The Generalized Fermat Equation

by **Nicolas Billerey*** and **Nuno Freitas****

Abstract.—In this note we will review the main steps in the proof of Fermat's Last Theorem and discuss Darmon's program to tackle the generalized Fermat equation $Ax^q + By^r = Cz^p$. Finally, we discuss how combining the classical approach with some ideas of Darmon led to recent results for equations of the form $x^r + y^r = Cz^p$.

1 Introduction

After Wiles' proof [27] of Fermat's Last Theorem (FLT) attention shifted towards the so-called generalized Fermat equation (GFE)

$$Ax^r + By^q = Cz^p \quad \text{with} \quad \mathcal{X} := \frac{1}{r} + \frac{1}{q} + \frac{1}{p} < 1, \quad (1.1)$$

where A, B, C are fixed non-zero coprime integers and $r, q, p \geq 2$ are integers. The triple (r, q, p) is called the *signature* of the GFE. A solution $(a, b, c) \in \mathbb{Z}^3$ to (1.1) is called *primitive* if $\gcd(a, b, c) = 1$ and *non-trivial* if $abc \neq 0$.

The condition $\mathcal{X} < 1$ is required to guarantee finiteness of solutions. More precisely, Darmon and Granville [13] proved that if we fix both the coefficients A, B, C and the exponents r, q, p satisfying $\mathcal{X} < 1$ then there are only finitely many primitive solutions to (1.1). But more is conjectured (see [4]): it is expected that the number of primitive solutions remains finite if we fix the coefficients but allow the three exponents to vary while still verifying $\mathcal{X} < 1$. On the other hand, if $\mathcal{X} > 1$ then the set of solutions is either empty or infinite by a result of Beukers [3] and, for $\mathcal{X} = 1$, the problem reduces to the determination of rational points on genus-1 curves. A very natural question is whether the strategy that proved FLT, which is now known as *the modular method*, can be used to establish more cases of the aforementioned conjecture.

As we shall see below, to apply the modular method to other instances of (1.1) one needs to start with the construction of a Frey curve. However, there are only a few choices of the exponents r, q, p in (1.1) for which Frey curves are known (see [10, p.14] for a complete list of rational Frey curves). To circumvent this issue, Darmon described in [11] a remarkable program to study (1.1) where he replaces Frey curves by higher dimensional abelian varieties. However, applying the rest of his program is very challenging because several of the main steps rely on open conjectures.

The objective of this expository note is to briefly discuss some recent results regarding the subfamily of (1.1) of the shape $x^r + y^r = Cz^p$ obtained by combining the classical approach with Frey curves and some of the ideas in the Darmon's program. For a brief introduction to Diophantine equations including a quick discussion of the modular method we refer the reader to [22].

2 Elliptic curves

For this section, the main reference is [24].

Let K be a field. An *elliptic curve* E over K is a smooth curve in $\mathbb{P}^2$ given by an equation

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3,$$

* Laboratoire de Mathématiques Blaise Pascal, Université Clermont Auvergne et CNRS, France
Email: Nicolas.Billerey@uca.fr

** Instituto de Ciencias Matemáticas, CSIC, Spain
Email: nuno.freitas@icmat.es

Nicolas Billerey is supported by the ANR-23-CE40-0006-01 Gaec project. We thank the referee for a careful reading and helpful remarks.

with $a_i \in K$. If the characteristic of K is not 2 or 3, then we can transform to a much simpler model given by the affine equation

$$Y^2 = X^3 + aX + b, \quad \Delta_E = -16(4a^3 + 27b^2) \neq 0,$$

where a and $b \in K$. There is a distinguished K-point, the 'point at infinity', which we denote by ∞. Given a field $L \supseteq K$, the set of L-points on E is

$$E(L) = \{(x, y) \in L^2 \ : \ y^2 = x^3 + ax + b\} \cup \{\infty\}.$$

It turns out that the set $E(L)$ has the structure of an abelian group with ∞ as the identity element. The group structure is easy to describe geometrically: three points $P_1, P_2, P_3 \in E(L)$ add up to the identity element if and only if there is a line ℓ defined over L meeting E in P_1, P_2, P_3 (with multiplicities counted appropriately). The classical Mordell–Weil Theorem states that for a number field K the group $E(K)$ is finitely generated.

Now suppose $K = \mathbb{Q}$. There is an integer N_E called *the conductor of* E with the following properties. There is an algorithm to compute N_E and, for all primes $p \nmid N_E$, the reduction modulo p of a minimal model for E gives an elliptic curve $\tilde{E}$ over $\mathbb{F}_p$. Moreover, if a prime $p \mid N_E$ then it divides the discriminant of any model for E so the reduced curve $\tilde{E}/\mathbb{F}_p$ is not an elliptic curve, and we can think of N_E as a measure of how 'complicated' these reduced curves are. Finally, for $p \nmid N_E$, the set $\tilde{E}(\mathbb{F}_p)$ is necessarily finite, and we define

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

3 Modular forms

For this section, the main reference is [14].

Let $N \in \mathbb{Z}_{\geq 1}$. A *modular form of weight* 2 *for* $\Gamma_0(N)$ is an analytic function on the complex upper half-plane H satisfying suitable growth conditions at the boundary as well as the transformations

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$$

for all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ satisfying $c \mid N$ and all $z \in \mathrm{H}$. Invariance under translation by 1 leads to a Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} a_n(f) q^n, \quad q = e^{2\pi i z}.$$

The group $\Gamma_0(N)$ acts on H via fractional linear transformations and the quotient $Y_0(N) = \Gamma_0(N)\backslash \mathrm{H}$ has the structure of a non-compact Riemann surface. This has a standard compactification denoted $X_0(N)$ and the difference $X_0(N) - Y_0(N)$ is a finite set of points called the *cusps*. To the modular forms that vanish at all the cusps we call *cusp forms*; in particular, they satisfy $a_0(f) = 0$.

The space of cusp forms $S_2(N)$ is a finite dimensional $\mathbb{C}$-vector space. There is a natural family of commuting operators $T_n \ : \ S_2(N) \to S_2(N)$ (with $n \geq 1$) called the *Hecke operators*. The *eigenforms* of level N are the cusp forms that are simultaneous eigenvectors for all the Hecke operators. An eigenform f is called *normalized* if $a_1(f) = 1$ and thus its Fourier expansion has the form

$$f = q + \sum_{n \geq 1} a_n(f) q^n.$$

Shimura-Taniyama-Weil Conjecture asserts that for every elliptic curve $E/\mathbb{Q}$ with conductor N_E there is a normalized eigenform f of weight 2 for $\Gamma_0(N_E)$, such that for every prime p the corresponding Fourier coefficient satisfies $a_p(f) = a_p(E)$. When this is the case we say that the curve E is *modular*. In his seminal paper [27] and its companion [26] (jointly with R. Taylor), Andrew Wiles proved the S-T-W Conjecture in the case of *semistable* elliptic curves, i.e. elliptic curves with square free conductor N_E. This groundbreaking theorem was also the final step to complete the proof of FLT.

4 Galois representations

For this section, the main references are [14, Chapter 9] and (for more advanced readers) [7].

Let $\overline{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$ inside $\mathbb{C}$. We write $G_{\mathbb{Q}} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ for the group of field automorphisms of $\overline{\mathbb{Q}}$ (fixing $\mathbb{Q}$). The group $G_{\mathbb{Q}}$ is called *the absolute Galois group of* $\mathbb{Q}$. The representations of $G_{\mathbb{Q}}$ are central objects in Arithmetic Geometry. Here we will work only with *residual* Galois representations, also known as *mod p* representations.

Definition 1.— A *mod p Galois representation* is defined to be a group homomorphism

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

which is continuous with respect to the profinite topology on the left and the discrete topology on the right. In particular, there is a finite extension $\mathbb{F}_q/\mathbb{F}_p$ such that the image of $\overline{\rho}$ lies in $\mathrm{GL}_2(\mathbb{F}_q)$.

Definition 2.— A mod p Galois representation $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is *unramified at a prime* $\ell \neq p$

if $\overline{\rho}(I_\ell) = \{1\}$, where I_ℓ is an inertia group at ℓ in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Otherwise, it is *ramified at* ℓ.

The reader unfamiliar with the inertia subgroups of $G_\mathbb{Q}$ should simply keep in mind that there is a unique (up to conjugation) inertia subgroup for each prime ℓ and that a representation $\overline{\rho}$ is easier to understand if it has little ramification. Further, there is a positive integer $N(\overline{\rho})$, called *the Serre level of* $\overline{\rho}$, that measures the ramification of $\overline{\rho}$ at all primes $\ell \neq p$. Moreover, by Galois theory, the kernel of a representation $\overline{\rho}$ as above corresponds to a field extension of finite degree which is ramified at a prime ℓ if and only if $\overline{\rho}$ is ramified at ℓ.

4.1 Representations from elliptic curves

Let E be an elliptic curve over $\mathbb{C}$. The structure of the abelian group $E(\mathbb{C})$ is particularly easy to describe. There is a discrete lattice $\Lambda \subset \mathbb{C}$ of rank 2 (that is, as an abelian group $\Lambda \simeq \mathbb{Z}^2$) depending on E, and an isomorphism

$$E(\mathbb{C}) \simeq \mathbb{C}/\Lambda.$$

Let p be a prime. By the p-torsion of $E(\mathbb{C})$ we mean the subgroup

$$E[p] = \{Q \in E(\mathbb{C}) \ : \ pQ = 0\}.$$

It follows that $E[p] \simeq (\mathbb{Z}/p\mathbb{Z})^2$ which can be viewed as a 2-dimensional $\mathbb{F}_p$-vector space. Now let E be an elliptic curve over $\mathbb{Q}$. Then we may view E as an elliptic curve over $\mathbb{C}$, and with the above definitions obtain an isomorphism $E[p] \simeq (\mathbb{Z}/p\mathbb{Z})^2$. However, in this setting the points of $E[p]$ have algebraic coordinates, and are acted on component-wise by $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Thus we obtain a 2-dimensional representation depending on $E/\mathbb{Q}$ and the prime p:

$$\overline{\rho}_{E,p} \ : \ G_\mathbb{Q} \to \mathrm{GL}_2(\mathbb{F}_p),$$

called *the mod p representation attached to* E. We say that $\overline{\rho}_{E,p}$ is *irreducible* if the image $\overline{\rho}_{E,p}(G_\mathbb{Q})$ cannot be conjugated into a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ consisting of upper triangular matrices.

4.2 Representations from modular forms

Let $f = \sum_{n\geq 1} a_n(f)q^n$ be a weight-2 normalized eigenform for $\Gamma_0(N)$ with $N \geq 1$. Denote by $K_f = \mathbb{Q}(\{a_n(f) \ : \ n \geq 1\})$ the field generated by the Fourier coefficients of f. It is a non-trivial theorem that $a_n(f)$ are algebraic integers and K_f is a number field, which we view as a subfield of $\overline{\mathbb{Q}}$. We denote by $\mathcal{O}_{K_f}$ the ring of integers of K_f, and we have $a_n(f) \in \mathcal{O}_{K_f}$ for all n; we refer to [14, §6.5] for details.

Let p be a prime number, and $\mathfrak{p}$ a prime in K_f above p. We write $\mathbb{F}_\mathfrak{p} = \mathcal{O}_{K_f}/\mathfrak{p}$ for the residue field at $\mathfrak{p}$. The following is a consequence of a deep result proved by Eichler and Shimura.

Theorem 3 (Eichler–Shimura).— Up to isomorphism, there is a unique semisimple mod p Galois representation

$$\overline{\rho}_{f,\mathfrak{p}} \ : \ \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_\mathfrak{p})$$

satisfying the following properties: it is unramified outside Np and for every prime $\ell \nmid Np$, the characteristic polynomial of $\overline{\rho}_{f,p}(\mathrm{Frob}_\ell)$ is the mod $\mathfrak{p}$ reduction of

$$X^2 - a_\ell(f)X + \ell. \tag{4.1}$$

Here Frob_ℓ denotes a choice of a Frobenius element at ℓ in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and by semisimple we mean that $\overline{\rho}_{f,\mathfrak{p}}$ is either irreducible or isomorphic to the sum of two characters.

Definition 4.— A mod p Galois representation

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

is said to be *modular of level* $N \geq 1$ if there exists a weight-2 eigenform f for $\Gamma_0(N)$ and a prime $\mathfrak{p} \mid p$ in K_f such that $\overline{\rho} \simeq \overline{\rho}_{f,\mathfrak{p}}$. In this case, we also say that $\overline{\rho}$ *arises from* f.

Building on the groundbreaking work of Wiles' and many others, Khare and Wintenberger [17, 18] have proved the following theorem known as Serre's Conjecture.

Theorem 5.— Let $\overline{\rho} \ : \ G_\mathbb{Q} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be an irreducible odd representation. Assume that $\overline{\rho}$ arises from a finite flat group scheme at p. Then $\overline{\rho}$ is modular of level $N(\overline{\rho})$ and weight 2.

The technical condition that $\overline{\rho}$ arises from a finite flat group scheme at p should, for simplicity, be thought informally as the restriction of $\overline{\rho}$ to an inertia subgroup at p being *well behaved*; recall that ramification at $\ell \neq p$ is measured by $N(\overline{\rho})$.

5 Proof of FLT

For this section, the main references are [9] and [12].

We have introduced the minimal set of tools to sketch the proof of FLT. We decided to organize the

proof in three main steps because these are the steps that we will focus on when presenting the Darmon program in the later sections.

Step 1—Construction: Suppose $p \geq 5$ is prime, and a, b and c are non-zero coprime integers satisfying $a^p + b^p = c^p$. We can reorder (a, b, c) so that

$$b \equiv 0 \pmod 2 \quad \text{and} \quad a^p \equiv -1 \pmod 4.$$

We consider the *Frey–Hellegouarch curve* which depends on (a, b, c):

$$E \ : \ Y^2 = X(X - a^p)(X + b^p). \tag{5.1}$$

From all the hypotheses on a, b, c, we compute the minimal discriminant and conductor of E:

$$\Delta = \frac{(abc)^{2p}}{2^8} \neq 0, \qquad N_E = \prod_{\ell \mid \Delta} \ell.$$

Note that the conductor is square-free and satisfies $2 \,||\, N$.

Step 2—Residual modularity: As $p \geq 5$, it follows from the work of Mazur [21] that $\overline{\rho}_{E,p}$ is irreducible. It is well known that $\overline{\rho}_{E,p}$ is odd and Hellegouarch showed that $\overline{\rho}_{E,p}$ arises on a finite flat group scheme at p. Computing the Serre level we obtain $N(\overline{\rho}_{E,p}) = 2$. Therefore, by Serre conjecture, we have that

$$\overline{\rho}_{E,p} \simeq \overline{\rho}_{g,\mathfrak{p}}$$

where g is an eigenform of level 2 and weight 2, and $\mathfrak{p} \mid p$ is a prime in K_g.

Step 3—Contradiction: There are no eigenforms of weight 2 and level 2, a contradiction.

Remark 1.— Note that the Frey curve construction applies for trivial solutions as well. However, in this case, it does not give rise to an elliptic curve (as it is singular), therefore, there are no modular representations associated with it. This is a fortunate feature of the classical Fermat equation. We will see below that this is no longer the case for the GFE which obstructs its resolution in many cases.

Remark 2.— The reader may be wondering where is Wiles' work used in the previous proof. Since the original proof of FLT predates the proof of Serre's conjecture, modularity of the residual representation $\overline{\rho}_{E,p}$ was instead derived as a corollary of modularity of the Frey curve E. Note that E has square-free conductor hence it is modular by the work of Wiles. We note also that the work of Wiles and all the ideas around it is heavily used in the proof of Serre's conjecture.

6 Darmon's program

As we see from the proof of FLT it is the modularity together with the little ramification of the 2-dimensional residual representation $\overline{\rho}_{E,p}$ that is key for the contradiction. The Frey curve E is simply a geometric object from which we know how to extract a 2-dimensional Galois representation with the right properties, namely $\overline{\rho}_{E,p}$.

There are higher dimensional generalizations of elliptic curves, called *abelian varieties*, in the sense that there is a group structure on the set of points of an abelian variety A. The main idea of Darmon's program is to put the focus directly on 2-dimensional mod p representations with the correct properties, and find the abelian varieties giving rise to them.

Definition 6.— Let $r, q, p \geq 2$ be integers. A *Frey representation* of signature (r, q, p) over a number field K in characteristic $\ell > 0$ is a Galois representation

$$\overline{\rho} = \overline{\rho}(t) \ : \ G_{K(t)} \to \mathrm{GL}_2(\mathbb{F})$$

where $\mathbb{F}$ is a finite field of characteristic ℓ such that the following conditions hold:

(i) The restriction of $\overline{\rho}$ to $G_{\overline{K}(t)}$ has trivial determinant and is irreducible.

(ii) The projectivization

$$\overline{\rho}^{\mathrm{geom}} \ : \ G_{\overline{K}(t)} \to \mathrm{PSL}_2(\mathbb{F})$$

of this representation is unramified outside $\{0, 1, \infty\}$.

(iii) It maps the inertia groups at 0, 1, and ∞ to subgroups of $\mathrm{PSL}_2(\mathbb{F})$ of order r, q, and p respectively.

Here $K(t)$ is the function field over K in the variable t and $\overline{K}$ is an algebraic closure of K, and $G_k := \mathrm{Gal}(\overline{k}/k)$ denotes the absolute Galois group of k for any field k.

In [11], Darmon counts the number of Frey representations up to some equivalence relation (introduced in *loc. cit.*) and describes (often not in an explicitly way) where they should arise. In particular, he proves the following classification result.

Theorem 7 (Hecke-Darmon).— Up to equivalence, there is only one Frey representation of signature (p, p, p). It occurs over $\mathbb{Q}$ in characteristic p and is associated with the Legendre family

$$L(t) \ : \ y^2 = x(x - 1)(x - t).$$

Example 1.— It is not difficult to check that the classical Frey–Hellegouarch curve

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

is obtained from $L(t)$ after *specialization* at

$$t_0 = \frac{a^p}{a^p + b^p}$$

and taking *quadratic twist* by $-(a^p + b^p)$.

A Frey representation $\overline{\rho}(t)$ should be seen as a family of representations where we can specialize the parameter t to obtain mod p representations of G_K as in the previous example. We are then interested in the modularity of the mod p representations obtained in this way.

From now on, we restrict ourselves to the case of K being *a totally real field*, i.e., a number field such that all embeddings into $\mathbb{C}$ have image in $\mathbb{R}$. This is a natural restriction, because modularity related objects are very poorly understood for fields with at least one complex embedding. In contrast, for a totally real K there is a well established theory of *Hilbert modular forms* (see [15]) which are the natural replacement for the modular forms over $\mathbb{Q}$; it is not our objective to discuss details of this theory here. The only thing to keep in mind is that they satisfy the analogous properties over K to those of modular forms over $\mathbb{Q}$. In particular, modularity of abelian varieties and their residual representations can be defined via a connection to representations arising from Hilbert eigenforms (see [25]). Therefore, we can state the following special case of Serre conjecture over totally real fields.

Conjecture 1 ([11, Conjecture 3.2]).— Let K be a totally real field. Let $\overline{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a totally odd and irreducible representation with determinant the mod p cyclotomic character.

Assume that $\overline{\rho}$ arises from a finite flat group scheme at all primes $\mathfrak{p}$ in K above p. Then there is a Hilbert eigenform g over K for $\Gamma_0(N(\overline{\rho}))$ of (parallel) weight 2 and a prime $\mathfrak{p} \mid p$ in the field of coefficients of g such that $\overline{\rho} \simeq \overline{\rho}_{g,\mathfrak{p}}$.

This conjecture is still open for all K, therefore when applying the Darmon program in the next section we need to derive residual modularity without it. Also, this conjecture is concerned with 2-dimensional representations whilst representations arising from abelian varieties of dimension n are naturally of dimension $2n$. We thus focus only on the subfamily of abelian varieties giving rise to 2-dimensional representations, as per the next definition and well known theorem.

Definition 8.— Let A be an abelian variety over a field L of characteristic 0. We say that A/L is of GL_2-type (or $\mathrm{GL}_2(F)$-type) if there is an embedding

$$F \hookrightarrow \mathrm{End}_L(A) \otimes_{\mathbb{Z}} \mathbb{Q}$$

where F is a number field with $[F : \mathbb{Q}] = \dim A$.

Theorem 9.— Let A/L be an abelian variety of $\mathrm{GL}_2(F)$-type. Let $\mathfrak{p}$ be a prime in F above p. Then there is a 2-dimensional mod p representation attached to A, denoted $\overline{\rho}_{A,\mathfrak{p}} : G_K \to \mathrm{GL}_2(\mathbb{F}_{\mathfrak{p}})$, unramified outside the primes where A has bad reduction and p.

Darmon also discusses the existence of Frey varieties $J(a, b, c)/\mathbb{Q}$ associated to solutions (a, b, c) of (1.1) for any choice of exponents, and explains how these give rise (after base changing to certain totally real number fields) to all the possible Frey representations. However, only the varieties for exponents (p, p, p) and (p, p, r) are explicit enough to work with. Finally, he finishes with the following extremely difficult conjecture [11, Conjecture 4.1].

Conjecture 2 (Large image conjecture).— Let K be totally real field. There exists a constant C_K such that, for any abelian variety A/K of GL_2-type with $\mathrm{End}_{\overline{K}}(A) \otimes \mathbb{Q} = K$, and all primes $\mathfrak{p}$ of K of norm $> C_K$, we have $\mathrm{SL}_2(\mathbb{F}_{\mathfrak{p}}) \subset \overline{\rho}_{A,\mathfrak{p}}(G_K)$.

We finish this section with the description of how the Darmon program is expected to work. We emphasize every step that we do not know how to do, or that depends on conjectures or relies on computations that are not possible in practice with current algorithms and hardware.

1. Let $a, b, c \in \mathbb{Z}$ satisfy $a^r + b^q = c^p$ and $\gcd(a, b, c) = 1$.
2. Let $J(a, b, c)/\mathbb{Q}$ be the associated Frey variety. Over a totally real field K it becomes of $\mathrm{GL}_2(K)$-type. We consider $J = J(a, b, c)/K$ and its mod $\mathfrak{p}$ representation $\overline{\rho}_{J,\mathfrak{p}}$ given by Theorem 9.
3. Assume $p > C_K$ where C_K is the constant in Conjecture 2. If (a, b, c) is non-trivial then $\mathrm{SL}_2(\mathbb{F}_{\mathfrak{p}})$ is *conjecturally* contained in the image of $\overline{\rho}_{J,\mathfrak{p}}$ by Conjecture 2. In particular, $\overline{\rho}_{J,\mathfrak{p}}$ is *conjecturally* irreducible.
4. The representation $\overline{\rho}_{J,\mathfrak{p}}$ is totally odd with cyclotomic determinant and *conjecturally* arises on a finite flat group scheme at all $\mathfrak{p} \mid p$ in K.
5. We *compute* the Serre level $N(\overline{\rho}_{J,\mathfrak{p}})$.

6. The representation $\overline{\rho}_{J,\mathfrak{p}}$ is *conjecturally* modular of level $N(\overline{\rho}_{J,\mathfrak{p}})$ and (parallel) weight 2 by Conjecture 1, that is $\overline{\rho}_{J,\mathfrak{p}} \simeq \overline{\rho}_{g,\mathfrak{p}}$ for some Hilbert eigenform g of level $N(\overline{\rho}_{J,\mathfrak{p}})$.

7. We *compute* the relevant space of eigenforms and *show* that $\overline{\rho}_{J,\mathfrak{p}} \not\simeq \overline{\rho}_{g,\mathfrak{p}}$ except for the eigenforms g_0 corresponding via modularity to the Frey varieties $J_0 := J(a,b,c)$ where (a,b,c) satisfies $abc = 0$ i.e. Frey varieties attached to trivial solutions.

8. *Conjecturally* the varieties J_0 have complex multiplication, thus $\mathrm{SL}_2(\mathbb{F}_{\mathfrak{p}})$ is not contained in the image of $\overline{\rho}_{g_0,\mathfrak{p}}$. Thus we also have $\overline{\rho}_{J,\mathfrak{p}} \not\simeq \overline{\rho}_{g_0,\mathfrak{p}}$, a contradiction with Step **6**.

In view of the three main steps in the proof of FLT, the previous bullet points are divided as follows: Step 1 corresponds to **1–2**, Step 2 corresponds to **3–6** and Step 3 corresponds to **7–8**.

To conclude this section, we note that the contradiction step which was trivial in the proof of FLT is quite challenging in this more general situation. As mentioned in Remark 1, the trivial solutions represent a major obstruction, but there are other issues. Namely, the space of revelant Hilbert modular forms might not be accessible with current software implementations (either because it is too large, or by lack of efficient algorithms in certain specific situations). Moreover, we miss a general method for discarding isomorphisms between residual Galois representations. In particular, it is an open problem to show that given two non-isogenous rational elliptic curves E, E', then for all large enough primes p, the representations $\overline{\rho}_{E,p}$ and $\overline{\rho}_{E',p}$ are not isomorphic.

7 Some recent results for signature (r,r,p)

We now discuss our contribution to the Darmon's program in the case of the generalized Fermat equation

$$x^r + y^r = Cz^n, \tag{7.1}$$

where r is a fixed prime ≥ 3, C is a fixed positive integer and $n \geq 2$ is an integer.

Throughout this paragraph, we fix the following notation.

- ζ_r primitive r-th root of unity
- $\omega_i = \zeta_r^i + \zeta_r^{-i}$, for every $i \geq 0$
- $h(X) = \prod_{i=1}^{(r-1)/2} (X - \omega_i) \in \mathbb{Z}[X]$
- $K = \mathbb{Q}(\zeta_r)^+ = \mathbb{Q}(\omega_1)$ maximal totally real subfield of $\mathbb{Q}(\zeta_r)$
- $\mathcal{O}_K$ integer ring of K
- $\mathfrak{p}_r$ unique prime ideal above r in $\mathcal{O}_K$ (totally ramified)

Let a, b be non-zero coprime integers such that $a^r + b^r \neq 0$. Following a construction of Kraus [19], we consider the curve $C_r(a,b)$ given by the equation

$$y^2 = (ab)^{\frac{r-1}{2}} xh\left(\frac{x^2}{ab} + 2\right) + b^r - a^r.$$

The discriminant of this model is

$$\Delta_r(a,b) = (-1)^{\frac{r-1}{2}} 2^{2(r-1)} r^r (a^r + b^r)^{r-1}$$

which is non-zero as $a^r + b^r \neq 0$. In particular, it defines a hyperelliptic curve of genus $(r-1)/2$.

Examples 1.— Here are explicit equations for Kraus' curve with $r = 3, 5, 7$.

$$\begin{aligned} r = 3: &\quad y^2 = x^3 + 3abx + b^3 - a^3 \\ r = 5: &\quad y^2 = x^5 + 5abx^3 + 5a^2b^2x + b^5 - a^5 \\ r = 7: &\quad y^2 = x^7 + 7abx^5 + 14a^2b^2x^3 + 7a^3b^3x + \\ &\qquad + b^7 - a^7. \end{aligned}$$

The Jacobian $J_r(a,b)$ of the curve $C_r(a,b)$ is thus an abelian variety of dimension $(r-1)/2$. In particular, when $r > 3$, it has dimension > 1 and hence there is no obvious way to attach 2-dimensional Galois representations to $J_r(a,b)$.

To circumvent this issue we use ideas from Darmon's program as explained in the previous section. In particular, the theorem below shows how to recover Kraus' Frey hyperelliptic curve in a similar way as the usual Frey-Hellegouarch elliptic curve (see Example 1). This result achieves Steps **1–2** from the description of Darmon's program given in Section 6 in the case of equation (7.1).

Theorem 10 ([6]).— There exists a hyperelliptic curve $C'_r(t)$ over $K(t)$ of genus $\frac{r-1}{2}$ such that $J'_r(t) = \mathrm{Jac}(C'_r(t))$ is of $\mathrm{GL}_2(K)$-type.

Moreover, for every prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ above a rational prime p, the representation

$$\overline{\rho}_{J'_r(t),\mathfrak{p}} : G_{K(t)} \to \mathrm{GL}_2(\mathcal{O}_K/\mathfrak{p})$$

is a Frey representation of signature (r,r,p).

The hyperelliptic curve $C_r(a,b)/K$ is obtained from $C'_r(t)$ after *specialization* at

$$t_0 = \frac{a^r}{a^r + b^r}$$

and taking the *quadratic twist* by

$$-\frac{(ab)^{\frac{r-1}{2}}}{a^r + b^r}.$$

In this result, it is crucial to notice that the prime p is arbitrary. In particular, if we choose $p = r$ (and hence $\mathfrak{p} = \mathfrak{p}_r$), then $\overline{\rho}_{J'_r(t),\mathfrak{p}_r}$ is a Frey representation of signature (r,r,r). According to Theorem 7, it arises in the Legendre family, allowing us to appeal to the stronger results available for the case of elliptic curves.

This is a key idea in Darmon's program that assuming an appropriate generalization of Serre's modularity conjecture for totally real fields (Conjecture 1), the mod $\mathfrak{p}_r$ representation is modular and plays the role of a 'seed' for modularity of all Frey varieties described by Darmon (see diagram in [11, p. 433]).

The result below makes this argument unconditional for the Kraus Frey variety - under some irreducibility assumption (which is proved to hold for many values of r such as $r = 7$ for instance) and parity conditions - hence completing Steps **3-6** in Darmon's program from Section 6 for equation (7.1).

Theorem 11.— Let (a,b,c) be a non trivial primitive solution to equation (7.1) for exponent $n = p$ prime such that $p \nmid 2rC$. Assume that

$$a \equiv 0 \pmod 2 \quad \text{and} \quad b \equiv 1 \pmod 4. \tag{7.2}$$

Let J_r be the Jacobian of $C_r(a,b)$ base changed to K. Suppose further that $\overline{\rho}_{J_r,\mathfrak{p}}$ is absolutely irreducible. Then, there is a Hilbert newform g over K satisfying the following properties:

(i) g is of parallel weight 2, trivial character and level $2^2\mathfrak{p}_r^2\mathfrak{n}_C$;

(ii) $\overline{\rho}_{J_r,\mathfrak{p}} \simeq \overline{\rho}_{g,\mathfrak{P}}$ for some $\mathfrak{P} \mid p$ in the field of coefficients K_g of g;

(iii) for all $\mathfrak{q}_2 \mid 2$ in K, we have $(\rho_{g,\mathfrak{P}} \otimes \overline{\mathbb{Q}}_p)|_{I_{\mathfrak{q}_2}} \simeq \delta \oplus \delta^{-1}$, where δ is a character of order r;

(iv) $K \subset K_g$.

Moreover, if $\mathfrak{n}_C \neq 1$ then g has no complex multiplication.

Note that, contrary to the case of Fermat's last theorem, the 2-adic assumptions (7.2) in Theorem 11 are not valid in general; indeed, from the symmetry of (7.1), we can only swap a and b, so the possibility of c being even is excluded in the above theorem. We shall explain in the next section how several 'Frey varieties' can complement each other to obtain a complete resolution of certain generalized Fermat equations (7.1) for specific values of r and C.

8 Diophantine applications

In this section, we discuss the Steps **7–8** from Section 6 for the case $r = 7$ and $C = 3$ in the generalized Fermat equations (7.1). In this situation, we achieve the following complete result.

Theorem 12 ([5, Theorem 1.1]).— For all integers $n \geq 2$, there are no non-trivial primitive solutions to

$$x^7 + y^7 = 3z^n. \tag{8.1}$$

First of all, we can reduce the problem of solving $x^7 + y^7 = 3z^n$ for $n \geq 2$ to the case where $n = p$ is prime and $p \geq 5$, $p \neq 7$, using simple arithmetic considerations and work of Bennett-Skinner [1] (for $n = 2$), Bennett-Skinner-Yazdani [2] (for $n = 3$) and Serre [23] (for $n = 7$).

In [5], we actually give three different proofs of Theorem 12 which rely on a 'multi-Frey' approach using a combination of Kraus' hyperelliptic curve $C_7(a,b)$ and two Frey elliptic curves $E/\mathbb{Q}$ and $F/\mathbb{Q}(\zeta_7 + \zeta_7^{-1})$ whose construction is due to Darmon and Freitas, respectively.

Our first proof uses the *classical* modular method outlined in the case of FLT in Section 5 with the two aforementioned Frey elliptic curves attached to equation (8.1).

- (Darmon, [20, §4.5.1.3]) A Frey curve over $\mathbb{Q}$:

$$E_{a,b} : y^2 = x^3 + a_2x^2 + a_4x + a_6$$

where

$$\begin{aligned} a_2 &= -(a-b)^2, \\ a_4 &= -2a^4 + a^3b - 5a^2b^2 + ab^3 - 2b^4, \\ a_6 &= a^6 - 6a^5b + 8a^4b^2 - 13a^3b^3 + 8a^2b^4 - \\ &\quad -6ab^5 + b^6. \end{aligned}$$

- (Freitas, [16, p. 619]) A Frey curve over the totally real cubic field $\mathbb{Q}(\zeta_7 + \zeta_7^{-1})$:

$$F_{a,b} : y^2 = x(x - A_{a,b})(x + B_{a,b}),$$

where for $i = 1, 2$, we have $\omega_i = \zeta_7^i + \zeta_7^{-i}$ and

$$\begin{aligned} A_{a,b} &= (\omega_2 - \omega_1)(a+b)^2 \\ B_{a,b} &= (2 - \omega_2)(a^2 + \omega_1 ab + b^2). \end{aligned}$$

We note here that Freitas' Frey elliptic curve $F = F_{a,b}$ is defined over a totally real field of degree > 1 and is not base change from $\mathbb{Q}$. In particular, its mod p representations are not explained by Darmon's classification of Frey representations of signature $(7, 7, p)$.

The total running time for this first proof is approximately 40 minutes with around 3/4 of this time devoted to computing the Hilbert newforms over $\mathbb{Q}(\zeta_7 + \zeta_7^{-1})$ of parallel weight 2 and level $\mathfrak{q}_2^3\mathfrak{q}_3\mathfrak{q}_7$ (with $\mathfrak{q}_i$ the unique prime ideal above i in $\mathbb{Q}(\zeta_7+\zeta_7^{-1})$) used to deal with the case where ab is even and $7 \mid a + b$. There are precisely 121 such newforms generating a space of dimension 818, with coefficient fields of degree up to 18.

Our second and third proofs of Theorem 12 add in the use of Kraus' Frey hyperelliptic curve

$$C_7(a, b) : y^2 = x^7+7abx^5+14a^2b^2x^3+7a^3b^3x+b^7-a^7$$

in two different ways: the second proof uses $C_7(a, b)$ *as much as possible* whilst the third and last proof is designed to minimize the computational time among all proofs we give. The total running times for these proofs are approximatively 10 minutes and 1 minute respectively.

Our second proof is much more involved and requires introducing many new elimination techniques [6, §9] to discard the isomorphism in Theorem 11((ii)). To illustrate the computational challenges we have faced, let us mention that we had to compute here in the space of Hilbert newforms of level $\mathfrak{q}_2^2\mathfrak{q}_3\mathfrak{q}_7^2$ which has dimension 698. This dimension is comparable in size with that of the space considered in the first proof, but it turns out to be much faster to initialize yielding only 61 newforms. Some of these forms have coefficient field of degree as large as 54 making the elimination procedure considerably more difficult. Fortunately, we are able to reduce the number of newforms to consider down to 25 using the condition $K \subset K_g$ from Theorem 11((iv)). As explained in [5] this 'instantaneous reduction' is only available when working with abelian varieties of dimension > 1. Moreover, we also developed a collection of techniques to speed up the elimination procedure resulting in a great saving in the total running time; see [5, §7]. While this approach a priori requires harder and lengthier computations, it ends up allowing for a faster proof than the previous one.

Our third and last proof builds on the two previous ones. Combining information about the Frey (hyper)elliptic curves introduced above and their twists we manage to lower down to 104 the dimension of the largest space we have to consider. Then we apply the techniques explained for the second proof to deal with the corresponding 19 newforms (whose coefficient fields are all of degree ≤ 15) yielding the most efficient proof in less than a minute. This illustrates how the additional structures carried by the Frey varieties of dimension > 1 can be exploited to reduce computations, despite the fact that we have to work with Jacobians of hyperelliptic curves.

Finally, let us point out that these methods have already been applied to other Fermat-type equations to obtain results not within reach of the classical approach with Frey elliptic curves. In the case of $r = 11$ in (7.1), we refer the reader to [6] and for signature $(p, p, 5)$ to the recent preprint of Chen and Koutsianas [8].

References

[1] Michael A. Bennett and Christopher Skinner. Ternary Diophantine Equations via Galois representations and modular forms. *Canad. J. Math.*, 56(1):23–54, 2004.

[2] Michael A. Bennett, Vinayak Vatsal, and Soroosh Yazdani. Ternary Diophantine equations of signature $(p, p, 3)$. *Compos. Math.*, 140(6):1399–1416, 2004.

[3] Fritz Beukers. The Diophantine equation $ax^p + by^q = cz^r$. *Duke Math. J.*, 91(1):61–88, 1998.

[4] Nicolas Billerey. Introduction to the modular method, 2024. `https://hal.science/hal-04421125v1`

[5] Nicolas Billerey, Imin Chen, Luis Dieulefait, and Nuno Freitas. On Darmon's program for the generalized Fermat equation, II. *Math. Comp.* To appear.

[6] Nicolas Billerey, Imin Chen, Luis Dieulefait, and Nuno Freitas. On Darmon's program for the generalized Fermat equation, I. *J. Reine Angew. Math.* To appear. ArXiv preprint v4, 2024. `https://arxiv.org/abs/2205.15861`

[7] Gebhard Böckle. Galois representations. In *Travaux mathématiques. Vol. XXIII*, volume 23 of *Trav. Math.*, pages 5–35. Fac. Sci. Technol. Commun. Univ. Luxemb., Luxembourg, 2013.

[8] Imin Chen and Angelos Koutsianas. A modular approach to Fermat Equations of signature

$(p, p, 5)$ using Frey hyperelliptic curves. ArXiv preprint, 2025. https://arxiv.org/abs/2210.02316

[9] Gary Cornell, Joseph H. Silverman, and Glenn Stevens, editors. *Modular forms and Fermat's last theorem*. Springer-Verlag, New York, 1997. Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995.

[10] H. Darmon. Faltings plus epsilon, Wiles plus epsilon, and the generalized Fermat equation. *C. R. Math. Rep. Acad. Sci. Canada*, 19(1):3–14, 1997.

[11] Henri Darmon. Rigid local systems, Hilbert modular forms, and Fermat's last theorem. *Duke Math. J.*, 102(3):413–449, 2000.

[12] Henri Darmon, Fred Diamond, and Richard Taylor. Fermat's last theorem. In *Current developments in mathematics, 1995 (Cambridge, MA)*, pages 1–154. Int. Press, Cambridge, MA, 1994.

[13] Henri Darmon and Andrew Granville. On the equations $z^m = F(x, y)$ and $Ax^p + By^q = Cz^r$. *Bull. London Math. Soc.*, 27(6):513–543, 1995.

[14] Fred Diamond and Jerry Shurman. *A first course in modular forms*, volume 228 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005.

[15] Eberhard Freitag. *Hilbert modular forms*. Springer-Verlag, Berlin, 1990.

[16] Nuno Freitas. Recipes to Fermat-type equations of the form $x^r + y^r = Cz^p$. *Math. Z.*, 279(3-4):605–639, 2015.

[17] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. I. *Invent. Math.*, 178(3):485–504, 2009.

[18] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. II. *Invent. Math.*, 178(3):505–586, 2009.

[19] Alain Kraus. On the equation $x^r + y^r = z^p$. Notes for a talk given at IEM, Universität Duisburg-Essen, 1998. Available at https://github.com/NicolasBillerey/xhyper

[20] Alain Kraus. On the equation $x^p + y^q = z^r$: a survey. *Ramanujan J.*, 3(3):315–333, 1999.

[21] B. Mazur. Rational isogenies of prime degree (with an appendix by D. Goldfeld). *Invent. Math.*, 44(2):129–162, 1978.

[22] Ariel Pacetti. A tour through some Diophantine equations. *CIM Bulletin*, 45:37–43, 2023.

[23] Jean-Pierre Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. *Duke Math. J.*, 54:179–230, 1987.

[24] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[25] Richard Taylor. On Galois representations associated to Hilbert modular forms. *Invent. Math.*, 98(2):265–280, 1989.

[26] Richard Taylor and Andrew Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995.

[27] Andrew Wiles. Modular elliptic curves and Fermat's Last Theorem. *Annals of Mathematics*, 144:443–551, 1995.