

Galois representations and the theorem of Fermat-Wiles *

Bas Edixhoven (translation by Albert Gunawan)

Universiteit Leiden

email: edix@math.leidenuniv.nl

October 2011

Abstract

The purpose of this text is to explain, to first year students in Mathematics, some things about the Galois representations that play an important role in the proof by Wiles of Fermat's theorem.

Prerequisite: Some mathematics language (sets, functions), complex numbers. The level will rise a bit, as we go along.

1 Automorphism of complex numbers

Our point of departure is the set \mathbb{C} of complex numbers. Recall that each complex number can be written in a unique way as $a + bi$ where a and b are real numbers, that is, with a and b in \mathbb{R} . The addition in \mathbb{C} is done coordinate by coordinate:

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

while the multiplication follows from the relation $i^2 = -1$:

$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i.$$

In \mathbb{C} we can also divide (except by 0 of course):

$$\frac{1}{a + bi} = \frac{a - bi}{(a + bi)(a - bi)} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2}i.$$

*This text is a translation of [Ed]. The author thanks Xavier Caruso for making the original text more accessible by simplifying the vocabulary and adding more explanatory passages. And he thanks Albert Gunawan for this translation from french to english.

For the notion of symmetry in number theory, which is the basis of Galois theory, we will forget the distance in \mathbb{C} and will only consider the *algebraic* properties: the addition and the multiplication. The symmetries of \mathbb{C} are called automorphisms and they are defined as follows.

Definition 1.1 The *automorphisms* of \mathbb{C} are the maps $\sigma: \mathbb{C} \rightarrow \mathbb{C}$ that satisfy:

- $\sigma(z + w) = \sigma(z) + \sigma(w)$,
- $\sigma(z \cdot w) = \sigma(z) \cdot \sigma(w)$,
- σ is bijective.

These properties imply that $\sigma(0) = \sigma(0 + 0) = \sigma(0) + \sigma(0)$, and then $\sigma(0) = 0$. They imply the equality $\sigma(1) = \sigma(1 \cdot 1) = \sigma(1) \cdot \sigma(1)$, from which we deduce that $\sigma(1) = 1$ or $\sigma(1) = 0$, and then finally $\sigma(1) = 1$ because $\sigma(1)$ is different from $\sigma(0)$.

The set of automorphisms of \mathbb{C} is denoted by $\text{Aut}(\mathbb{C})$. If σ_1 and σ_2 are elements of $\text{Aut}(\mathbb{C})$, then so are their composition $\sigma_2 \circ \sigma_1$ and inverses: mathematicians say that $\text{Aut}(\mathbb{C})$ is a *group*.

We know two automorphisms of \mathbb{C} : the identity $\text{id}_{\mathbb{C}}: z \mapsto z$, and the complex conjugation: $\sigma(a + bi) = a - bi$, written as $z \mapsto \bar{z}$. It is a good exercise (left to the reader) to show these are the only automorphisms that are continuous. We can show that $\text{Aut}(\mathbb{C})$ is very large (as large as the set of subsets of \mathbb{R}). For example, knowing that $\sqrt{2}$ is irrational (see [Du-Le] for 65 proofs of this statement), we can consider the map $a + b\sqrt{2} \mapsto a - b\sqrt{2}$ (which is defined on the set K of numbers of the form $a + b\sqrt{2}$ with a and b in the set \mathbb{Q} of rational numbers) and show that this extends to an automorphism of \mathbb{C} . This last property of extension generalizes to all automorphisms $\sigma: K \rightarrow K$ where K is a subset of \mathbb{C} stable under addition, subtraction, multiplication and division (we call this a *subfield* of \mathbb{C}). This permits us to obtain a lot of automorphisms of \mathbb{C} .

A lot of questions that we can ask about the algebraic properties of complex numbers are open. For example, we know that the numbers e and π are transcendental, meaning that they are not a root of a polynomial $x^n + a_{n-1}x^{n-1} + \dots + a_0$ with coefficients a_i in \mathbb{Q} . This implies that there exists a σ in $\text{Aut}(\mathbb{C})$ such that $\sigma(e) = \pi$, and there exists a σ in $\text{Aut}(\mathbb{C})$ with $\sigma(\pi) = e$. It is believed that e and π are algebraically independent (meaning that there exists no non-zero polynomial f in $\mathbb{Q}[x, y]$ such that $f(e, \pi) = 0$). However, we don't know, at present, how to prove this. We also don't know if there exists a σ in $\text{Aut}(\mathbb{C})$ that exchanges e and π .

2 Galois symmetry in number theory

Let us start with Galois theory. Let $f = x^n + a_{n-1}x^{n-1} + \dots + a_0$ be a polynomial with coefficients a_i in \mathbb{Q} . The fundamental theorem of Algebra says that the equation $f(z) = 0$ has exactly n solutions in \mathbb{C} , counting with multiplicity. Let us write $\text{Roots}(f)$ for the set of these

roots. For σ in $\text{Aut}(\mathbb{C})$ and z in $\text{Roots}(f)$ we have

$$\begin{aligned} 0 &= \sigma(0) = \sigma(f(z)) = \sigma(z^n + \cdots + a_1z + a_0) \\ &= \sigma(z^n) + \cdots + \sigma(a_1z) + \sigma(a_0) \\ &= \sigma(z)^n + \cdots + \sigma(a_1)\sigma(z) + \sigma(a_0) \\ &= \sigma(z)^n + \cdots + a_1\sigma(z) + a_0 = f(\sigma(z)) \end{aligned}$$

hence $\sigma(z)$ is again in $\text{Roots}(f)$. We conclude that each σ in $\text{Aut}(\mathbb{C})$ permutes the elements of $\text{Roots}(f)$. The *Galois group* of f is defined as the set $\text{Gal}(f)$ of permutations of $\text{Roots}(f)$ induced by the elements of $\text{Aut}(\mathbb{C})$.

Let us consider some examples. Let f be the polynomial $x^2 - 4$. Then $\text{Roots}(f)$ is the set $\{-2, 2\}$. The two elements of this set are fixed by all the σ in $\text{Aut}(\mathbb{C})$; indeed, we have already seen that $\sigma \in \text{Aut}(\mathbb{C})$ necessarily sends 1 to 1, it sends also $1+1 = 2$ to $\sigma(1)+\sigma(1) = 1+1 = 2$. The same reasoning shows that it sends -2 to itself. Finally, $\text{Gal}(f)$ only contains the identity permutation $2 \mapsto 2, -2 \mapsto -2$.

For $g = x^2 + 4$, we have $\text{Roots}(g) = \{2i, -2i\}$. This time, there exists an element of $\text{Aut}(\mathbb{C})$ that exchanges these two roots: the complex conjugation. Therefore $\text{Gal}(g)$ contains the two possible permutations of two roots: the identity, and the one that exchanges the two roots: $2i \mapsto -2i, -2i \mapsto 2i$.

For $h = x^2 - 2$, the roots are $\sqrt{2}$ and $-\sqrt{2}$ and we have already said that there exists a σ in $\text{Aut}(\mathbb{C})$ that permutes these two numbers. Hence $\text{Gal}(h)$ contains the two possible permutations of $\{\sqrt{2}, -\sqrt{2}\}$.

Let us go back to the general theory. This theory has initially been invented for solving the problems of solution by radicals of algebraic equations. Nowadays, however, it has many other applications, both in theory and in computation.

Readers interested in the possibility of calculating Galois groups are invited to install the freely-available software PARI/GP (see [PARI]) and to try the command `polgalois`, which compute $\text{Gal}(f)$ for polynomials of degree at most 11.

We propose in the rest of this article to sketch how Galois theory is used in the proof of Fermat's Last Theorem by Wiles and Taylor-Wiles .

3 Galois representations of dimension one and two

Important examples of polynomials are those attached to the division of the circle in n equal parts or, equivalently, to regular polygons.

Let $n > 1$, $f = x^n - 1$, and $z = e^{2i\pi/n} = \cos(\frac{2\pi}{n}) + i \sin(\frac{2\pi}{n})$. The map

$$\{0, 1, \dots, n-1\} \rightarrow \text{Roots}(f), \quad a \mapsto z^a$$

is then bijective; we call it *labeling* of the roots of f . Figure 1 above shows the case $n = 5$.

In term of the labeling $\{0, 1, \dots, n-1\} \rightarrow \text{Roots}(f)$, $a \mapsto z^a$, the group $\text{Gal}(f)$ then exactly consists of the permutations of $\{0, 1, \dots, n-1\}$ of the form $a \mapsto ka \pmod n$, with $0 \leq k < n$

and $\gcd(n, k) = 1$. Here $b \bmod n$ denotes the remainder obtained from the euclidean division of b by n . For example, the complex conjugation corresponds to the multiplication by $k = \overline{-1}$ (or rather $n-1$, but that has the same effect on the remainder modulo n), because $\bar{z} = e^{2\pi i/n} = e^{-2\pi i/n} = z^{-1}$.

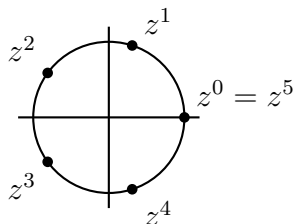


Figure 1: $n = 5$

In this example, the way the Galois group permutes the roots (we often say the *action* of Galois group) is then very simple: if we number the roots properly, the action is by multiplication. Such data is called a *Galois representation of dimension 1*.

We can say that for a long time mathematicians had tried to prove Fermat’s theorem by studying these Galois representations (roots of unity, cyclotomic field), and that finally it is through the study of Galois representations of dimension 2 that they finally succeeded.

What are these representations? The basic idea is instead of labeling the roots by integers from 0 and $n - 1$, we label the roots by pair of integers, the notion of multiplication in this new situation is slightly different and explained precisely below.¹

Definition 3.1 A Galois representation of dimension 2 is a polynomial

$$f = x^{n^2} + \dots + a_1x + a_0$$

of degree n^2 , with the a_i in \mathbb{Q} , together with a labeling of $\text{Roots}(f)$ with the pairs $v = (v_1, v_2)$ with v_1 and v_2 in $\{0, 1, \dots, n-1\}$, such that the Galois group $\text{Gal}(f)$ is consists of permutations of the form

$$(v_1, v_2) \mapsto ((av_1 + bv_2) \bmod n, (cv_1 + dv_2) \bmod n)$$

where a, b, c and d are integers such that $ad - bc$ is relatively prime to n .

Remark. To remember the fixed integer n in this story, mathematicians usually say that the Galois representation has coefficients in $\mathbb{Z}/n\mathbb{Z}$.

These objects play very important role in the proof of Fermat’s Last Theorem by Wiles (see [Wi]). In the last 40 years the study of such representations is central in arithmetic (that was the beginning of the famous “Langlands program”). The essential difference with the dimension 1 case is the *non-commutativity*: when we compose two permutations taking the form given by above definition, the order of factors is important.

4 Elliptic curves

We have seen that the polynomial $x^n - 1$ gives a Galois representation of dimension 1. We will see that certain polynomials associated to elliptic curves provide Galois representations of dimension 2.

¹Those who knows matrices will recognize the multiplication a vector by 2×2 matrices.

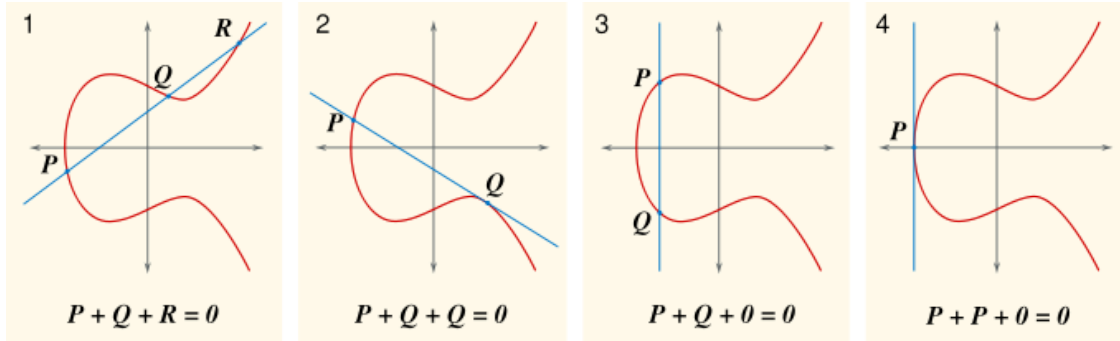


Figure 2: An elliptic curve and its addition operation

An elliptic curve (over the field \mathbb{Q} of rational numbers) is a curve in the plane given by one equation of the form $y^2 = x^3 + ax + b$ with a and b in \mathbb{Q} , such that the polynomial $x^3 + ax + b$ does not have double roots in \mathbb{C} , meaning that $4a^3 + 27b^2$ is non-zero. In Figure 2 below (found on Wikipedia) we see the equation of the curve $y^2 = x^3 - x + 1$ in the real plane (as well as some of the intersections with lines).

As the equation of the curve is of degree 3, each line in the plane has at most 3 intersection points with the curve. In fact, if we count these intersection points with multiplicity and if we accept points whose coordinates are complex numbers, each line given by an equation $y = cx + d$ with c non-zero has exactly 3 intersection points, and each vertical line (the equation $x = c$) has 2 intersection points. For having 3 intersection points in all cases, we usually add to the elliptic curve a “point at infinity” denoted by 0 (zero), at which all vertical lines meet. We denote by E the complete complex curve, meaning the set $\{(x, y) \in \mathbb{C}^2 : xy^2 = x^3 + ax + b\} \cup \{0\}$.

If now we have two points P and Q on the curve E , we take the line that passes through P and Q (the tangent line if $P = Q$, and forget the case $P = Q = 0$). This gives us the third intersection point, say R . The point symmetric to R with respect to the x -coordinate axis is then denoted by $P + Q$. One can show that for all P, Q and R in E , we have $P + 0 = P$ (this is immediate), $(P + Q) + R = P + (Q + R)$ (this is difficult), $P + Q = Q + P$ (this is immediate), and that for all P there exists a Q such that $P + Q = 0$ (take for Q the point symmetric to P with respect to the x -coordinate axis). This justifies the notation $P + Q$ (as well as) with the notation 0 for the point in infinity: E with the operation $+$: $E \times E \rightarrow E$ is a *commutative group*.

Another (difficult) classical result says that for each integer $n \geq 1$, the set $E[n]$ of points P of E such that the sum $P + \dots + P$ with n terms is 0 consists of n^2 elements. Better than that: there exists a bijection (labeling) between this set $E[n]$ and the set of pairs (v_1, v_2) with v_1 and v_2 in $\{0, \dots, n-1\}$, and this bijection is *compatible* with the addition operation of the two sides, in the sense if P corresponds to (v_1, v_2) and Q to (w_1, w_2) , then $P + Q$ corresponds to $((v_1 + w_1) \bmod n, (v_2 + w_2) \bmod n)$. For the proof, we can say that this result is obtained by studying E in the terms of Weierstrass’s doubly periodic complex meromorphic functions.

On the other hand, as the curve E is defined by an equation with coefficients in \mathbb{Q} , the addition operation on E is compatible with all σ in $\text{Aut}(\mathbb{C})$: for all P and Q on E we have $\sigma(P + Q) = \sigma(P) + \sigma(Q)$ where we define, for $P = (x, y)$ with x and y in \mathbb{C} , $\sigma(P) = (\sigma(x), \sigma(y))$.

This implies that for σ in $\text{Aut}(\mathbb{C})$ and P in $E[n]$, we have $\sigma(P)$ in $E[n]$. From this, we can deduce that the points of $E[n]$ are in bijection with the roots of a polynomial f in $\mathbb{Q}[x]$ of degree n^2 . For obtaining such a bijection, we choose a point C in \mathbb{Q}^2 that does not lie on a line containing two points in $E[n]$, and whose second coordinate is distinct from all second coordinates of points in $E[n]$. Then we associate to each P in $E[n]$ the first coordinate of the intersection point of the line passing through C and P with the x -coordinate axis. We denote this coordinate by x_P . The map $P \mapsto x_P$ is compatible with the action of $\text{Aut}(\mathbb{C})$: for all P and all σ we have $\sigma(x_P) = x_{\sigma(P)}$. The polynomial f is then the product, indexed by P in $E[n]$, of the $(x - x_P)$. Indeed, this product is invariant under $\text{Aut}(\mathbb{C})$, hence it is in $\mathbb{Q}[x]$.

If we go back to the definition of a Galois representation of dimension 2 and we put together all that has been said above, we see that the polynomial that we have constructed defines a Galois representation of dimension 2.

Let us finish with a small example. Let E be the completed curve given by $y^2 = x^3 - x + 1$. Take $n = 2$. Then $E[2]$ is the set of 4 points $0, (z_1, 0), (z_2, 0)$ and $(z_3, 0)$ where z_1, z_2 and z_3 are the roots of $x^3 - x + 1$. If we take $C = (0, 1)$, then the numbers x_P are $0, z_1, z_2$ et z_3 and the polynomial with these roots is $f = x(x^3 - x + 1)$. The Galois group of this polynomial consists of the permutations of the set $\text{Roots}(f) = \{0, z_1, z_2, z_3\}$ that fix 0: it is clear that if a permutation of the roots of f comes from an automorphism of \mathbb{C} , then it leaves 0 fixed, and to show the equality, we can use PARI/GP for assuring that the Galois group of $x^3 - x + 1$ contains all the permutations $\{z_1, z_2, z_3\}$. Finally, if we label the roots as follows

$$0 \leftrightarrow (0, 0) \quad ; \quad z_1 \leftrightarrow (1, 0) \quad ; \quad z_2 \leftrightarrow (0, 1) \quad ; \quad z_3 \leftrightarrow (1, 1)$$

then we can verify, by looking separately all the case for example, that the 6 elements of $\text{Gal}(f)$ can be put in the form imposed in the definition a Galois representation of dimension 2. For example the permutation that sends z_1 to z_2, z_2 to z_3 and z_3 to z_1 can be written $(v_1, v_2) \mapsto (v_2, (v_1 + v_2) \bmod 2)$.

5 Fermat's last theorem

Fermat's last theorem says that if n is an integer greater or equal than 3, then the equation $x^n + y^n = z^n$ has no solution in positive integers. It appeared for the first time in a 17th century in the manuscript of Pierre de Fermat, without proof. Since then, many generations of mathematicians have tried to prove this result. Their work has led, over the centuries, to significant progress, including the development of the theory of algebraic numbers. The first complete proof was given by Wiles and Taylor-Wiles in 1995. It used Galois representations of dimension 2 as the main idea, especially those related to elliptic curves.

The discovery of a link between Fermat's last theorem and elliptic curves seems to date from 1969. It appeared in a lecture by Hellegouarch in Bordeaux (see [He2], and appendix [He1]). The question that is asked by Hellegouarch was to know if an elliptic curve E (over \mathbb{Q}) can have a point $P = (x, y)$ in $E[n]$ with x and y in \mathbb{Q} and n "big". Under certain hypotheses, he showed that if that happened, then Fermat's last theorem would fail in degree n .

In 1985, Frey studied the question in other direction: can we prove Fermat’s last theorem by showing that the elliptic curves associated by Hellegouarch to a counter example to Fermat’s last theorem can not exist? To support his idea, he remarked that the existence of such an elliptic curve would contradict with a the so-called *modularity conjecture* (on which we will say a few words in the following), that is very important in mathematics.

After that, things went quickly. In an article in 1987, Serre clarified and generalized the ideas of Frey placing them in the context of Galois representations, and formulated a precise conjecture. Then, Ribet showed a sufficient portion of Serre’s conjecture to conclude that the modularity conjecture implies Fermat’s last theorem. Finally, Wiles and Taylor-Wiles showed enough of modularity conjecture to deduce Fermat’s last theorem. We don’t give more detail of the whole story (see section 7). Our goal is rather to point out the role played by Galois representations.

6 The role of Galois representations

To prove Fermat’s last theorem, it is sufficient to suppose that there exists a solution of the Fermat equation $x^n + y^n = z^n$ with $n \geq 3$ and deduce a contradiction. As we already know (by the work of Fermat himself and of Gauss) that the theorem is true for degrees 3 and 4, we can suppose that n is a prime number $p \geq 5$. So let $p \geq 5$ be a prime, and a, b and c in \mathbb{Z} such that $a^p + b^p + c^p = 0$, with $abc \neq 0$. We can suppose that a, b and c are relatively coprime, and by symmetry that b is even and $a + 1$ is divisible by 4. Let $A = a^p, B = b^p$, and $C = c^p$ and, by following Hellegouarch and Frey, consider the elliptic curve $E_{A,B,C}$ over \mathbb{Q} given by:

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

It turns out that this curve $E_{A,B,C}$ has properties that make its existence impossible. The first of these is the *non-ramification* outside 2 and p of the Galois representation $E_{A,B,C}[p]$. This means that for all prime numbers ℓ different from 2 and p , in suitable coordinates, the points P of $E_{A,B,C}[p]$ remain distinct after reduction modulo ℓ (if the coordinates of these points are integers, this means simply “after coordinate-wise taking their remainders by Euclidean division by ℓ ”, except that the notion is more complicated). The proof of this property uses “standard” tools: the discriminant of the polynomial $x(x - A)(x + B)$ (which appeared in the definition of elliptic curve), which is equal to $A^2B^2(A + B)^2 = (ABC)^2 = (abc)^{2p}$ and has remarkable property of being the p -th power. Using an analogous argument, one shows also a slightly weaker property for the reduction modulo $\ell = p$.

The second property that we use is the *modularity* of elliptic curves arising from (part of) the modularity conjecture proven by Wiles and Taylor-Wiles. It would be too long to explain exactly here the modularity conjecture. For the purpose of this article, it is important to remember that this conjecture predicts in particular that the Galois representation $E_{A,B,C}[p]$ can be obtained by completely different ways involving so-called *modular forms* (which are in nature analytic objects having *a priori* no relation with elliptic curves). The reader wishing to know more may consult [Di-Sh].

Finally, by studying modular forms, it can be shown that a Galois representation which is constructed from modular forms cannot verify the property of non-ramification that was mentioned above. The contradiction that appears here proves Fermat's last theorem.

In fact, the modularity conjecture, which relates elliptic curves to modular forms, can be stated without Galois representations. However, on the one hand, as we have explained, it is through Galois representations that the contradiction (of which results Fermat's last theorem) appears, and on the other hand, the proof of Wiles and Taylor-Wiles could not be done without Galois representations. For example, in this proof, we pass first from the elliptic curve E to system of Galois representations $(E[3^n])_{n \geq 1}$. If we were physicists, we would say that it is Galois representations that carry forces between Fermat's last theorem, elliptic curves and modular forms.

To finish, since the most recent results by Khare-Wintenberger and Kisin, who proved Serre's conjecture mentioned above, there is a more direct proof of Fermat's last theorem, but again it proceeds via Galois representations.

7 Read more

Good references for reading more on this topic (the story of the proof of Fermat's last theorem, theory of modular forms and modularity of elliptic curves over \mathbb{Q}) are the books by Hellegouarch [He1, He2] and the book by Diamond and Shurman [Di-Sh], and the expositions at Séminaire Bourbaki [Se] and [Oe].

References

- [Di-Sh] F. Diamond and J. Shurman, *A first course in modular forms* GTM 228, Springer, Berlin, 2005
- [Du-Le] Ludmila Duchêne et Agnès Leblanc, *Rationnel mon \mathbb{Q}* , Hermann éditeurs, 2010
- [Ed] S.J. Edixhoven. *Représentations galoisiennes et théorème de Fermat-Wiles*.
<http://images.math.cnrs.fr/Representations-galoisiennes-et.html>
- [He1] Yves Hellegouarch, *Invitation aux mathématiques de Fermat-Wiles*, Enseignement des Mathématiques. Masson, Paris, 1997. ISBN : 2-225-83008-8
- [He2] Yves Hellegouarch, *Points d'ordre fini des variétés abéliennes de dimension un.*, Colloque de Théorie des Nombres (Univ. Bordeaux, Bordeaux, 1969). Bull. Soc. Math. France, Mem. 25, Soc. Math. France, Paris, 1971
- [Oe] J. Oesterlé, *Travaux de Wiles (et Taylor...)*. II., Séminaire Bourbaki, Vol. 1994/95, Astérisque No. 237 (1996), Exp. No. 804, 5, 333–355

- [PARI] C. Batut, K. Belabas, D. Bernardi, H. Cohen, and M. Olivier, *User's guide to PARI/GP (version 2.3.1)*, <http://pari.math.u-bordeaux.fr>
- [Se] J-P. Serre, *Travaux de Wiles (et Taylor...)*. I., Séminaire Bourbaki, Vol. 1994/95, Astérisque No. 237 (1996), Exp. No. 803, 5, 319–332
- [Ta-Wi] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) 141 (1995), no. 3, 553–572
- [Wi] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) 141 (1995), no. 3, 443–551