The Langlands-Tunnell theorem and its application to proving Fermat's Last Theorem

Ravi Fernando – fernando@berkeley.edu

April 30, 2015*

1 Introduction

Both historically and in this class, great effort has been put into proving modularity lifting theorems. But in order to use these, we need some starting point. In Wiles's (and Taylor-Wiles's) proof of Fermat's Last Theorem, the main starting point is the Langlands-Tunnell theorem. The following is a very rough schematic of how modularity is generated and propagated around.

$$\begin{aligned}\text{Langlands-Tunnell} &\overset{(a)}{\Longrightarrow} \overline{\rho}_{E,3} \text{ is modular} && (1)\\ &\overset{(b)}{\Longrightarrow} \text{some lift } \rho_0 : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O}_L) \text{ of } \overline{\rho}_{E,3} \text{ is modular} && (2)\\ &\overset{(c)}{\Longrightarrow} \rho_{E,3} \text{ is modular} && (3)\\ &\overset{(d)}{\Longrightarrow} E \text{ is modular.} && (4)\end{aligned}$$

To explain the statements above: $E/\mathbb{Q}$ is an elliptic curve, $\overline{\rho}_{E,p}$ denotes the representation $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ given by the Galois action on the p-torsion of $E(\overline{\mathbb{Q}})$, and $\rho_{E,p}$ is the representation $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}_p)$ given by the action on the Tate module, which lifts $\overline{\rho}_{E,p}$.

One definition of modularity of $\overline{\rho}_{E,3}$ is that it is the reduction mod 3 of a Galois representation coming from a modular form; we'll work with another definition in the next section. Modularity of E means that there exists a weight two modular form $f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$ with $a_q = q + 1 - \#E(\mathbb{F}_q)$ for almost all primes q. Equivalently, this means that at least one, and equivalently all, p-adic representations of $G_{\mathbb{Q}}$ given by T_pE (or $H^1_{\text{ét}}(E, \mathbb{Q}_p)$), agree with the representations $\rho_{f,p}$ coming from f.

The implication (a) holds under the hypothesis that $\overline{\rho}_{E,3}$ is irreducible. This involves a mostly elementary argument, taking one theorem of Deligne and Serre as an input. We will mostly focus on this step today.

*Notes for a talk given in a shadow seminar for Sug Woo Shin's course on Galois representations at Berkeley. Main reference: Gelbart, *Three lectures on the modularity of $\overline{\rho}_{E,3}$ and the Langlands Reciprocity Conjecture.*

Aside: for the case where $\overline{\rho}_{E,3}$ is reducible, we need a different argument. Wiles filled this gap using the 3/5 switch.

Implication (b) holds more or less definitionally. (Modularity of $\overline{\rho}_{E,3}$ means that $\overline{\rho}_{E,3}$ is the reduction mod 3 of some Galois representation $\rho_{f,3}$ induced by a modular form, so we can take ρ_0 to be $\rho_{f,3}$ itself.) Implication (c) involves modularity lifting, which obviously takes a lot of work. (Wiles uses a modularity lifting theorem that requires the restriction of $\overline{\rho}_{E,3}$ to $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{-3}))$ to be absolutely irreducible, which follows from irreducibility of $\overline{\rho}_{E,3}$ in this case.) Implication (d) is also basically definitional, depending on which version of the definition you use.

2 Langlands-Tunnell and modularity of $\overline{\rho}_{E,3}$

We would like to show the following:

Theorem 2.1. *Given an elliptic curve $E/\mathbb{Q}$, if $\overline{\rho}_{E,3}$ is irreducible, then it is modular. That is, there exists a normalized eigencuspform $f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$ of weight two and a prime λ of $\overline{\mathbb{Q}}$ such that $a_q \equiv \mathrm{tr}(\overline{\rho}_{E,3}(\mathrm{Frob}_q)) \pmod{\lambda}$ for almost all primes q.*

Langlands-Tunnell gives us this:

Theorem 2.2. *Suppose $\sigma : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$ is an odd irreducible representation with solvable image in $\mathrm{PGL}_2(\mathbb{C})$. Then there exists a normalized eigencuspform $g(z) = \sum_{n=1}^{\infty} b_n e^{2\pi i n z}$ such that $b_q = \mathrm{tr}(\sigma(\mathrm{Frob}_q))$ for almost all primes q.*

Remark: Since the image in $\mathrm{GL}_2(\mathbb{C})$ is automatically finite, and we have a nice classification of the finite subgroups of $\mathrm{PGL}_2(\mathbb{C})$, there are essentially only three cases: the projective image can be dihedral, tetrahedral (A_4), or octahedral (S_4). (The icosahedral case doesn't have solvable image.) The first case is essentially due to Hecke and Maass, the second to Langlands, and the third to Tunnell.

In order to prove (2.1) using (2.2), we must do three things:

1. Use $\overline{\rho}_{E,3}$ to produce a σ (with a compatible trace) satisfying the hypotheses of (2.2).
2. Apply (2.2).
3. Tweak the resulting modular form to make it weight-two instead of weight-one.

We'll sketch the proof, taking one result of Deligne and Serre as a black box in part (3).

Step 1: Let $\overline{\rho}_{E,3} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_3)$ be given. Produce a representation $\sigma : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$ by composing $\overline{\rho}_{E,3}$ with the injection $\Psi : \mathrm{GL}_2(\mathbb{F}_3) \to \mathrm{GL}_2(\mathbb{Z}[\sqrt{-2}]) \subset \mathrm{GL}_2(\mathbb{C})$ defined by

$$\Psi \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}, \tag{5}$$

$$\Psi \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -\sqrt{-2} & -1+\sqrt{-2} \end{pmatrix}. \tag{6}$$

Notice that if we reduce $\mathbb{Z}[\sqrt{-2}]$ modulo $1+\sqrt{-2}$, Ψ becomes the identity map $\mathrm{GL}_2(\mathbb{F}_3) \to \mathrm{GL}_2(\mathbb{Z}[\sqrt{-2}]/(1+\sqrt{-2}) \equiv \mathbb{F}_3)$. In particular, Ψ respects traces and determinants modulo $(1+\sqrt{-2})$.

To see that $\sigma = \Psi \circ \overline{\rho}_{E,3}$ satisfies the hypotheses of Langlands-Tunnell, first observe that the "solvable image" criterion comes for free, as $\mathrm{GL}_2(\mathbb{F}_3)$ itself is solvable. (This is the reason why it is necessary to use the prime 3: $\mathrm{GL}_2(\mathbb{F}_q)$ is only solvable for $q = 2, 3$, and 2 is bad for many reasons.) Next recall that $\overline{\rho}_{E,3}$ is odd. (This follows from the fact that $\det \overline{\rho}_{E,p} \equiv \mu_p$ as Galois representations, which can be proved using a Weil pairing on $E[p]$.) Since Ψ respects determinants mod $1+\sqrt{-2}$, and $\det \sigma(\tau)$ is a priori ± 1 (where $\tau \in G_{\mathbb{Q}}$ is complex conjugation), we get that $\det \sigma(\tau) = -1$; that is, σ is odd as well.

Finally, we need to show that σ is irreducible, assuming that $\overline{\rho}_{E,3}$ is. To do this, suppose the contrary. Then, since σ is a representation in characteristic 0 factoring through a finite group, it is completely reducible, and thus it must be a sum of two characters. It follows that σ has abelian image, so $\overline{\rho}_{E,3}$ does too, as Ψ is injective. Now recall that because complex conjugation has order 2, $\overline{\rho}_{E,3}(\tau)$ must have eigenvalues ± 1, and in fact the two eigenvalues must have opposite signs by oddness. So we can write $\overline{\rho}_{E,3}(\tau) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ after some conjugation in $\mathrm{GL}_2(\mathbb{F}_3)$. Since $\overline{\rho}_{E,3}$ is irreducible over $\mathbb{F}_3$, its image contains a non-diagonal matrix. Such a matrix cannot commute with $\overline{\rho}_{E,3}(\tau)$, so we have a contradiction, and σ is indeed irreducible.

Step 2: Applying Langlands-Tunnell, we get a normalized eigencuspform $g(z) = \sum_{n=1}^{\infty} b_n e^{2\pi i n z}$ of weight 1 such that $b_q = \mathrm{tr}(\sigma(\mathrm{Frob}_q))$ for almost all primes q. Note that since Ψ preserves traces mod $1+\sqrt{-2}$, this is the same as $\mathrm{tr}(\overline{\rho}_{E,3}(\mathrm{Frob}_q))$ mod $1+\sqrt{-2}$. If g had weight 2, we would be done. But it doesn't, so we need to work some more.

Step 3: To increase the weight of our form, we will multiply it by another modular form. Specifically, we multiply it by the Eisenstein form $E(z) = 1 + 6\sum_{n=1}^{\infty} \sum_{d|n} \chi(d) e^{2\pi i n z}$, where $\chi = \chi_3$ is the Dirichlet character mod 3. This is a weight-1 modular form, albeit not cuspidal. Since the non-constant Fourier coefficients of $E(x)$ are all divisible by 3, we see that $g(z)E(z)$ is a weight-2 cuspform whose Fourier coefficients agree with $g(z)$ mod 3. Finally, a theorem of Deligne and Serre tells us that in such a situation (where $g(z)E(z)$ is congruent to the eigenform $g(z)$ mod 3), we can "deform" $g(z)E(z)$ modulo a specified prime over 3 to obtain a weight-2 normalized eigencuspform $f(z)$. By construction, the Fourier coefficients of $f(z)$ agree with those of $g(z)$ modulo some prime over 3, so we have proved Theorem 2.1.

3 Reformulation of Langlands-Tunnell

In order to prove Langlands-Tunnell, it is useful to reformulate it in a more Langlandsy way.

Theorem 3.1. *Given an irreducible representation $\sigma : W_F \to \mathrm{GL}_2(\mathbb{C})$ with solvable image in $\mathrm{PGL}_2(\mathbb{C})$, there exists a corresponding automorphic cuspidal representation $\pi(\sigma) = \otimes' \pi_v$ of $\mathrm{GL}_2(\mathbb{A}_F)$ with central character $\det \sigma$, such that $\mathrm{tr}\, t_{\pi_v} = \mathrm{tr}\, \sigma(\mathrm{Frob}_v)$ for almost all v. (Conjecturally, the solvable image assumption is unnecessary.)*

Before attempting to prove it, or even explain how it relates to the earlier statement, let's first understand the statement. The Langlands class t_{π_v} is a conjugacy class in $\mathrm{GL}_2(\mathbb{C})$ that corresponds naturally to the unramified representation π_v of $\mathrm{GL}_2(F_v)$. We have seen this before: one representative of this conjugacy class is just the diagonal matrix whose entries are the two Satake parameters of π_v. So $\operatorname{tr} t_{\pi_v}$ is the sum of the two.

Another issue: we have worked a fair amount with Weil groups of p-adic fields, but what is the Weil group of a number field? Tate's *Number theoretic background* gives a general definition for F any local or global field. A Weil group of F is a group W_F equipped with a homomorphism $W_F \to G_F$ with dense image, and equipped with an isomorphism of topological groups $W_E^{\mathrm{ab}} \cong C_E$ for all E/F finite. Here, W_E is the preimage of G_E in W_F, "ab" denotes quotienting by the closure of the commutator subgroup, and C_E denotes $E^\times$ in the local case and $\mathbb{A}_E^\times/E^\times$ in the global case. This data is required to satisfy various compatibility conditions, which turn out to determine W_F uniquely.

Here's what we'll need to know about Weil groups of number fields: for F a number field, W_F is a topological group equipped with a natural surjection onto G_F. If E is a finite Galois extension of F, then W_E is an open subgroup of W_F with quotient isomorphic to $\mathrm{Gal}(E/F)$ in the obvious way.

We'll only need a very special case of the theorem above: assume that $F = \mathbb{Q}$ and that σ factors through $G_{\mathbb{Q}}$ via an odd representation $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$. But stating the theorem as we did is useful, partly for the sake of further generalizations, but also because the proof crucially relies on inducting from one number field to another.

Why does Theorem 3.1 imply our earlier version of Langlands-Tunnell, Theorem 2.2? We need to know that an automorphic cuspidal representation of weight one gives rise to a normalized eigencuspform of weight one. In fact, this is true:

Fact 3.2. *There is a bijective correspondence between normalized newforms $f(z) \in S_1(\Gamma_0(N), \psi)$ and automorphic cuspidal representations $\pi = \otimes' \pi_p$ of weight one. For all $p \nmid N$, this correspondence identifies the Hecke eigenvalues a_p of f with the sum of the Satake parameters $\mu_1(p)+\mu_2(p)$ of π_p.*

(This generalizes, in some sense, to all weights, but we only need this version.) As a result of this correspondence, it suffices to prove the reformulated Langlands-Tunnell theorem, working with automorphic representations instead of modular forms.

4 Proof idea of Langlands-Tunnell

Believe it or not, some of the ideas of this proof were sketched in a guest lecture way back on February 11. The biggest tool here is global base change. I'll sketch the construction of $\pi(\sigma)$, but omit the proof of correctness, including some matrix calculations as well as more serious arguments involving automorphic forms, L-functions, and so on.

We're given a representation $\sigma : W_F \to \mathrm{GL}_2(\mathbb{C})$, and we want to find a "corresponding" automorphic representation $\pi(\sigma)$ of $\mathrm{GL}_2(\mathbb{A}_F)$. If E is a cyclic extension of F, then the theory of base change tells us we want $\mathrm{BC}_{E/F}(\pi(\sigma))$ to "correspond" to $\pi(\sigma|_{W_E})$. If we have already constructed $\pi(\sigma|_{W_E})$, then we can hope to construct $\pi(\sigma)$ as one of the π that base change to it. This is plausible, because we can choose E such that $\sigma|_{W_E}$ has a smaller projective image than σ itself.

As mentioned earlier, there are three cases: dihedral, tetrahedral, and octahedral images in $\mathrm{PGL}_2(\mathbb{C})$. The dihedral case was essentially done by Hecke and Maass long ago. Let's try to identify $\pi(\sigma)$ in the tetrahedral case. The projective image of σ in this case is A_4, which has a convenient normal subgroup V_4, the Klein four-group. So we can build the diagram

$$\begin{tikzcd}
1 \arrow[r] & W_E \arrow[r] \arrow[d] & W_F \arrow[r] \arrow[d] & \mathrm{Gal}(E/F) \arrow[r] \arrow[d, "\sim"] & 1 \\
1 \arrow[r] & V_4 \arrow[r] & A_4 \arrow[r] & \mathbb{Z}/3 \arrow[r] & 1
\end{tikzcd}$$

where E is whatever cubic extension of F has the correct Weil group. Now $\sigma|_{W_E}$ has dihedral ($V_4 = D_4$) projective image, so $\pi(\sigma|_{W_E})$ is already known to exist. By base change theory, there are exactly three cuspidal representations π of $\mathrm{GL}_2(\mathbb{A}_F)$, each a twist of the others. One can show that exactly one of the three has the correct central character $\det \sigma$, and this is the one we choose.

In the octahedral case, we do much the same thing. We define the extension E/F by the diagram

$$\begin{tikzcd}
1 \arrow[r] & W_E \arrow[r] \arrow[d] & W_F \arrow[r] \arrow[d] & \mathrm{Gal}(E/F) \arrow[r] \arrow[d, "\sim"] & 1 \\
1 \arrow[r] & A_4 \arrow[r] & S_4 \arrow[r] & \mathbb{Z}/2 \arrow[r] & 1
\end{tikzcd}$$

Now $\sigma|_{W_E}$ is tetrahedral, so the previous case proved the existence of $\pi(\sigma|_{W_E})$, at least modulo the actual proof. So we just need to choose $\pi(\sigma)$ from among the cuspidal automorphic representations π base changing to $\pi(\sigma|_{W_E})$. But there's a problem: there are two of these, and they have exactly the same central character! Which one do we use?

Langlands couldn't resolve this issue in generality, and this is where Tunnell made his contribution. Tunnell used a newer type of base change due to Jacquet, Piatetski-Shapiro, and Shalika, which is valid for cubic extensions that are not necessarily Galois. Roughly speaking, this allowed him to induct from D_8 to S_4 instead of from A_4.