Capítol 4

The Modular Approach to some Generalized Fermat Equations

Nuno Freitas

4.1 Introduction

In the middle of the 17th century, Fermat wrote that for $n \geqslant 3$ the equation $a^n + b^n = c^n$ had no solution in the set of strictly positive integers. This sentence became known as Fermat's Last Theorem (FLT). This problem proved to be unexpectedly difficult and a global solution was not found for 350 years. It was only in the 60's that Hellegouarch noticed that non-trivial solutions of the Fermat equation were related to the existence of torsion points in some elliptic curves. On the other hand in the 50's Taniyama formulated a precise conjecture saying that *All rational elliptic curves arise from modular forms*, and it was only in 1985 that Frey suggested that the elliptic curve $y^2 = x(x - a^p)(x + b^p)$ constructed from a solution of the Fermat equation should not be modular. It was along this ideas and with deep results from Serre, Mazur and Ribet on elliptic curves, Ga-

Under the supervision of Luis Dieulefait

lois representations and modular forms that FLT was reduced to the proof of the Taniyama conjecture. Finally, in 1995 in papers from Wiles and Taylor-Wiles the conjecture was proved for semi-stable elliptic curves, establishing the FLT. Wiles' theorem, now known by Modularity theorem, was improved by Breuil, Conrad, Taylor, and Diamond and states that the Taniyama conjecture is indeed true for all rational elliptic curves.

Among the important consequences of the Modularity theorem and the theory around it is the possibility of using and generalizing some key ideas in the proof of FLT in order to study other Diophantine equations. For example, the generalized Fermat equation $x^p + 2^\alpha y^p = z^p$ has been solved by Ribet, and equations of the form $\phi(x, y) = dz^p$, where ϕ is a degree-3 separable homogeneous form had been extensively studied by Billerey.

The interplay between elliptic curves, modular forms and Galois representations given by the Modularity theorem and the Ribet-Mazur theorem is the central point in the modern strategy to solve Diophantic equations, in particular the FLT. The purpose of this work is to introduce some tools and techniques used to prove the FLT and see how they generalize to other Diophantic equations. Precisely, we study Ribet's paper [6] where he solves the generalized Fermat equation $x^p + 2^\alpha y^p = z^p$.

4.2 Elliptic Curves

In this section we start by recalling some basic facts about Galois representations associated with elliptic curves, then we study some properties of $E_{A,B,C}$ curves and finally we introduce the Tate curve and use it to prove a result due to Hellegouarch.

4.2.1 Galois Representation

Let $\bar{\mathbb{Q}} \subset \mathbb{C}$ be the integral closure of $\mathbb{Q}$. It is known that $\bar{\mathbb{Q}}/\mathbb{Q}$ is a Galois extension and we denote its Galois group by $G_{\mathbb{Q}}$.

Let E be an elliptic curve defined over $\mathbb{Q}$, $n \geqslant 1$ and set $V = E(\bar{\mathbb{Q}})[n]$.

Since V is a free $\mathbb{Z}/n\mathbb{Z}$-module of rank 2, we see that if P_1, P_2 is a basis of V, we have

$$(\sigma(P_1), \sigma(P_2)) = (P_1, P_2) \begin{bmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{bmatrix}.$$

4.2.1 Theorem *The action of $G_{\mathbb{Q}}$ on $E[n]$ defines a representation*

$$G_{\mathbb{Q}} \xrightarrow{\rho_n} GL_2(\mathbb{Z}/n\mathbb{Z}).$$

The image is isomorphic to the Galois group of the extension: newline $\mathbb{Q}(E[n])/\mathbb{Q}$.

About this representation there are two important theorems due to Serre and Mazur. The version that follows of the theorem from Mazur is simplified and stated as it will be of use later.

4.2.2 Theorem *(Serre) Let E be an elliptic curve defined over $\mathbb{Q}$ which is not isomorphic over $\bar{\mathbb{Q}}$ to any curve having complex multiplication. Then there exists an integer $N \geqslant 1$, depending only on E, such that for every integer n prime to N, the representation ρ_n is surjective.*

4.2.3 Theorem *(Mazur) Let $p \geqslant 5$ be a prime and E a semi-stable elliptic curve over $\mathbb{Q}$. Then, the representation ρ_p as above is irreducible.*

Idea of proof: If ρ_p is reducible, meaning that there exists a subspace invariant for all $\rho_p(\sigma)$, then there exists a subgroup C of order p invariant under $G_{\mathbb{Q}}$. From the semistability hypothesis it is possible to deduce that there exists some curve over $\mathbb{Q}$ isogenous to E with a group of rational points isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2l\mathbb{Z}$. This contradicts a result of Mazur in [4].

□

Now fix a prime l. Considering the action of $G_{\mathbb{Q}}$ on the l^n-torsion for all $n \in \mathbb{N}$ and the asociated representations we can put them together to obtain an action on the Tate module $T_l(E) \cong \mathbb{Z}_l \oplus \mathbb{Z}_l$ and a continuous representation

$$\rho_{E,l} : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}_l) \subset GL_2(\mathbb{Q}_l).$$

We call $\rho_{E,l}$ the 2-dimensional Galois representation associated to E at l.

4.2.4 Theorem *Let l be a prime and E be an elliptic curve over $\mathbb{Q}$ with conductor N. The Galois representation $\rho_{E,l}$ is unramified at every prime $p \nmid lN$. For any such p let $\mathfrak{p} \subset \bar{\mathbb{Z}}$ be any maximal ideal over p. Then the characteristic equation of $\rho_{E,l}(Frob_{\mathfrak{p}})$ is*

$$x^2 - a_p(E)x + p = 0.$$

The Galois representation $\rho_{E,l}$ is irreducible.

4.2.2 $E_{A,B,C}$ curves

Now we introduce Frey's idea which allowed to relate solutions of the Fermat equation to particular elliptic curves. We want to associate to each point (a, b, c) (with a,b,c relatively prime) on the curve $x^p + y^p = z^p$ a cubic curve $E_{A,B,C}$ such that it is an elliptic curve if and only if the point (a, b, c) is a non-trivial solution ($abc \neq 0$). Going into this direction it is natural to search for curves of the form

$$y^2 = (x - \alpha)(x - \beta)(x - \gamma)$$

with the following conditions

$$\beta - \gamma = a^p, \qquad \gamma - \alpha = b^p, \qquad \alpha - \beta = c^p,$$

making the discriminant of the right-hand side is $(abc)^{2p} \neq 0$. Thus, putting $\gamma = 0$ we have

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

More generally, let A, B, C be three relatively prime non-zero integers. We say that the equation $A+B+C = 0$ is an **ABC relation.**

4.2.5 Definition *Given an ABC relation, we set*

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

These curves satisfy $\Delta = 2^4(ABC)^2$ and $j(E_{A,B,C}) = 2^8(BC + CA + AB)^3(ABC)^{-2}$. Thus saying that $ABC \neq 0$ is equivalent to saying that $E_{A,B,C}$ is smooth. Also, if we make a circular permutation of (A, B, C), the new curve $E_{B,C,A}$ is isomorphic to $E_{A,B,C}$ over $\mathbb{Q}$. In the case of Fermat equation we have $A = a^p, B = b^p$ and $C = -(a^p + b^p) = -c^p$.

A subset of these curves which is going to be of use in the proof of Fermat's last theorem is that of those curves such that

$$A \equiv 3 \bmod 4 \qquad B \equiv 0 \bmod 32.$$

In this situation the following holds.

4.2.6 Theorem *Let l be a prime.*

1. *If l does not divide ABC, the curve $E_{A,B,C}$ has good reduction modulo l.*
2. *If $l \neq 2$ divides ABC, the reduction of $E_{A,B,C}$ modulo l is a curve of genus zero and multiplicative type.*
3. *If $l = 2$, and if 2 divides ABC, the reduction modulo l of a minimal model of $E_{A,B,C}$ is a curve of genus zero and multiplicative type.*

Proof:(1.)If l does not divide ABC then it does not divide Δ, and the reduced curve modulo l is smooth.

(2.) If an odd prime l divides A or B, the equation of the reduced curve is of the type

$$Y^2 = X^2(X + \tilde{c}), \text{ with } \tilde{c} \in \mathbb{F}_l \text{ different from zero.}$$

Then the tangent lines at $(0, 0)$ are given by

$$Y^2 - \tilde{c}X^2 = (Y - \sqrt{\tilde{c}}X)(Y + \sqrt{\tilde{c}}X).$$

Thus we have distinct tangents over $\bar{\mathbb{F}}_l$ and the reduction is multiplicative. If $l \mid C$ we take a circular permutation and apply the previous case.

(3.) For $l = 2$, we consider the change of variables $X = 4x$ and $Y = 4x + 4y$ leading to the minimal model equation

$$y^2 + xy = x^3 + cx^2 + dx$$

with $c = (B-1-A)/4$ and $d = -AB/16$. It follows that the reduced modulo 2 equation is

$$y^2 + xy = \begin{cases} x^3 \text{ if } A \equiv 7 \bmod 8 \\ x^3 + x^2 \text{ if } A \equiv 3 \bmod 8, \end{cases}$$

Hence we can see that the tangents at $(0, 0)$ are given by

$$\begin{cases} y(x + y) \text{ if } A \equiv 7 \bmod 8 \\ y^2 + xy + x^2 \text{ if } A \equiv 3 \bmod 8, \end{cases}$$

In the first case it is clear that the tangents are distinct; for the second case we need to consider the extension $\mathbb{F}_2[u]$, where u is a root of the polynomial $z^2 - z + 1$, to factorize into two distinct tangents. Hence the reduction is multiplicative.

□

4.2.1 Corol·lari. *When $A \equiv 3$ mod 4 and $B \equiv 0$ mod 32, then $E_{A,B,C}$ is semi-stable and its conductor is $rad(ABC)$, the product of the primes dividing ABC.*

4.2.3 The Tate Curve E_q

To finish with elliptic curves we will introduce two theorems due to Tate. The following theorems are essential in the proof of Hellegouarch Theorem which allow to study the ramification of the Galois representation $\rho_p = \bar{\rho}_{E,p}$, the reduction mod p of the representation in the Tate module at p associated to the curve $E = E_{a^p,b^p,c^p}$.

Let $\mathbb{Q}_p$ be the p-adic integers and $|.|_p$ its p-adic absolute value. It is known that every elliptic curve over $\mathbb{C}$ is of the form $\mathbb{C}/\Lambda$ with Λ a lattice. Although $\mathbb{Q}_p$ is a complete field, a similar result for $\mathbb{Q}_p$ has no chances of success because there are no discrete subgroups in $\mathbb{Q}_p$. However, the multiplicative group $\mathbb{Q}_p^*$ has a lot of discrete subgroups, namely those of the form $q^{\mathbb{Z}}$, for $|q|_p \neq 1$. In fact, Tate has constructed a curve E_q for every $q \in \mathbb{Q}_p^*$ such that $|q|_p < 1$ and achieved to prove an uniformization theorem for all elliptic curves over $\mathbb{Q}_p$ with $|j(E)|_p > 1$. That is the content of the two following theorems.

4.2.7 Theorem *(Tate) Let $q \in \mathbb{Q}_p^*$ satisfy $|q|_p < 1$ and let,*

$$s_k(q) = \sum_{n \geqslant 1} \frac{n^k q^n}{1-q^n}, \qquad a_4(q) = -s_3(q),$$
$$a_6(q) = -\frac{5s_3(q) + 7s_5(q)}{12}$$

(a) The series $a_4(q)$ and $a_6(q)$ converge in $\mathbb{Q}_p$ to elements in $\mathbb{Z}_p$ and allow to define the **Tate curve** *E_q over $\mathbb{Q}_p$ by the equation*

$$E_q : y^2 + xy = x^3 + a_4(q)x + a_6(q)$$

with discriminant and j-invariant given by

$$\Delta(q) = q\prod_{n \geqslant 1}(1-q^n)^{24} \text{ and } \quad j(E_q) = \frac{1}{q} + 744 + 196884q + \ ...$$

(b) There is an isomorphism $\phi : \bar{\mathbb{Q}}_p^*/\langle q \rangle \xrightarrow{\sim} E_q(\bar{\mathbb{Q}}_p)$ where $\langle q \rangle \subset \mathbb{Q}_p^*$ is the multiplicative subgroup generated by q.

(c) *The map ϕ on (b) is compatible with the action of the Galois group of $\bar{\mathbb{Q}}_p/\mathbb{Q}_p$ in the sense that*

$$\phi(u^\sigma) = \phi(u)^\sigma \text{ for all } u \in \bar{\mathbb{Q}}_p^*,\ \sigma \in G_{\bar{\mathbb{Q}}_p/\mathbb{Q}_p}.$$

In particular, for any algebraic extension L/K ϕ induces an isomorphism

$$\phi : L^*/\langle q \rangle \xrightarrow{\sim} E_q(L)$$

The reduction modulo p of E_q gives the curve

$$\tilde{E}_q : y^2 + xy = x^3$$

and elementary calculus shows that there is a double point at $(0,0)$ and the tangents at this point are $y = 0$ and $x + y = 0$. Thus the Tate curve has multiplicative split reduction at p. Note that $|j(E_q)|_p = 1/|q|_p > 1$ thus the uniformization theorem can not work for curves with $j(E) \in \mathbb{Z}_p$. Fortunately, this is the only constraint.

4.2.8 Theorem *(Tate) Let $E/\mathbb{Q}_p$ be an elliptic curve with $|j(E)|_p > 1$.*

A. *There is a unique $q \in \mathbb{Q}_p^*$ with $|q|_p < 1$ such that E is isomorphic over $\bar{\mathbb{Q}}_p$ to the Tate curve E_q.*

B. *Furthermore, the isomorphism is over $\mathbb{Q}_p$ if and only if E has split multiplicative reduction at p. If E does not have split multiplicative reduction at p then the isomorphism is over a (unique) quadratic extension of $\mathbb{Q}_p$. This quadratic extension is unramified if and only if E has (non-split) multiplicative reduction.*

Let $E = E_{a^p,b^p,c^p}$ be the ABC curve associated to a solution of the Fermat equation with $p \geqslant 5$. As mentioned in the beginning of this section we want to study the ramification of the representation $\bar{\rho}_{E,p}$. We know that if $\rho_{E,p}$ ramifies at l then there is a $\sigma \in I_\mathfrak{l}$ acting non-trivially in $T_p(E)$. As we will see ramification can disappear when reducing mod p. So we do not know if the action of $I_\mathfrak{l}$ on $E[p]$ is non-trivial, that is if $\bar{\rho}_{E,p}$ ramifies at l. The action of $I_\mathfrak{l}$ on $E[p]$ is

non-trivial if and only if the field extension $K_p = \mathbb{Q}(E[p])/\mathbb{Q}$ has non-trivial inertia subgroup at l, that is K_p ramifies at l. Then we want to understand the ramification of the field K_p. Since ramification is a local property we can suppose for each prime l that E is defined over $\mathbb{Q}_l$ and that $K_p = \mathbb{Q}_l(E[p])$. For l dividing abc we prove the following theorem.

4.2.9 Theorem *(Hellegouarch) Let l be a prime dividing abc. Then the field K_p associated to the curve E_{a^p,b^p,c^p} can be considered as a subfield of $\mathbb{Q}_l(\zeta_p, 2^{1/p})$ (or $\mathbb{Q}_l(\beta^{1/2})(\zeta_p, 2^{1/p})$ with $\mathbb{Q}_l(\beta^{1/2})$ unramified).*

Proof: We start by showing that K_p always contain a primitive p-root of unity ζ_p. One can show that $E[p]$ is equipped with the Weil form

$$e_p : E[p] \times E[p] \to \mu_p(\bar{\mathbb{Q}}),$$

where $\mu_p(\bar{\mathbb{Q}})$ are the p-roots of unity. This form turns out to be bilinear, alternating ($e_p(T,T) = 1$), non-degenerate and compatible with the action of $G_{\bar{\mathbb{Q}}/K}$ in the sense that $e_p(S,T)^\sigma = e_p(S^\sigma, T^\sigma)$ for any extension $K/\mathbb{Q}$ and all $\sigma \in G_{\bar{\mathbb{Q}}/K}$. Now, since $E[p] \subset K_p$ we have for each pair of points (S,T)

$$e_p(S,T)^\sigma = e_p(S^\sigma, T^\sigma) = e_p(S,T) \text{ for all } \sigma \in G_{\bar{\mathbb{Q}}/K_p}$$

Hence $e_p(S,T) \in K_p$ thus $\mu_p(\bar{\mathbb{Q}}) \subset K_p$.

Replacing (A,B,C) by (a^p, b^p, c^p) in the formula for the j-invariant of $E_{A,B,C}$ we get that $j = -2^8(a^pc^p + b^pc^p + a^pb^p)^3(abc)^{-2p}$. Recall that $mdc(a,c) = mdc(a,b) = mdc(b,c) = 1$ and let l be a prime dividing abc. Then $\nu_l(j) = -2p\nu_l(abc)$ if $l \neq 2$ and $\nu_l(j) = 8 - 2p\nu_l(abc)$ if $l = 2$. That is, $|j|_l > 1$ for all l if $p \geqslant 5$. From Tate's uniformization theorem and corollary 4.2.1 we know that E_{a^p,b^p,c^p} is equivalent to the curve E_q over the field $\mathbb{Q}_l$ or over an unramified quadratic extension of $\mathbb{Q}_l$.

Now we suppose that the isomorphism is over $\mathbb{Q}_l$ and let $L = \mathbb{Q}_l(\zeta_p, 2^{1/p})$. From Tate's theorem we have $E_q(L)$ isomorphic to $L^*/\langle q \rangle$. Since j is up to an unit a p-th power in L it follows from the discriminant formula of E_q that the same is true for the parameter q. Hence there exists $q' \in L$ such that $q = unit * (q')^p$. Thus, $\langle \zeta_p, q' \rangle / \langle q \rangle$ is

contained in $L^*/\langle q\rangle$. Since $\langle \zeta_p, q'\rangle/\langle q\rangle$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$ we conclude that $L^*/\langle q\rangle \sim E_q(L)$ already contains all the p-torsion, implying $K_p \subset L$. In the case that the isomorphism is over the unramified quadratic extension $\mathbb{Q}_l(\beta^2)$ we take $L = \mathbb{Q}_l(\beta^{1/2})(\zeta_p, 2^{1/p})$ and repeat the reasoning.

□

4.2.10 Theorem *(Néron-Ogg-Shafarevich) Let $E/\mathbb{Q}$ be an elliptic curve. E has good reduction at l if and only if $\rho_{E,p}$ is unramified at l for some prime $p \neq l$ if and only if $\rho_{E,p}$ is unramified at l for all primes $p \neq l$.*

4.2.2 Corol·lari. *For $p \geqslant 5$, the representation $\bar{\rho}_{E,p} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_p)$ is unramified outside $2p$.*

Proof: Let $l \neq p$. If $l \nmid abc$ then $l \nmid \Delta$ and E has good reduction at l. By theorem 4.2.10 $\rho_{E,p}$ is unramified at l, hence $\bar{\rho}_{E,p}$ also is. If $l|abc$ then Hellegouarch theorem implies that K_p does not ramify at l if $l \neq 2, p$. Then $\bar{\rho}_{E,p}$ is unramified outside $2p$.

□

4.3 Modular Representations

In this section we recall results about cusp forms and their associated representations and we will also introduce the definition of modular representation.

Let $\mathcal{S}_k(\Gamma_0(N))$ denote the $\mathbb{C}$-vector space of the weight k cusp forms respect to the congruence subgroup $\Gamma_0(N)$. From the theory surrounding the Riemann-Roch theorem it is possible to derive formulas for the dimension of these spaces. An important corollary that will be of use later is the following.

4.3.1 Corol·lari. *$\mathcal{S}_2(\Gamma_0(2^t)) = \{0\}$ for $t \in \{0,1,2,3,4\}$ and $\mathcal{S}_2(\Gamma_0(32))$ has dimension 1.*

Now we will need some notation. Let K be any number field (i.e. a finite extension of $\mathbb{Q}$) and $\mathcal{O}_K$ its ring of integers. Let l be a prime

number and λ any maximal ideal lying over l. Denote by K_λ the λ-adic field obtained by taking fractions of

$$\mathcal{O}_{K,\lambda} = \varprojlim_n \{\mathcal{O}_K/\lambda^n\}.$$

We may view $\mathbb{Z}_l$ as a subring of $\mathcal{O}_{K,\lambda}$ and $\mathbb{Q}_l$ as a subfield K_λ. For a modular form f denote by $\mathbf{K}_f$ the field generated by its Fourier coefficients. It can be shown that $\mathbf{K}_f$ is a number field.

It is possible to construct from the modular curves $X_1(N)$ an abelian variety $J_1(N)$, the **Jacobian** of the modular curve $X_1(N)$. Similarly to what happen with elliptic curves, to the torsion points of $J_1(N)$ there is an associated Galois representation. This representation decomposes into 2-dimensional representations associated to modular forms. The next theorem is a consequence of the mentioned procedure and says that there are Galois representations arising from weight 2 cusp forms.

4.3.1 Theorem *Let $f \in \mathcal{S}_2(\Gamma_0(N))$ be a normalized eigenform with number field $\boldsymbol{K}_f$. Let l be a prime. For each maximal ideal λ of $\mathcal{O}_{\boldsymbol{K}_f}$ lying over l there is a 2-dimensional Galois representation*

$$\rho_{f,\lambda} : G_{\mathbb{Q}} :\to GL_2(\boldsymbol{K}_{f,\lambda}).$$

This representation is unramified at every prime $p \nmid lN$. For any such p let $\mathfrak{p} \subset \overline{\mathbb{Z}}$ be any maximal ideal lying over p. Then $\rho_{f,\lambda}(Frob_{\mathfrak{p}})$ satisfies the polynomial equation

$$x^2 - a_p x + p = 0.$$

For the converse phenomenon we make the following definition.

4.3.2 Definition *An irreducible Galois representation*

$$\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{Q}_l)$$

such that $det\rho = \chi_l$ is ***modular of weight 2*** *if there exists a newform $f \in \mathcal{S}_2(\Gamma_0(M))$ such that $\boldsymbol{K}_{f,\lambda} = \mathbb{Q}_l$ for some maximal ideal λ of $\mathcal{O}_{K_f}$ lying over l and such that $\rho_{f,\lambda} \sim \rho$.*

4.3.2 Remarca. Note that the representations $\rho_{E,l}$ associated to elliptic curves as in chapter 1 are good candidates to be modular.

We can extend this ideas for mod l representations. Let $f \in \mathcal{S}_2(\Gamma_0(M))$ be a newform and let $\lambda \subset \mathcal{O}_{\mathbf{K}_f}$ lie above l. It can be shown that up to similarity we may assume that the representation $\rho_{f,\lambda}$ maps to $GL_2(\mathcal{O}_{K_f,\lambda})$. So it reduces modulo l to a representation

$$\bar{\rho}_{f,\lambda} : G_{\mathbb{Q}} \to GL_2(\mathcal{O}_{K_f,\lambda}/\lambda\mathcal{O}_{K_f,\lambda}).$$

More generally we consider continuous mod l representations $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\overline{\mathbb{F}}_l)$. Since $G_{\mathbb{Q}}$ is compact this means that the image is finite and therefore lies in $GL_2(\mathbb{F}_{l^r})$ for some r. The notion of modularity has a mod l analogous.

4.3.3 Definition *An irreducible representation $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\overline{\mathbb{F}}_p)$ is* ***modular of weight*** 2 ***and level*** N *if there exists a newform $f \in \mathcal{S}_2(\Gamma_0(N))$ and a maximal ideal $\lambda \subset \mathcal{O}_{\mathbf{K}_f}$ lying over p such that $\bar{\rho}_{f,\lambda} \sim \bar{\rho}$*

4.3.3 Remarca. Also for $f \in \mathcal{S}_k(\Gamma_0(N))$ a normalized eigenform of weight $k > 2$ there is attached to it (by a result of Deligne) a Galois representations where the trace of Frobenius agree with the values $a_p(f)$. Thus, definitions 4.3.2 and 4.3.3 also generalize to any weight $k > 2$.

4.4 The Big Theorems

In this section we present the final ingredients for the study of the Fermat equation: the Modularity theorem and the Mazur-Ribet theorem.

4.4.1 Wiles' Theorem

Wiles proved that every semi-stable elliptic curve over $\mathbb{Q}$ is modular and later the result was generalized for all elliptic curves. This general version of Wiles Theorem is known as Modularity Theorem and there are several equivalent versions of it. Here we will state three versions: the first one is the more arithmetic and the other two use the Galois representations for elliptic curves and Modular forms. The

last version will directly take part in the proof of Fermat last theorem.

Let E be an elliptic curve defined over $\mathbb{Q}$, and let

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

be a minimal Weirstrass model for E. For a prime p of good reduction, i.e. primes not dividing the conductor of E, we define the quantities

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

4.4.1 Theorem *(Modularity Theorem, Version a_p) Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E. Then for some newform $f \in \mathcal{S}_2(\Gamma_0(N_E))$,*

$$a_p(f) = a_p(E) \quad \textit{for all primes } p \nmid N_E.$$

4.4.2 Theorem *(Modularity Theorem, Version R) Let E be an elliptic curve over $\mathbb{Q}$. Then $\rho_{E,l}$ is modular for some l.*

4.4.3 Theorem *(Modularity Theorem, strong Version R) Let E be an elliptic curve over $\mathbb{Q}$ with conductor N. Then for some newform $f \in \mathcal{S}_2(\Gamma_0(N))$ with number field $K_f = \mathbb{Q}$,*

$$\rho_{f,l} \sim \rho_{E,l} \textit{ for all } l.$$

4.4.4 Proposition *Let E be an elliptic curve over $\mathbb{Q}$. Then if $\rho_{E,l}$ is modular for some l then $\rho_{E,l}$ is modular for all l.*

4.4.2 Mazur-Ribet's Theorem

The last ingredient for the proof of the FLT is a deep and technical fact about representations. The Ribet-Mazur theorem allows to lower the level of modularity of representations and we will see that this has powerful consequences.

We will call **odd** to a representation ρ such that $\det \rho(conj) = -1$, where *conj* is the complex conjugation. There is a very important conjecture (now is a theorem) regarding modularity of mod l representations due to Serre. In its formulation Serre gives a recipe for obtain a minimal level $N_{\bar{\rho}}$ in terms of the ramification of $\bar{\rho}$.

1 Conjecture *(Serre) Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\bar{\mathbb{F}}_p)$ be irreducible and odd. Then $\bar{\rho}$ is modular of level $N_{\bar{\rho}}$ (the **Artin conductor of** $\bar{\rho}$) and some weight $k \geqslant 2$. For example, a prime $l \neq p$ divides $N_{\bar{\rho}}$ if and only if $\bar{\rho}$ is ramified at p.*

4.4.1 Remarca. For our purposes, we consider the Artin conductor outside of p, that is $p \nmid N_{\bar{\rho}}$.

4.4.5 Theorem *(Mazur-Ribet) Let $p \geqslant 3$ be a prime. Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_p)$ be a representation irreducible over $\bar{\mathbb{F}}_p$ and modular of level N.*

If $\bar{\rho}$ is finite at p then we can take N to be the Artin conductor of the representation and $k = 2$. In other words if $\bar{\rho}$ is modular then it is modular of level $N_{\bar{\rho}}$ predicted by Serre and weight 2.

We will not explain the meaning of '$\bar{\rho}$ is finite at p', because it is too technical. For our considerations it is enough to know that for an elliptic curve E, semi-stable at a prime p such that $p|\nu_p(\Delta)$, then the representation $\rho_p : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_p)$ is finite at p. Also, if E has good reduction at p then ρ_p is finite at p. The requirement '$\bar{\rho}$ is finite at p' is only needed to remove the prime p from the level of modularity. The other primes can be removed with less hypothesis.

4.5 The Equation $x^p + 2^\alpha y^p = z^p$

In this section we will use all the machinery from the previous sections to study the equation $x^p + 2^\alpha y^p = z^p$. Although the Modularity theorem will be used three times in this chapter, there are weaker results that would be enough for this equations. We first study the case $\alpha = 0$ which only needs Wiles result on semi-stable elliptic curves; then we proceed to the cases $\alpha > 1$ and $\alpha = 1$. For both cases results due to Diamond on the modularity of $E_{A,B,C}$ curves are enough.

4.5.1 Case $\alpha = 0$

Since we are considering $\alpha = 0$ our equation is the Fermat equation. To a solution (a, b, c) of the equation $x^p + y^p = z^p$ we will call it

primitive if a, b, c are relatively prime and *non-trivial* if $(abc \neq 0)$. Now we state and prove Fermat's Last Theorem.

4.5.1 Theorem *(Fermat-Wiles) Let $p \geqslant 5$ be a prime. There are no non-trivial primitive solutions of*

$$x^p + y^p = z^p.$$

Proof: Let (a, b, c) be a non-trivial primitive solution of Fermat's equation for $p \geqslant 5$. Since the solution is primitive it is easy to see that we can suppose that b is even and a, c are odd and also that $a \equiv -1 \pmod 4$ (if $a \equiv 1 \pmod 4$ we take the solution $(-a, -b, -c)$).

Now consider the curve $E = E_{a^p, b^p, c^p}$ (which is semistable by corollary 4.2.1) and the representation $\bar{\rho}_{E,p} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_p)$ induced by the action of $G_{\mathbb{Q}}$ on $E[p]$. The modularity theorem says that $\rho_{E,p}$ must be modular of level N, hence $\bar{\rho}_{E,p}$ is also modular. This representation is irreducible by theorem 4.2.3 and finite at p, hence it satisfies the hypothesis of Mazur-Ribet theorem. Thus we can take N to be the Artin conductor of $\bar{\rho}_{E,p}$. Since $\bar{\rho}_{E,p}$ only ramifies at $2p$ its Artin conductor equals 2. But $\mathcal{S}_2(\Gamma_0(2)) = \{0\}$ by corollary 4.3.1 so $\bar{\rho}_{E,p}$ is not modular and so $\rho_{E,p}$ is not modular, reaching a contradiction.

□

4.5.2 Case $\alpha > 1$

In this section we solve the equation $x^p + 2^\alpha y^p = z^p$ for the cases $\alpha > 1$ but first we make some considerations which are valid also when $\alpha = 1$.

Suppose that (a, b, c) is a solution of $0 = a^p + 2^\alpha b^p + c^p$, then also $a^p + 2^{\alpha - kp}(2^k b)^p + c^p = 0$ thus there exists a solution for an equation with α satisfying $1 \leqslant \alpha < p$. Let $1 \leqslant \alpha < p$. If the solution is primitive, then it is immediate that a and c are odd, meaning that $A = a^p$, $B = 2^\alpha b^p$ and $C = c^p$ are relatively prime. As before we normalize the solutions to $a \equiv -1 \pmod 4$. Because of the normalization we have $A \equiv -1 \equiv 3 \pmod 4$ and B even. Consider

the Frey curve

$$E : y^2 = x(x - A)(x + B),$$

with minimal discriminant of the form $\Delta_E = 2^s(ABC)^2$ by the proof of part 3 of theorem 4.2.6. Looking at the proof of parts 1 and 2 of the same theorem we see that we only needed the hypothesis $A \equiv 3$, hence E is semi-stable at every prime $p \neq 2$. Moreover, calculations show that the conductor N_E of E has the form $2^t rad'(ABC)$ with $t \in \{0, 1, 3, 5\}$, where $rad'(ABC)$ is the product of the odd primes in $rad(ABC)$. Furthermore, 4 divides B if and only if $t \leqslant 3$; E is semi-stable at 2 ($t = 0, 1$) if and only if $16|B$, more precisely from part 3 of theorem 4.2.6 follows that if 32 divides B then the reduction at 2 is multiplicative ($t = 1$); and $t = 5$ if and only if $ord_2(B) = 1$. Now, keeping $p \geqslant 5$ we will use the same ideas of the previous section to prove the following theorem.

4.5.2 Theorem *Let $p \geqslant 5$. The equation $a^p + 2^\alpha b^p + c^p = 0$ has no solutions in nonzero integers a, b, c if $\alpha > 1$.*

Proof: By the modularity theorem $\rho_{E,p}$ is modular of level N_E and so $\bar{\rho}_{E,p}$ is also modular of level N_E. It is not possible to apply theorem 4.2.3, because E is not semi-stable at 2. For this case Ribet achieves to prove irreducibility working with information about the conductor of $\bar{\rho}_{E,p}$ for this specific curve. Since every prime $l \neq 2$ is semi-stable and $\Delta(E)$ is a p-th power times a power of 2, $\bar{\rho}_{E,p}$ is finite at p. Hence, by the Mazur-Ribet theorem, $\bar{\rho}_{E,p}$ is modular of level equal to its Artin conductor. With an argument similar to Hellegouarch it is possible to show that the Artin conductor is 2^t. By corollary 4.3.1 we see that $t = 5$, that is $ord_2(B) = 2$. Since $B = 2^\alpha b^p$ we have a contradiction with $\alpha > 1$.

□

4.5.3 Case $\alpha = 1$

In this section we treat the hardest case, $\alpha = 1$. We say it is the hardest not only because we only cover it for values of $p \equiv 1 \bmod 4$, but also because it makes use of tools that have not been introduced so far. In view of this will only give the guideline of the proof by sating new results at each step. We aim to the following theorem.

4.5.3 Theorem *Let $p \geqslant 17$ and $p \equiv 1$ mod 4. Let (a, b, c) be a solution in non-zero integers of the equation $x^p + 2y^p + z^p = 0$. If (a, b, c) are coprime and we use the normalization $a \equiv -1$ mod 4, then the only possible solution is $(a, b, c) = (-1, 1, -1)$.*

Since (a, b, c) are coprime, it is clear that a and c need to be odd. But it is an immediate corollary of the previous section that b must also be odd.

4.5.1 Corol·lari. *The equation $a^p + 2b^p + c^p = 0$ has no integer solutions with b even.*

Proof: From the proof of theorem 4.5.2 we see that we must have $ord_2(B) = 2$. This is not possible when b is even since $B = 2b^p$.

□

Let E_0 be the elliptic curve associated with the trivial solution $(-1, 1, -1)$ (i.e. the elliptic curve with complex multiplication $y^2 = x^3 - x$) and E the elliptic curve associated to a solution (a, b, c). Let also $\bar{\rho}_{E_0,p}$ and $\bar{\rho}_{E,p}$ be the associated 2-dimensional mod p representations. The following holds.

4.5.4 Proposition *The 2-dimensional mod p representations of $G_{\mathbb{Q}}$ defined by E and E_0 are isomorphic.*

Idea of proof: We have seen in the previous section that $\bar{\rho}_{E,p}$ is associated with a eigenform coming from $\Gamma_0(32)$. This is a one dimensional space, that is $J_0(32)$ is an elliptic curve, hence $\bar{\rho}_{E,p}$ arises from $J_0(32)[p]$. In particular, the isomorphism class of $\bar{\rho}_{E,p}$ does not depend on the solution. Then $\bar{\rho}_{E_0,p} \sim \bar{\rho}_{E,p}$.

□

Now we can use information about the elliptic curve E_0.

4.5.5 Proposition *The image of $\bar{\rho}_{E_0,p}$ is contained in the normalizer of a Cartan subgroup of $GL_2(\mathbb{F}_p)$. If $p \equiv 1$ mod 4 (or $p \equiv -1$ mod 4) then it is the normalizer of a Cartan split (or non-split) subgroup of $GL_2(\mathbb{F}_p)$.*

The next theorem puts strong constraints on the set of primes at which E does not have potential good reduction.

4.5.6 Theorem *(Mazur-Momose) Let $p \geqslant 17$. If the image of $\bar{\rho}_{E,p}$ is contained in the normalizer of a Cartan split subgroup of $GL_2(\mathbb{F}_p)$ then E has potential good reduction at all primes $l \neq 2$.*

Finally, we prove theorem 4.5.3. Since $p \equiv 1 \bmod 4$, from the above propositions we see that $\bar{\rho}_{E,p}$ is under the hypothesis of Mazur-Momose theorem. From the previous section we know that E has multiplicative reduction at all odd primes dividing abc, then it can not have potential good reduction at these primes. Then Mazur-Momose theorem imply that there is no such a prime, hence abc is 2^n with $n \geqslant 0$. Since all of them are odd we must have $n = 0$. Hence the only normalized solution is $(-1, 1, -1)$.

□

4.6 More Equations

In this section, in order to illustrate that the techniques of the previous sections also work with curves that are not ABC we will give two more examples of Diophantine equations. In the examples bellow we will make explicit the associated Frey curve, the discriminant Δ, the conductor N and the Artin conductor N_ρ, but we will not solve the equation.

As we already mentioned, Diamond proved without using the full generality of the Modularity theorem (MT) that the curves $E_{A,B,C}$ are modular. Furthermore, he also proved a weaker version of the MT for curves with restricted ramification. In the first of the following two examples modularity of the Frey curve follows from the work of Diamond, but in the second example the full power of the Modularity theorem is needed. Let (a, b, c) denote a non-trivial primitive solution.

Example 1: The equation $a^p + b^p = c^2$. If ab is even, we can assume $c \equiv 1 \pmod 4$ and consider the elliptic curve

$$y^2 + xy = x^3 + \frac{c-1}{4}x^2 + \frac{a^p}{2^6}x.$$

This curves satisfies

$$\Delta = \frac{1}{2^{12}}(a^2b)^p, \quad N = rad(ab), \quad N_\rho = 2.$$

If ab is odd, we can assume $a \equiv -1 \pmod 4$ and consider

$$y^2 = x^3 + 2cx^2 + a^p x$$

which satisfies

$$\Delta = 2^6(a^2b)^p, \quad N = 2^5 rad(ab), \quad N_\rho = 32.$$

Example 2: The equation $a^p + b^p = c^3$. If ab is even consider

$$y^2 = x^3 - 3(a^p + 9b^p)cx - 2(a^{2p} - 18a^pb^p - 27b^{2p})$$

with discriminant $\Delta = 2^{12}3^3(a^3b)^p$. If $c = 2c_0$ is even let

$$y^2 + b^p y = x^3 - 3(c_0^3 + b^p)c_0 x - c_0^3(2c_0^3 - 5b^p)$$

which have discriminant $\Delta = 3^3(a^3b)^p$. In both cases

$$N = rad(ab), \quad N_\rho = 3 \quad \text{if } 3|ab$$

or

$$N = 3^3 rad(ab), \quad N_\rho = 27 \quad \text{if } 3 \nmid ab.$$

Bibliografia

[1] H Darmon and L. Merel. Winding quotients and some variants of fermat's last theorem. *http://people.math.jussieu.fr/ merel/winding.pdf*.

[2] F. Diamond and J. Shurman. *A First Course in Modular Forms*. Springer, 2005.

[3] Y. Hellegouarch. *Invitation to the Mathmatics of Fermat-Wiles*. Academic Press, 2002.

[4] B. Mazur. Rational isogenies of prime degree. *Invent. Math*, 44:129–162, 1978.

[5] J.S. Milne. *Elliptic Curves*. BookSurge Publishers, 2006.

[6] K. Ribet. On the equations $a^p + 2^\alpha b^p + c^p = 0$. *Acta Arithmetica*, LXXIX.1:7–16, 1997.

[7] J. H. Silverman. *The Arithmetic of Elliptic Curves*. Springer-Verlag, 1986.

[8] J. H. Silverman. *Advances Topics in the Arithmetic of Elliptic Curves*. Springer-Verlag, 1994.

Nuno Freitas
Departament d'Àlgebra i Geometria
Facultat de Matemàtiques
Universitat de Barcelona
Gran via de les Corts Catalanes 585, 08007 Barcelona
nunobfreitas@gmail.com

NOTES DEL SEMINARI

MONOGRÀFIC SOBRE TREBALLS DE KENNETH RIBET

Barcelona 2010

19

Notes del Seminari de Teoria de Nombres
(UB-UAB-UPC)

MONOGRÀFIC SOBRE TREBALLS DE KENNETH RIBET

Edició a cura de

M. Alsina N. Vila

Amb contribucions de

X. Guitart L. Terracini B. Plans N. Freitas

M. Alsina
Dept. Matemàtica Aplicada III
E P Superior d'Enginyeria de Manresa
Universitat Politècnica de Catalunya
Agda Bases de Manresa, 61-73
08242 Manresa
montserrat.alsina@upc.edu

N. Vila
Facultat de Matemàtiques,
Universitat de Barcelona
Gran Via de les Corts Catalanes, 585
08007 Barcelona
nuriavila@ub.edu

Classificació AMS
Primària: 11G18, 11F33
Secundària: 11D41, 11G05, 11G30, 14G35, 14H25, 14H52

Barcelona, 2010
Amb suport parcial de MTM2006-04895 i MTM2009-07024.
ISBN: 978-84-934244-9-7