Recent work on Serre's conjectures

Kenneth A. Ribet
University of California, Berkeley

June 4, 2005

Canadian Math Society Summer Meeting

This talk concerns a new chapter in a story that began in the late 1960s.

For a copy of these slides, consult `http://math.berkeley.edu/~ribet/cms.pdf`.

In his 1967–1968 DPP seminar on modular forms ("Une interprétation des congruences relatives à la a fonction τ de Ramanujan"), J-P. Serre proposed the possibility of linking Galois representations to holomorphic modular forms that are eigenforms for Hecke operators.

Soon after, P. Deligne constructed the representations whose existence was conjectured by Serre.

For example, let

$$\Delta = q\prod_{n=1}^{\infty}(1-q^n)^{24}, \qquad q = e^{2\pi iz}$$

and write Δ as a sum $\sum_{n=1}^{\infty}\tau(n)q^n$ with $\tau(n) \in \mathbf{Z}$. For each prime p, there is a continuous representation

$$\rho_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_p)$$

whose arithmetic is tied up with that of the $\tau(n)$.

The representation ρ_p is unramified at all primes different from p. If ℓ is such a prime and Frob_ℓ is a Frobenius element for ℓ in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then the matrix $\rho_p(\mathrm{Frob}_\ell)$ has trace $\tau(\ell)$ mod p and determinant ℓ^{11} mod p.

These constraints determine ρ_p up to isomorphism once we agree to replace ρ_p by its semisimplification in the rare situation where it is not already simple.

Serre and Swinnerton-Dyer studied the ρ_p for Δ and some other modular forms in the early 1970s. They showed that the numbers $\tau(n)$ satisfy no congruences other than those that had been established by Ramanujan and others in the early part of the 20th century.

Around 1975, I extended their study to modular forms on the full modular group $\mathbf{SL}(2, \mathbf{Z})$ whose coefficients were not necessarily *rational* integers.

Meanwhile, Serre asked whether there might be a converse to Deligne's construction. Suppose that $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is a continuous homomorphism, where $\mathbf{F}$ is a finite field. To fix ideas, suppose that ρ is simple (i.e., that there is no basis in which the image of ρ is upper-triangular). Assume also that ρ is unramified at all primes other than p, the characteristic of $\mathbf{F}$. Might we dare guess that ρ is isomorphic to a representation associated with a form on $\mathbf{SL}(2, \mathbf{Z})$?

There is a salient necessary condition involving parity. Each non-zero form on $\mathbf{SL}(2,\mathbf{Z})$ has a unique weight that describes the behavior of the form under fractional linear transformations; for example, the weight of Δ is 12. Since forms are trivially invariant under $z \mapsto \frac{-z}{-1} = z$, it turns out that the weight of a form on $\mathbf{SL}(2,\mathbf{Z})$ is always even.

If ρ is a mod p representation associated to a form of weight k, the the determinant of $\rho(\mathrm{Frob}_\ell)$ is ℓ^{k-1} mod p; this is an odd power of ℓ mod p.

It follows from the Cebotarev density theorem that $\det \rho = \chi_p^{k-1}$, where χ_p is the mod p cyclotomic character: the homomorphism $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_p^*$ giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the pth roots of unity in $\overline{\mathbf{Q}}$.

What's important is that $\det \rho$ is always odd—it's an odd power of the mod p cyclotomic character in fact.

Serre's conjecture for modular forms of level 1 (i.e., those on $\mathbf{SL}(2,\mathbf{Z})$) states: *if a continuous odd irreducible representation* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2,\mathbf{F})$ *is unramified outside* p*, then it arises from a modular form on* $\mathbf{SL}(2,\mathbf{Z})$*.*

This conjecture was advanced as a question to Swinnerton-Dyer in 1972 and was discussed, for instance, in Serre's article for the Journées Arithmétiques de Bordeaux ("Valeurs propres des opérateurs de Hecke modulo l").

Using discriminant bounds, Tate proved the conjecture for mod 2 representations. Later, Serre proved the conjecture for $p = 3$ in a similar manner. In 1999, S. Brueggeman treated the case $p = 5$ modulo the generalized Riemann hypothesis.

For those values of p, there simply are no representations ρ of the type contemplated by the conjecture! As Serre explained in Bordeaux, the conjecture predicts in fact that there are no irreducible ρ as in the conjecture for $p < 11$.

Serre's 1987 article "Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$" conjectured the modularity of odd continuous irreducible representations $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ that are permitted to be ramified outside p.

The level and weight of the form associated to ρ are predicted from local behavior: The level depends on the restriction of ρ to inertia subgroups of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ for primes $\neq p$; the weight depends on the restriction to the inertia subgroup at p.

My "level-lowering" theorem, plus work of Edixhoven, Carayol, Diamond, Buzzard and others, shows that if a representation comes from some modular form, it actually comes from a form of the predicted weight and level. (If $p = 2$, we still need to make a technical hypothesis.)

See "Lectures on Serre's conjectures" by Ribet and W. A. Stein for details. This is the work that reduced Fermat's Last Theorem to the modularity theorem for (semistable) elliptic curves.

If a continuous odd irreducible representation $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ *is unramified outside* p*, then it arises from a modular form on* $\mathbf{SL}(2, \mathbf{Z})$*.*

This statement is proved in a current preprint of Chandrashekhar Khare that builds on an earlier manuscript of Khare and Jean-Pierre Wintenberger.

Some related work has been done by Luis Dieulefait. In particular, Dieulefait treated the special case of forms of level 1 and weight 2.

arXiv Citations

• Khare and Jean-Pierre Wintenberger, "On Serre's reciprocity conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$," December 3, 2004

• Luis Dieulefait, "The level 1 weight 2 case of Serre's conjecture," December 5 and 8, 2004

• Chandrashekhar Khare, "On Serre's modularity conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ unramified outside p," April 5, 2005

Review of FLT

Start with $a^p + b^p = c^p$ with $p \geq 5$. Examine the mod p representation $\rho_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_p)$ that gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the p-division points on $E : y^2 = x(x - a^p)(x + b^p)$.

This representation is odd and it is irreducible (Mazur). According to Serre, ρ_p comes from a space of modular forms that happens to be 0. By level-lowering, once ρ_p can be associated to some modular form, we can conclude with a contradiction.

Thus we need "only" show that ρ_p is modular. The main theme here is that E gives rise to a family of ℓ-adic Galois representations $\tilde{\rho}_\ell$, one for each prime ℓ. Each $\tilde{\rho}_\ell$ has a reduction ρ_ℓ; this is the mod ℓ representation attached to E.

Because the $\tilde{\rho}_\ell$ form a compatible system of representations, all $\tilde{\rho}_\ell$ are modular as soon as a single $\tilde{\rho}_\ell$ is modular. In particular, if one $\tilde{\rho}_\ell$ is modular, then ρ_p is modular, and we are done.

The powerful *Modularity lifting theorems* of Wiles, Taylor, Breuil, Diamond, Conrad, Fujiwara, Kisin, Savitt, Skinner and others (list copied from Khare's paper) tend to show that $\tilde{\rho}_\ell$ is modular if its reduction ρ_ℓ is modular. (There are hypotheses to check. In particular, $\tilde{\rho}_\ell$ must look locally as if it is modular.)

Thus ρ_p is modular if it has a lifting $\tilde{\rho}_p$ that fits in a family $(\tilde{\rho}_\ell)$ such that one of the reductions ρ_ℓ is modular.

In Wiles's case, ρ_3 was known be modular because its image is contained in a solvable $\mathbf{GL}(2)$. (Langlands and Tunnell had essentially proved the modularity of two-dimensional Galois representations with solvable image using Langlands's base-change machine that he constructed after seeing the work of Saito and Shintani.)

Amazingly, one proves the modularity of ρ_p (and thus FLT) from the modularity of ρ_3.

A key step in the work of Khare–Wintenberger and Khare is to find $(\tilde{\rho}_\ell)$ starting from ρ_p.

This is not easy. As a first step, after Serre's conjecture was formulated, mathematicians attempted to establish the corollary of the conjecture that every mod p Galois representation ρ_p has a p-adic lift with the "same" arithmetic properties as ρ_p.

In the late 1990s, R. Ramakrishna lifted $\rho_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ to $\tilde{\rho}_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, W)$ where W is the ring of Witt vectors of $\mathbf{F}$, but only at the expense of allowing $\tilde{\rho}_p$ to be ramified at some primes where ρ_p is unramified.

In their recent work, K–W use results of Böckle, Taylor and others to find a $\tilde{\rho}_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathcal{O})$ with "minimal" ramification properties. Here $\mathcal{O}$ is the ring of integers of a (possibly ramified!) finite extension of the field of fractions of W.

The next step is to insert $\tilde{\rho}_p$ into a family $(\tilde{\rho}_\ell)$ of representations that are compatible with $\tilde{\rho}_p$. D and K–W use powerful theorems in articles that R. Taylor wrote around 1999 ("Remarks on a conjecture of Fointaine and Mazur," "On the meromorphic continuation of degree two L-functions"). A sample result is that there are totally real number fields F so that ρ_p becomes modular after restriction to $\mathrm{Gal}(\overline{\mathbf{Q}}/F)$. One can select F so that p is unramified in $F/\mathbf{Q}$, so that the group $\rho_p(\mathrm{Gal}(\overline{\mathbf{Q}}/F))$ is not too different from the image of ρ_p, and so on.

A guiding philosophy (due to Nigel Boston, I think) is that one should be able to manipulate deformations of mod p Galois representations in a way that mimics mod p congruences among modular forms. K–W establish a number of results in this direction. For example, if ρ_p is unramified outside p and has Serre weight $p+1$, they can choose the family $(\tilde{\rho}_\ell)$ so it seems to come from a modular form of weight 2 on the subgroup $\Gamma_0(p)$ of $\mathbf{SL}(2,\mathbf{Z})$. Here one thinks of Serre's theorem relating mod p forms of weight $p+1$ with forms of weight 2 on $\Gamma_0(p)$.

A technical comment

About ps and ℓs: The representations $\tilde{\rho}_\ell$ are actually indexed by prime ideals of the ring of integers of a number field, so we need to reduce mod $\wp$ or mod λ instead of mod p or mod ℓ. It is helpful to regard number fields inside of a fixed $\overline{\mathbf{Q}}$ and to choose once and for all a place of $\overline{\mathbf{Q}}$ over each rational prime (=prime number). With fixed choices in place, we can return to the mod p and mod ℓ terminology.

Khare's proof of Serre's conjecture for representations that are unramified outside p is an elaborate induction on p.

If we know the conjecture for a given prime p, then we know it for representations mod all primes q whenever the weights of the representations are $\leq p+1$. Indeed, if we have a representation ρ_q mod q of small weight, we lift and get a family $(\tilde{\rho}_\ell)$. The representation $\tilde{\rho}_p$ is modular because ρ_p is modular and because of the modularity lifting theorems. (We can apply them because the weight for $\tilde{\rho}_p$ is $\leq p+1$.) Hence $\tilde{\rho}_q$ is modular, so ρ_q is modular.

A consequence is that the conjecture for p implies the conjecture for all smaller primes. If q is less than p, the conjecture for q can be verified by checking weights $\leq q+1$, but $q+1 \leq p+1$.

The idea of the induction is to start knowing the conjecture for a given prime p (and hence for all smaller primes) and to deduce the conjecture for some prime number P larger than p. Khare's arguments enable him to do this only when P is not a Fermat prime (which it isn't likely to be!) and when P is $\leq$ a certain bound which is roughly $2p$. To know there is a P of the right size is to find estimates like those proving Bertrand's postulate. These are due essentially to Chebyshev.

Suppose that we know the conjecture mod p and want to deduce it mod P, where (roughly)

$$p < P < 2p.$$

Take a mod P representation ρ_P. We can assume that its weight is $\leq P + 1$. If its weight is $\leq p + 1$, we already have what we want. Hence we imagine that its weight k satisfies $p + 1 < k \leq P + 2$.

The Chebyshev estimates don't kick in until p is large enough. Khare verifies things more or less prime by prime until $p = 31$.

Suppose for example that we know the conjecture mod $p = 5$ and want to prove it mod $P = 7$. The only even k satisfying $p + 1 < k \leq P + 2$ is $k = 8$. Take a $\rho = \rho_7 : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$, where $\mathbf{F}$ has characteristic 7. This representation is assumed to be ramified only at 7 and to have Serre weight 8. We want to prove that it is modular.

If ρ comes from a form of weight 8 on $\mathbf{SL}(2, \mathbf{Z})$, then it comes from a form of weight 2 on $\Gamma_0(7)$. Hence we expect that ρ is the reduction of a 7-adic representation $\tilde{\rho}_\lambda$ in a system $(\tilde{\rho}_\lambda)$ with the local properties that would correspond to a weight-2 form on $\Gamma_0(7)$. The work of Khare–Wintenberger builds such a system (as we noted before).

We consider an odd prime dividing $P - 1 = 6$; let's choose 3. (In general, there will be such a prime because P is not a Fermat prime.)

Take a λ of residue characteristic 3 and let ρ_3 be the reduction of $\tilde{\rho}_\lambda$. This is a mod 3 Galois representation that may be ramified at 3 and at 7; it is not known to be modular by our induction hypothesis because of the likely ramification at 7.

Reminder: it should correspond to a weight-2 form with trivial Nebentypus character on $\Gamma_1(7)$ (i.e., to a form on $\Gamma_0(7)$).

Guided by the work on Carayol, we expect that there should be a weight-2 modular form with cubic Nebentypus character on $\Gamma_1(7)$ that gives the same mod 3 representation ρ_3 as the weight-2 form on $\Gamma_0(7)$. K–W build a system of representations $(\tilde{\rho}'_\lambda)$ that look locally as if they come from a form with a cubic character.

The link between $(\tilde{\rho}'_\lambda)$ and $(\tilde{\rho}_\lambda)$ is that they share a common mod 3 reduction.

It turns out that ρ_7' has weight smaller than 8: the weight will be 4 or 6. By induction, ρ_7' is modular. By the lifting theorems, the system $(\tilde{\rho}_\lambda')$ is modular. Hence $\rho_3' = \rho_3$ is modular. Invoking lifting theorems again, we see that the system $(\tilde{\rho}_\lambda)$ is modular. Hence $\rho = \rho_7$ is modular.

The passage from 5 to 7 gives one a very good idea of how to move from p to P. I have oversimplified things because one needs to separate out cases in which mod p representations have reducible image and when representations are less ramified than expected.

For a copy of these slides, consult `http://math.berkeley.edu/~ribet/cms.pdf`.