Serre's Conjectures

Henri Darmon

September 9, 2007

Contents

This article explains Serre's conjectures relating mod p Galois representations of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ to modular forms mod p, with special emphasis on the aspects related to Wiles' recent breakthrough on the Shimura-Taniyama conjecture.

It is really impossible to improve on Serre's original exposition, given in [Se7]. The reader is urged to consult [Se7] before reading this article.

It is a pleasure to thank Fred Diamond and Eric Liverance for many useful discussions over the last year related to the topics of this paper, and the anonymous referee for making a careful and thorough review of the manuscript.

1 Statement of Serre's conjecture

We begin with a statement of Serre's conjectures. For generalities on modular forms, see [Sh] or the paper by Diamond and Im in this volume.

Let

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{F})$$

be an irreducible two-dimensional representation of $G_{\mathbf{Q}} = \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ over a finite field $\mathbf{F}$ of characteristic p. Asume that ρ is *odd*, i.e., $\det \rho : G_{\mathbf{Q}} \longrightarrow \mathbf{F}^*$ is an odd character. This means that if c is a complex conjugation, then $\rho(c)$ has eigenvalues 1 and -1.

Note that, if ρ is unramified at l, and Frob_l is a Frobenius element at l, then $\rho(\mathrm{Frob}_l)$ is a well-defined conjugacy class in $\mathbf{GL}_2(\mathbf{F})$; in particular, its characteristic polynomial is well defined.

If R is any subring of $\mathbf{C}$, let $S_k(N, \epsilon, R)$ be the space of cusp forms of weight k, level N, and character ϵ with Fourier coefficents in R. These are the functions on the upper half plane which vanish at the cusps, satisfy the transformation property

$$f(\frac{a\tau+b}{c\tau+d}) = (c\tau+d)^k \epsilon(d) f(\tau),$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, and can be written in the form

$$f(\tau) = \sum_{n \geq 1} a_n q^n, \qquad q = e^{2\pi i \tau}, \quad a_n \in R.$$

If R contains the ring $\mathbf{Z}[\epsilon]$ generated by the values of the character ϵ, then we have (q-expansion principle)

$$S_k(N,\epsilon,R) = S_k(N,\epsilon,\mathbf{Z}[\epsilon]) \otimes R.$$

For any ring R equipped with a map $\phi : \mathbf{Z}[\epsilon] \longrightarrow R$, we can thus consistently define

$$S_k(N,\epsilon,R) := S_k(N,\epsilon,\mathbf{Z}[\epsilon]) \otimes_\phi R.$$

The Hecke operators T_n with $\gcd(n,N) = 1$ and U_q with $q|N$ act on the spaces $S_k(N,\epsilon,R)$ and on the subspace $S_k^{new}(N,\epsilon,R)$ of newforms. A simultaneous eigenform for this commuting algebra of operators will simply be called an *eigenform*, and will be said to be *normalized* if its first Fourier coefficient a_1 is equal to 1.

We say that ρ is *modular* if there exists a normalized eigenform f (of some weight $k \geq 2$, level N, and character ϵ) with Fourier coefficients in $\mathbf{F}$,

$$f = \sum_{n\geq 1} a_n q^n, \qquad a_1 = 1, \quad a_n \in \mathbf{F},$$

such that for all l which are unramified for ρ and do not divide Np, $\rho(\mathrm{Frob}_l)$ has characteristic polynomial

$$x^2 - a_l x + l^{k-1}\epsilon(l).$$

In this case, we say that ρ and f are *associated*. A construction of Eichler and Shimura for weight 2, and Deligne in weight $k \geq 2$, shows that any eigenform f gives rise to an associated (not necessarily irreducible) representation ρ. In [Se7], Serre conjectures that the converse holds as well. In some sense, this is an analogue of the Shimura-Taniyama conjecture for mod p representations.

Conjecture 1.1 (Serre's conjecture, vague form) *Any odd irreducible representation ρ as above is modular.*

Serre's conjecture is much more precise than this; that is what accounts for its usefulness and importance. In fact, Serre gives a *precise recipe* (described in sec. 2) for assigning to ρ a weight $k(\rho) > 0$, a level $N(\rho) > 0$, and an $\mathbf{F}$-valued character $\epsilon(\rho)$. He then conjectures that

Conjecture 1.2 (Serre's conjecture, precise form) *There exists a normalized mod p eigenform of level $N(\rho)$, weight $k(\rho)$, and (when char $\mathbf{F} > 3$) character $\epsilon(\rho)$ which is associated to ρ.*

Remark: When char $\mathbf{F} = 2$ or 3, it is not always possible to find an eigenform of the correct weight, level, *and* character associated to ρ. The difficulty is due to the possible presence of elliptic points of order 2 or 3 on the modular curves $X_0(N)$. The "naive" definition of modular forms mod p that we are using (following Serre) is not quite adequate in this context, and Katz's definition of modular forms mod p is more appropriate for dealing with these situations. (cf., for example, [Ed].)

The outline of this paper is as follows. Section 2 partly explains Serre's recipe for $N(\rho)$, $k(\rho)$, and $\epsilon(\rho)$, completing the statement of conj. 1.2.

Section 3 presents some of the evidence for Serre's conjecture. The evidence that exists is of three types. Firstly, computational evidence has been amassed, mostly by Mestre, in support of conj. 1.2. Secondly, a great deal of work has been done in the direction of proving that conjecture 1.1 implies conj. 1.2. Thirdly, (and this is a key point in Wiles' general attack on the Shimura-Taniyama conjecture) the Serre conjecture is largely known to be true when $\mathbf{F}$ is the field $\mathbf{F}_2$ with two elements, when the image of ρ is dihedral, or when $\mathbf{F}$ is the field $\mathbf{F}_3$ with three elements, the last thanks to the work of Langlands and Tunnell on base change. Much of the computational evidence is summarized in [Se7], §5, and we will not say more on this, focusing instead on the theoretical evidence.

Sec. 4 is devoted to various applications of the Serre conjectures to Fermat's Last theorem and other Diophantine questions (see also the article by Liem Mai in this volume), and to the Shimura-Taniyama conjecture.

We conclude in section 5 by briefly mentioning the relation between Wiles' work and the Serre conjectures.

2 Serre's recipe for $N(\rho)$, $k(\rho)$ and $\epsilon(\rho)$

The invariant $N = N(\rho)$ attached to ρ is the *Artin conductor* of the representation ρ, with the possible factors of p removed; see [Se7], §1.2 for the precise definition. In particular, the level $N(\rho)$ is divisible only by the primes $l \neq p$ where ρ is ramified, and the value of $N(\rho)$ depends only on the restriction of ρ to the decomposition groups D_l at these ramified places. Note that, by definition, the level $N(\rho)$ is always prime to p, although the representation ρ may be ramified at p (and, in fact, typically is).

The character $\epsilon(\rho)$ is read off from the determinant character $\det\rho$:

$G_{\mathbf{Q}} \longrightarrow \mathbf{F}^*$ associated to ρ, as follows: a direct calculation shows that the conductor of $\det \rho$ divides Np, so that $\det \rho$ can be identified with an $\mathbf{F}$-valued Dirichlet character $(\mathbf{Z}/Np\mathbf{Z})^* \longrightarrow \mathbf{F}^*$, or, by the Chinese remainder theorem, with a pair of characters:

$$\epsilon : (\mathbf{Z}/N\mathbf{Z})^* \longrightarrow \mathbf{F}^*, \qquad \phi : (\mathbf{Z}/p\mathbf{Z})^* \longrightarrow \mathbf{F}^*.$$

We set $\epsilon(\rho) := \epsilon$.

We can write

$$\phi(x) = x^{k_0 - 1}, \qquad 2 \leq k_0 \leq p.$$

The integer k_0 determines the value of $k(\rho)$ mod $p-1$, namely, we have:

$$k(\rho) \equiv k_0 \pmod{p-1}.$$

For the precise value of $k(\rho)$, one needs to shift the value of k_0 up by a certain multiple of $(p-1)$. This corresponds to predicting the *filtration* (cf. [Se2]), and not just the weight, of the corresponding modular form mod p.

The recipe for the precise value of $k(\rho)$ depends on the value of certain *exponents* a and b associated to the restriction of ρ to the inertia group I_p at p. For simplicity we confine ourselves to the case where p is odd, referring the reader to [Se7] for the complete recipe. Let W_p denote the wild inertia subgroup (which is the maximal pro-p-subgroup of I_p), and let V be the two-dimensional $\mathbf{F}$-vector space which realizes ρ. The quotient $I_t = I_p/W_p$ is isomorphic to $\lim_{\leftarrow} \mathbf{F}_{p^n}^*$, where the inverse limit is taken with respect to the norm maps. It can be shown that W_p acts trivially on the semi-simplification V^{ss} of V, so that I_p acts on V^{ss} via its tame quotient I_t. Since I_t is abelian, this action is reducible and corresponds to two characters ϕ and $\phi^{'}$ of I_t with values in $\bar{\mathbf{F}}_p^*$. The fact that the representation ρ extends to the full decomposition group D_p shows that the characters ϕ and $\phi^{'}$ are stable under the action of Frobenius $x \mapsto x^p$. Hence we can distinguish two cases:
Case 1: $\phi^p = \phi^{'}$, $\phi^{'p} = \phi$. Then we can write

$$\phi = \Psi^{a+pb} = \Psi^a \Psi^{'b},$$

where $\Psi : I_t \longrightarrow \mathbf{F}_{p^2}^*$ is one of the two natural projections. We normalize the exponents a and b so that

$$0 \leq a, b \leq p-1.$$

Case 2: $\phi^p = \phi$, $\phi'^p = \phi'$. Then we can write

$$\rho|_{I_p} = \begin{pmatrix} \chi^a & * \\ 0 & \chi^b \end{pmatrix},$$

where $\chi : I_t \longrightarrow \mathbf{F}_p^*$ is the natural map (i.e., the cyclotomic character). The exponents a and b are well-defined modulo $p-1$, and we normalize them so that

$$0 \le a \le p-2 \text{ if } \rho|_{I_p} \text{ is semisimple}, \quad 1 \le a \le p-1 \text{ otherwise.}$$

$$0 \le b \le p-2.$$

Now the formula for k is

$$k = 1 + a + b + (p-1)\min(a,b) + (p-1)\delta,$$

where $\delta = 0$ or 1, the case $\delta = 1$ arising when $(a,b) = (0,0)$ (i.e., ρ is unramified at p) or when $\rho|_{I_p}$ is "très ramifié", c.f. [Se7], §2.4. For more details the reader is invited to consult §2 of [Se7] or §4 of [Ed]. Serre's precise recipe for the weight took shape through an exchange of letters with J-M. Fontaine; Fontaine's ideas have been crucial in elucidating the relationship between the weight and restriction to the inertia group at p of a modular mod p Galois representation.
Remark: There is a certain amount of flexibility in defining $N(\rho)$ and $k(\rho)$. For example, if f is an eigenform of weight k and level $N = Mp^r$, then the mod p representation ρ associated to f by Deligne's construction also arises from an eigenform of level M and weight k' for some k' (cf. §2 of [Ri4]). For an example where this occurs, see for instance [Se4], th. 11. The representation ρ also arises from a form of weight 2 and level $Mp^{r'}$ for some r'; cf. §6 of [Di2] and [Wi1].

An example: the Galois representations associated to a semi-stable elliptic curve over $\mathbf{Q}$*:* Let E be a semi-stable elliptic curve over $\mathbf{Q}$ and let $\rho_{E,p}$

$$\rho_{E,p} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}\,(E_p) \simeq \mathbf{GL}_2(\mathbf{F}_p)$$

be the Galois representation associated to E_p. Let N_E be the conductor of E; since E is semi-stable, N_E is simply the product of the primes of bad reduction of E. Let Δ_E be the minimal discriminant of E.

Proposition 2.1 *The invariants $N(\rho_{E,p})$, $k(\rho_{E,p})$ and $\epsilon(\rho_{E,p})$ are given by*

1. *$N(\rho_{E,p})$ is the product of all the primes $l \neq p$ such that $ord_l(\Delta_E) \neq 0 \pmod{p}$.*
2. *$k(\rho_{E,p}) = 2$, if $ord_p(\Delta_E) \equiv 0 \pmod{p}$, and is equal to $p+1$ otherwise.*
3. *$\epsilon(\rho_{E,p}) = 1$.*

Proof: See [Se7], prop. 5.
Remark: In general, we say that ρ is *finite* at a prime l if, when $l \neq p$, ρ is unramified, and if, when $l = p$, ρ comes from a finite flat group scheme over $\mathbf{Z}_p$. The condition $\mathrm{ord}_p(\Delta_E) \equiv 0 \pmod{p}$ implies that $\rho_{E,p}$ is finite at p.

3 Evidence for Serre's conjecture

3.1 Proofs of Serre's epsilon-conjecture

Conjecture 1.2 appears very difficult to attack except in all but a few very special cases. A more manageable problem has been to prove conj. 1.2, assuming that conj. 1.1 is satisfied, which is expressed in the following conjecture (known as Serre's "epsilon conjecture").

Conjecture 3.1 (Serre's epsilon-conjecture) *If ρ is modular (i.e., is associated to an eigenform mod p of some level, weight, and character), then it is associated to a modular form mod p of level $N(\rho)$, weight $k(\rho)$, and (if char$\mathbf{F} > 3$) character $\epsilon(\rho)$.*

Alot has been proved in this direction, thanks to the work and ideas of many people, including N. Boston [BLR], H. Carayol [Ca], R. Coleman [CV] F. Diamond [Di1], [DT1], [DT2], [Di2], B. Edixhoven [Ed], G. Faltings, J-M. Fontaine, B. Gross [Gr], B. Jordan [JL], H.W. Lenstra, R. Livné, B. Mazur, K. Ribet [Ri2], [Ri4], J-P. Serre, R. Taylor, J. Tilouine, F. Voloch, and A. Wiles.

3.1.1 Theorems of Mazur and Ribet

The first result in the direction of Serre's epsilon-conjecture was proved by B. Mazur:

Theorem 3.2 (Mazur) *Suppose that $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{F})$ is absolutely irreducible and arises from an eigenform of weight* 2*, level N, and trivial character. If $l||N$ but ρ is finite at l, and if*

$$l \not\equiv 1 \pmod{p},$$

then ρ arises from a mod p eigenform on $X_0(M)$, $M = N/l$.

The argument is reproduced in [Ri2].

The next major breakthrough came with the work of Ribet, who showed how to remove all primes l at which ρ is finite (and not just the $l \not\equiv 1 \pmod{p}$) from the level of the mod p representation.

Theorem 3.3 (Ribet) *Suppose that $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{F})$ is absolutely irreducible and arises from an eigenform of weight* 2*, level N, and trivial character. Suppose also that char($\mathbf{F}$) is odd. If $l||N$ but ρ is finite at l, then ρ arises from a modular form of level $M = N/l$.*

This result is enough (cf. 4.1, or the articles by Mai and Prasad in this volume) to show that the Shimura-Taniyama conjecture implies Fermat's Last theorem. Ribet proved his result by a very ingenious argument, exploiting a deep interplay between the arithmetic of modular curves and Shimura curves associated to indefinite quaternion algebras. For the details on the proof, see Prasad's article in this volume, (or [Ri2], [Ri3], and [Ri4]).

The following examples give some illustrations of the theorems of Mazur and Ribet. We follow the notations of Cremona's book [Cr], which extends the classical Antwerp tables [MF].

Examples: 1. Let $E = X_0(11)$ be the elliptic curve with equation

$$y^2 + y = x^3 - x^2 - 10x - 20$$

having conductor $N = 11$ and discriminant $\Delta = -11^5$. Let ρ be the mod 5 representation associated to E. Prop. 2.1 gives $N(\rho) = 1$, $k(\rho) = 2$, and $\epsilon(\rho) = 1$. By Ribet's theorem (note that $11 \equiv 1 \pmod{5}$), if ρ were irreducible it would arise from an eigenform of weight 2, level 1, and trivial character. Since there are no such forms, the representation ρ must be reducible. This, of course, is well known, and one does not require the full power of Ribet's deep theorem to prove it! In fact, one knows that

$$E_5 \simeq \mathbf{Z}/5\mathbf{Z} \oplus \mu_5$$

as a Galois module.

2. The curve $57C$ (or $57F$ in the Antwerp tables)

$$57C : y^2 = x^3 + x^2 + 20x - 127/4 = f(x).$$

of conductor $57 = 3 \cdot 19$ has discriminant $\Delta = -3^{10}19$, and its mod 2 representation is irreducible. Hence the mod 2 representation $\rho_{E,2}$ arises from the unique cusp form of weight 2 on $X_0(19)$. There is a unique isogeny class of curves of conductor 19, represented by the curve $19C$ with equation

$$19C : y^2 = x^3 + x^2 + x + 1/4 = g(x).$$

It is not hard to check that these two curves define the same mod 2 representation, by showing that the polynomials $f(x)$ and $g(x)$ have the same splitting field. If α denotes the real root of $f(x)$ and β the real root of $g(x)$, then

$$\alpha = 4\beta^2 - 3\beta.$$

In this example, the conclusion of Ribet's theorem 3.3 holds, even though the hypothesis $\mathrm{char}(\mathbf{F}) \neq 2$ is not satisfied. It is likely that this assumption can be removed, especially when $\mathbf{F} = \mathbf{F}_2$.

3. The curve

$$33A : y^2 + xy = x^3 + x^2 - 11x$$

has conductor $N = 33$ and discriminant $\Delta = 3^6 11^2$. Its mod 3 representation $\rho_{E,3}$ is irreducible, and finite at 3, hence by Mazur's result it arises from a form of weight 2 and level 11. There is only one such eigenform, corresponding to the curve $X_0(11)$, and one checks that their Fourier coefficients a_l are the same mod 3 when $l \neq 3, 11$, at least for $l \leq 43$:

l	2	3	5	7	11	13	17	19	23	29	31	37	41	43
$33A$	1	-1	-2	4	1	-2	-2	0	8	-6	-8	6	-2	0
$11A$	-2	-1	1	-2	1	4	-2	0	-1	0	7	3	-8	-6

4. The curve

$$46A : y^2 + xy = x^3 - x^2 - 10x - 12$$

is a modular elliptic curve of conductor $N = 2 \cdot 23$ and discriminant $\Delta = -2^{10} \cdot 23$. Since the mod 5 representation associated to E is irreducible, it must be associated to a mod 5 eigenform of weight 2 on $X_0(23)$. There are no rational eigenforms of weight 2 on $X_0(23)$, but the theta functions

$$\begin{aligned}\theta_1 &= \sum_{m,n\in\mathbf{Z}} q^{m^2+mn+6n^2} (= 1 + 2q + 2q^4 + 4q^6 + 4q^8 + 2q^9 + \cdots)\\ \theta_2 &= \sum_{m,n\in\mathbf{Z}} q^{2m^2+mn+3n^2} (= 1 + 2q^2 + 2q^3 + 2q^4 + 2q^6 + 2q^8 + 2q^9 + \cdots)\end{aligned}$$

associated to the two classes of binary quadratic forms of discriminant -23 give modular forms of weight 1 on $X_0(23)$ with character $(\frac{\cdot}{23})$ (cf. [Hc]). (Note that

$$\frac{1}{2}(\theta_1 - \theta_2) = \eta(\tau)\eta(23\tau) = q\prod(1-q^n)(1-q^{23n})$$

is a *cusp form* of weight 1 on $X_0(23)$ with character.) Setting

$$\begin{aligned}F &= \frac{1}{2}(\theta_1 - \theta_2)\theta_1 = q + q^2 - 3q^3 - 2q^4 + 2q^5 - q^6 + 4q^7 - 3q^8 + 2q^9 + \cdots\\ G &= \frac{1}{2}(\theta_1 - \theta_2)\theta_2 = q - q^2 + q^3 - 2q^5 - 3q^6 + q^8 + 2q^9 + 4q^{10} + \cdots\end{aligned}$$

gives a $\mathbf{Q}$-basis for the space of cusp forms of weight 2 on $X_0(23)$. The action of the Hecke operator T_2 can be computed explicitly and is given by:

$$T_2F = \frac{1}{2}(F+G), \quad T_2G = \frac{1}{2}(F-3G).$$

By diagonalizing T_2, (letting $\omega = \frac{1+\sqrt{5}}{2}$ be the golden ratio, and $\bar{\omega}$ its conjugate) we find that the form

$$g = \frac{1}{2}(\omega F + (1+\bar{\omega})G) = q - \bar{\omega}q^2 - \sqrt{5}q^3 - \omega q^4 - 2\bar{\omega}q^5 - (\omega+2)q^6 + \cdots$$

is an eigenform of weight 2 for $X_0(23)$ with trivial character. (The other eigenform, of course, is merely the Galois conjugate.) One can check that the first few Fourier coefficients $a_l(f)$ of f are congruent to the coefficients $a_l(g)$ of g modulo the ideal $(\sqrt{5})$ when $l \neq 2, 23$:

l	2	3	5	7	11	13	17	19
$a_l(f)$	-1	0	4	-4	2	-2	-2	-2
$a_l(g)$	$-\bar{\omega}$	$-\sqrt{5}$	$-1+\sqrt{5}$	$1+\sqrt{5}$	$-3-\sqrt{5}$	3	$3-\sqrt{5}$	-2

5. The modular elliptic curve

$$988B : y^2 = x^3 - 362249x + 165197113$$

listed in Cremona's tables has conductor $N = 988 = 2^2 \cdot 13 \cdot 19$ and discriminant

$$\Delta = -2^4 \cdot 13 \cdot 19^{13}.$$

The Galois representation ρ_{13} acting on the 13-division points of E is unramified at 19, and Mazur's theorem says that ρ_{13} is associated to a modular form mod 13 of level 52. There is a unique rational eigenform of level 52, given by the curve

$$52A : y^2 = x^3 + x - 10,$$

and one finds that the Fourier coefficients a_l (l not dividing N) associated to these two curves agree mod 13, at least for $l \leq 43$:

l	2	3	5	7	11	13	17	19	23	29	31	37	41	43
$988B$	0	0	2	-2	-2	-1	-7	1	-5	2	-3	7	7	-9
$52A$	0	0	2	-2	-2	-1	6	-6	8	2	10	-6	-6	4

This gives an example of two elliptic curves over $\mathbf{Q}$ whose mod 13 Galois representations are isomorphic. It would be interesting to see how often such pairs occur. (See the discussion is sec. 4.1.)

6. This example examines what happens when one replaces mod p representations by mod p^n representations.

The curves

$$142A : y^2 + xy + y = x^3 - x^2 - 12x + 15,$$

$$142E : y^2 + xy = x^3 - x^2 - 2626x + 52244,$$

(denoted $142F$ and $142G$ respectively in the Antwerp tables) have discriminant $2^9 \cdot 71$ and $2^{27} \cdot 71$ respectively. By using Tate's analytic description of these curves over $\mathbf{Q}_2$ (cf. Liem Mai's article in this volume), one can see that the mod 9 representation ρ_1 associated to $142A$, and the mod 27 representation ρ_2 associated to $142E$, are unramified at 2. A natural extension of Ribet's theorem, replacing mod p representations with mod p^n representations, would lead us to expect that ρ_1 (resp. ρ_2) is realized on the points of order 9 (resp. 27) of the Jacobian of the modular curve $X_0(71)$.

The genus of $X_0(71)$ is 6. One constructs modular forms of weight 2 and level 71 as in example 4, by letting

$$\begin{aligned}\theta_1 &= \sum_{m,n\in\mathbf{Z}} q^{m^2+mn+18n^2}(= 1 + 2q + 2q^4 + 2q^9 + 2q^{16} + 4q^{18} + \cdots),\\ \theta_2 &= \sum_{m,n\in\mathbf{Z}} q^{4m^2+3mn+5n^2}(= 1 + 2q^4 + 2q^5 + 2q^6 + 2q^{12} + 2q^{15} + 2q^{16} + \cdots),\end{aligned}$$

be the theta functions corresponding to the two classes of quadratic forms of discriminant -71, and setting

$$F_1 = \theta_1(\theta_1 - \theta_2)/2, \quad F_{i+1} = T_2(F_i).$$

One checks from the q-expansions that $F_1, \ldots, F_6$ are linearly independent, and hence generate the space of cusp forms of weight 2 on $\Gamma_0(71)$. Furthermore,

$$T_2(F_6) = 9F_1 - 3F_2 - 23F_3 + 5F_4 + 9F_5 - F_6.$$

By diagonalizing the Hecke operators T_2, one finds that the eigenforms for the Hecke algebra are given by the forms:

$$(15-3\alpha^2)F_1+(20-3\alpha-4\alpha^2)F_2+(\alpha^2-4\alpha-8)F_3+(\alpha^2+\alpha-9)F_4+(\alpha+1)F_5+F_6,$$

$$(6-3\alpha^2)F_1+(2\alpha^2-3\alpha+2)F_2+(5\alpha^2+5\alpha-17)F_3-(\alpha^2+3)F_4-(\alpha^2+\alpha-4)F_5+F_6,$$

where α is one of the three roots of the equation $\alpha^3 - 5\alpha + 3 = 0$. By normalizing these forms so that the coefficient of q is 1, one obtains the q expansions:

$$f = q + \alpha q^2 + (-\alpha^2 + 3)q^3 + (\alpha^2 - 2)q^4 + (-\alpha - 1)q^5 + (-2\alpha + 3)q^6 + \cdots,$$

$$g = q + (-\alpha^2 - \alpha + 3)q^2 + (\alpha^2 + \alpha - 3)q^3 + (\alpha + 1)q^4 + (-\alpha^2 - 2\alpha + 5)q^5 + \cdots$$

The eigenforms f and g, together with their Galois conjugates, give the 6 normalized newforms for $X_0(71)$.

The ideal $(3, \alpha)$ is the unique prime ideal of degree 1 in the ring $\mathbf{Z}[\alpha]$ lying above 3. Let $R \simeq \mathbf{Z}_3$ be the completion of $\mathbf{Z}[\alpha]$ at this prime ideal, and let $\bar{f}$ and $\bar{g}$ be the images of the forms f and g in $S_2(71, R)$. Their Fourier coefficients for the first few primes are listed modulo 81, (i.e., with a 3-adic accuracy of 3^{-4}) in the following table:

l	2	3	5	7	11	13	17	19	23	29
$\bar{f}$	60	48	20	24	57	4	24	73	68	70
$\bar{g}$	69	12	11	24	39	40	15	37	77	34

From the table it appears that $a_l(\bar{f}) \equiv a_l(\bar{g}) \pmod{9}$. Assuming that this is true, one can see that the forms

$$f_A = \frac{7\bar{f} - 4\bar{g}}{3} \pmod{9}, \quad f_E = 2\bar{f} - \bar{g} \pmod{81}$$

are modular forms in $S_2(71, \mathbf{Z}/9\mathbf{Z})$ and $S_2(71, \mathbf{Z}/81\mathbf{Z})$ which are *eigenforms* for the Hecke operators. From the following table, one checks that the Fourier coefficients of the form corresponding to 142A are congruent to those of f_A mod 9, at least for $l \neq 2$, $l \leq 29$:

l	2	3	5	7	11	13	17	19	23	29
f_A	3	6	5	6	0	1	0	4	2	1
$142A$	1	-3	-4	-3	0	1	0	-5	-7	-8

Likewise, one checks that the Fourier coefficients of the form corresponding to $142E$ are congruent to those of f_E mod 27, at least for $l \neq 2$, $l \leq 29$:

l	2	3	5	7	11	13	17	19	23	29
f_E	51	84	29	24	75	49	33	28	59	25
$142E$	-1	3	2	-3	-6	-5	6	1	5	-2

This example suggests that the philosophy of Serre's conjectures, and of the ϵ-conjecture, extends to mod p^n representations. Wiles has proved a number of precise statements in this direction, and used them to bound the order of the Selmer group of the symmetric square under certain conditions; cf. [Wi1], or [Wi2].

3.1.2 The latest word

The work of Mazur and Ribet alluded to before was mainly concerned with modular forms of weight 2, and trivial character. A great number of mathematicians have extended the scope of these results, to cover more general cases involving arbitrary weights, levels, and characters, so that now the full epsilon conjecture is almost proved. For a good summary of these results, together with an explanation of the techniques involved in proving them, and an extensive bibliography, see [Ri4] and [Di2].

The latest result, which is the culmination of all these efforts, is proved in [Di2]: say that the irreducible representation ρ is an *exceptional case* if $char\mathbf{F} = 3$ and ρ is induced from a character of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(\sqrt{-3}))$, or if $char\mathbf{F} = 2$ and ρ is induced from a character of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}(i))$.

Theorem 3.4 *Assume $\mathbf{F}$ is a field of odd characteristic. If $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{F})$ is a representation arising from an eigenform, then ρ is associated to an eigenform of level $N(\rho)$, weight $k(\rho)$, and, if ρ is not an exceptional case, character $\epsilon(\rho)$.*

For more details, see [Di2].

3.1.3 Raising the level

The results alluded to so far in this section have to do with "lowering the level" of a modular Galois representation; i.e., showing that if it arises from a modular form of some level it also arises from the "optimal" level $N(\rho)$ predicted by the Serre conjectures. There is a considerable amount of literature on congruences between modular forms which is devoted to the problem of listing the possible levels of newforms which are congruent mod p to a given eigenform f. For example, one has the following result which is a corollary of [Ri1] and [Ca]:

Theorem 3.5 *Let f be a newform of weight 2, trivial character, and level N, and assume that the associated representation ρ is irreducible and not an exceptional case. Suppose that l is a prime not dividing N and that $(l-1)(a_l(f)^2-(l+1)^2)$ is divisible by a prime $\mathcal{P}$ over p. Then there is a newform g of weight 2, trivial character and level dl for some $d|N$ such that g is congruent to f mod $\mathcal{P}$.*

For more precise results in this direction, see [DT1].

There are also more precise quantitative measures of the "amount" of mod p congruences that arise between f and newforms of level Nl, involving the notion of the "congruence ideal" of a Hecke ring. For precise definitions, see Kumar Murty's article in this volume. This shows that (in some "sophisticated" sense) the "number" of eigenforms of some level M divisible by N

which are congruent to f mod p can be described by a simple formula. This remark plays a key role in Wiles' proof of the Shimura-Taniyama conjecture for infinitely many j-invariants.

3.2 Cases where the Serre conjecture is known

In spite of the spectacular success in establishing more and more cases of Serre's epsilon-conjecture, very little is known about conj. 1.2 without first assuming conj. 1.1. There are a few notable exceptions, which play an important role in Wiles' work on the Shimura-Taniyama conjecture.

3.2.1 Cases where ρ has dihedral image

Suppose that the image of ρ is isomorphic to a dihedral group D_{2n} with $(n, p) = 1$.

The group D_{2n} can be embedded in $\mathbf{GL}_2(\mathbf{C})$, and hence ρ gives rise to an Artin representation

$$\rho^{'} : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{C}).$$

Since ρ is odd, we may assume that $\rho^{'}$ is odd. This representation is associated to a cusp form g of weight 1, which can be constructed explicitly in terms of theta-functions associated to (definite or indefinite) binary quadratic forms, as was known already to Hecke.

The construction works as follows. The field L cut out by ρ is an abelian extension, with Galois group a cyclic group G of order n, of a quadratic field $K/\mathbf{Q}$, and we can write

$$\rho^{'} = \mathrm{Ind}_{K/\mathbf{Q}}\chi,$$

where $\chi : \mathrm{Gal}(L/K) \longrightarrow \mathbf{C}^*$ is a non-trivial one-dimensional character, which can be viewed as a character on the ideals of $\mathcal{O}_K$ by class field theory.

Let τ be a reflection in $\mathrm{Gal}(L/\mathbf{Q})$, and let $D = \mathrm{Disc}(L^{\tau})$. Let

$$\theta = \sum_J \chi(J) q^{\mathbf{N}J}$$

be the theta function, which is a cusp form of weight 1 and level $|D|$.

By multiplying θ by an Eisenstein series of weight 1, level p and character ω^{-1}, where ω is the Teichmuller character, one obtains a form f of weight 2 which is an eigenform for the Hecke operators T_l mod p. Hence we have:

Proposition 3.6 *If ρ has dihedral image, then ρ is associated to a modular form, i.e., conj. 1.1 is true for ρ.*

Furthermore, thm. 3.4 tells us that:

Corollary 3.7 *If ρ is dihedral and char$\mathbf{F}$ is odd, then ρ is associated to a modular form of level $N(\rho)$, weight $k(\rho)$, and, if ρ is not an exceptional case, character $\epsilon(\rho)$.*

An interesting special case is the one where $\mathbf{F} = \mathbf{F}_2$. Here the image is contained in S_3 which is a dihedral group, so conj. 1.1 is known for ρ, but Serre's epsilon-conjecture remains unproved. The status of conj. 1.2 for $\mathbf{F} = \mathbf{F}_2$ is therefore unclear at present, although it may be quite accessible, since in some cases the epsilon conjecture in the dihedral case can be proved without appealing to thm. 3.3.

3.2.2 Cases where $\mathbf{F} = \mathbf{F}_3$

When $\mathbf{F} = \mathbf{F}_3$, the group $\mathbf{GL}_2(\mathbf{F}_3)$ can be embedded into $\mathbf{GL}_2(\mathbf{C})$, allowing one to lift the mod 3 representation to a characteristic 0 representation $\rho^{'}$. Moreover, the image of $\rho^{'}$ is a solvable group: the group $\mathbf{GL}_2(\mathbf{F}_3)$ is isomorphic to a double cover of the alternating group A_4. A deep result of Langlands and Tunnell [La], [Tu] shows that $\rho^{'}$ is associated to a modular form of weight 1. By the same trick of multiplying this form by an appropriate Eisenstein series, one can exhibit a modular form of higher weight associated to ρ, and show that ρ satisfies conj. 1.1. In light of thm. 3.4, we therefore have:

Theorem 3.8 *If $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{F}_3)$ is absolutely irreducible, then it is associated to an eigenform of level $N(\rho)$, weight $k(\rho)$, and, if ρ is not in the exceptional case, character $\epsilon(\rho)$.*

This theorem is at the center of Wiles' very compelling strategy for proving the Shimura-Taniyama conjecture for semi-stable elliptic curves. See for example Kumar Murty's article in this volume.

4 Applications

4.1 Diophantine applications: Fermat's Last Theorem and some variants

Let

$$a^p + b^p = c^p, \qquad abc \neq 0, \quad p \geq 5$$

be a solution to Fermat's equation. Assume without loss of generality that $a \equiv -1 \pmod 4$ and that b is even, and let

$$E : y^2 = x(x - a^p)(x + b^p)$$

be the elliptic curve first considered by Hellegouarch [He]. It can be shown that E is a semi-stable elliptic curve and that its discriminant Δ_E is

$$\Delta_E = -2^{-8}(abc)^{2p}.$$

(cf. [Se7], p. 200). Consider the mod p representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}\,(E_p) \simeq \mathbf{GL}_2(\mathbf{F}_p)$$

associated to E_p. By a theorem of Mazur [Ma1], ρ is irreducible. Prop. 2.1 implies that $N(\rho) = 2$, $k(\rho) = 2$, and $\epsilon(\rho) = 1$, contradicting conj. 1.2, since there are no non-trivial eigenforms of weight 2 on $\Gamma_0(2)$. Thus, Serre's conjecture implies Fermat's Last Theorem. (For more details, see Mai's article in this volume, or [Se7], §4.)

In fact, thanks to Ribet's work on the epsilon-conjecture 3.1, conj. 1.1 applied to the mod p Galois representation ρ is already enough to imply Fermat's Last Theorem. If E is a modular elliptic curve, then ρ satisfies conj. 1.1. Hence Fermat's Last Theorem follows from the Shimura-Taniyama conjecture (for semi-stable elliptic curves), confirming a remarkable insight of G. Frey.

The reader should consult §4 of [Se7] for more examples where the Serre conjectures are used to study certain variants of the Fermat equation, for example, equations of the form

$$Ax^p + By^p = Cz^p.$$

It is interesting to explore the limits of the Serre conjectures in studying Diophantine equations of the above type. Having come tantalizingly close to Fermat's Last Theorem, a good testing ground for further applications of Serre's conjecture (as well as a nice source of concrete Diophantine questions) is given by the following "generalized Fermat conjecture" [DG].

Conjecture 4.1 *The equation*

$$x^p + y^q = z^r, \qquad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1, \quad \gcd(x,y,z) = 1, \quad xyz \neq 0,$$

has no integer solutions except for

$$1 + 2^3 = 3^2, \;\; 2^5 + 7^2 = 3^4, \;\; 7^3 + 13^2 = 2^9, \;\; 2^7 + 17^3 = 71^2, \;\; 3^5 + 11^4 = 122^2,$$

$$17^7 + 76271^3 = 21063928^2, \;\; 1414^3 + 2213459^2 = 65^7, \;\; 9262^3 + 15312283^2 = 113^7,$$

$$43^8 + 96222^3 = 30042907^2, \;\; 33^8 + 1549034^2 = 15613^3.$$

So far, very little is known about this conjecture, which combines the difficulties of the Fermat and Catalan conjecture. However, one does have the following fragments:

Proposition 4.2 *If Serre's conjecture 1.2 holds, then the equations*

$$x^p + y^p = z^2, \qquad p \geq 13, \quad p \equiv 1 \pmod 4,$$

$$x^p + y^p = z^3, \qquad p \geq 13, \quad p \equiv 1 \pmod 3,$$

$$x^4 - y^4 = z^p, \qquad p \geq 13, \quad p \equiv 1 \pmod 4,$$

have no solutions (x,y,z) *with* $xyz \neq 0$ *and* $\gcd(x,y,z) = 1$.

To prove these propositions one constructs the appropriate Frey curves associated to solutions to the above equations, and uses Serre's conjecture to prove that the associated mod p representation does not exist. Since these representations come from elliptic curves, the Shimura-Taniyama conjecture is enough to deduce the result, in light of Ribet's thm. 3.3. For the details, see [Da1] and [Da2].

A more interesting example is that of the equation

$$x^4 + y^4 = z^p, \quad \gcd(x,y,z) = 1,$$

which generalizes the equation $x^4 + y^4 = z^2$ originally considered by Fermat. Given a solution $a^4 + b^4 = c^p$ to this equation, one considers the elliptic curve over $\mathbf{Q}(i)$

$$E : y^2 = x^3 + 4(1+i)bx^2 + 4i(b^2 + ia^2)x$$

which has discriminant

$$\Delta = 2^{12}c^p(a^2 - ib^2).$$

The map η given by

$$\eta(x,y) = \left(\frac{iy^2}{2x^2}, \frac{-4y(b^2+ia^2) - iyx^2}{(2i-2)x^2}\right)$$

is a 2-isogeny from E to its Galois conjugate $E^{'}$ which is defined over $\mathbf{Q}(i)$. The curve E is a $\mathbf{Q}$-*curve*, i.e., it is an elliptic curve defined over a number field which is isogenous to all of its Galois conjugates. Even though E is not defined over $\mathbf{Q}$, it can be used to construct a 2-dimensional representation of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ as follows: Let $\eta^{'} : E^{'} \longrightarrow E$ be the Galois conjugate of η over $\mathbf{Q}(i)$. One checks that $\eta\eta^{'}$ and $\eta^{'}\eta$ are the endomorphisms of $E^{'}$ and E respectively given by multiplication by 2. If we set $V = E_p \times E^{'}_p$, for p an odd prime, then V is equipped with a natural action of $\mathbf{F}_p[\phi] \simeq \mathbf{Z}[\sqrt{2}] \otimes \mathbf{F}_p$, where $\phi : V \longrightarrow V$ is the endomorphism defined by

$$\phi(P,Q) = (\eta^{'}Q, \eta P).$$

In this way V is a module over $\mathbf{F}_p[\phi]$ of rank 2. The natural action of $G_{\mathbf{Q}(i)}$ on V can be extended to an action of $G_{\mathbf{Q}}$, by defining

$$\sigma(P,Q) = (P^\sigma, Q^\sigma) \text{ if } \sigma i = i,$$

$$\sigma(P,Q) = (Q^\sigma, P^\sigma) \text{ if } \sigma i = -i.$$

This $G_{\mathbf{Q}}$-action commutes with the scalars in $\mathbf{F}_p[\phi]$, and hence gives rise to a two-dimensional representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}_{\mathbf{F}_p[\phi]}(V) \simeq \mathbf{GL}_2(\mathbf{Z}[\sqrt{2}] \otimes \mathbf{F}_p).$$

Because $\mathrm{ord}_{\mathcal{Q}}(\Delta) \equiv 0 \pmod{p}$ for all primes $\mathcal{Q}$ of $\mathbf{Z}[i]$ which are not above 2, and because E is semistable at those primes, one sees that the representation ρ is unramified for all primes $\neq 2, p$, and that it is finite at p. A more careful analysis shows that

$$N(\rho)|2^{14}, \qquad k(\rho) = 2, \qquad \epsilon(\rho) = 1.$$

One can hope to use the Serre conjecture to show that a mod p representation with such a small conductor does not exist; this would give a Diophantine application of the Serre conjectures which does not pass through the Shimura-Taniyama conjecture.

Note that this application does follow from a somewhat stronger version of the Shimura-Taniyama conjecture, which characterizes the elliptic curves over $\bar{\mathbf{Q}}$ which are modular. See the next section for details.

Although the Serre conjectures have striking Diophantine consequences for the Fermat equation and some variants, it seems that they do not yield sweeping Diophantine results applying, say, to all of the equations of the type $Ax^p + By^p = Cz^p$. For example (cf. [Se7], p. 204) the Serre conjectures do not allow one to show (or at least, not in an obvious way!) that the equation

$$x^n + y^n = 31z^n, \quad \gcd(x, y, z) = 1, n \geq 4.$$

has finitely many solutions (x, y, z, n). (Although this is certainly expected to be true, and follows, for example, from the abc-conjecture.) The problem here is that there is at least one non-trivial solution, $(-1, 2, 1, 5)$, so that the methods based on Serre's conjecture, which tend to prove non-existence of such solutions, are bound to fail here.

To reap further Diophantine results from Serre's conjecture, one needs more knowledge about the Galois representations arising from elliptic curves. For example, Frey [Fr3] has made the following conjecture:

Conjecture 4.3 (Frey) *Let A be an elliptic curve over a number field K. There are only finitely many pairs (E, p) consisting of an elliptic curve E over K which is not isogenous to A and a prime number $p > 5$, such that*

$$E_p \simeq A_p \quad \textit{as } G_K - \textit{modules.}$$

Remarks:
1. For fixed $p > 5$, the conjecture is true, by Falting's proof of the the Mordell conjecture. This is because pairs (E, p) as above correspond to rational points on a twist of the modular curve $X(p)$, which has genus greater than 1.
2. The obvious analogue of the above conjecture with (E, p) replaced by pairs (f, p) where f is a modular form of weight 2 is of course false, as follows from thm. 3.5.

One can propose even more ambitious conjectures. Say that an integer n has the *isogeny property* (relative to a number field K) if the implication

$$A_n \simeq B_n \text{ as } G_K - \text{modules} \quad \Rightarrow \quad A \text{ is isogenous to } B \tag{1}$$

holds for all pairs of elliptic curves A, B over K. It is not known whether there are any integers satisfying the isogeny property (over $\mathbf{Q}$, say), and example 5 of sec. 3.1.1 shows that 13 does not have the isogeny property. It is tempting, however, to conjecture the following:

Conjecture 4.4 *Given any global field K, there exists a constant M_K such that all $n \geq M_K$ have the isogeny property.*

This conjecture, which can be viewed as a "mod p" analogue of Tate's isogeny conjecture proved by Faltings, seems very difficult to prove.

Remark: Say that n satisfies the *weak isogeny property* if the implication (1) holds, with at most finitely many exceptional pairs (A, B). A strengthening of conj. 4.4 is

Conjecture 4.5 *There exists an absolute constant M such that all $n \geq M$ have the weak isogeny property over all number fields K.*

It would be very interesting to formulate a convincing guess about the precise value of M.

As Frey has observed, we have:

Proposition 4.6 *If Serre's conjecture 1.2 and Frey's conjecture 4.3 hold, then the equation*

$$Ax^n + By^n = Cz^n, \qquad n > 3, \quad \gcd(x, y, z) = 1,$$

has only finitely many integer solutions (x, y, z, n).

Sketch of proof: We argue by contradiction. Suppose that there are infinitely many solutions (x_i, y_i, z_i, n_i). We can assume without loss of generality that the n_i are distinct primes which do not divide $2ABC$. Now, let

$$E_i : Y^2 = X(X - Ax_i^{n_i})(X + Bx_i^{n_i})$$

be the Frey curve associated to the solution (x_i, y_i, z_i, n_i), and let ρ_i be its associated mod n_i representation. The level of ρ_i can be shown to divide $32(ABC)^2$. By the Serre conjecture, each ρ_i arises from a mod n_i eigenform f_i of level dividing $32(ABC)^2$. Since there are finitely many such eigenforms, there is an eigenform f such that $a_l(f) \equiv a_l(f_i) \pmod{\mathcal{N}_i}$ for infinitely many i, where $\mathcal{N}_i$ is a place of $\bar{\mathbf{Q}}$ above n_i. We claim that f has integer Fourier coefficients, contradicting Frey's conjecture 4.3. For, let l be a prime not dividing $2ABC$, and let $p(x)$ be the minimal polynomial of $a_l(f)$. The curve E_i has either semistable or good reduction at l. In the former case, $a_l(f) \equiv \pm(l+1) \pmod{\mathcal{N}_i}$, and in the latter, $a_l(f) \equiv a_l(f_i) \pmod{\mathcal{N}_i}$. By the Hasse bound, $a_l(f_i)$ is an integer of absolute value less than $2\sqrt{l}$. Hence there exists a in the finite set

$$\{0, \pm 1, \pm 2, \ldots, \pm[2\sqrt{l}], \pm(l+1)\}$$

such that

$$a_l(f) \equiv a \pmod{\mathcal{N}_i} \text{ for infinitely many } i.$$

Hence $p(a) \equiv 0 \pmod{n_i}$ for infinitely many i, so that $p(a) = 0$. This shows that $a_l(f) = a$ is rational, and concludes the proof.

4.2 Relation with the Shimura-Taniyama conjecture

Let

$$f(\tau) = \sum_{n \geq 1} a_n q^n, \quad q = e^{2\pi i \tau}, \quad a_1 = 1$$

be a normalized eigenform of weight 2 on $\Gamma_0(N)$ with trivial nebentypus character. The differential $f(\tau)d\tau$ is invariant under the action of $\Gamma_0(N)$; if the Fourier coefficients a_n belong to $\mathbf{Z}$, then the function

$$\phi_f(\tau) := 2\pi i \int_{i\infty}^{\tau} f(z)dz \left(= \sum_{n \geq 1} \frac{a_n}{n} q^n \right)$$

defines a complex-analytic map from the upper half plane $\mathcal{H} \cup \{\text{cusps}\}$ to $\mathbf{C}/\Lambda_f$, where Λ_f is a rank 2 lattice, generated by the modular symbols $\phi_f(\frac{a}{Nb})$, for $a, b \in \mathbf{Q}$ with $(a, Nb) = 1$. The elliptic curve $E = \mathbf{C}/\Lambda_f$ can in fact be defined over $\mathbf{Q}$. It has good reduction at p for all $p \not| N$, and

$$\#E(\mathbf{F}_p) = p + 1 - a_p.$$

The conjecture of Shimura Taniyama states that this beautiful connection between eigenforms with rational Fourier coefficients and elliptic curves over $\mathbf{Q}$ also goes in the other direction, namely:

Conjecture 4.7 (Shimura-Taniyama) *If E is an elliptic curve over $\mathbf{Q}$ of conductor N, then there exists a normalized eigenform f such that E is isogenous to $\mathbf{C}/\Lambda_f$.*

By the work of Eichler and Shimura, if E satisfies the Shimura-Taniyama conjecture, then $\rho_{E,p}$ satisfies conj. 1.1; we have already used this fact (more or less implicitely) several times so far. In fact, it is also true that conj. 1.2 implies the Shimura-Taniyama conjecture.

Theorem 4.8 *Serre's conj. 1.2 implies conj. 4.7.*

The proof is explained in [Se7], §4.6, but the reader may find it instructive to work it out on her own.

The Serre conjectures also imply generalizations of the Shimura-Taniyama conjecture which say that every abelian variety with real multiplications is a quotient of $J_0(N)$ for some N (cf. [Se7], §4.7).

The following is also worth mentioning: suppose E is an elliptic curve over $\bar{\mathbf{Q}}$, which is a $\mathbf{Q}$-curve. (cf. 4.1.) It is not hard to see that if E is modular (i.e., is a quotient of $J_0(N)$ for some N) then E is a $\mathbf{Q}$-curve. Conversely:

Proposition 4.9 (Ribet) *If Serre's conjecture 1.2 is true, then an elliptic curve E over $\bar{\mathbf{Q}}$ is modular if and only if it is a $\mathbf{Q}$-curve.*

The proof is explained in [Ri5].

5 Wiles' work and the Serre conjectures

We finish with some brief comments about the relation between Wiles' work and the Serre conjectures, following [Wi1].

The Shimura-Taniyama conjecture states that the map

$$\left\{ \begin{array}{l} \text{Newforms of weight 2 on } X_0(N) \\ \text{with rational Fourier coefficients.} \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Isogeny classes of} \\ \text{elliptic curves} \\ \text{ over } \mathbf{Q} \text{ of conductor } N \end{array} \right\}$$

is a bijection. One might try to tackle such a conjecture by showing that the two sets above have the same number of elements. One difficulty is that the rationality condition on the Fourier coefficients is a very subtle one over which one has little control. On the other hand, it is easy to count the number of *all* eigenforms of weight 2 on $X_0(N)$. Such eigenforms (with not necessarily rational coefficients) do not correspond to elliptic curves in general, but they do give rise to p-adic Galois representations which generalize the Tate modules $T_p(E)$ of an elliptic curve, by the work of Eichler and Shimura. Assume for simplicity that p^2 does not divide N. It can be shown that the p-adic representation $\tilde{\rho}$ arising from an eigenform f of weight 2 and level N has the following properties:

1. $\tilde{\rho}$ is unramified outside Np.

2. (Weight condition) If $l \not| Np$, the eigenvalues α_l and $\bar{\alpha}_l$ of the Frobenius element $\tilde{\rho}(\mathrm{Frob}_l)$ are the roots of a polynomial with coefficients in $\bar{\mathbf{Z}} \subset \bar{\mathbf{Z}}_p$, and, when viewed as complex numbers, they have absolute value $\sqrt{l}$.

3. (Determinant condition) We have $\det(\tilde{\rho}) = \chi$, where $\chi : G_{\mathbf{Q}} \longrightarrow \mathbf{Z}_p^*$ is the cyclotomic character.

4. (Condition at p) The restriction of the representation $\tilde{\rho}$ to the inertia group I_p at p satisfies the condition of being ordinary or flat in Wiles' terminology. (For the definition of these terms, see Kumar Murty's article in this volume.)

5. (Condition at ∞) The representation $\tilde{\rho}$ is odd, i.e., if c denotes complex conjugation, then $\tilde{\rho}(c)$ has eigenvalues 1 and -1.

We call irreducible representations satisfying properties 1–5 *admissible* (in a non-standard terminology). One can define the conductor of $\tilde{\rho}$ as Np^δ, where N is the Artin conductor of $\tilde{\rho}$ and $\delta = 0$ if $\tilde{\rho}$ comes from a p-divisible group over $\mathbf{Z}_p$, and $\delta = 1$ otherwise.

One can generalize the Shimura-Taniyama conjecture to state that the map

$$\left\{ \begin{array}{l} \text{Newforms of weight 2} \\ \text{ on } X_0(N) \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Admissible } p\text{-adic representations} \\ \text{ of } G_{\mathbf{Q}} \text{ of conductor } N \end{array} \right\}$$

is a bijection. This conjecture is due, essentially, to Mazur.

Remark: By the isogeny conjecture of Tate which was recently proved by Faltings the functor $E \mapsto T_p(E)$ is a *fully faithful* functor from the category of isogeny classes of elliptic curves over $\mathbf{Q}$ to the category of admissible p-adic representations of $G_{\mathbf{Q}}$. Hence, one does not lose any information in passing from the category of elliptic curves to the "larger" category of p-adic Galois representations.

Now the problem of proving the generalized Shimura Taniyama conjecture above can be broken into two parts:

Problem 1: (Serre's conjecture) Show that the map

$$\left\{ \begin{array}{l} \text{mod } p \text{ newforms of weight } 2 \\ \text{ on } X_0(N) \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Odd, irreducible} \\ \text{mod } p \text{ representations} \\ \text{ of } G_{\mathbf{Q}} \text{ of conductor } N \end{array} \right\}$$

is a bijection. Here the conductor of a mod p representation ρ is defined using Serre's recipe for $N(\rho)$, explained in §2, except when $k(\rho) > 2$, in which case one multiplies $N(\rho)$ by an appropriate power of p.

Problem 2: (Lifting conjecture) Assuming that problem 1 is solved for a *specific* mod p representation ρ_0, i.e., there is an eigenform f_0 such that $f_0 \pmod{p}$ is associated to ρ_0. Show that the map:

$$\left\{ \begin{array}{l} \text{Newforms } f \text{ of weight } 2 \\ \text{ on } X_0(N) \text{ such that} \\ f \equiv f_0 \pmod{p} \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Admissible} \\ p\text{-adic representations} \\ \tilde{\rho} \text{ of } G_{\mathbf{Q}} \\ \text{ of conductor } N \text{ such that} \\ \tilde{\rho} \equiv \rho_0 \pmod{p} \end{array} \right\}$$

is a bijection.

Wiles has made substantial inroads into problem 2, showing that the lifting conjecture is satisfied when p is odd and, for example:

1. The image of ρ_0 is dihedral, or

2. ρ_0 arises from a semistable, modular elliptic curve A and p does not divide the degree of the modular parametrization $X_0(N) \longrightarrow A$.

When $p = 3$ or 5, results of this kind are particularly interesting because then for a specific ρ_0 there are infinitely many non-isomorphic elliptic curves E satisfying $\rho_{E,p} \simeq \rho_0$, and so this gives infinitely many distinct $\bar{\mathbf{Q}}$-isomorphism classes of elliptic curves which are modular.

Remarks:
1. The counting argument which was alluded to in the remark after thm. 3.5 plays an important role in proving the second statement, by allowing Wiles to reduce to the case where $N = N(\rho_0)$ in the statement of the lifting conjecture.
2. The case $p = 2$ of Wiles' program, which is not covered by the above results, is also quite interesting: the desired upper bound on the Selmer group of the symmetric square for $p = 3$, which is the major unresolved issue in Wiles' strategy, is known for the dihedral case, thanks to the work of Karl Rubin on the two variable main conjecture for quadratic imaginary fields. Can one show that every (semistable, say) elliptic curve with a point of order 2 over a non-cyclic cubic extension is modular? There seem to be no conceptual barriers in doing this, only technical difficulties (which could still make the task quite arduous!)

Wiles' compelling strategy for proving the Shimura Taniyama conjecture brilliantly avoids proving any new cases of the Serre conjecture. Rather, it uses the few cases where conj. 1.2 is known as a very tenuous foothold ("une prise d'ongles", in the words of Serre) from which to mount an impressive attack on the Shimura Taniyama conjecture.

This means that the Serre conjectures remain wide open. These fascinating conjectures, which represent a first step in the direction of a "Langlands philosophy mod p", will probably keep number theorists busy in years to come – perhaps long after the Shimura-Taniyama conjecture has been completely proved.

References

[AS] A. Ash and G. Stevens, *Modular forms in characteristic ℓ and special values of their L-functions*, Duke Math. J. **53** (1986), 849–868.

[BLR] N. Boston, H.W. Lenstra and K. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sci. Paris, Série I **312** (1991), 323–328.

[Ca] H. Carayol, *Sur les représentations galoisiennes modulo l attachées aux formes modulaires*, Duke Math. J. **59** (1989), 785–801.

[CV] R.F. Coleman and J.F. Voloch, *Companion forms and Kodaira-Spencer theory*, Invent. Math. **110** (1992), 263–281.

[Cr] Cremona, J., *Algorithms for modular elliptic curves*, Cambridge University Press, 1992.

[Da1] Darmon, H., *The equations $x^n + y^n = z^2$ and $x^n + y^n = z^3$*, Int. Math. Res. Not. 10 (1993), 263-274.

[Da2] Darmon, H., *The equation $x^4 - y^4 = z^p$*, C.R. Math. Rep. Acad. Sci. Canada, vol. XV, no. 6, 286-291, Dec. 1993.

[DG] Darmon, H., Granville, G., *On the equations $z^m = f(x,y)$ and $Ax^p + By^q = Cz^r$*, to appear, in Bulletin of the London Math. Soc.

[DR] P. Deligne and M. Rapoport, *Les schémas modulaires de courbes elliptiques*, Lecture Notes in Math. **349** (1973), 143–316.

[DS] Deligne, P. and Serre, J.-P., *Formes modulaires de poids* 1, Ann. Sci. Ec. Norm. Sup. 7, 507-530 (1974)

[Di1] F. Diamond, *Congruence primes for cusp forms of weight $k \geq 2$*, Astérisque 196-197 (1991), pp. 205-213.

[Di2] F. Diamond, *The refined conjecture of Serre*, to appear.

[DT1] F. Diamond and R. Taylor, *Non-optimal levels for mod l modular representations of Gal*($\bar{\mathbf{Q}}/\mathbf{Q}$), Inv. Math. 115, 435-462, (1974)

[DT2] F. Diamond and R. Taylor, *Lifting modular mod l representations*, Duke Math. J. **74** (1994), 253–269.

[Dr] V.G. Drinfeld, *Coverings of p-adic symmetric regions* (in Russian), Funkts. Anal. Prilozn. **10** (1976), 29–40. Translation in Funct. Anal. Appl. **10** (1976), 107–115.

[Ed] B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), 563–594.

[Fa1] Faltings, G., *p-adic Hodge theory*, J. of the A. M. S. **1** (1988) 255–299.

[Fa2] Faltings, G., *Crystalline cohomology and p-adic Galois representations*, in Algebraic analysis, geometry and number theory. Proceedings of the JAMI Inaugural Conference, J. I. Igusa, ed., Johns Hopkins University Press, Baltimore (1989) 25–80.

[Fr1] Frey G., *Links between stable elliptic curves and certain diophantine equations*, Ann. Univ. Saraviensis, **1** (1986), 1–40.

[Fr2] Frey, G., *Links between solutions of $A - B = C$ and elliptic curves*, in Number theory, Ulm 1987, Proceedings, Lecture Notes in Math. **1380**, Springer-Verlag, New York (1989) 31–62.

[Fr3] Frey, G., Oral communication, Chinese University of Hong Kong, Hong Kong, Dec. 1993.

[Gr] B.H. Gross, *A tameness criterion for Galois representations associated to modular forms mod p*, Duke math. J. **61** (1990), 445–517.

[Hc] Hecke, E., *Zur Theorie der elliptischen Modulfunktionen*, (no. 23 in Mathematische Werke, Vandenhoeck and Ruprecht, Göttingen 1970).

[He] Hellegouarch, Y., *Points d'ordre $2p^h$ sur les courbes elliptiques*, Acta. Arith. **26** (1974/75) 253–263.

[Hi] H. Hida, *Galois representations into $GL_2(\mathbf{Z}_p[[X]])$ attached to ordinary cusp forms*, Invent. Math. **85** (1986), 545–613.

[JL] B. Jordan and R. Livné, *Local diophantine properties of Shimura curves*, Math. Ann. **270** (1985), 235–248.

[La] Langlands, R. *Base Change for GL(2)*, Princeton Univ. Press, 1980.

[Ma1] Mazur, B. *Modular curves and the Eisenstein ideal*, Publ. Math. IHES 47, 33-186 (1977)

[Ma2] Mazur, B., *Rational isogenies of prime degree*, Inv. Math. 44, 129-162 (1978)

[MF] Birch, B., Kuyk, W., eds., Modular functions of one variable IV, vol. 476, Springer-Verlag, New York (1975) 74–144.

[Ri1] Ribet, K., *Congruence relations between modular forms*, Proc. I.C.M. (1983), 503–514.

[Ri2] Ribet, K., *On modular representations of Gal*$(\bar{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), 431–476.

[Ri3] Ribet, K., *From the Taniyama-Shimura conjecture to Fermat's Last Theorem*, Ann. de la Fac. des Sci. de l'Univ. de Toulouse, 11:116-139.

[Ri4] Ribet, K., *Report on mod l representations of Gal*$(\bar{\mathbf{Q}}/\mathbf{Q})$, in Motives, Proc. Symp. Pure Math. **55**:2 (1994), 639–676.

[Ri5] Ribet, K., *Abelian varieties over* $\mathbf{Q}$ *and modular forms*, to appear.

[Se1] Serre, J-P. *Abelian l-adic representations and elliptic curves*, New York: W.A. Benjamin 1968.

[Se2] Serre, J-P., *Congruences et formes modulaires* (d'apres H.P.F. Swinnerton-Dyer), Sém. Bourbaki, 1971/72, exposé 416.

[Se3] Serre, J.-P., *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inv. Math. 15, 259-331 (1972).

[Se4] Serre, J.-P., *Formes modulaires et fonctions zeta p-adiques*, Lect. Notes in Math. 350, 191-268, Springer-Verlag, 1973.

[Se5] Serre, J.-P., *Valeurs propres des opérateurs de Hecke modulo l*, Journées arith., Bordeaux, 1974, Astérisque 24-25 (1975) 109-117.

[Se6] Serre, J.-P., *Modular forms of weight one and Galois representations*, Algebraic Number Fields, (A. Frölich, ed.), Acad. Press, 1977, 193-268.

[Se7] Serre, J.-P., *Sur les représentations modulaires de degré* 2 *de Gal*$(\bar{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. Vol. 54, no. 1, 179-230 (1987).

[Sh] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Iwanami Shoten and Princeton University Press, 1971.

[Tu] Tunnell, J., *Artin's conjecture for representations of octahedral type*, Bull. A.M.S. **5** (1981) 173–175.

[Wi1] A. Wiles, Course at Princeton University, February–April, 1994.

[Wi2] A. Wiles, Private communication, May 1994.