Introduction to Serre's conjecture.

Bas Edixhoven

September 23, 2005

Abstract

The conjecture will be stated, and put in its historical context and in the wider context of the Langlands program. Serre's level and weight of a 2-dimensional mod p Galois representation will be defined. Khare's result will be stated. An overview will be given of what will be treated in the seminar.

1 Historical context of Serre's conjecture

Let us start by saying that Chapter 1 of the book [5] provides a very good introduction to this matter, with much more details than this lecture.

The conjecture in question, in a rough qualitative form, dates from the early 1970's, as one can read in Serre's article [6]. The conjecture is about 2-dimensional representations over finite fields of the absolute Galois group of $\mathbb{Q}$.

We let $\mathbb{Q} \to \overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$, for example the one that consists of all algebraic numbers in $\mathbb{C}$. Then $\overline{\mathbb{Q}}$ is the union of all its finite dimensional Galois subextensions $\mathbb{Q} \to K$, and therefore $G_{\mathbb{Q}} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) = \mathrm{Aut}(\overline{\mathbb{Q}})$ is the projective limit of the $\mathrm{Gal}(K/\mathbb{Q})$. We consider $G_{\mathbb{Q}}$ with the topology that comes with the projective limit: it is the coarsest topology for which all projections to the discrete $\mathrm{Gal}(K/\mathbb{Q})$ are continuous. With this topology, $G_{\mathbb{Q}}$ is a compact totally disconnected topological group. The open subgroups of $G_{\mathbb{Q}}$ are precisely the stabilisers of elements of $\overline{\mathbb{Q}}$; they form a basis of neighborhoods of 1.

Since Kronecker and Weber it is known that the maximal abelian extension of $\mathbb{Q}$ in $\overline{\mathbb{Q}}$ is the cyclotomic extension, obtained by adjoining all roots of unity to $\mathbb{Q}$. Equivalently, the action of $G_{\mathbb{Q}}$ on $\overline{\mathbb{Q}}^{\times}_{\mathrm{tors}} \cong \mathbb{Q}/\mathbb{Z}$ gives a surjection $G_{\mathbb{Q}} \to \mathrm{Aut}(\overline{\mathbb{Q}}^{\times}_{\mathrm{tors}}) = \hat{\mathbb{Z}}^{\times}$ which turns out to be the maximal abelian Hausdorff quotient.

Class field theory gives a description of the maximal abelian Hausdorff quotient of the absolute Galois group G_K for K any number field (or function field over a finite field). So in this

case the Galois group of the maximal abelian extension of K is described, but we point out that we do not know an explicit description of the maximal abelian extension itself.

The Langlands program, that was started in the end of the 1960's, aims to generalise CFT to the description of suitable representations of $G_{\mathbb{Q}}$ on finite dimensional vector spaces, mainly over $\mathbb{C}$ and over $\mathbb{Q}_l$. We will see more of this is this seminar, in the case of 2-dimensional representations of G_K with K totally real. I find Taylor's article [7] an excellent introduction into this matter. One of the aims of this seminar is also to learn more about the Langlands program, as more or less all of it that is known in the GL_2-case is used in Khare's work (and there has been really a lot of progress, recently).

The context of Serre's conjecture is that of explicit descriptions of 2-dimensional continuous representations of $G_{\mathbb{Q}}$ over finite fields. In the case of CFT for $\mathbb{Q}$ the object to consider is $\mathbb{G}_{\mathrm{m}}(\overline{\mathbb{Q}})_{\mathrm{tors}}$. Typical objects that give 2-dimensional representations of $G_{\mathbb{Q}}$ are $E(\overline{\mathbb{Q}})_{\mathrm{tors}}$, for E an elliptic curve over $\mathbb{Q}$. The objects that play a role in Serre's conjecture are modular forms, corresponding to torsion points on jacobian varieties of curves of the form $\overline{\Gamma\backslash\mathbb{H}}$ with Γ a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, or of étale cohomology of certain locally constant sheaves on such curves.

As we said above, a qualitive form of Serre's conjecture was made in the beginning of the 1970's. Motivated by an application to Fermat's Last Theorem, Serre formulated a very precise conjecture in his article [6] of 1987. Let us recall this application (also described in detail in Serre's article).

Suppose that $p \geq 5$ is prime, that $a^p + b^p + c^p = 0$ and $abc \neq 0$, with a, b and c integers that are relatively prime. Then one writes $A = a^p$, $B = b^p$ and $C = c^p$, permutes them such that $2|b$, and $a = -1 \pmod 4$, and defines the elliptic curve:

$$E_{A,B,C}: \quad y^2 = x(x-A)(x+B).$$

This elliptic curve is often called the "Frey elliptic curve" associated to the solution (a, b, c, p) of Fermat's equation. But in fact it had been considered before Frey by Hellegouarch in 1969, and even (a bit implicitly) by Fricke and Klein (as mentioned in the book by Kubert and Lang on modular units).

The elliptic curve E has good reduction at all primes l not dividing ABC, and it has multiplicative reduction at the primes dividing ABC. The discriminant of a minimal Weierstrass equation for E is $2^{-8}(abc)^{2p}$.

Frey's new idea was to relate $E_{A,B,C}$ to the conjecture that all elliptic curves over $\mathbb{Q}$ are *modular*, i.e., can be mapped with a finite kernel to the jacobian of a suitable modular curve. Frey's exact idea of showing that $E_{A,B,C}$ cannot be modular, by using groups of connected components of Néron models, turned out to be wrong. But Frey's idea motivated Serre to make his rough conjecture more precise, and to have sufficiently many tests done by Mestre in order to be convinced

of what this precise version should be. The Galois representation in question here is $E(\overline{\mathbb{Q}})[p]$. The very special property that it has is that it is unramified outside $2p$, semi-stable at 2, and is "finite at p". We will see in the next section what consequences this has.

2 Statement of the conjectures

The rough form of the conjecture is quite easy to state.

2.1 Conjecture. (Serre, weak form) *Let p be a prime number and $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous* odd *representation (odd meaning that $\det \rho c = -1$, where c is any complex conjugation). Then ρ can be obtained from a modular form.*

In the next lecture modular forms will be defined, and a precise statement will be given about modular forms giving rise to Galois representations.

In order to state the precise version, giving the minimal type of the modular form of which ρ can be obtained, we need more terminology.

Let p be prime, and $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ continuous. As the image of ρ is finite, we take a suitable power q of p such that ρ lands in $\mathrm{GL}_2(\mathbb{F}_q)$. Serre then associates two positive integers $N(\rho)$ and $k(\rho)$ to ρ, called the *level* and the *weight* of ρ, and a character $\varepsilon(\rho)\colon (\mathbb{Z}/N(\rho)\mathbb{Z})^\times \to \overline{\mathbb{F}}_p^\times$ called the *character* of ρ.

The level $N(\rho)$ is by definition prime to p, and given by the usual formula for the Artin conductor for a Galois representation. Let $G := \rho G_{\mathbb{Q}} = \mathrm{Gal}(K/\mathbb{Q})$, and let $V := \overline{\mathbb{F}}_p^2$ with its $G_{\mathbb{Q}}$-action via ρ. Let $l \neq p$ be a prime, and choose a place of K above l, i.e., a maximal ideal m of the ring of integers O_K. The decomposition group G_{-1} of G is the stabiliser of m, and the ramification group G_i ($i \geq 0$) is the kernel of $G_{-1} \to \mathrm{Aut}(O_K/m^{i+1})$. The subgroup G_0 is called the inertia group at l, and the p-group G_1 is called the *wild* inertia subgroup. We have $G_1/G_0 = \mathrm{Aut}_{\mathbb{F}_l}(O_K/m)$; this group is cyclic, generated by the Frobenius element Frob_p that sends x to x^p. In this notation, we have:

$$N(\rho) = \prod_{l \neq p} l^{n(l,\rho)}, \quad n(l,\rho) = \sum_{i \geq 0} \frac{1}{\#(G_0/G_i)} \dim(V/V^{G_i}).$$

We note that from the formula it is not clear at all that $n(l,\rho)$ is an integer (it is clear if ρG_1 is trivial, in which case we have $n(l,\rho) = \dim(V/V^{G_0})$). We note that $n(l,\rho) = 0$ if and only if ρ is unramified at l, i.e., if ρG_0 is trivial. We have $N(\rho) = 1$ if and only if ρ is unramified outside p. Although the formula for $N(\rho)$ is just a formula, and not a very nice one, I want to mention that the conductor of a continuous representation $G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{C})$ shows up naturally in the functional equation of its L-function.

The next step is to define $\varepsilon(\rho)$, and the class of $k(\rho)$ modulo $p-1$. For this, we consider the character $\det\rho$ from $G_{\mathbb{Q}}$ to $\overline{\mathbb{F}}_p^{\times}$. This character factors via the cyclotomic character : $G_{\mathbb{Q}} \to \hat{\mathbb{Z}}^{\times}$ that we have seen above. Comparing the formula for $n(l,\rho)$ with that for $n(l,\det\rho)$, and noting that $n(p,\det\rho) \leq 1$ (because it is tame and of dimension one) it follows that $\det\rho$ factors through $(\mathbb{Z}/N(\rho)p\mathbb{Z})^{\times} = (\mathbb{Z}/N(\rho)\mathbb{Z})^{\times} \times \mathbb{F}_p^{\times}$. Then $\varepsilon(\rho)$ is by definition the character induced on the first factor, and $k(\rho)$ mod $p-1$ is by definition the element of $\mathbb{Z}/(p-1)\mathbb{Z}$ such that the character induced on the second factor is raising to the power $k(\rho)-1$. A simple characterisation of this is that for l not dividing $N(\rho)p$ we have:

$$\det\rho\mathrm{Frob}_p = \varepsilon(l)l^{k(\rho)-1}.$$

As the level $N(\rho)$ reflects the ramification of ρ outside p, the weight $k(\rho)$ reflects the ramification at p, but the definition is more involved, and is given in terms of a recipe that as far as I know can only be motivated by looking at the properties of Galois representations associated to modular forms. For the best understanding as of today of this matter the reader is invited to look this up in the recent work of Fred Diamond on weights in the context of Hilbert modular forms.

Let $\overline{\mathbb{Z}}$ be the integral closure of $\mathbb{Z}$ in $\overline{\mathbb{Q}}$. Let us choose a surjection $\overline{\mathbb{Z}} \to \overline{\mathbb{F}}_p$ (these are permuted transitively by $G_{\mathbb{Q}}$). Then we have a sequence of subgroups:

$$I_p \lhd I \lhd G_p \subset G_{\mathbb{Q}},$$

where G_p is the decomposition subgroup, I the ramification subgoup, and I_p the wild ramification subgroup. We can identify G_p with $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$, G_p/I with $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$, and $I_t := I/I_p$ (the tame ramification group) with $\mathrm{Gal}(\mathbb{Q}_p^{\mathrm{tr}}/\mathbb{Q}_p^{\mathrm{unr}})$ and with $\varprojlim \mathbb{F}_{p^n}^{\times}$. A character $\phi\colon I_t \to \overline{\mathbb{F}}_p^{\times}$ is said to be of level n if n is the smallest integer such that ϕ factors through $\mathbb{F}_{p^n}^{\times}$. The n characters $I_t \to \mathbb{F}_{p^n}^{\times} \to \overline{\mathbb{F}}_p^{\times}$ that are induced by embeddings of fields $\mathbb{F}_{p^n} \to \overline{\mathbb{F}}_p$ are called the fundamental characters of level n. We let $\chi : G_{\mathbb{Q}} \to \overline{\mathbb{F}}_p^{\times}$ denote the cyclotomic character giving the action on the pth roots of unity: $\sigma(\zeta) = \zeta^{\chi(\sigma)}$.

Let $\rho_p : G_p \to \mathrm{GL}(V)$ be any continuous 2-dimensional representation on a $\overline{\mathbb{F}}_p$-vector space. Let V^{ss} be the semi-simplification of V with respect to the action of G_p. Then I_p acts trivially on V^{ss} and the action of I_t on V^{ss} is given by two characters ϕ, $\phi' : I_t \to \overline{\mathbb{F}}_p^{\times}$. Since $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$ acts by conjugation on I_t it follows that $\{\phi^p, \phi'^p\} = \{\phi, \phi'\}$. This means that there are two cases:

1. ϕ and ϕ' are both of level 1, ρ_p is reducible.

2. ϕ and ϕ' are both of level 2, $\phi^p = \phi'$, $\phi'^p = \phi$, and ρ_p is irreducible.

The representation ρ_p is called *finite at* p if it comes from the representation of G_p of the form $W(\overline{\mathbb{Q}}_p)$ where W is a finite locally free groupscheme over $\mathbb{Z}_p$ (for more details see [6], or section 8 of [2]).

Serre now associates an integer $k(\rho_p)$ to ρ_p as follows.

1. Suppose that ϕ and ϕ' are of level 2. We have:

$$\rho_p|_I \cong \begin{pmatrix} \phi & 0 \\ 0 & \phi' \end{pmatrix}$$

After interchanging ϕ and ϕ' if necessary, we have (uniquely) $\phi = \psi^{a+pb} = \psi^a \psi'^b$ and $\phi' = \psi'^a \psi^b$ with $0 \leq a < b \leq p-1$. We set $k(\rho_p) = 1 + pa + b$.

2. Suppose that ϕ and ϕ' are of level 1.

   (a) If $\rho_p|_{I_p}$ is trivial then we have:

$$\rho_p|_I \cong \begin{pmatrix} \chi^a & 0 \\ 0 & \chi^b \end{pmatrix}$$

   with $0 \leq a \leq b \leq p-2$. We set $k(\rho_p) = 1 + pa + b$ if $(a, b) \neq (0, 0)$ and $k(\rho_p) = p$ if $(a, b) = (0, 0)$.

   (b) Suppose that $\rho_p|_{I_p}$ is not trivial. We have:

$$\rho_p|_I \cong \begin{pmatrix} \chi^\beta & * \\ 0 & \chi^\alpha \end{pmatrix}$$

   for unique α and β with $0 \leq \alpha \leq p-2$ and $1 \leq \beta \leq p-1$. We set $a = \min(\alpha, \beta)$, $b = \max(\alpha, \beta)$. If $\chi^{\beta-\alpha} = \chi$ and $\rho_p \otimes \chi^{-\alpha}$ is not finite at p then we set $k(\rho_p) = 1 + pa + b + p - 1$ if $p \neq 2$ and $k(\rho_p) = 4$ if $p = 2$; otherwise we set $k(\rho_p) = 1 + pa + b$.

2.2 Remark. In [2] I defined a variant $k'(\rho_p)$ of $k(\rho_p)$ that is adapted to a slightly more general kind of modular forms mod p (so-called Katz modular forms, defined geometrically in characteristic p, and not by reduction from characteristic zero). By comparing the definitions of $k(\rho)$ and $k'(\rho)$ one sees that always $k'(\rho) \leq k(\rho)$ and that they differ only in two cases. In both cases ϕ and ϕ' are of level 1. In the first case $\rho_p|_{I_p}$ is trivial and $a = 0 = b$: then $k'(\rho) = 1$ and $k(\rho) = p$. In the second case $p = 2$, $\rho_p|_{I_p}$ is non-trivial, $\alpha = 0$, $\beta = 1$ and ρ_p is not finite at p: then $k'(\rho) = 3$ and $k(\rho) = 4$.

We can now state the precise version of Serre's conjecture. We state it in what is called the "epsilon" form, the strong version of Serre's conjecture is then the conjunction of the weak and the epsilon conjectures.

2.3 Conjecture. (Serre, epsilon conjecture) *Let p be a prime number and $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be an irreducible continuous odd representation. Suppose that ρ can be obtained from a modular form. Then it can be obtained from a modular form of level $N(\rho)$ and weight $k(\rho_p)$.*

2.4 Remark. The reader will notice that now ρ is required to be irreducible. Reducible ρ are known to come from Eisenstein series (easy, if one knows enough), but formulating a correct epsilon conjecture for them seems to be more involved. Anyway, one would then first have to choose if one wants to consider arbtitrary irreducible representations, or only semi-simple ones (i.e., direct sums of two characters). Maybe one can find more on this in the work of Skinner and Wiles.

2.5 Remark. One can also try to specify the character $\varepsilon(\rho)$ of the modular form. This can be done but leads to some problems. See [3], or [1], or the comments on [6] in volume 4 of Serre's collected works.

3 Known cases and some applications

The first application is of course the obvious one:

modularity plus epsilon implies Fermat.

More precisely, let $E_{A,B,C}$ be the elliptic curve associated to a solution of Fermat's equation, as above. Then it is known by work of Mazur that the odd representation ρ of $G_{\mathbb{Q}}$ on $E_{A,B,C}(\overline{\mathbb{Q}})[p]$ is irreducible. We have $N(\rho) = 2$ and $k(\rho) = 2$. So if $E_{A,B,C}$ is modular (STW conjecture!) then the epsilon conjecture above implies that ρ comes from a modular form of level 2 and weight 2, quod non.

In 1987, when [6] appeared, already enough of epsilon had been proved by Mazur and Ribet (see [4]) so that Fermat was a consequence of the modularity conjecture for elliptic curves over $\mathbb{Q}$ with semistable reduction over $\mathbb{Z}$.

The epsilon conjecture has also played an important role in the work of Wiles, because it allowed him to reduce his problem to the special case where the level is minimal.

For a proof of the following result, see the appendix by Buzzard in [5] and the references in there.

3.1 Theorem. (many people) *The epsilon conjecture as stated above is true except possibly in the case where $p = 2$, ρ is unramified at 2 and $\rho(\mathrm{Frob}_2)$ scalar.*

So, apart from these remaining cases, the strong and weak form of Serre's conjecture are equivalent. Much less is known about the remaining conjecture: are all representations as above modular?

Let us note for example, that as all elliptic curves E over $\mathbb{Q}$ are known to be modular, all representations of the form $E(\overline{\mathbb{Q}})[p]$ are modular.

The result of Khare that is the subject of this seminar is about this question. It says the following.

3.2 Theorem. (Khare) *Serre's conjecture is true for ρ such that $N(\rho) = 1$, i.e., for odd continuous $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ that are unramified outside p.*

An interesting consequence of this is that there exist no irreducible finite flat group schemes of type (p, p) over $\mathbb{Z}$; this corresponds to the fact that the space of cupsidal modular forms of weight two and level one is zero.

Khare's result gives me hope that the general case of Serre's conjecture will be proved in the not so far future. This would mean a big step in our understanding of 2-dimensional representations of $G_{\mathbb{Q}}$. Of course, Khare's result should be seen in the context of all developments in this field since 20 years. It is fair to say that his methods use all techniques that have been developed in that period: deformation theory, Wiles's "R=T" type theorems, Taylor's potential modularity etc. The study of its proof is a very good opportunity to learn more about this subject.

4 Program of the seminar

How can this be described better than giving the list of lectures (of course, the complete program is on my homepage):

- September 23, Utrecht
  - Bas Edixhoven: Introduction to Serre's conjecture.
  - Johan Bosman: Galois representations associated to modular forms.
  - Gunther Cornelissen: Theorem G of Taylors article Remarks on a conjecture of Fontaine and Mazur.
- October 7, Leiden
  - Sander Dahmen: Lower bounds for discriminants.
  - Frits Beukers: Upper bounds for discriminants.

  - Bas Edixhoven: Overview of Khare's proof.
  - Johan de Jong: Kloosterman lecture, Rational points and rational connectivity.

- October 14, Amsterdam (UvA)
  - Gerard van der Geer: Hilbert modular forms and Hilbert modular varieties.
  - Bart de Smit: General theory of deformations of Galois representations.
  - Johan de Jong: Rational points and rational connectivity, II.

- November 11, Leiden
  - Gebhard Boeckle: Deformation rings of Galois representations in nice cases
  - Fabio Mainardi: Automorphic theory for GL_2.
  - Theo van den Bogaart: Construction of Galois representations in cohomology of Shimura curves.
  - Johan de Jong: Rational points and rational connectivity, III.

- November 25, Nijmegen
  - Johan de Jong: Strictly compatible families of Galois representations.
  - ????: Overview of modularity lifting.
  - ????: Taylor's potential modularity, I.
  - Johan de Jong: Rational points and rational connectivity, IV.

- December 9, Utrecht
  - ????: Taylor's potential modularity, II
  - Luis Victor Dieulefait: Existence of minimal lifts and the proof in the weight 2 case.
  - Jean-Pierre Wintenberger: Khare's proof in the general case.

References

[1] F. Diamond. *The refined conjecture of Serre.* In: Elliptic Curves, Modular Forms and Fermat's Last Theorem, J. Coates, S.T. Yau, eds., International Press, Cambridge, pages 22–37 (1995).

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

[3] S.J. Edixhoven. *Serre's conjecture.* In: Modular Forms and Fermat's Last Theorem (Gary Cornell, Joseph Silverman and Glenn Stevens, editors). Springer-Verlag, 1997. Pages 209–242.

[4] K.A. Ribet. *On modular representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms.* Invent. math. 100, 431–476 (1990).

[5] K.A. Ribet and W.A. Stein. *Lectures on Serre's conjectures.* Available at: `http://modular.fas.harvard.edu/papers/serre/ribet-stein.ps`

[6] J-P. Serre. *Sur les représentations de degré 2 de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Duke Mathematical Journal 54, No. 1, (1987), 179–230.

[7] R. Taylor. *Galois representations.* Extended version of his ICM lecture Available on the author's webpage: `http://abel.math.harvard.edu/~rtaylor/longicm02.ps`