Mod p Galois representations attached to modular forms

Ken Ribet
UC Berkeley

April 7, 2006

After Serre's article on elliptic curves was written in the early 1970s, his techniques were generalized and extended in different directions. In particular, Serre and Swinnerton-Dyer began to study the mod ℓ representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ attached to the cusp form Δ of weight 12 on $\mathbf{SL}(2, \mathbf{Z})$.

These representations weren't always with us: It was only in his 1967–1968 DPP seminar on modular forms ("Une interprétation des congruences relatives à la a fonction τ de Ramanujan") that Serre proposed the possibility of linking Galois representations to holomorphic modular forms that are eigenforms for Hecke operators. Almost immediately afterwards, P. Deligne constructed the representations whose existence was conjectured by Serre.

If you want the full ℓ-adic representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ that are associated to a modular form, then you need to understand Deligne's construction. A forthcoming book by Brian Conrad is recommended. ["My book on Galois representations and modular forms is still undergoing revisions. (It is now shorter than it was before, with much better proofs; if you have an earlier version, please burn it.) The following link has been disabled."] For mod ℓ representations, it's enough to look in $J_1(Nl)[l]$.

Recall that

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

For each prime ℓ, let $\rho_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\ell)$ be the mod ℓ representation associated with Δ. Then ρ_ℓ is unramified at all primes different from ℓ. If p is such a prime and Frob_p is a Frobenius element for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then the matrix $\rho_\ell(\mathrm{Frob}_p)$ has trace $\tau(p)$ mod ℓ and determinant p^{11} mod ℓ.

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

Serre and Swinnerton-Dyer proved that the image G_ℓ of ρ_ℓ is "as large as possible" except for an explicit list of prime numbers ℓ, namely 2, 3, 5, 7, 23 and 691.

Because the determinant of ρ_ℓ is the 11th power of the mod ℓ cyclotomic character χ_ℓ, we have

$$G_\ell \subseteq A_\ell := \{\, M \in \mathbf{GL}(2, \mathbf{F}_\ell) \mid \det M \in \mathbf{F}_\ell^{*11} \,\}.$$

We say that G_ℓ is *as large as possible* if $G_\ell = A_\ell$. An equivalent condition is that G_ℓ contains $\mathbf{SL}(2, \mathbf{F}_\ell)$. For this, there are two necessary and sufficient conditions: (1) irreducibility of ρ_ℓ; (2) divisibility by ℓ of $|G_\ell|$.

The group G_ℓ is reducible when $\ell = 2, 3, 5, 7, 619$ and irreducible but of order prime to ℓ when $\ell = 23$.

Note that Δ is of weight $k = 12$. Looking at the list of exceptional primes, we observe that 2, 3, 5, and 7 are small primes (they're less than the weight) and that $23 = 2k - 1$. The prime 691 has become "famous" as the first numerator > 1 of a Bernoulli number: if $\dfrac{x}{e^x - 1} = \sum B_n \dfrac{x^n}{n!}$, then $B_{12} = -\frac{691}{2 \cdot 3 \cdot 5 \cdot 7 \cdot 13}$.

One can replace Δ by a normalized cuspidal eigenform of weight k, level N and coefficients that are not necessarily rational integers. If $f = \sum a_n q^n$ is such a form, then $f|T_n = a_n f$ for $n \geq 1$, and the a_n are algebraic integers. Moreover the field $E = \mathbf{Q}(\ldots, a_n, \ldots)$ has finite degree over $\mathbf{Q}$; it is either a totally real or a CM field.

While preparing this lecture, I decided that it would already be interesting to stick with $N = 1$, so the form f will be on $\mathbf{SL}(2, \mathbf{Z})$.

Here, the only known cases where $E = \mathbf{Q}$, i.e., $a_n \in \mathbf{Z}$, are the analogues of Δ for weights $16, 18, 20, 22$ and 26; these were treated by Serre and Swinnerton-Dyer. They uncovered the example of the mod 59 representation associated with the form of weight 16: the projective image of G_{59} is isomorphic to the exceptional group $\mathbf{S}_4$.

When the weight is 24, $E = \mathbf{Q}(\sqrt{144169})$, as some of us recalled on Wednesday while walking in Custer National Forest.

Suppose now that f is a normalized weight-k cuspidal eigenform on $\mathbf{SL}(2,\mathbf{Z})$ and let $\mathcal{O} = \mathbf{Z}[\ldots, a_n, \ldots] \subseteq \mathbf{C}$ be the ring generated by the coefficients of f. Then $\mathcal{O}$ is a subring of finite index in the ring of integers of E. The analogue of the family (ρ_ℓ) is a set of representations $\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\lambda)$, one for each maximal ideal λ of $\mathcal{O}$. Here we have written $\mathbf{F}_\lambda = \mathcal{O}/\lambda$. The determinant of ρ_λ is the $(k-1)$st power of the mod ℓ cyclotomic character where ℓ is the characteristic of $\mathbf{F}_\lambda$. For $p \neq \ell$, the trace of $\rho_\lambda(\mathrm{Frob}_p)$ is a_p mod λ.

Let G_λ be the image of ρ_λ. In 1975, I proved that G_λ contains $\mathbf{SL}(2, \mathbf{F}_\lambda)$ for all but finitely many λ. Note that G_λ is a subgroup of

$$A_\lambda := \{\, M \in \mathbf{GL}(2, \mathbf{F}_\lambda) \mid \det(M) \in \mathbf{F}_\ell^{*k-1} \,\}$$

that maps via $\det$ onto $\mathbf{F}_\ell^{*k-1}$. Hence $G_\lambda = A_\lambda$ if and only if G_λ contains $\mathbf{SL}(2, \mathbf{F}_\lambda)$.

Today's theme is to revisit the proof from a reasonably modern point of view while tacitly contemplating questions of effectivity.

First consider the irreducibility of ρ_λ. Because of the "almost all" situation, we can and will suppose that $k \leq \ell + 1$. Assume that ρ_λ is reducible, say of the form $\begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$, where α and β are characters $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\lambda^*$. Because α and β are unramified outside ℓ, each of these characters is a power of the mod ℓ cylotomic character, say $\alpha = \chi_\ell^n$, $\beta = \chi_\ell^m$. The integers n and m are determined by the restriction of ρ_λ to the inertia group for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

When f is *supersingular* at λ in the sense that we have $a_\ell \equiv 0 \bmod \lambda$, a theorem of Fontaine asserts, in particular, the irreducibility of the restriction of ρ_λ to the decomposition group for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. In particular, of course, ρ_λ is (globally) irreducible. Hence the reducible case is necessarily ordinary, i.e., non-supersingular.

For the theorem of Fontaine, and the next theorem of Deligne, see B. Edixhoven, "The weight in Serre's conjectures on modular forms."

The theorem of Deligne in question is the one that tells us (in the ordinary case) that the restriction of ρ_λ to the decomposition group for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ has an unramified 1-dimensional quotient. It follows that α and β are (perhaps up to permutation) the trivial character and the $(k-1)$st power of the cyclotomic character. This gives the congruence $a_p \equiv 1 + p^{k-1}$ mod λ for all primes $p \neq \ell$.

Finally, an argument of Serre and Swinnerton-Dyer shows that the numerator of B_k is divisible by ℓ: one compares f with the weight-k Eisenstein series on $\mathbf{SL}(2, \mathbf{Z})$, whose pth Fourier coefficient is $1 + p^{k-1}$ and whose constant term is $-\frac{B_k}{2k}$.

Thus ρ_λ is irreducible for ℓ large and we know explicitly what "large" means in this case.

If we had forms of level $N > 1$, we would still prove that ρ_λ is irreducible for ℓ large, but the estimate for "large" would be less tight.

Next, think what happens if G_λ is irreducible but of order prime to ℓ. If the projective image of G_λ is cyclic, then G_λ is contained in a Cartan subgroup of $\mathbf{GL}(2, \mathbf{F}_\lambda)$. This implies that ρ_λ becomes reducible if we replace $\mathbf{F}_\lambda$ by a finite extension of $\mathbf{F}_\lambda$, but the argument that we have given rules out this case except when ℓ is small or a divisor of B_k. Hence the projective image is either dihedral or one of the three exceptional groups $\mathbf{S}_4$, $\mathbf{A}_4$, $\mathbf{A}_5$.

The theorems of Deligne and Fontaine show that the projective image of inertia at p has order either $\frac{\ell-1}{\gcd(k-1,\ell-1)}$ or $\frac{\ell+1}{\gcd(k-1,\ell+1)}$. When ℓ is large relative to k, these quotients exceed the orders all elements of the exceptional groups. Hence these groups cannot occur as projective images for large ℓ.

Suppose that the projective group $D = \overline{G}_\lambda$ is dihedral, so that it has a cyclic subgroup C of index 2. The group D/C is then the Galois group of a quadratic extension of $\mathbf{Q}$. Because ρ_λ is ramified only at ℓ, this extension is ramified only at ℓ. Thus it *is* ramified at ℓ, which implies that the projective image of the inertia subgroup for ℓ in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is not contained in C. It follows that this image has order 2: the elements of D not in C all have order 2. The argument we have just given (with $\ell \pm 1$) rules out this case for large ℓ.

Let's focus in now on the case where G_λ is irreducible and of order divisible by ℓ. A theorem of L. E. Dickson implies that $G_\lambda \subseteq \mathbf{GL}(2, \mathbf{F}_\lambda)$ contains (after a possible change of basis) the group $\mathbf{SL}(2, F)$ for some subfield F of $\mathbf{F}_\lambda$.

Because ℓ is bigger than k, a short argument shows that $\mathbf{F}_\lambda$ is generated by the images of the a_p with $p \neq \ell$. Thus $\mathbf{F}_\lambda$ is generated by the traces of the elements of G_λ. This suggests that G_λ contains $\mathbf{SL}(2, \mathbf{F}_\lambda)$. We aim to establish this fact.

Fixing λ, we will simplify notation and write $\mathbf{F}$ for $\mathbf{F}_\lambda$, ρ for ρ_λ, G for G_λ. Note that G is a subgroup of

$$A := \{\, M \in \mathbf{GL}(2, \mathbf{F}) \mid \det(M) \in \mathbf{F}_\ell^{*k-1} \,\}$$

and that G maps onto $\mathbf{F}_\ell^{*k-1}$ via the determinant.

We seek to show that $G = A$. This statement is true if and only if G contains $\mathbf{SL}(2, \mathbf{F})$.

Let $\overline{G}$ be the image of G in $\mathbf{PGL}(2, \mathbf{F})$. Let $\overline{\mathbf{F}}$ be an algebraic closure of $\mathbf{F}$ and let K be a subfield of $\overline{\mathbf{F}}$. *I claim that $\overline{G}$ is contained in $\mathbf{PGL}(2, K)$ if and only if G is contained in $\mathbf{GL}(2, K)$.* One direction is obvious. For the converse, we assume that $G \subseteq \overline{\mathbf{F}}^* \cdot \mathbf{GL}(2, K)$ and seek to show that G lies in $\mathbf{GL}(2, K)$.

The composite map

$$\alpha : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to G \to \overline{\mathbf{F}}^* \cdot \mathbf{GL}(2,K)/\,\mathbf{GL}(2,K)$$

may be viewed as taking values in $\overline{\mathbf{F}}^*/K^*$, a torsion abelian group whose elements have order prime to ℓ. It is unramified outside ℓ. We want to that α is trivial. For this, it suffices to show that $\alpha(\sigma) = 1$ for some $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ whose image in $\mathrm{Gal}(\mathbf{Q}(\mu_\ell)/\mathbf{Q})$ is a generator of this cyclic group, i.e., for some σ that is mapped to a generator of $\mathbf{F}_\ell^*$ by the mod ℓ cyclotomic character.

It is elementary that $\alpha(\sigma) = 1$ for all $\sigma \in \mathrm{Gal}$ such that $\mathrm{tr}(\rho(\sigma))$ is a non-zero element of K; in particular, $\alpha(\sigma) = 1$ if $\mathrm{tr}(\rho(\sigma))$ happens to be a non-zero element of $\mathbf{F}_\ell$.

In the non-supersingular case, let σ be an element of the inertia group for p such that $t = \chi(\sigma)$ generates $\mathrm{Gal}(\mathbf{Q}(\mu_\ell)/\mathbf{Q}) = \mathbf{F}_\ell^*$. By Deligne's theorem, $\mathrm{tr}(\rho(\sigma)) = 1 + t^{k-1}$. When ℓ is large relative to k, this quantity is non-zero, as required.

The supersingular case is quite similar.

To summarize: (1) $G \subseteq \mathbf{GL}(2, \mathbf{F})$; (2) $\mathrm{tr}(G)$ generates $\mathbf{F}$ over $\mathbf{F}_\ell$; (3) G contains $\mathbf{SL}(2, F)$ for some subfield F of $\mathbf{F}$. Further, G has the property in the claim: for $K \subseteq \overline{\mathbf{F}}$, G is contained in $\mathbf{PGL}(2, K)$ if and only if G is contained in $\mathbf{GL}(2, K)$.

Assuming that ℓ is at least 5, we will deduce that G contains $\mathbf{SL}(2, \mathbf{F})$, as desired, by invoking more results of Dickson.

Let L be a finite extension of $\mathbf{F}$ inside $\overline{\mathbf{F}}$ for which $[L : \mathbf{F}_\ell]$ is even. Then $\overline{G} \subseteq \mathbf{PSL}(2, L)$. By Dickson's theory, we have (after a change of basis) either $\overline{G} = \mathbf{PSL}(2, K)$ or $\overline{G} = \mathbf{PGL}(2, K)$, for some subfield K of L. In both cases, $\overline{G} \subseteq \mathbf{PGL}(2, K)$. By the claim, $G \subseteq \mathbf{GL}(2, K)$.

Because $\mathbf{F}$ is generated by $\mathrm{tr}(G)$, we get $\mathbf{F} \subseteq K$. Since $\overline{G}$ contains $\mathbf{PSL}(2, K)$, $\overline{\mathbf{F}}^* \cdot G$ contains $\mathbf{SL}(2, K)$. Taking commutators, we get that the commutator subgroup of G contains $\mathbf{SL}(2, K)$ and hence $\mathbf{SL}(2, \mathbf{F})$. Comparing with the original set-up in which G was a subgroup of

$$A := \{\, M \in \mathbf{GL}(2, \mathbf{F}) \mid \det(M) \in \mathbf{F}_\ell^{*k-1} \,\}$$

that maps onto $\mathbf{F}_\ell^{*k-1}$, we see that $G = A$.