MF Overview Notes

Niven Achenjang

Spring 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.

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 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

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

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 2 (not important). Let $\mathcal{Y}(2)$ denote the (fine) moduli space of elliptic curves equipped with a (naive) full level 2 structure.[1] 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),[2] 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 $a+b=c$. ○

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:

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

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

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

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

[2] 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.

(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

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

(I've also seen $y^2 = x(x-a)(x-b)$ before, but the above is what appears in [CSS97] and [DDT07]).

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

- 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 $A \not\equiv -1 \mod 4$, but I was too lazy to check this.

- If $a \equiv -1 \mod 4$ and $b \equiv 0 \mod 16$, 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

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

Now we specialize to the case of a Fermat triple.

Remark 3. 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 4. 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 every semistable reduction, and its minimal discriminant and conductor are given by

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

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

Remark 5. 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 6 (Frey, Serre)**.** *Let $\overline{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[p]) \simeq \mathrm{GL}_2(\mathbb{F}_p)$ be the Galois representation on $E[p]$. Then,*

TODO: Figure out how to prove this

(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.*

Remark 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 1, Definition (2.9)] correctly. ○

Theorem 6 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 8 (Serre)**.** *Every odd, absolutely irreducible Galois representation $\rho_0 : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_q)$ is modular.*

Because $\overline{\rho}_{E,p}$ is unramified outside $2p$ and flat at p, a more precise form a Serre's conjecture 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 $X_0(2)$ is genus 0.[3]). As far as I know, Serre's conjecture is still open, so it won't directly be the route by which we obtain a contradiction. Instead, one uses the following "level-lowering" result of Ribet.

Theorem 9 (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 6 with (several applications of) Theorem 9, 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 of 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 10. 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 11. *$\overline{\rho}_{E,3}$ is modular.*

Proof Sketch, stated roughly. One uses a few coincidences to deduce 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 $G_{\mathbb{Q}} \xrightarrow{\overline{\rho}_{E_3}} \mathrm{GL}_2(\mathbb{F}_3) \xrightarrow{\psi} \mathrm{GL}_2(\mathbb{C})$.

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

- Using that $\overline{\rho}_{E,3}$ is odd and that $\mathrm{GL}_2(\mathbb{F}_3)$ is solvable, Langlands-Tunnel 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.
- 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 $b_\ell \cong \mathrm{Tr}(\overline{\rho}_{E,3}(\mathrm{Frob}_\ell)) \mod 3$.

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 not a witness to $\overline{\rho}_{E,3}$'s modularity. ■

TODO: Figure out a sketch of this

Theorem 12. *$\rho_{E,3}$ (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 more than that.

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

Setup 13. 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 connected modularity of E to modularity of some other elliptic curve E'.

Theorem 14. *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 $\overline{\rho}_{E',5} \cong \overline{\rho}_{E,5}$.*

TODO: Add sketch of this

Given such an E', Theorem 11 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 12 now shows that E is modular. Finally,

Lemma 15. *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). ■

References

[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, 3, 4, 5

[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, 5

[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