Overview of the proof

Kenneth A. Ribet

UC Berkeley

CIRM
16 juillet 2007

Saturday: Berkeley → CDG

Sunday: CDG → MRS → Gare Saint Charles → CIRM

Monday: Jet lag

Jet lag $\Longrightarrow$ Slides

Basic setup and notation

$G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$

We deal with 2-dimensional representations of G. These representations are always continuous. They are *odd* in the sense that $\det(\rho(\text{complex conjugation})) = -1$.

The symbol "ρ" is used primarily for mod p (or mod $\ell\ldots$) representations, while "$\tilde{\rho}$" is used for p-adic representations.

My contribution to the subject: For each prime ℓ, we choose and fix one prime λ over ℓ in $\overline{\mathbf{Q}}$. We refer to λ-adic representations as "ℓ-adic representations."

Modularity

A p-adic representation $\tilde{\rho}$ is *modular* if it associated with a modular form $\sum a_n q^n$. Recall that this means that $\operatorname{tr} \tilde{\rho}(\mathrm{Frob}_\ell) = a_\ell$ for almost all primes ℓ. (There's a similar formula for the determinant.)

Warning: a mod p representation is referred to as “modular” if it is either associated with a modular form or is *reducible*.

Serre's conjecture asserts the modularity of all mod p (odd, 2-dimensional, irreducible if you prefer) representations of G.

Given an irreducible, odd $\rho : G \rightarrow \mathbf{GL}(2, \mathbf{F})$ (where F is a finite field of characteristic p), the aim is to show that ρ arises from a modular form f of level $N(\rho)$ (the prime-to-p conductor of ρ) and weight $k(\rho)$. The integer $k(\rho)$ satisfies $2 \leq k(\rho) \leq p^2 - 1$, but we can and usually do twist ρ by a power of the mod p cyclotomic character to ensure

$$2 \leq k(\rho) \leq p + 1.$$

Consider the example coming from Fermat's Last Theorem. Start with a purported non-trivial solution to the Fermat equation of exponent *p*. Construct the Frey curve *E*, and let ρ be defined by *E*[*p*].

How do we know that ρ modular?

The short answer (in most cases) is that ρ is the mod p reduction of the p-adic representation $\tilde{\rho}_p$ associated with E. The curve E furnishes a rich structure: the family $(\tilde{\rho}_\ell)$ of ℓ-adic representations associated to E. One obvious principle is that one of these representations is modular if and only if they all are.

Meanwhile, by base change (Saito–Shintani, Langlands, Tunnell) we know that ρ_3 is modular. Finally, the Modularity Lifting Theorem of Taylor–Wiles proves that $\tilde{\rho}_3$ is modular if ρ_3 is modular and irreducible.

The proof depends on having the appropriate infrastructure, namely the full package $(\tilde{\rho}_\ell)$. Once we have this package, the curve E is no longer essential.

Given ρ, we want to lift ρ to a $\tilde{\rho}_p$ and insert $\tilde{\rho}_p$ into a compatible system of ℓ-adic representations $(\tilde{\rho}_\ell)$. The system $(\tilde{\rho}_\ell)$ should display local properties that reflect the modular form f whose existence is asserted by Serre's conjecture.

Conjecturally, we have

$$\rho \overset{?}{\mapsto} f \mapsto (\tilde{\rho}_\ell).$$

A wrinkle is that there can be several types of f that give rise to the initial representation ρ:

An initial f might have weight $k = k(\rho)$ and level $N = N(\rho)$. If $k \leq p + 1$, we could find a second form g of weight 2 and level pN that gives rise to ρ.

Also, if f has character ϵ and ϵ' is the product of ϵ and a character of p-power order, then Carayol taught us that there's an f' with character ϵ' that also gives ρ.

A major achievement (Dieulefait, Khare–Wintenberger) was achieved by combining representation theory, Brauer induction and the “potential modularity” results proved by R. Taylor around 1999. The bottom line is that Khare–Wintenberger can produce custom-made $(\tilde{\rho}_\ell)$s for a given ρ that look as if they come from the forms f, g, f', ...

Richard Taylor will lecture about “potential modularity and applications” on Wednesday and Thursday.

Once we have in place some mechanism for producing systems of ℓ-adic representations, we can characterize the proof of Serre's conjecture as a 2-stage induction in which we first vary p and then N (the level).

The induction begins with the observation that we know Serre's conjecture in certain extreme situations where one proves that all irreducible ρ are modular by showing that there are no ρ at all. Specifically, Tate made this observation when $N = 1$ and $p = 2$; Serre used similar ideas for $N = 1$, $p = 3$. Other relevant results pertain to semistable abelian varieties with good reduction outside specific small sets of primes (Fontaine, Schoof, Brumer–Kramer) and to p-adic representations with implausibly small ramification.

An essential tool in the induction are the *modularity lifting theorems* of Taylor–Wiles, Skinner–Wiles, Kisin and others. We can think of these theorems as a bridge between Serre's conjecture and the Fontaine–Mazur conjecture. Here is a plausible generalization:

Modularity Lifting Conjecture (MLC): *Let* $\tilde{\rho} : G \to \mathbf{GL}(2, E)$ *be an irreducible odd, continuous p-adic representation that is ramified only at finitely many primes and is deRham (i.e., potentially semistable) at p. If the reduction* ρ *of* $\tilde{\rho}$ *is modular (or reducible), then* $\tilde{\rho}$ *is modular (perhaps up to a twist by an integral power of the p-adic cyclotomic character?).*

MLC $\implies$ Serre's conjecture

First assume that ρ is given with $N(\rho) = 1$. Because ρ is irreducible, we have $p \geq 5$ (Tate, Serre). We can and might as well assume that we have $k = k(\rho) \leq p + 1$. Build a system $(\tilde{\rho}_\ell)$ where each individual $\tilde{\rho}_\ell$ is ramified only at ℓ (and is furthermore crystalline at ℓ with Hodge–Tate weights 0, $k - 1$). The reduction ρ_3 of $\tilde{\rho}_3$ is reducible by Serre; thus, $\tilde{\rho}_3$ is modular (visibly of level 1) by MLC. It follows that $\tilde{\rho}_p$ is modular and then finally that $\rho = \rho_p$ is modular.

It appears that there is no level-lowering going on. When we learn that ρ is modular, we discover more precisely that it is modular of level 1. The level-lowering is now done by the deformation-theory. The idea that this might be possible perhaps originated with Nigel Boston.

Having “done” the case $N = 1$, we contemplate the case where N is greater than 1. Choose a prime q dividing N; note $q \neq p$. Starting with ρ, make $(\tilde{\rho}_\ell)$ so that it appears to come from a form f of level N and weight k. The mod q representation ρ_q has some level N' dividing the prime-to-q part of N. By induction, ρ_q is modular of level N', and hence $\tilde{\rho}_q$ is modular. It follows again that $\tilde{\rho}_p$ is modular and then that $\rho = \rho_p$ is modular.

MLC $\neq$ MLT

The modularity lifting theorems that are now available are nowhere near as strong as the conjecture that we used in the previous arguments. A typical restriction occurs on the weight. If $\tilde{\rho}$ is a *p*-adic representation that appears to come from a form of weight *k* and level prime to *p*, then at present we are unable to deduce the modularity of $\tilde{\rho}$ from the modularity of its reduction ρ without assuming either that $k \leq p + 1$ or $\tilde{\rho}$ is "ordinary."

Following Khare–Wintenberger again, we will pretend that modularity can be established under the modified hypothesis $k \leq 2p$ (which is less restrictive than $k \leq p+1$) and deduce Serre's conjecture when $N = 1$:

We do an induction on the set of prime numbers! Suppose that Serre's conjecture has been established for level-1 mod p representations, and let P be the next prime after p. Let ρ be a mod P representation with $N(\rho) = 1$. After twisting ρ by a power of the mod P cyclotomic character we can and will assume that $k = k(\rho)$ is at most $P + 1$. By Bertrand's postulate (!), $k \leq 2p$. Construct a family $(\tilde{\rho}_\ell)$ so that $\rho_P = \rho$ and so that the family appears to come from a weight-k cusp form on $\mathbf{SL}(2, \mathbf{Z})$. By induction, ρ_p is modular. By the modularity lifting theorem that we are admitting, $\tilde{\rho}_p$ is modular. The argument finishes as before: $\tilde{\rho}_P$ is modular, so $\rho = \rho_P$ is modular.

Now let's get more precise about the theorems that one can prove. For simplicity, I assume $p > 2$. (There's also a statement for $p = 2$.) We consider a p-adic representation $\tilde{\rho}$, making the standard assumptions that it is odd, continuous, irreducible and ramified only at finitely many primes. Let ρ be its reduction; we want a result of the sort

$$\rho \text{ modular } \implies \tilde{\rho} \text{ modular.}$$

There are really two distinct cases, the case where ρ is "modular" because it is reducible and the case where ρ is irreducible and comes from a modular form.

Here are a few remarks about the first case: One wishes to apply Skinner–Wiles in this case; for this, one needs to know that $\tilde{\rho}$ is ordinary. Fortunately, one does tend to know in this case that $\tilde{\rho}$ is ordinary: First, by a theorem of Berger–Li–Zhu, it is typically true that $\tilde{\rho}$ is ordinary if ρ is ordinary. Second, if ρ is non-ordinary and not of a weird weight, its restriction to the decomposition group for *p* in *G* is already irreducible.

Pretend that the reducible case can be relegated to the background. Thus, ρ will be assumed to be irreducible. Unfortunately, the lifting theorem that I will quote requires the more stringent assumption that ρ remain absolutely irreducible even after restriction to $H := \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\mu_p))$. The possibility that ρ might be irreducible but somehow have small image is a technical pain that engenders many parenthetical comments in Khare–Wintenberger.

The following theorem is Theorem 4.1 in the first part of Khare–Wintenberger, “Serre’s modularity conjecture”: *Assume that ρ is modular and that the restriction of ρ to H is absolutely irreducible. Suppose that, locally at p, the representation $\tilde{\rho}$ is either crystalline of weight k with $2 \leq k \leq p+1$ or “potentially semistable of weight 2.” Then $\tilde{\rho}$ is modular.*

For the remainder of this talk, I propose to speak in some more detail about the “level 1” case where ρ is ramified only at *p*.

A first remark is that you need only prove the level-1 Serre conjecture for an infinite set of primes. If you know the conjecture for a prime, you know it automatically for all smaller primes. Indeed, suppose you want to prove the modularity of a mod p representation ρ and that you know the conjecture for mod P representations where P is some prime bigger than (or equal to) ρ. By twisting, we can assume as usual that ρ has weight $k \leq p + 1 \leq P + 1$. Build the system $(\tilde{\rho}_\ell)$ so that it appears to come from a weight-k form on $\mathbf{SL}(2, \mathbf{Z})$. The representation $\tilde{\rho}_P$ is modular because its reduction ρ_P is known to be modular and because the modularity lifting theorem applies to $\tilde{\rho}_P$. One concludes as usual.

In fact, this argument proves is that Serre's conjecture mod P implies Serre's conjecture for all representations ρ (mod arbitrary primes) whose weights $k(\rho)$ are at most $P+1$.

Now adopt the optic that we want to prove the conjecture modulo an infinite set of primes. Assume that we know the conjecture mod p, can we prove the conjecture mod some prime $P > p$? We will in fact look for primes P with $p < P < 2p$, and for each P, we will examine odd prime divisors q of $P - 1$.

In other words, we look at pairs (P, q) with $p < P < 2p$ and $q | (P - 1)$ with q odd.

If P is a Fermat prime, there is no q for P. For example, if $p = 3$, there are simply no pairs (P, q). For this reason (among others), K–W need to find alternative arguments to treat the case $P = 5$; in fact, they provide specialized arguments for $k = 2$, $k = 4$ and $k = 6$, which are the only relevant weights. For $p \geq 5$, there will always be a pair (P, q) for which what I am about to describe actually works.

Here's the argument. Take $P > p$ and $q|(P-1)$. Let ρ be a level 1 Serre-type mod P representation. Twist it by powers of the cyclotomic character so that its weight is as low as possible. After replacing ρ by the twist, we have $2 \leq k(\rho) \leq P+1$. If $k \leq p+1$, we are done. It will be the case that P is less than $2p$, so one might say that we have at least a 50% chance of being done already. In fact, using the definition of the Serre weight, one can show that we will be done already if ρ is supersingular.

Suppose now that we are in the unfavorable case

$$p+1 < k \leq P+1 \leq 2p, \quad k = k(\rho).$$

If ρ comes from a weight-k form on $\mathbf{SL}(2,\mathbf{Z})$, then it also comes from a weight-2 form f on $\Gamma_1(P)$ with character $\epsilon = \omega^{k-2}$. Here, ω is the mod P Teichmüller character (of order $P-1$) with values in $\overline{\mathbf{Q}}^* \subseteq \overline{\mathbf{Q}}_P^*$. While, for the moment, we can't produce f, we produce a system $(\tilde{\rho}_\ell)$ that looks as if it came from f. In particular, the reduction ρ_P of $\tilde{\rho}_P$ will be (isomorphic to) ρ.

Now we bring in the divisor q of $P-1$. Suppose that in fact q^r divides $P-1$; we might as well take r maximal. The character $\theta = \omega^{(P-1)/q^r}$ has q-power order and so is the identity modulo q. For each i mod $(P-1)/q^r$, $\epsilon\theta^i$ is congruent to ϵ mod q. If there were a form f, for each i there would be a form f'_i with character $\epsilon\theta^i$ that agrees with f mod q (Carayol). Fix i (in addition to P and q) and write f' for f'_i.

We can't build f', but we can build a system $(\tilde{\rho}'_\ell)$ that looks as if it came from f'. The congruence between f and f' translates as an isomorphism $\rho_q \approx \rho'_q$. The representation ρ'_P has no direct contact with $\rho = \rho_P$.

The lack of dependence of ρ'_P on ρ means that ρ'_P starts out with a clean slate with respect to the question of whether some twist of this representation has a Serre weight $\leq p+1$. K–W observe in fact that this question will have a positive answer if certain inequalities involving p, P, q, r, and i are satisfied. Moreover, they show, given $p \geq 5$, that P, q, r, and i can be chosen so that the inequalities *are* satisfied. In this case, ρ'_P is modular.

Using a weight-2 modularity lifting theorem, they deduce that $\tilde{\rho}'_P$ is modular and then that ρ_q is modular. Using the weight-2 theorem again, they deduce that $\tilde{\rho}_q$ is modular, and we then get, as required, that ρ_P is modular.