Serre's modularity conjecture

Chandrashekhar Khare

These are notes for talks given in Boston in April 2006 at:

(i) Number Theory Seminar, Boston University

(ii) Eigenvariety semester, Harvard University

(iii) MIT colloquium

Various subsets of these slides were used for each talk!

The goal of these talks is to report on recent progress on Serre's conjecture. All of this is joint work with J-P. Wintenberger.

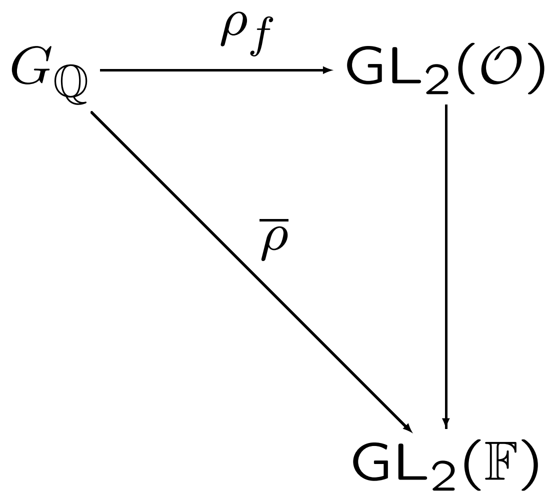
Statement of the conjecture

Let $\overline{\rho}$: Gal$(\overline{\mathbb{Q}}/\mathbb{Q}) \to$ GL$_2(\mathbb{F})$ be a continuous, absolutely irreducible, two-dimensional, odd (det$\overline{\rho}(c) = -1$ for c a complex conjugation), mod p representation, with $\mathbb{F}$ a finite field of characteristic p. We say that such a representation is of *Serre-type*, or S-type, for short.

We denote by $N(\overline{\rho})$ the (prime to p) Artin conductor of $\overline{\rho}$, and $k(\overline{\rho})$ the weight of $\overline{\rho}$ as defined in Serre's 1987 Duke paper. The invariant $N(\overline{\rho})$ is made out of $(\overline{\rho}|_{I_\ell})_{\ell \neq p}$, and is divisible exactly by the primes ramified in $\overline{\rho}$ that are $\neq p$, while $k(\overline{\rho})$ is such that $2 \leq k(\overline{\rho}) \leq p^2 - 1$ if $p \neq 2$ ($k(\overline{\rho}) = 2$ or 4 if $p = 2$), and is made from information of $\overline{\rho}|_{I_p}$.

It is an important feature of the weight $k(\overline{\rho})$, for $p > 2$, that if $\overline{\chi}_p$ is the mod p cyclotomic character, then for some $i \in \mathbb{Z}$, $2 \leq k(\overline{\rho} \otimes \overline{\chi}_p{}^i) \leq p+1$ (as this range almost falls within the range of Fontaine-Lafaille theory!).

Serre has conjectured (1973–1987) that such a $\overline{\rho}$ *arises* (with respect to some fixed embedding $\iota : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}_p}$) from a newform f of weight $k(\overline{\rho})$ and level $N(\overline{\rho})$. We fix embeddings $\iota : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}_p}$ for all primes p hereafter, and when we say (a place above) p, we will mean the place induced by this embedding. By *arises* from we mean that the reduction of an integral model of the p-adic representation ρ_f associated to f, which is valued in $GL_2(\mathcal{O})$ for $\mathcal{O}$ the ring of integers of some finite extension of $\mathbb{Q}_p$, modulo the maximal ideal of $\mathcal{O}$ is isomorphic to $\overline{\rho}$:


$G_{\mathbb{Q}}$
ρ_f
$\mathsf{GL}_2(\mathcal{O})$
$\overline{\rho}$
$\mathsf{GL}_2(\mathbb{F})$

Two forms of the conjecture

It is convenient to split the conjecture into 2 parts:

1. *Qualitative form:* In this form, one only asks that $\overline{\rho}$ arise from a newform of unspecified level and weight.

2. *Refined form:* In this form, one asks that $\overline{\rho}$ arise from a newform $f \in S_{k(\overline{\rho})}(\Gamma_1(N(\overline{\rho})))$.

A large body of difficult, important work of a large number of people, Ribet, Mazur, Carayol, Gross, Coleman-Voloch, Edixhoven, Diamond et al., proves that (for $p > 2$) the qualitative form implies the refined form. We focus only on the qualitative form of the conjecture.

In fact the conjecture, especially in its qualitative form, is much older and dates from the early 1970's. The *qualitative form* of the level 1 (i.e., $N(\overline{\rho}) = 1$) conjecture/question was officially formulated in an article that Serre wrote for the Journées Arithmétiques de Bordeaux in 1975. Perhaps the restriction to level 1 is merely for simplicity. The fact that it is only in a qualitative form, in the weight aspect, is for a more substantial reason. The definition of the weight $k(\overline{\rho})$ is delicate and probably came later.

Some early results

Serre wrote about the level one conjectures to Tate on May 1st, 1973. Tate replied to Serre first on June 11, and then on July 2, 1973: in the second letter, he proved the conjecture for $p = 2$. The method of Tate for $p = 2$ was later applied to the case of $p = 3$ by Serre. The case of $p = 5$ was treated under the GRH by Sharon Bruegemann. In these cases one has to prove that there is no representation

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}})$$

that is odd, irreducible and unramified outside the residue characteristic of $\mathbb{F}$ where $\mathbb{F} = \mathbb{F}_2, \mathbb{F}_3, \mathbb{F}_5$. The method of Tate was to bound the root discriminant of $K_{\overline{\rho}}$, the fixed field of the kernel of $\overline{\rho}$, from above and play this off against lower bounds given by Minkowski (and then strengthenings of this of Odlyzko).

Historical backdrop

Serre made his conjecture when Deligne had attached Galois representations to higher weight newforms and Swinnerton-Dyer and Serre had been studying the properties of these representations.

For instance, consider the Ramanujan Δ function, which is a cusp form of weight 12 on $\mathrm{SL}_2(\mathbb{Z})$:

$$\Delta(z) = q\Pi(1-q^n)^{24} = \Sigma\tau(n)q^n$$

with $q = e^{2\pi iz}$. This is an eigenform for the Hecke operators T_ℓ for each prime ℓ, where the action of T_ℓ is given by:

$$\Delta(z)|T_\ell = \Sigma\tau(n\ell)q^n + \ell^{11}\Sigma\tau(n)q^{n\ell},$$

and the fact that Δ is an eigenfunction means that

$$\Delta(z)|T_\ell = \tau(\ell)\Delta(z).$$

For each prime p there is an attached Galois representation

$$\rho_{\Delta,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Z}_p)$$

which is irreducible for all p, and which is unramified outside p. This is characterised by the property that the characteristic polynomial of $\rho_{\Delta,p}(\mathrm{Frob}_\ell)$ for all primes $\ell \neq p$ is $X^2 - \tau(\ell)X + \ell^{11}$.

Swinnerton-Dyer and Serre proved that this representation has large image, i.e., $\rho_{\Delta,p}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))$ contains $\mathrm{SL}_2(\mathbb{Z}_p)$ for all p different from

$$2, 3, 5, 7, 23, 691.$$

This they did by first proving that the mod p image contains $\mathrm{SL}_2(\mathbb{F}_p)$ for these primes. This sufficed as Swinnerton-Dyer had proved:

Lemma: For $p > 3$ a closed subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$ that contains $\mathrm{SL}_2(\mathbb{F}_p)$ in its reduction mod p, contains $\mathrm{SL}_2(\mathbb{Z}_p)$.

Congruences between modular forms

They also initiated the study of congruences between modular forms which has been an intense focus of research ever since.

For instance they observed the following which is a particular case of a result of Serre:

$$\Delta(z) = q\Pi(1-q^n)^{24} = q\Pi(1-q^n)^2\Pi(1-q^n)^{22}$$

$$\cong q\Pi(1-q^n)^2\Pi(1-q^{11n})^2 \pmod{11}$$

and the latter is the q-expansion of the unique cusp form in $S_2(\Gamma_0(11))$.

It is in this context of the study of Galois representations attached to newforms, and congruences, that Serre asked for the converse diretion.

Measures of the complexity of $\overline{\rho}$ and known cases

The difficulty in proving the modularity of a $\overline{\rho}$ can be either measured in terms of the ramification properties of $\overline{\rho}$, for example $N(\overline{\rho}), k(\overline{\rho})$, or in terms of $\mathrm{im}(\overline{\rho})$.

The results proven by Tate and Serre early on were for the cases of residue characteristic $p = 2, 3$ and $N(\overline{\rho}) = 1$, i.e., cases which were more tractable in terms of the first measure.

The results of Langlands, and Tunnell in the lates 70's and early 80's, implied Serre's conjecture when $\mathrm{im}(\overline{\rho})$ is solvable, i.e., cases which were more tractable in terms of the second measure. In our proof the "complexity" of $\overline{\rho}$ is again, like in the results of Tate and Serre, measured in terms of the ramification properties of $\overline{\rho}$.

Taylor's potential version of Serre's conjecture

A very important breakthrough was when Taylor proved a potential version of Serre's conjecture, i.e., there is a totally real Galois extesnsion $F/\mathbb{Q}$ such that $\overline{\rho}|_{G_F}$ arises from a Hilbert modular form. This is one of the main inputs into the proof of our theorem. To make the result easier to use Taylor also ensures that F is unramified at p.

In some cases Taylor, and following Taylor's method, Manoharmayum and Ellenberg, were able to control F to be solvable and thus prove some (non-solvable) cases of Serre's conjecture when the image was contained in $\mathrm{GL}_2(\mathbb{F}_5)$, $\mathrm{GL}_2(\mathbb{F}_7)$ or $\mathrm{GL}_2(\mathbb{F}_9)$. The case when the image is contained in $\mathrm{GL}_2(\mathbb{F}_4)$ had been addressed by Shepherd-Baron and Taylor by another method. We see again that the conjecture is proven by this method when im($\overline{\rho}$) is relatively small.

Theorem

Theorem(joint with Wintenberger)

(i) For $p > 2$ Serre's conjecture is true for odd conductors, i.e., for $\overline{\rho}$ unramified at 2.

(ii) For $p = 2$ Serre's conjecture is true when $k(\overline{\rho}) = 2$.

We expect that the general case of Serre's conjecture is not far away.

Part of the proof of this theorem, *Serre's modularity conjecture: the odd conductor case (I)* is available at `http://www.math.utah.edu/~shekhar`: this part proves the theorem modulo 2 technical results which are proved in the part that is being written.

In the latter part of this talk, when we give some proofs, we will focus on the level one case: $N(\overline{\rho}) = 1$.

Main tool

Modularity lifting: This is a result of the following type:

Given a continuous 2-dimensional representation $\overline{\rho} : G_{\mathbf{Q}} \to GL_2(\mathbb{F})$ that *arises* from a newform then any continuous representation ρ that lifts $\overline{\rho}$:

$$\begin{array}{ccc} & & GL_2(\mathcal{O}) \\ & \overset{\rho}{\nearrow} & \downarrow \\ G_{\mathbf{Q}} & \xrightarrow{\bar{\rho}} & GL_2(k) \end{array}$$

which is finitely ramified, and satisfies a local condition at p of potential semistablity in the sense of Fontaine, again arises from a newform, i.e., is *modular*.

This can be viewed as a *relative version* of the Fontaine-Mazur conjecture in a particular situation. This was the type of theorem proved by

Wiles and Taylor in their work on the Shimura-Taniyama conjecture.

This is one of the main tools available to prove modularity of Galois representations. Its limitation is that it assumes residual modularity. Somewhat paradoxically, the proof of Serre's conjecture which addresses residual modularity, uses these *modularity lifting techniques* combined with other developments arising from it that are due principally to Taylor (potential version of Serre's conjecture), and Skinner-Wiles, Diamond, Fujiwara and Kisin (developements of the modularity lifting machinery).

Where does Serre's conjecture fit?

To answer the question, it seems appropriate to state other related conjectures.

Conjecture (Artin) *Suppose* $\rho_{\mathbf{C}} : G_{\mathbf{Q}} \to GL_2(\mathbf{C})$ *is an irreducible continuous odd representation. Then* $\rho_{\mathbf{C}}$ *arises from a newform of weight 1.*

Conjecture (Serre) *Let* $p > 2$ *be a prime. Suppose* $\overline{\rho} : G_{\mathbf{Q}} \to GL_2(\overline{\mathbf{F}}_p)$ *is an irreducible continuous odd representation. Then* $\overline{\rho}$ *arises from a newform.*

Conjecture (Fontaine-Mazur) *If an absolutely irreducible* $\rho : G_{\mathbf{Q}} \to GL_2(\mathcal{O})$*, with* $\mathcal{O}$ *the ring of integers of a finite extension of* $\mathbf{Q}_p$*, is odd* *and ramified at only finitely many primes and* $\rho|_{D_p}$ *is potentially semistable, then up to twist* ρ *arises from a newform.*

These conjectures are related to the **Langlands program**.

There are relationships between these conjectures:

Serre implies Fontaine-Mazur (Wiles, Taylor, ...)

Serre implies Artin (we will explain this below)

Fontaine-Mazur implies Serre (Ramakrishna)

The actual statements, especially in the first stated implication, are more hedged. But the recent work of Kisin and Emerton, building on the p-adic Langlands program of Berger, Breuil and Colmez, might even make this literally true (when the Hoodge-Tate numbers of the lift are unequal)!

2-dimensional compatible systems

Serre's conjecture implies modularity of many motives of rank 2 over $\mathbb{Q}$. We prefer to formulate this implication for compatible systems instead. In the result below we will pretend that all of Serre's conjecture has been proven. (Thus the conservative reader might just want to assume that the compatible systems below are "uunramified" at 2.)

We fix embeddings ι_p, ι_∞ of $\overline{\mathbb{Q}}$ in its completions $\overline{\mathbb{Q}_p}$ and $\mathbb{C}$ for each prime p.

Definition For a number field E, an E-rational, 2-dimensional strictly compatible system of representations (ρ_λ) of $G_\mathbb{Q}$ is the data of:

(i) for each finite place λ of E, $\rho_\lambda : G_{\mathbb{Q}} \to \mathrm{GL}_2(E_\lambda)$ is a continuous, absolutely irreducible representation of G_F,

(ii) for all finite places q of $\mathbb{Q}$, if λ is a place of E of residue characteristic different from q, the Frobenius semisimplification of the Weil-Deligne parameter of $\rho_\lambda|_{D_q}$ is independent of λ, and for almost all primes q this parameter is unramified.

(iii) For almost all primes λ of E, $\rho_\lambda|_{D_\ell}$ is crystalline at ℓ of integral Hodge-Tate weights (a, b) with $a \geq b$ that are independent of ℓ, where ℓ is the residue characteristic of λ.

When $a \neq b$, then the system is said to be regular, otherwise irregular.

When we say that for some number field E, a E-rational compatible system of 2-dimensional representations (ρ_λ) of $G_{\mathbb{Q}}$ lifts $\overline{\rho}$ we mean that for the place λ of E fixed by ι_p, the residual representation arising from ρ_λ is isomorphic to $\overline{\rho}$ (i.e., ρ_λ lifts $\overline{\rho}$).

By abusing notation, for a prime ℓ we denote by ρ_ℓ the ℓ-adic representation ρ_λ of the compatible system (ρ_λ) for λ the place above ℓ we have fixed. We denote by $\overline{\rho}_\ell$ the (semisimplification of) residual representation that arises from ρ_ℓ. We say that the compatible system is *irreducible* if each ρ_λ is absolutely irreducible.

A consequence of Serre's conjecture

Proposition:

(i) A regular strictly compatible system that is *odd* arises up to twist from a newform of weight ≥ 2.

(ii) An irregular strictly compatible system that is *odd* arises up to twist from a newform of weight 1.

Proof of modularity of compatible systems

The proofs are similar to the argument used by Serre in his Duke paper to deduce modularity of elliptic curves over $\mathbb{Q}$ from his conjecture

Firstly after twisting we may assume that the Hodge-Tate numbers (a, b) of the compatible system are such that $b = 0$ and $a \geq 0$.

In the case of (i), when $a > 0$, we see that Serre's conjectures apply to $\overline{\rho}_\lambda$ for infinitely many λ and that these arise from a fixed newform $f \in S_k(\Gamma_1(N), \mathcal{O})$ for some integers k, N with $k > 1, N > 6$, and with $\mathcal{O}$ the ring of integers of a number field E. From this it follows, comparing the characteristic polynomials of Frobenii that arise from the compatible system attached to f and those attached to (ρ_λ), that (ρ_λ) arises from f.

In the case of (ii), when $a = b = 0$, the complication is that mod ℓ weight 1 forms need not lift to characteristic 0. The argument to circumvent this, gives a way of going from results about "regular" Galois representations to "irregular" ones.

By a theorem of Sen and Fontaine for all but finitely many λ, ρ_λ is unramified at $\ell(\lambda)$ which is the residue characteristic of the residue field arising from λ.

From this we can at first only conclude that $\overline{\rho}_\lambda$ for almost all λ arises from $S_{\ell(\lambda)}(\Gamma_1(N), \mathcal{O})$ $(N > 6)$ where $\ell(\lambda)$ is the residue charateristic of the residue field arising from λ. We cannot conclude as the dimensions of $S_{\ell(\lambda)}(\Gamma_1(N))$ tend to infinity.

But a result of Gross, Coleman-Voloch (this is the "non-formal" step of this extra argument in the irregular case) yields that in fact $\overline{\rho}_\lambda$ arises from a much smaller space $S_1(\Gamma_1(N), \mathbb{F}_\lambda)_{\text{Katz}} \subset S_{\ell(\lambda)}(\Gamma_1(N), \mathcal{O}) \otimes \mathbb{F}_\lambda$. The dimensions of

$$S_1(\Gamma_1(N), \mathbb{F}_\lambda)_{\text{Katz}}$$

are bounded independently of ℓ. On the other hand the reduction map

$$S_1(\Gamma_1(N), \mathcal{O})_{\text{classical}} \to S_1(\Gamma_1(N), \mathbb{F}_\lambda)_{\text{Katz}}$$

may still not be surjective.

But after throwing out a further "sporadic" finite set of primes (primes in the support of the torsion of $H^1(X_1(N)_{\mathcal{O}}, \underline{\omega}_{\mathcal{O}})$), this map is indeed surjective and thus one sees that for almost all λ, $\overline{\rho}_\lambda$ does indeed arise from $S_1(\Gamma_1(N), \mathcal{O})_{\text{classical}}$ allowing us to conclude as earlier.

Non-liftable forms

There are example of mod ℓ weight 1 forms that cannot be lifted to characteristic 0.

(i) Mestre has an example of a weight 1 mod 2 form whose attached Galois representation has image $\mathrm{SL}_2(\mathbb{F}_8)$: this cannot be lifted to characteristic 0 as the subgroups of $\mathrm{GL}_2(\mathbb{C})$ are limited.

(ii) Consider a 2-dimensional mod 2 representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ that is unramified at 2 and has projective image A_5 and is even. This arises from a weight 1 mod 2 form which cannot be lifted.

Abelian varieties of GL_2-type and Artin's conjecture

Definition: An abelian variety A over $\mathbb{Q}$ is said to be of GL_2-type if it is simple, and if there is a number field L such that $[L:\mathbb{Q}] = \dim(A)$, and an order $\mathcal{O}$ of L such that $\mathcal{O} \hookrightarrow \mathrm{End}_{\mathbb{Q}}(A)$.

A corollary of part (i) of the above proposition is the generalised Shimura-Taniyama-Weil conjecture:

Corollary: For an abelian variety A over $\mathbb{Q}$ of GL_2-type with conductor N, there is a non-constant morphism $\pi : X_1(N) \to A$.

The corollary follows from the part (i) of the earlier proposition (in the case of $|a-b| = 1$)

and Faltings' isogeny theorem. For $|a-b|>1$, we do not seem to be able to make a "motivic" statement for lack of a Tate conjecture for modular forms of weight $k>2$.

Part (ii) yields Artin's conjecture as an Artin representation $\rho_{\mathbb{C}}$ gives rise to a compatible system of the type in part (ii).

Liftings of residual $\overline{\rho}$

Consider a residual representation $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F})$ such that $\overline{\rho}|_{\mathbb{Q}(\mu_p)}$ is irreducible. Fix a large finite extension $E/\mathbb{Q}_p$, and its ring of integers $\mathcal{O}$.

Consider *lifting data* of the following kind:

For each prime $q \neq p$ a lifting $\tilde{\rho}_q$ of $\overline{\rho}|_{D_q}$ with values in $\mathrm{GL}_2(\mathcal{O})$, such that the lifts are unramified for almost all q. For $q = p$ fix a potentially semistable lift $\tilde{\rho}_p$ of $\overline{\rho}|_{D_p}$ with values in $\mathrm{GL}_2(\mathcal{O})$: attached to this there is a Weil-Deligne parameter (τ, N) with τ a complex representation of I_p and N a nilpotent matrix in $\mathrm{GL}_2(\mathbb{C})$, and Hodge-Tate numbers (a, b).

Also assume that there is an odd Hecke character ψ which matches with the determinants of $\tilde{\rho}_q|_{I_q}$ (this is only a condition for $p = 2$).

One of the main technical tools in the proof of Serre's conjecture is the following kind of result:

LT: After enlarging E, there is a lift $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathcal{O})$ of $\overline{\rho}$ with determinant arising from ψ, such that for all primes $q \neq p$, $\rho|_{I_q}$ is isomorphic to $\tilde{\rho}_q$ (as E-valued representations), and $\rho|_{D_p}$ is potentially semistable such that its inertial Weil-Deligne parameter is isomorphic to $(\tau|_{I_p}, N)$ and with Hodge-Tate numbers (a, b).

We do not quite prove this result (especially one needs this, and we have proved this, with many more restrictions at p) but prove enough of it to suffice for our needs. Such liftings with the properties above ensured for almost all primes q is a result of Ramakrishna. For our applications it is important to produce the above more calibrated type of liftings with the properties ensured at *all* primes q

When $\overline{\rho}$ is modular this is a result of Diamond and Taylor building on work of Ribet and Carayol.

How does one produce the liftings?

We consider a deformation ring R (a complete, Noetherian, local (CNL) $\mathcal{O}$-algebra), defined in terms of the lifting data, which has the property such that whenever there is a morphism $\pi : R \to \mathcal{O}'$, with $\mathcal{O}'$ the ring of integers of a finite extension $E'/\mathbb{Q}_p$, then there is a lifting $\rho_\pi : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathcal{O}')$ with the desired local properties (and conversely). Thus it will be enough to prove the existence of such a morphism π. The definition of this ring R is after Mazur, and also needs Kisin's idea of framed deformations.

Using Taylor's results on potential modularity of $\overline{\rho}$, and modularity lifting theorems ($R = T$ theorems) over totally real fields, one first proves that R is a finite $\mathcal{O}$-module. Using obstruction-theory arguments of the type used by Mazur, Boeckle and Kisin, one shows that the Krull dimension of R is at least 1. These 2 facts together yield that $p \in R$ is not nilpotent and hence there is a prime ideal I of R with $p \notin I$, and thus the fraction field of R/I is a finite extension E' of $\mathbb{Q}_p$. Thus the map $R \to R/I \hookrightarrow E'$ produces the desired lifting.

Existence of compatible systems

Taylor and Dieulefait, using Taylor's potential modularity result, Brauer's theorem and solvable base change, prove that given a lift ρ as above, it is part of a strictly compatible system (ρ_λ). In practise at times one needs finer properties of $\rho_\ell|_{I_\ell}$, that are not part of the definition of a compatible system. These I will spell out as and when needed.

Applications of minimal liftings to low levels and weights

A minimal lift is a particular case of the lifts constructed which are roughly characterised by the fact that, assuming $2 \leq k(\overline{\rho}) \leq p+1$, $N(\rho) = N(\overline{\rho})$, and $\rho|_{D_p}$ is crystalline of weight k, i.e., of Hodge-Tate weights $(k(\overline{\rho})-1, 0)$.

This yields the following result:

Proposition:

(i) There are no S-type $\overline{\rho}$ with $k(\overline{\rho}) = 2, N(\overline{\rho}) = 1$. As a consequence there are no irreducible finite flat group schemes $\mathcal{G}$ of type (p,p) over Spec($\mathbb{Z}$).

(ii) There are no S-type representations $\overline{\rho}$ with $k(\overline{\rho}) = 2$, and $N(\overline{\rho}) = q = 2, 3, 5, 7, 13$. A S-type representation $\overline{\rho}$ with $k(\overline{\rho}) = 2$, and $N(\overline{\rho}) = q = 11$ arises from $J_0(11)$ (and also from the Δ function).

(iii) There are no S-type representations $\overline{\rho}$ such that $N(\overline{\rho}) = 1$, and $2 \leq k(\overline{\rho}) \leq 8$, or $k(\overline{\rho}) = 14$. Any S-type representation $\overline{\rho}$ such that $N(\overline{\rho}) = 1$, and $k(\overline{\rho}) = 12$, arises from the Ramanujan Δ-function.

Low levels and weights: proofs

(i) The fact that (i) follows from minimal liftings is an observation that is independently due to Dieulefait and Wintenberger. Let $\overline{\rho}$ be in the statement. Consider a minimal lifting ρ and the compatible system (ρ_λ) that it is part of, and consider ρ_7. Results proved by Fontaine and Abrashkin show that ρ_7 is reducible, hence so is ρ which contradicts the irreducibility of $\overline{\rho}$.

The statment about finite flat group schemes follows by an argument of Serre in his Duke paper. This uses results of Raynaud (to restrict behavior of I_p in the Galois representations that $\mathcal{G}$ gives rise to), and the fact that $\mathbb{Q}$ has no unramified extensions!

(ii) Again one uses a minimal lifting ρ and Taylor's results yield that ρ arises from an abelian variety over $\mathrm{Spec}(\mathbb{Z})[\frac{1}{q}]$, with semistable reduction at q. All these by results of Brumer-Kramer and Schoof are isogenous to powers of $J_0(q)$.

(iii) This uses (ii) and the fact that weights considered are one more than a prime. Lets do the case:

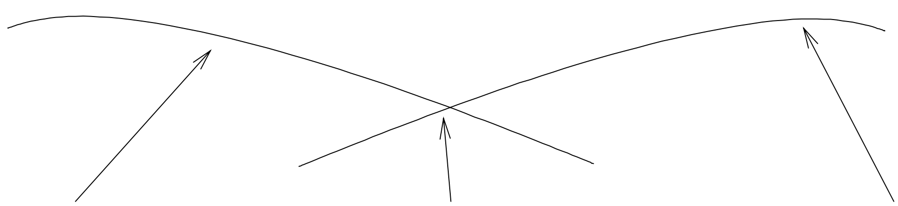
$k(\overline{\rho}) = 6, N(\overline{\rho}) = 1$: Suppose we have an irreducible $\overline{\rho}$ with $N(\overline{\rho}) = 1, k(\overline{\rho}) = 6$ as in the theorem (and we may assume $p > 3$ by Serre's result for $p = 3$). We use minimal lifts to get a compatible system (ρ_λ) of weight 6, i.e., Hodge-Tate of weights $(5, 0)$, and with good reduction everywhere.

Assume the representation $\overline{\rho}_5$ arising from the compatible system is reducible. Then we deduce that ρ_5 is modular by Skinner-Wiles lifting results. Otherwise, we get a minimal lift ρ' of

$\overline{\rho}_5$ that is unramified outside 5, and semistable (and not crystalline) of weight 2 at 5.

$$\begin{pmatrix} \overline{\chi}_5 & * \\ 0 & 1 \end{pmatrix}.$$

Such a ρ' by results of Taylor arises from an abelian variety A over $\mathbb{Q}$ that is semistable and with good reduction outside 5. But by results of Brumer-Kramer such an A does not exist.



Two linked compatible systems: the case of weight 6

The following definition is useful:

Definition of linked compatible systems:

Let E be a number field. We say that two E-rational, compatible systems of representations of $G_{\mathbb{Q}}$ are **linked** (at λ) if for some finite place λ of E the semisimplifications of the corresponding residual mod λ representations arising from the two systems are isomorphic up to a twist by a (one-dimensional) character of $G_{\mathbb{Q}}$.

Applications of minimal liftings to Fermat's Last Theorem

One can give a variation on Wiles' proof of Fermat's last theorem.

Recall that in Serre's Duke paper it is shown how the Frey construction of a semistable elliptic curve E_{a^p,b^p,c^p} over $\mathbb{Q}$, associated to a Fermat triple (a,b,c), i.e., $a^p+b^p+c^p=0$, a,b,c coprime and $abc \neq 0$, and where we may assume a is -1 mod 4, b even, and p a prime >3, leads to a S-type representation $\overline{\rho}$ (the irreducibility is a consequence of a theorem of Mazur) with $k(\overline{\rho})=N(\overline{\rho})=2$. Wiles proved Fermat's Last Theorem by showing that E is modular and hence $\overline{\rho}$ is modular and thus by Ribet's level-lowering results, $\overline{\rho}$ arises from $S_2(\Gamma_0(2))$

which gave a contradiction as the latter space is empty.

A possible way to prove Fermat's Last Theorem is to show that $\overline{\rho}$ arises from a semistable abelian variety A over $\mathbb{Q}$ with good reduction outside 2 (note that E has conductor the radical of abc). This would give a contradiction as by the results of Brumer-Kramer such an A does not exist.

Minimal liftings enable one to give such a proof. The minimal lifting result produces a lift of ρ that is Barsotti-Tate at p and has semistable reduction at 2, and is unramified everywhere else. Taylor's results towards the Fontaine-Mazur conjecture prove that such a ρ arises from A which gives a contradiction.

This proof while different in appearance from that of Wiles, uses all the techniques he developed in his original proof. The "simplifications" in this slightly different approach are:

(i) We do not need to use the results of Langlands-Tunnell, that prove Serre's conjecture for a $\overline{\rho}$ with solvable image, that Wiles had needed;

(ii) This altered proof does not make use of the most difficult of the level-lowering results which is due to Ribet, but instead uses the level lowering up to base change results of Skinner-Wiles. This step does use the solvable base change of Langlands.

Modularity lifting results

Consider a 2-dimensional mod $p > 2$ representation $\overline{\rho}$ of $G_{\mathbf{Q}}$ which is odd and with $2 \leq k(\overline{\rho}) \leq p+1$ with $p > 2$. We do not assume that $\overline{\rho}$ is irreducible, but we do assume that $\overline{\rho}$ is modular, which in the reducible case simply means odd. The following theorem is the work of many people, Wiles, Taylor, Breuil, Conrad, Diamond, Flach, Fujiwara, Guo, Kisin, Savitt, Skinner et al. and is absolutely vital to us.

Theorem: (ML)

1. Let ρ be a lift of $\overline{\rho}$ to a p-adic representation that is unramified outside p and crystalline of weight k, with $2 \leq k \leq p+1$, at p. Then ρ is modular.

2. Let ρ be a lift of $\overline{\rho}$ to a p-adic representation that is unramified outside p and either

semistable of weight 2, or Barsotti-Tate over $\mathbb{Q}_p(\mu_p)$. Then ρ is modular.

3. Let ρ be a lift of $\overline{\rho}$ to a p-adic representation that is unramified outside a finite set of primes and is Barsotti-Tate at p. Then ρ is modular.

An inductive hook

The following result follows immediately from minimal liftings and modularity lifting results and is crucial for our inductive proof of the level one conjecture.

Theorem: (i) Given an odd prime p, if all 2-dimensional, mod p, odd, irreducible representations $\overline{\rho}$ of a given weight $k(\overline{\rho}) = k \leq p+1$, and unramified outside p, are known to be modular, then for any odd prime $q \geq k-1$, all mod q representations $\overline{\rho}'$ of S-type, of level one, and of weight $k(\overline{\rho}') = k$, are modular.

(ii) If the level one case of Serre's conjecture is known for a prime $p > 2$, then for any prime q, Serre's conjectures is known for all mod q representations $\overline{\rho}$ which are of S-type, of level one, and of weight $k(\overline{\rho}) \leq p+1$.

Chebyshev-type estimates on primes

In the arguments below we will use, that for each prime $p \geq 5$, there is a non-Fermat prime $P > p$ (for example P the smallest non-Fermat prime $> p$) and an odd prime power divisor $\ell^r || (P-1)$ so that

$$\frac{P}{p} \leq \frac{2m+1}{m+1} - (\frac{m}{m+1})(\frac{1}{p}) \qquad (1)$$

where we have set $\ell^r = 2m+1$ with $m \geq 1$.

The level 1 case: Proof

We know the level 1 case for $p = 5$ as we have proved the level 1 conjecture for weights up to 6. The inductive hook ensures that if we prove the level one case of Serre's conjecture for a prime $p > 2$, then we know the level one case of Serre's conjecture for all primes $\leq p$. The number of non-Fermat primes is infinite. Thus it will suffice to prove the level 1 conjecture for each non-Fermat prime $p > 5$. We do this by induction on primes.

Assume we have proved the level one case of Serre's conjecture for a (not necessarily non-Fermat) prime $p \geq 5$. Consider the smallest non-Fermat prime $P > p$.

Consider a S-type representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ with $\mathbb{F}$ a finite field of residue characteristic $P > 5$, and such that $2 \leq k(\overline{\rho}) \leq P + 1$.

By the estimates on primes, there is an odd prime power divisor $l^r := 2m + 1$ of $P - 1$, and such that the primes p and P satisfy the bound given earlier:

$$\frac{P}{p} \leq \frac{2m+1}{m+1} - (\frac{m}{m+1})(\frac{1}{p}) \qquad (2)$$

We produce a weight 2 irreducible compatible system (ρ_λ) lifting $\overline{\rho}$. By this we mean that ρ_p is unramified outside p, and either semistable of weight 2 at p, or with WD parameter $(\omega_p^{k(\overline{\rho})-2} \oplus 1, 0)$.

Consider the residual (semisimplified) mod ℓ representation $\overline{\rho}_\ell$ that arises from ρ_ℓ. If $\overline{\rho}_\ell$ is reducible, or is unramified at P, and hence is reducible by the weight 2 level 1 case, or $\overline{\rho}_\ell$ is dihedral (and hence modular), we are done by applying ML. Thus we may assume $\overline{\rho}_\ell$ is *not* dihedral, is *not* reducible, and *is* ramified at P.

Observe that there is an

$$i \in [\frac{m}{2m+1}(P-1), (\frac{m+1}{2m+1})(P-1)]$$

such that the character ω_P^i is congruent modulo our chosen place above ℓ to the character $\omega_P^{k(\overline{\rho})-2}$. Using lifting techniques LT we see there is an irreducible, strictly compatible system (ρ'_λ) lifting $\overline{\rho}_\ell$ with the following property: the irreducible ℓ-adic lifting ρ'_ℓ is unramified outside $\{\ell, P\}$, is Barsotti-Tate at ℓ, and $\rho'_\ell|_{I_P}$ is of the form

$$\begin{pmatrix} \omega_P^i & * \\ 0 & 1 \end{pmatrix},$$

with $i \in [\frac{m}{2m+1}(P-1), (\frac{m+1}{2m+1})(P-1)]$.

Consider the P-adic representation ρ'_P in this system (ρ'_λ), and the corresponding (semisimplified) residual representation $\overline{\rho}'_P$. By a result of Breuil, Mezard and Savitt, we get that $k(\overline{\rho}'_P) = i+2$ or $k(\overline{\rho}'_P \otimes \overline{\chi}_P{}^{-i}) = P+1-i$. Using the Chebyshev-type estimates on primes we see that $2 \leq k(\overline{\rho}'_P \otimes \overline{\chi}_P{}^j) \leq p+1$ for some $j \in \mathbb{Z}$.

Thus either $\overline{\rho}'_P$ is reducible, or $\overline{\rho}'_P$ is irreducible and by our induction hypothesis and the inductive hook we again see that $\overline{\rho}'_P$ is modular.

As ρ'_P is irreducible, unramified away from P, $\rho'_P|_{D_P}$ is Barsotti-Tate over $\mathbb{Q}_P(\mu_P)$, *and* $\overline{\rho}'_P$ is modular, ML implies that ρ'_P, and hence (ρ'_λ), is modular. Now the proof concludes by using that the compatible systems (ρ_λ) and (ρ'_λ) are linked (at ℓ): we spell this out.

The proof concludes

As (ρ'_λ) is modular, ρ'_ℓ is modular. Therefore by definition $\overline{\rho}_\ell$, being the reduction of ρ'_ℓ, is modular. As ρ_ℓ is an irreducible lifting of $\overline{\rho}_\ell$ that is finitely ramified and Barsotti-Tate at ℓ, *and* its reduction $\overline{\rho}_\ell$ is modular, we conclude by ML that ρ_ℓ is modular, Therefore (ρ_λ) is modular, and hence ρ_P is modular. Hence $\overline{\rho}$ is modular.

Where to from here?

One might wonder whether the methods of the proof of Serre's conjecture can work in greater generality to prove modularity of more complicated Galois representations.

For instance one might try and extend it to prove an analog of Serre's conjecture for 2-dimensional totally odd irreducible representations of G_F with F a totally real field.

But even for real quadratic F we run into a serious problem because of the following reason.

The proof of Serre's conjecture in retrospect can be viewed as a method to exploit an accident which occurs in a few different guises:

1. (Fontaine, Abrashkin) There are no non-zero abelian varieties over Spec($\mathbb{Z}$).

2. (Tate, Serre) There are no irreducible representations $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}})$ with $\overline{F}$ the algebraic closure of either $\mathbb{F}_2$ or $\mathbb{F}_3$ that are unramified outside 2 and 3 respectively.

3. $H^*_{\mathrm{cusp}}(\mathrm{SL}_2(\mathbb{Z}), \mathbb{R}) = 0$.

These accidents allow one to prove Serre's conjecture for very small invariants $(p, k(\overline{\rho}), N(\overline{\rho}))$ attached to $\overline{\rho}$ (for example $p = 2, N(\overline{\rho}) = 1, k(\overline{\rho}) = 2$).

Given these accidents, there is a fairly systematic method to exploit them to prove Serre's conjecture for all other triplets of invariants by linking different invariants by a series of *linked* compatible systems.

Such accidents do not happen even for a general real quadratic field F.

But another direction is more promising to apply the methods above: we consider higher dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

This is especially so in the light of the recent breakthrough of Taylor that proves modularity lifting and potential automorphy in higher dimensions completing his work with Clozel, Harris and Shepherd-Barron.

Here the accidents persist for a while:

(a) (Fontaine) There is no 7-adic representation of $G_{\mathbb{Q}}$ that is unramified outside 7, and crystalline at 7 with Hodge-Tate weights $\in [0,3]$, that is irreducible.

(b) (Fermigier, Miller) $H^*_{\mathrm{cusp}}(\mathrm{SL}_n(\mathbb{Z}), \mathbb{R}) = 0$ for $n \leq 23$. Irreducible n-dimensional p-adic representations which are expected to be attached to cusp forms which contribute to

$H^*_{\mathsf{cusp}}(\mathsf{SL}_n(\mathbb{Z}), \mathbb{R})$ are supposed to have the property that they are unramified outside p, and crystalline at p with Hodge-Tate numbers $0, 1, \cdots, n-1$.

This suggests that for $n \leq 23$ there are no irreducible n-dimensional p-adic representations of $G_{\mathbb{Q}}$ unramified outside p and crystalline at p with Hodge-Tate weights $0, 1, \cdots, n-1$. Fontaine's result proves this for $n \leq 4$.

This holds out promise that one can prove an analog of Serre's conjecture for certain 3-dimensional and 4-dimensional representations of $\mathsf{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. For instance consider the following very vague conjecture which might be accessible because of Fontaine's result, and the methods used to prove Serre's conjecture:

Conjecture: Let $\mathsf{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GSp_4(\mathbb{F})$ be an irreducible representation that is *odd* (i.e., the similitude character is odd). Then $\overline{\rho}$ arises from an automorphic representation of $GSp_4(\mathbb{A})$.

Still there is a lot of work to be done to fulfill the promise.