Remark 3.6. The third bullet point of Theorem 2.2 produced a weight 2 cusp form, there called Eg, which was congruent to a normalized (cuspidal) eigenform mod 3 ($Eg \equiv g \pmod 3$). Hence, this Eg gives rise to a ring homomorphism $\mathbb{T}_2(N) \to \overline{\mathbb{F}}_3$, $T \mapsto a_1(T \cdot Eg)$ mod 3. ○

Lemma 3.7 (Deligne-Serre)**.** *Fix $k \geq 2$, $N \geq 1$, and a prime p. Let $\psi : \mathbb{T}_k(N) \to \overline{\mathbb{F}}_p$ be a ring homomorphism. Then, there exists some cusp form $g = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma_0(N))$, which is an*

[8] You'll need to use that $p > 2$. Consider the eigenspaces of complex conjugation.

eigenvector for all Hecke operators, with coefficients in an order $\mathscr{O}$ of some number field $K = \operatorname{Frac}\mathscr{O}$ such that there exists a prime $\mathfrak{p}$ of $\mathscr{O}$ for which

$$a_n \equiv \psi(T_n) \pmod{\mathfrak{p}}$$

for all $n \geq 1$.

Proof. Consider the maximal ideal $\mathfrak{m} := \ker\psi \subset \mathbb{T}_k(N)$. Since $p \in \mathfrak{m}$ is not nilpotent (after all, $\mathbb{T}_k(N)$ is contained in a $\mathbb{C}$-algebra), there must exist some prime $\mathfrak{q} \subset \mathfrak{m}$ not containing p. Hence, $\mathfrak{q} \cap \mathbb{Z} = (0)$ (otherwise, $\mathfrak{m}$ would contain two prime numbers and so be the unit ideal), so $\mathscr{O} := \mathbb{T}_k(N)/\mathfrak{q}$ is a domain of characteristic 0. Note that $\mathscr{O}$ is an order in the number field $K = \operatorname{Frac}\mathscr{O}$. Let $\mathfrak{p} = \mathfrak{m}/\mathfrak{q}$ be the image of $\mathfrak{m}$ in $\mathscr{O}$, and write $\varphi : \mathbb{T}_k(N) \to \mathscr{O}$ for the quotient map. Then, $\mathfrak{p} \subset \mathscr{O}$ is a prime ideal and $\varphi(T_n) \equiv \psi(T_n) \pmod{\mathfrak{p}}$ by construction. Our desired cusp form is simply $g := \sum_{n\geq 1} \varphi(T_n)q^n \in S_k(\Gamma_0(N))$. ■

3.3 Torsion in semistable curves

Proposition 3.8 ([Ser72, Proposition 21 ii)])**.** *Let $E/\mathbb{Q}$ be a semistable elliptic curve, and assume that $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ is not surjective. Then, E either contains a point of order p or E is p-isogenous to an elliptic curve w/ a point of order p.*

The first part of Serre's argument is showing that $\overline{\rho}_{E,p}$ is reducible if it is not surjective. I am now sure how he shows this, but it is my understanding that one often knows a priori that E has some subgroup of order p when applying Proposition 3.8, so we will only prove this proposition under the assumption that $\overline{\rho}_{E,p}$ is reducible.

Assumption. Fix a semistable elliptic curve $E/\mathbb{Q}$ as well as a prime p. Assume that there exists some exact sequence

$$0 \longrightarrow G_1 \longrightarrow E[p] \longrightarrow G_2 \longrightarrow 0$$

of (non-trivial) finite $\mathbb{Q}$-group schemes. Equivalently, there is an $\overline{\rho}_{E,p}$ is an extension

$$0 \longrightarrow \chi_1 \longrightarrow \overline{\rho}_{E,p} \longrightarrow \chi_2 \longrightarrow 0$$

of characters $\chi_i : G_{\mathbb{Q}} \to \mathbb{F}_p^\times$.

Remark 3.9. Before getting into a more detailed argument, let's say broadly what one does to prove Proposition 3.8. The main point is to show that one of the characters χ_1, χ_2 must be unramified everywhere. Since $\mathbb{Q}$ has no non-trivial unramified extensions, this will mean that $\chi_1 = 1$ or $\chi_2 = 1$, equivalently, that $G_1 = \mathbb{Z}/p\mathbb{Z}$ or $G_2 = \mathbb{Z}/p\mathbb{Z}$. In the first case E has a point of order p, and in the second case, E/G_1 has a point of order p. Ramification of χ_1, χ_2 is analyzed mostly by looking at p-torsion in Tate curves. ○

Lemma 3.10. *Let $\ell \neq p$ be prime. Then, both χ_1 and χ_2 are both unramified at ℓ.*

Proof. If ℓ is a prime of good reduction, then $E[p]$ is unramified at ℓ, so χ_1, χ_2 are both unramified as well. If ℓ is a prime of bad reduction, then (possibly after an unramified extension $K/\mathbb{Q}_\ell$) $E_{\mathbb{Q}_\ell}$ is a Tate curve. Hence, one has an extension $0 \to \mu_p \to E[p] \to \mathbb{Z}/p\mathbb{Z} \to 0$ of $\mathbb{Q}_\ell$-group schemes. Thus, χ_1, χ_2 are the characters attached to $\mu_p, \mathbb{Z}/p\mathbb{Z}$ (i.e. one is trivial and one is the mod p cyclotomic character), so both of them are unramified at ℓ. ■

Lemma 3.11. *If E does not have good supersingular reduction at p, then one of χ_1 and χ_2 is unramified at p (and the other is ramified).*

Proof. By assumption, E either has bad (in which case, it is multiplicative) or good reduction at E. If E has bad reduction, one says the phrase 'Tate curve' and then deduces the existence of an extension $0 \to \mu_p \to E[p] \to \mathbb{Z}/p\mathbb{Z} \to 0$. Whichever of χ_1, χ_2 corresponds to $\mathbb{Z}/p\mathbb{Z}$ is unramified at p (and the other, corresponding to μ_p, is ramified at p). So, suppose that E has good ordinary reduction at p. By [Sil09, Proposition VII.2.1], the $G_{\mathbb{Q}_p}$-module $E[p](\overline{\mathbb{Q}}_p)$ has $E[p](\overline{\mathbb{F}}_p)$ as a quotient. Now, the inertial subgroup $I_p \le G_{\mathbb{Q}_p}$ acts trivially on $E[p](\overline{\mathbb{F}}_p)$ (since $G_{\mathbb{Q}_p}$ acts on the $\overline{\mathbb{F}}_p$-points via its map down to $G_{\mathbb{F}_p}$), so I_p acts on the kernel $K := \ker\big(E[p](\overline{\mathbb{Q}}_p) \longrightarrow E[p](\overline{\mathbb{F}}_p)\big)$ via the mod p cyclotomic character (which is ramified at p). Thus, exactly one of χ_1, χ_2 is ramified at p. ■

Lemma 3.12. *E does not have good supersingular reduction at p. In fact, if $E/\mathbb{Q}$ is any elliptic curve with good supersingular reduction at p, then $\overline{\rho}_{E,p}$ is irreducible.*

Proof. Suppose that $E/\mathbb{Q}$ is an elliptic curve with good supersingular reduction at p such that $\overline{\rho}_{E,p}$ is reducible. Hence, we have an order p subgroup $G_1 \le E[p]$. Let $\mathcal{E}/\mathbb{Z}_p$ be E's Néron model, an elliptic scheme by assumption. Let $\mathcal{G}_1/\mathbb{Z}_p$ be the scheme-theoretic closure of G_1 in $\mathcal{E}[p]$, so $\mathcal{G}_1$ is a finite, flat $\mathbb{Z}_p$-group scheme[9] with generic fiber G_1. Because E has supersingular reduction at p, the special fiber $\mathcal{E}[p]_{\mathbb{F}_p}$ is an extension $0 \to \alpha_p \to \mathcal{E}[p]_{\mathbb{F}_p} \to \alpha_p \to 0$ of α_p by α_p. Thus, the special fiber $\mathcal{G}_{1,\mathbb{F}_p}$ must be isomorphic to α_p, so, to finish the proof, it suffices to show that no commutative finite flat $\mathbb{Z}_p$-group scheme has α_p as its special fiber. We sketch a proof of this based on the following result of Raynaud:

This is possibly more complicated than need be. If you know an easier way to show α_p doesn't prolongate over $\mathbb{Z}_p$, please let me know.

Fact (Raynaud, [CSS97, Theorem V.4.5.1])**.** Let R be a dvr of mixed characteristic $(0, p)$. Assume that its absolute ramification index satisfies $v(p) < p - 1$. Let G be a commutative finite flat R-group scheme of p-power order. Then, G is, up to isomorphism, the unique prolongation of its generic fiber.

Raynaud's theorem tells us that any finite, flat commutative $\mathbb{Z}_p$-group scheme of order p is determined by its generic fiber. Thus, it suffices to classify finite $\mathbb{Q}_p$-group schemes of order p and see that none of them specialize to α_p over $\mathbb{F}_p$ by simply writing down some prolongation of each one to a $\mathbb{Z}_p$-scheme.

To classify (commutative) finite $\mathbb{Q}_p$-group schemes (of order p), one can first use Cartier's theorem [Ach21, Theorem 5.2] in order to know that any such group scheme is étale. Thus, commutative finite $\mathbb{Q}_p$-group schemes of order p are nothing other than $G_{\mathbb{Q}_p}$-modules of order p. That is, they

[9] See here for how to argue that $\mathcal{G}_1$ is a flat group scheme. It is finite since it's closed in the finite group scheme $\mathcal{E}[p]$.

are all given by actions of $G_{\mathbb{Q}_p}$ on some finite group M of order p (necessarily, $M = \mathbb{Z}/p\mathbb{Z}$), so they're simply homomorphisms $G_{\mathbb{Q}_p} \to \mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) \cong \mathbb{Z}/(p-1)\mathbb{Z}$. These can be classified using either class field theory or Kummer theory.[10] ■

If you combine the statements of the above three lemmas (along with the remark preceding them), then you can prove Proposition 3.8, at least under the (a priori) stronger assumption that $\overline{\rho}_{E,p}$ is reducible.

References

[Ach21] Niven Achenjang. 18.737 notes. https://www.mit.edu/~NivenT/assets/pdf/18_737_Notes.pdf, 2021. 12

[Ces15] Kestutis Cesnavicius. Selmer groups as flat cohomology groups. https://arxiv.org/abs/1301.4724, 2015. 9

[Con] Brian Conrad. Semistable reduction for abelian varieties. http://virtualmath1.stanford.edu/~conrad/mordellsem/Notes/L13.pdf. 9

[CSS97] Gary Cornell, Joseph H. Silverman, and Glenn Stevens, editors. *Modular forms and Fermat's last theorem*. Springer-Verlag, New York, 1997. Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995. 1, 5, 6, 9, 12

[DDT07] Henri Darmon, Fred Diamond, and Richard Taylor. Fermat's last theorem. https://www.math.mcgill.ca/darmon/pub/Articles/Expository/05.DDT/paper.pdf, Sept 2007. 1, 3, 6, 8

[KW09a] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. I. *Invent. Math.*, 178(3):485–504, 2009. 5

[KW09b] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. II. *Invent. Math.*, 178(3):505–586, 2009. 5

[Ser72] Jean-Pierre Serre. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. *Invent. Math.*, 15(4):259–331, 1972. 11

[Ser87] Jean-Pierre Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987. 5, 8, 9, 10

[Sil94] Joseph H. Silverman. *Advanced topics in the arithmetic of elliptic curves*, volume 151 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994. 3, 5, 9

[10] In either case, the key is to realize that $\mathbb{Q}_p^\times \cong \mathbb{Z}_p^\times \times p^{\mathbb{Z}}$. In impressionistic terms, the $p^{\mathbb{Z}}$ factor will give rise to μ_p and anything coming from the $\mathbb{Z}_p^\times$ factor will extend to an *étale* $\mathbb{Z}_p$-group scheme. Admittedly, it wasn't until just now that I realized you could have a homomorphism non-trivial on both factors, but probably these work out to not give you α_p as well...

[Sil09] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009. 12

4 List of Marginal Comments