Non-abelian descent and the generalized Fermat equation

Hugo Chapdelaine

August 2008

Contents

1 The main result

This Chapter gives some finiteness results for the set of primitive solutions of the generalized Fermat equation

$$x^p + y^q = z^r \tag{1}$$

where the exponents p, q, r satisfy the inequality $1/p+1/q+1/r < 1$. The very special "shape" of the surface defined by (1) allows us to use some geometry to reduce its study to the study of non-abelian unramified covers of $\mathbb{P}^1\backslash\{0, 1, \infty\}$ of *signature* (p, q, r) in the sense of Definition 1.1. Therefore the study of the arithmetic of equation (1) can be transferred to the setting of algebraic curves. The main ingredients in the proof are a variant of the Chevalley-Weil theorem, and the finiteness theorems of Hermite-Minkowski and Faltings. This finiteness result for (1) which was proved in [8] can be viewed as an illustrative special case of the Campana program which was presented in Dan Abramovich's lecture series at this summer school.

The author would like to thank Henri Darmon for a careful proofreading of this article which led to many improvements.

A solution $(a,b,c) \in \mathbb{Z}^3$ of (1) is called *nontrivial* if $abc \neq 0$ and *primitive* if $gcd(a,b,c) = 1$. When the exponents p, q and r are pairwise coprime, the following exercise shows that (1) has infinitely many nontrivial but not necessarily primitive solutions.

Exercise 1 Let p, q and r be pairwise coprime. Show that the affine surface defined by $x^p + y^q = z^r$ in $\mathbb{A}^3_{\mathbb{Q}}$ is rational, i.e., the field $qt(\mathbb{Q}[x,y,z]/(x^p + y^q - z^r))$ is purely transcendental of degree 2 over $\mathbb{Q}$.

From now on we are only interested in studying the set of nontrivial primitive solutions of (1). The study of (1) can be split into three cases:

(1) The *spherical case*: $1/p+1/q+1/r > 1$. The possibilities are $\{p,q,r\} = \{2,2,k\}$ with $k \geq 2$, $\{2,3,3\}$,$\{2,3,4\}$ and $\{2,3,5\}$.

(2) The *euclidean case*: $1/p + 1/q + 1/r = 1$. The possibilities are $\{p,q,r\} = \{3,3,3\}$, $\{2,4,4\}$ and $\{2,3,6\}$.

(3) The *hyperbolic case*: $1/p + 1/q + 1/r < 1$.

This division is reminiscent of the classification of algebraic curves which also falls into 3 cases depending on the genus or the sign of the Euler characteristic. Here is the main theorem that we wish to prove.

Theorem 1.1 *(Darmon, Granville) If* $1/p+1/q+1/r < 1$ *then* (1) *has only finitely many nontrivial primitive solutions.*

Note that the statement of this theorem concerns the existence of integral points on a surface. We would like to reduce the study of integral solutions of (1) to the study of K-rational points on an auxiliary projective curve X/K where K is a suitable number field. We consider the map

$$\{\text{Set of nontrivial primitive solutions of equation (1)}\} \to \mathbb{P}^1(\mathbb{Q}) \subseteq \mathbb{P}^1(\mathbb{C})$$
$$(a,b,c) \mapsto \frac{a^p}{c^r},$$

which allows us to reduce the study of (1) to the study of certain branched coverings of $\mathbb{P}^1(\mathbb{C})$. We define the set

$$\Sigma_{p,q,r} := \left\{ \frac{a^p}{c^r} \in \mathbb{Q} : a^p + b^q = c^r, abc \neq 0, gcd(a,b,c) = 1 \right\} \subseteq \mathbb{P}^1(\mathbb{Q}).$$

Exercise 2 Show that $\#\Sigma_{p,q,r} < \infty$ if and only if (1) has finitely many primitive solutions.

Now let us explain the main ideas of Theorem 1.1.

Proof of Theorem 1.1 We want to show that the set of nontrivial primitive solutions of (1) is finite. By Exercise 2, it is enough to show that $\Sigma_{p,q,r}$ is finite when $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} < 1$. The proof can be broken into four steps.

First step: The existence of a Galois branched covering.

Definition 1.1 *A Galois covering $\pi : X \to \mathbb{P}^1$ is said to be of signature (p, q, r) if its ramification indices above $0, 1$ and ∞ are equal to p, q and r respectively, and if π is unramified everywhere else.*

The first stage of the proof consists in constructing a Galois covering of $\mathbb{P}^1$ of signature (p, q, r) defined over a suitable number field K and Galois over that field (The construction of such a cover will be done in detail in Section 2). The Riemann-Hurwitz formula then determines the genus $g(X)$ of X in terms of the degree d of π :

$$\begin{aligned} 2g(X) - 2 &= d(2g(\mathbb{P}^1(\mathbb{C})) - 2) + \frac{d}{p}(p-1) + \frac{d}{q}(q-1) + \frac{d}{r}(r-1) \\ &= d(1 - 1/p - 1/q - 1/r). \end{aligned}$$

Since $1 - 1/p - 1/q - 1/r > 0$ we conclude that $g(X) > 1$.

Second step: A Chevalley-Weil theorem for branched coverings.

Given $t \in \mathbb{P}^1(K)$, let L_t be the smallest field of definition of the closed points in $\pi^{-1}(t)$. As is explained in Section 3, the field L_t is a Galois extension of K with Galois group isomorphic (non-canonically) to a subgroup of $Gal(X/\mathbb{P}^1)$. The Chevalley-Weil theorem for branched coverings (see Theorem 3.2) shows that the ramification of L_t, for $t \in \Sigma_{p,q,r}$, is bounded *independently* of t, in light of the following elementary property of $\Sigma_{p,q,r}$:

Lemma 1.1 *Let $t = \frac{a^p}{c^r} \in \Sigma_{p,q,r}$ then for all prime numbers ℓ we have*

(1) $v_\ell(Numerator(t)) \equiv 0 \ (mod\, p)$,

(2) $v_\ell(Numerator(t-1)) \equiv 0 \ (mod\, q)$,

(3) $v_\ell(Numerator(\frac{1}{t})) \equiv 0 \ (mod\, r)$,

where for $x \in \mathbb{Q}$, $v_\ell(x)$ stands for the valuation of x at the prime ℓ.

Note that proof of Lemma 1.1 uses in a crucial way the primitivity of the solution (a, b, c) corresponding to $t = \frac{a^p}{c^r}$ and the fact that $t - 1 = -\frac{b^q}{c^r}$.

Third step: Hermite-Minkowski.

By the Hermite-Minkowski theorem (cf. Theorem 1.1 in Section 1.1 of [5]) the compositum L of all the number fields L_t, for $t \in \Sigma_{p,q,r}$, is a finite extension of K.

Fourth step: Faltings' Theorem.

By definition of L we have $\pi^{-1}(\Sigma_{p,q,r}) \subseteq X(L)$. Since $g(X) > 1$, we deduce by Faltings' theorem that $X(L)$ is a finite set and therefore $\pi^{-1}(\Sigma_{p,q,r})$ and $\Sigma_{p,q,r}$ are also finite sets. This concludes the sketch of the proof of Theorem 1.1. $\square$

Remark 1.1 The conclusion of Theorem 1.1 remains the same if we replace the equation $x^p + y^q = z^r$ by the more general equation $Ax^p + By^q = Cz^r$ for nonzero fixed integers A, B and C. For a further discussion of the equation $Ax^p + By^q = Cz^r$, see [8].

Remark 1.2 In some special cases, for example when $(p, q, r) = (n, n, n)$ with $n \geq 3$ we know by the work of Wiles and Taylor (see [22] and [20]), that (1) has no nontrivial solutions. Using similar techniques, Darmon and Merel (see [6] and [9]) could also treat the case (p, p, r) where $r = 2$ or 3 and p is a prime number larger or equal to $6 - r$ to conclude that (1) has no nontrivial primitive solutions.

For the rest of the paper, we would like first to explain in details the construction of the auxiliary branched covering $(X_K, \pi, \mathbb{P}^1_K)$ of signature (p, q, r) above $\{0, 1, \infty\}$ which was needed in the first step of the proof of Theorem 1.1. Secondly, we would like to give a more detailed discussion about the variant of the Chevalley-Weil theorem that we have used to control the ramification of the number field L_t over K for the special elements $t \in \Sigma_{p,q,r}$. We won't say anything about Steps 3 and 4 which are discussed in [5]. Sections 2 and 3 are devoted to a discussion of Steps 1 and 2 respectively.

2 Construction of the branched covering

In this section we will use the theory of Riemann surfaces in order to construct certain *analytic* Galois branched coverings over $\mathbb{P}^1(\mathbb{C})$ unramified outside $\{0, 1, \infty\}$.

For every triple of integers (p, q, r) with $p, q, r \geq 2$ we define the Hecke triangle group by the abstract presentation

$$\Gamma_{p,q,r} := \langle \gamma_0, \gamma_1, \gamma_\infty | \gamma_0^p = \gamma_1^q = \gamma_\infty^r = \gamma_0\gamma_1\gamma_\infty = 1 \rangle.$$

It is convenient to allow the exponents p, q and r to be infinite which will be taken to mean that the order of the corresponding element is infinite.

One has that $\pi_1(\mathbb{P}^1(\mathbb{C})\backslash\{0, 1, \infty\}) \simeq \Gamma_{\infty,\infty,\infty} = \langle l_0, l_1, l_\infty | l_0 l_1 l_\infty = 1 \rangle$ which is isomorphic to the free group on two generators. We have the short exact sequence

$$1 \to N_{p,q,r} \to \Gamma_{\infty,\infty,\infty} \xrightarrow{\varphi} \Gamma_{p,q,r} \to 1,$$

where $\varphi(l_0) = \gamma_0$, $\varphi(l_1) = \gamma_1$ and $N_{p,q,r} = ker(\varphi)$. The universal covering space of $\mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\})$ is the upper-half plane, see for example Theorem 6.4.3 of [18]. Let us denote by

$$\theta : \mathcal{H} \to \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\} \tag{2}$$

a choice of such a universal covering map. From the theory of covering spaces one has a (non-canonical) isomorphism between the group of Deck transformations of (2) and the fundamental group of $\mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\})$, see for example §80 of [14]. Such an isomorphism allows us to define an action of $\Gamma_{\infty,\infty,\infty} \simeq \pi_1(\mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\})$ on $\mathcal{H}$. From this, one may deduce the following diagram

$$\begin{array}{ccc} \mathcal{H} & & \\ \downarrow{\scriptstyle \theta_1} & \searrow^{\theta} & \\ U := \mathcal{H}/N_{p,q,r} & & \\ \downarrow{\scriptstyle \theta_2} & & \\ \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\} \simeq \mathcal{H}/\Gamma_{\infty,\infty,\infty} & & \end{array}$$

where θ_1 (resp. θ_2) is the covering map induced by the action of $N_{p,q,r}$ on $\mathcal{H}$ (resp. $\Gamma_{\infty,\infty,\infty}/N_{p,q,r}$ on U).

Note that U is a connected Riemann surface such that $\pi_1(U) \simeq N_{p,q,r}$ and that θ_2 is a Galois covering map with Galois group isomorphic to $\Gamma_{\infty,\infty,\infty}/N_{p,q,r} \simeq \Gamma_{p,q,r}$. One can show that θ_2 is of finite degree if and only if $1/p + 1/q + 1/r > 1$ (see Exercise 4). Since in our setting we work under the assumption that $1/p + 1/q + 1/r < 1$ we see that in this case the map θ_2 is never of finite degree. The pair (U, θ_2) is universal among all Galois coverings over $\mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\}$ of signature (p,q,r) in the following sense: Let $\pi : X \to \mathbb{P}^1(\mathbb{C})$ be a Galois branched covering unramified outside $\{0,1,\infty\}$ with ramification index p above 0, q above 1 and r above ∞. Then π factors through θ_2, i.e., there exists a covering map $\widetilde{\theta}_2 : U \to X\backslash\pi^{-1}(\{0,1,\infty\})$ which makes the following diagram commutative:

$$\begin{array}{ccc} & & U \\ & \swarrow^{\widetilde{\theta}_2} & \\ X\backslash\pi^{-1}(\{0,1,\infty\}) & & \downarrow{\scriptstyle \theta_2} \\ & \searrow_{\pi} & \\ & & \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\} \end{array}$$

Note that $\widetilde{\theta}_2$ is onto and unramified since all the ramification happens already in π. Let us assume that π is finite of degree d then, in this case, X is a compact Riemann

surface. Using the Riemann Hurwitz formula one gets that

$$\begin{aligned} 2g(X)-2 &= d(2g(\mathbb{P}^1(\mathbb{C}))-2)+\frac{d}{p}(p-1)+\frac{d}{q}(q-1)+\frac{d}{r}(r-1) \\ &= d(1-1/p-1/q-1/r). \end{aligned}$$

We thus see that

(1) $g(X)=0$ if $1/p+1/q+1/r>1$,

(2) $g(X)=1$ if $1/p+1/q+1/r=1$,

(3) $g(X)\geq 2$ if $1/p+1/q+1/r<1$.

Again using Theorem 6.4.3 of [18], one may deduce that the universal covering space of X is $\mathbb{P}^1(\mathbb{C})$ if $1/p+1/q+1/r>1$, $\mathbb{C}$ if $1/p+1/q+1/r=1$, and $\mathcal{H}$ if $1/p+1/q+1/r<1$. This explains the trichotomy for the study of (1).

We would like to give a geometrical realization of the universal pair (U,θ_2) in the case where $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}<1$. This will be used to understand the set of elliptic elements of $\Gamma_{p,q,r}$ (see Exercise 3). Since $1/p+1/q+1/r<1$, there exists a hyperbolic triangle in the Poincaré unit disc with angles $\pi/p,\pi/q,\pi/r$, see Figure 1. Let σ_P be the

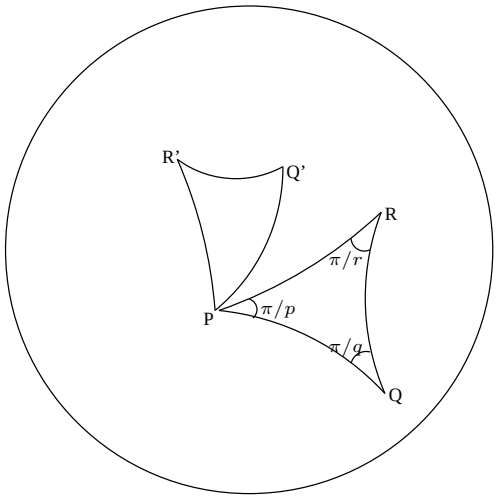


Figure 1: Hyperbolic triangle inside the Poincaré disc.

symmetry with respect to the geodesic passing through QR, σ_Q the symmetry with respect to the geodesic passing through PR and σ_R the symmetry with respect to the geodesic passing through PQ. Let $\gamma_P=\sigma_Q\sigma_R$ be the rotation around P with angle $\frac{2\pi}{p}$, $\gamma_Q=\sigma_R\sigma_P$ be the rotation around Q with angle $\frac{2\pi}{q}$ and $\gamma_R=\sigma_P\sigma_Q$ be the rotation around R with angle $\frac{2\pi}{r}$. We have drawn the image of the triangle PQR

under the rotation γ_P in Figure 1. Since the open unit disc $D(0,1)$ is biholomorphic to $\mathcal{H}$ we can identify the group $\langle \gamma_P, \gamma_Q, \gamma_R \rangle$ as a subgroup of $PSL_2(\mathbb{R}) \simeq Aut(\mathcal{H})$. We have an isomorphism between $\langle \gamma_P, \gamma_Q, \gamma_R \rangle$ and $\Gamma_{p,q,r}$ given by $\gamma_P \mapsto \gamma_0$, $\gamma_Q \mapsto \gamma_1$ and $\gamma_R \mapsto \gamma_\infty$ (prove it by using Figure 1). In particular, we can think of $\Gamma_{p,q,r}$ as a subgroup of $PSL_2(\mathbb{R})$. The group $\Gamma_{p,q,r}$, when applied to the triangle PQR, gives a "half tessellation"of $D(0,1)$. A fundamental domain for the action of $\Gamma_{p,q,r}$ on $D(0,1)$ is given for example by the geodesic quadrilateral $PQRQ'$ where the geodesic RQ' is identified with the geodesic RQ and the geodesic PQ with the geodesic PQ'. It thus follows that the quotient $\mathcal{H}/\Gamma_{p,q,r}$ is isomorphic to $\mathbb{P}^1(\mathbb{C})$. Let

$$\widetilde{\pi} : \mathcal{H} \to \mathcal{H}/\Gamma_{p,q,r} \simeq \mathbb{P}^1(\mathbb{C}).$$

Since $PSL_2(\mathbb{C})$ acts triply transitively on $\mathbb{P}^1(\mathbb{C})$ we can assume that $\widetilde{\pi}(P) = 0, \widetilde{\pi}(Q) = 1$ and $\widetilde{\pi}(R) = \infty$. Therefore the Galois branched covering $\widetilde{\pi}$ has signature (p,q,r) above $\{0,1,\infty\}$. Unfortunately $\widetilde{\pi}$ has infinite degree but the next lemma takes care of this difficulty.

Exercise 3 Define $U := \mathcal{H} \backslash \widetilde{\pi}^{-1}\{0,1,\infty\}$. Show that the map

$$\widetilde{\pi}|_U : U \to \mathbb{P}^1(\mathbb{C}) \backslash \{0,1,\infty\}$$

corresponds to the universal map associated to Galois branched coverings over $\mathbb{P}^1(\mathbb{C})$ of signature (p,q,r). It thus gives a geometrical realization of U as the unit disc minus the vertices of all the $\Gamma_{p,q,r}$-translates of the triangle PQR. Conclude that an element $\gamma \in \Gamma_{p,q,r}$ is elliptic if and only if it fixes a vertex of a $\Gamma_{p,q,r}$-translate of the triangle PQR. Remember that an elliptic element in $PSL_2(\mathbb{R})$ is by definition a matrix which fixes a point in $\mathcal{H}$.

Exercise 4 Show that $\Gamma_{p,q,r}$ is finite if and only if $1/p + 1/q + 1/r > 1$. Show that $\Gamma_{p,q,r}$ is infinite and non abelian if and only if $1/p + 1/q + 1/r \leq 1$.

Lemma 2.1 *There exists a normal subgroup $H \leq \Gamma_{p,q,r}$ such that $[\Gamma_{p,q,r} : H] < \infty$ and such that H acts without fixed point, i.e., H contains no elliptic elements.*

Remark 2.1 Note that the set of all elliptic elements of $\Gamma_{p,q,r}$ consists of the union of the conjugacy classes in $\Gamma_{p,q,r}$ of $\gamma_0^{\mathbb{Z}}$, $\gamma_1^{\mathbb{Z}}$ and $\gamma_\infty^{\mathbb{Z}}$. Moreover, if H is as in Lemma 2.1 then the order of $\overline{\gamma}_0$, $\overline{\gamma}_1$ and $\overline{\gamma}_\infty$ in $\Gamma_{p,q,r}/H$ is equal to p, q and r respectively.

Proof of lemma 2.1 We follow essentially the proof of Proposition 4.4 of [3]. Let us construct an abstract group homomorphism of $\Gamma_{p,q,r}$ onto a certain subgroup of $PSL_2(\mathbb{C})$ for which all its matrices have algebraic entries. Consider the matrices

$$A = \begin{pmatrix} 0 & -\zeta_{2p}^{-1} \\ \zeta_{2p} & \zeta_{2p} + \zeta_{2p}^{-1} \end{pmatrix} \qquad C = \begin{pmatrix} 0 & \zeta_{2p}^{-1}\zeta_{2q}^{-1} \\ -\zeta_{2p}\zeta_{2q} & \zeta_{2r} + \zeta_{2r}^{-1} \end{pmatrix} \qquad B = AC^{-1}$$

where $\zeta_n = e^{2\pi i/n}$. One can verify that the order of A, B and C in $PSL_2(\mathbb{C})$ are p, q and r respectively. For example to show that A has order p one can use the observation that $(-1, 1)$ and $(-\zeta_p^{-1}, 1)$ are eigenvectors with eigenvalues ζ_{2p} and ζ_{2p}^{-1}. A similar argument can be used for B and C. We thus have an onto group homomorphism

$$\rho : \Gamma_{p,q,r} \to \langle A, B, C\rangle =: \mathcal{N} \subseteq PSL_2(R)$$

given by $\rho(\gamma_0) = A^{-1}, \rho(\gamma_1) = B, \rho(\gamma_\infty) = C$ where $R = \mathbb{Z}[\zeta_{2p}, \zeta_{2q}, \zeta_{2r}]$. Note that ρ sends an elliptic element of $\Gamma_{p,q,r}$ to an elliptic element of $\mathcal{N}$ and all elliptic elements of $\mathcal{N}$ are contained in a conjugacy class of $A^{\mathbb{Z}}$, $B^{\mathbb{Z}}$ or $C^{\mathbb{Z}}$. Let π be some prime ideal of R. Note that $A = P \begin{pmatrix} \zeta_{2p} & 0 \\ 0 & \zeta_{2p}^{-1} \end{pmatrix} P^{-1}$ for some matrix $P \in PSL_2(R)$. Therefore if $A(mod\,\pi) \equiv I(mod\,\pi)$ then $\begin{pmatrix} \zeta_{2p} & 0 \\ 0 & \zeta_{2p}^{-1} \end{pmatrix} \equiv I(mod\,\pi)$, where I stands for the identity matrix. This implies that $\pi|(1 - \zeta_{2p})$. We have a similar thing for B and C. Let us choose a prime ideal π such that π does not divide $1 - \zeta_n^k$ for $1 \leq k \leq n-1$ and $n \in \{p, q, r\}$. Finally define the group

$$H := \{g \in \Gamma_{p,q,r} | \rho(g) \equiv I(mod\,\pi)\}. \tag{3}$$

The group H satisfies the property of Lemma 2.1. $\square$

We can finally define the auxiliary curve that was used in the course of the proof of Theorem 1.1. Define

$$X := \mathcal{H}/H,$$

where H is as in Lemma 2.1 and let π be the natural map

$$\pi : X \to \mathcal{H}/\Gamma_{p,q,r} \simeq \mathbb{P}^1(\mathbb{C}). \tag{4}$$

By construction π is a finite complex analytic Galois branched covering over $\mathbb{P}^1(\mathbb{C})$ of signature (p, q, r). Since π has finite degree and $\mathbb{P}^1(\mathbb{C})$ is compact we deduce that X is compact Riemann surface. Note that the complex structure of X is inherited from the complex structure of $\mathcal{H}$, where some care should be taken in order to define local charts around fixed points of elliptic elements of $\Gamma_{p,q,r}$.

There is a dictionary between non singular projective curves over $\mathbb{C}$ and compact Riemann surfaces

Theorem 2.1 *Any compact Riemann surface S is algebraic.*

Let us sketch a proof of this important result in the special case where S is the compact Riemann surface X that was previously constructed as a quotient of the upper half-plane.

Sketch of the proof We will break the proof in three steps.

Step 1: X admits a large supply of non-constant meromorphic functions.

We first show that X admits a large supply of non-constant meromorphic functions in the sense that for every pair of points $P, Q \in X$ with $P \neq Q$ there exists a meromorphic function f on X such that $f(P) \neq f(Q)$ (separates points) and for every $P \in X$ there exists a meromorphic function g on X such that g is a local chart in a small neighborhood of P (separates tangents).

Let G be the preimage of H under the natural projection $SL_2(\mathbb{R}) \to PSL_2(\mathbb{R})$. Note that G is a discrete subgroup of $SL_2(\mathbb{R})$ which contains the element $-I$. For every pair of points $P, Q \in \mathcal{H}$ consider the Poincaré series (modular form)

$$f_m(P, Q, z) = \sum_{g \in G} r_{P,Q}(gz) j(g, z)^{-m} \tag{5}$$

where m is any fixed *even integer* larger or equal to 4, $r_{P,Q}(z) = \frac{z-P}{z-Q}$, $gz = \frac{az+b}{cz+d}$ and $j(g, z) = (cz + d)$ for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$. The infinite sum (5) converges absolutely since $m \geq 3$. Therefore the function $f_m(P, Q, z)$ is meromorphic on $\mathcal{H}$ and satisfies the important transformation formula

$$f_m(P, Q, gz) = (cz + d)^m f_m(P, Q, z) \quad \forall\, g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G. \tag{6}$$

Note that when m is odd, the transformation formula (6) applied to the matrix $-I \in G$ implies that $f_m(P, Q, z)$ is identically equal to 0. Let $P, Q \in \mathcal{H}$ be arbitrary points such that GP and GQ are distinct right orbits. Now choose a third point $R \in \mathcal{H}$ such that $f_m(R, Q, z)$ does not vanish at $z = P$ (It is easy to see that such a point R always exists by considering for example the function $\omega \mapsto f_m(\omega, Q, P)$). A simple calculation reveals that the function $f_m(R, Q, z)$ has a pole of order one at every elements of GQ (this uses the fact that m is even). It thus follows that $f_m(R, Q, z)$ is a non-constant meromorphic function on $\mathcal{H}$. Now let us consider the quotient

$$F_m(z) = \frac{f_m(R, P, z)}{f_m(R, Q, z)}.$$

Using (6), one readily sees that $F_m(z)$ descends to a meromorphic function on X. Moreover, by construction, it has a zero of order one at the point $GQ \in X$ and a pole of order one at the point $GP \in X$. This shows that X has a set of meromorphic functions that separates points and tangents.

Step 2: Riemann-Roch.

Let D be a divisor of X and let $\mathcal{L}_D$ be the locally free $\mathcal{O}_X$-module of rank 1 associated to D where for every open set $U \subseteq X$

$$\mathcal{L}_D(U) = \{ f : U \to \mathbb{C} : f \text{ is meromorphic and } div(f) \geq -D|_U \}.$$

Then the famous theorem of Riemann-Roch says

Theorem 2.2 *(Riemann-Clebsch-Roch)*

$$dim_{\mathbb{C}} H^0(X, \mathcal{L}_D) - dim_{\mathbb{C}} H^1(X, \mathcal{L}_D) = deg(D) + 1 - g,$$

where g stands for the genus of X and $deg(D)$ for the degree of the divisor D.

For an elementary proof of Theorem 2.2 which uses only Step 1, see chapter IV of [13].

Step 3: Construction of a planar parametrization of X.

Proposition 2.1 *Let z be a non-constant meromorphic function on X. Then there exists a meromorphic function f and an irreducible algebraic equation $P(z, f)$ defined over $\mathbb{C}$ such that the map*

$$x \mapsto (z(x), f(x))$$

is a conformal bijection of X onto the compact Riemann surface associated to the irreducible equation $P(z, f) = 0$.

Proposition 2.1 is a nice application of Riemann-Roch and the analytic continuation principle for germs of holomorphic functions. For a detailed proof, see for example the discussion on p. 242 of [11]. This concludes the sketch of the proof of the algebraicity of X. $\square$

Remark 2.2 Historically, Riemann proved the inequality $dim_{\mathbb{C}} H^0(X, \mathcal{L}_D) \leq deg(D) + 1$ by constructing meromorphic differential forms with prescribed poles at points appearing in D, see [16]. His construction appealed to the so-called Dirichlet's principle which back then was not rigorously proved. An inequality going in the other direction was proved by Clebsch (see [4]) and then refined by Roch (see [17]). In general, the construction of non-constant meromorphic functions (or non-zero meromorphic differential forms) on an abstract compact complex manifold of dimension one (i.e., a compact Riemann surface) is a highly non-trivial fact. When the dimension is higher than one, it is the lack of non-constant meromorphic functions which prevents compact complex manifolds to be algebraic. In dimension one, the construction of such functions can done abstractly by the use of harmonic analysis, see for example Section 5.2 of [11]. Note that in Step 1 of the previous argument, we could get around this

non-trivial fact by taking advantage of the description of X as a certain quotient of $\mathcal{H}$. This allowed us to define directly Poincaré series which are meromorphic $\frac{m}{2}$-fold differential forms. The idea of constructing meromorphic $\frac{m}{2}$-fold differential forms by averaging over the elements of a fuchsian group is due to Poincaré. Poincaré was the first one to announce that for every algebraic curve $P(x,y)=0$ (of genus ≥ 2) there exists two non-constant fuchsian functions $f(z)$ and $g(z)$ such that $P(f(z),g(z)) \equiv 0$, see [15]. Finally, we should mention a more recent way of proving Theorem 2.1 under the *additional assumption* that the compact Riemann surface S admits a single non-constant meromorphic function f, i.e., a non-constant holomorphic function $f : S \to \mathbb{P}^1(\mathbb{C})$. This alternative approach is a special case of a general equivalence between analytic and algebraic coherent sheaves on smooth projective algebraic varieties. Very often, this equivalence is quoted under the acronym "GAGA principle", see Section 6.1 of [18] and [19]. The key point is that this holomorphic function $f : S \to \mathbb{P}^1(\mathbb{C})$ gives rise to a coherent analytic sheaf $\mathcal{F}$ on $\mathbb{P}^1(\mathbb{C})$ (which is an algebraic curve) and therefore, by GAGA, $\mathcal{F}$ is an algebraic sheaf. From this we may conclude that S is algebraic.

Now we recall that we have constructed previously a branched covering $\pi : X \to \mathbb{P}^1(\mathbb{C})$ of signature (p,q,r). Now armed with Proposition 2.1, we know that there exists a meromorphic function f on X and a polynomial $P(x,y) \in \mathbb{C}[x,y]$ such that $P(\pi, f) = 0$. Let M be the subfield of $\mathbb{C}$ generated by the coefficients of $P(x,y)$. Note that M is a finitely generated field over $\mathbb{Q}$. In general the field M will not be an algebraic extension over $\mathbb{Q}$. Nevertheless we have the following key proposition:

Proposition 2.2 *There exists a smooth projective algebraic curve $\widetilde{X}$ defined over a number field K such that the following diagram commutes*

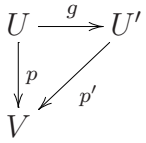
$$\begin{array}{ccc} X & \xrightarrow{\widetilde{g}} & \widetilde{X} \\ \downarrow{\scriptstyle \pi} & \swarrow{\scriptstyle \widetilde{\pi}} & \\ \mathbb{P}^1 & & \end{array}$$

where $\widetilde{g} : X(\mathbb{C}) \to \widetilde{X}(\mathbb{C})$ is an isomorphism defined over $\mathbb{C}$ and where $\widetilde{\pi}$ is a branched covering defined over K.

Proposition 2.2 is a direct application of the following general result

Theorem 2.3 *Let V be an algebraic variety defined over an algebraically closed field L of characteristic 0, and let L' be an algebraically closed field extension of L. Then every covering $p : U \to V$ defined over L' comes from a covering $p' : U' \to V$*

defined over L in the sense that there exists a commutative diagram



where g is an isomorphism of varieties defined over L' and p' is a covering defined over L.

Proof See the proof of Theorem 6.3.3 of [18]. $\square$

Let us explain how the existence of $\widetilde{f}$ and $\widetilde{\pi}$ follows from Theorem 2.3. Let Y be the algebraic curve over $\mathbb{C}$ defined by $X \backslash \pi^{-1}\{0,1,\infty\}$. Note that $\pi|_Y : Y \to \mathbb{P}^1 \backslash \{0,1,\infty\}$ is a covering defined over $\mathbb{C}$ and that $\mathbb{P}^1 \backslash \{0,1,\infty\}$ is an algebraic curve defined over $\overline{\mathbb{Q}}$ (in fact over $\mathbb{Q}$!). From Theorem 2.3, we know that there exists a covering $\pi' : Y' \to \mathbb{P}^1 \backslash \{0,1,\infty\}$ defined over $\overline{\mathbb{Q}}$ and an isomorphism $g : Y \to Y'$ defined over $\mathbb{C}$ such that $\pi' \circ g = \pi$. Let K be the field generated by the coefficients of the equations defining the algebraic curve Y'. Note that K is finitely generated over $\mathbb{Q}$ and therefore it is a number field. The open Riemann surfaces $Y(\mathbb{C})$ and $Y'(\mathbb{C})$ admit natural compactifications X and $\widetilde{X}$ (just add the deleted points) where $\widetilde{X}$ can be chosen to be defined over K. Finally, note that the map g (resp. π') extends uniquely to a map $\widetilde{g} : X \to \widetilde{X}$ defined over $\mathbb{C}$ (resp. $\widetilde{\pi} : \widetilde{X} \to \mathbb{P}^1$ defined over K).

Remark 2.3 Unfortunately, the proof of Theorem 2.3 doesn't give any control on the number field K which appears in Proposition 2.2. For a different proof which gives some control on the number field K, see [12].

Remark 2.4 Note that Proposition 2.2 implies the "if part"of the famous *Belyi*'s theorem which states that a compact Riemann surface X admits a model over $\overline{\mathbb{Q}}$ if and only if there exists a branched covering $\pi : X \to \mathbb{P}^1(\mathbb{C})$ which is unramified outside $\{0,1,\infty\}$. Historically, this direction is due to Weil; see [21]. The "only if part"is not really longer to prove, in fact it is shorter. Its proof is completely algorithmic and is due to Belyi; see [2].

Remark 2.5 In general, for higher dimensional complex varieties one has the following criterion which characterizes varieties which admit a model over a number field

Theorem 2.4 *(Gonzàlez-Diez) An irreducible complex projective variety X can be defined over a number field if and only if the family of all its conjugates X^σ , where σ is any field automorphism of $\mathbb{C}$, contains only finitely many isomorphism classes of complex projective varieties.*

For a proof of this criterion see [10].

Combining Theorem 2.1, Proposition 2.2 and our discussion on branched coverings we see that every finite index normal subgroup $H \leq \Gamma_{p,q,r}$, which contains no elliptic elements, gives rise to an algebraic Galois branched covering over $\mathbb{P}^1$ of signature (p, q, r) defined over a suitable number field K where the number field K depends on H. Such covers turn out to be extremely useful since they can be used to study the set of integral solutions of (1). From the previous observation one may deduce the following general principle:

Principle 2.1 *There is a dictionnary between the distinct strategies for studying $x^p + y^q = z^r$ and the finite quotients of the Hecke triangle group $\Gamma_{p,q,r}$.*

This principle is slightly imprecise but at least, from the author's point of view has the virtue of being inspiring. We won't say more about it and we encourage the reader to look at [7] where Principle 2.1 is explained in greater details.

3 A Chevalley-Weil theorem for branched coverings

In this section we would like to present a variant of the Chevalley-Weil theorem that allowed us, during the second step of the proof of Theorem 1.1, to control the ramification of the field extension L_t over K for the special elements $t \in \Sigma_{p,q,r}$. Let us first recall the Chevalley-Weil theorem in the context of curves (see also Section 1.2 of [5]).

Theorem 3.1 *(Chevalley-Weil) Let X and Y be smooth schemes of relative dimension one defined over the ring of S-integers $\mathcal{O}_{L,S}$ of a number field L where S is a finite set of places of L. Let $f : X \to Y$ be a morphism of schemes defined over $\mathcal{O}_{L,S}$ which is unramified over the generic fiber. Then there exists a finite extension L'/L such that*

$$f^{-1}(Y(\mathcal{O}_{L,S})) \subseteq X(\mathcal{O}_{L',S'}),$$

where the set of places S' extend the set of places of S.

Remark 3.1 In the statement it was important to work with integral models of X and Y in order to make sense to the notion of integral points, i.e., $\mathcal{O}_{L,S}$-valued points. In general, the notion of integral points and rational points differ since the set $X(\mathcal{O}_{L,S})$ could be smaller that the set $X(L)$. For example, consider the *affine* curve E defined by the equation $y^2 - x^3 - 73x = 0$. By Siegel's Theorem one has that $\#E(\mathbb{Z}) < \infty$. On the other hand, since the Mordell-Weil group of $E/\mathbb{Q}$ has positive

rank, one has that $\#E(\mathbb{Q}) = \infty$. However, there is an important situation where the two notions coincide namely in the special case where the curve X is projective.

Remark 3.2 At this point we can't resist to give an nice application of the Chevalley-Weil theorem when combined with Faltings' theorem. Consider the affine complex curve embedded in $\mathbb{A}^4(\mathbb{C})$ defined by the zero locus

$$Z(u+v-1, uw-1, vt-1) = \{(u,v,w,t) \in \mathbb{A}^4(\mathbb{C}) : u+v-1 = uw-1 = vt-1 = 0\}.$$

It is easy to see that the map

$$\begin{aligned} \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\} &\to Z(u+v-1, uw-1, vt-1) \\ [u,1] &\mapsto (u, 1-u, 1/u, 1/(u-1)) \end{aligned}$$

is an isomorphism of complex curves. From this, we deduce that the coordinate ring of $\mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\}$, which is $\mathbb{C}[u, \frac{1}{u}, \frac{1}{u-1}]$, is isomorphic to $\mathbb{C}[u,v,w,t]/(u+v-1, uw-1, vt-1)$. Now choose a *covering* (so unramified)

$$\pi : Y(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\}$$

where $Y(\mathbb{C})$ is an open Riemann surface of genus larger or equal to 2 (there are infinitely many possibilities for π). Finally, combining Faltings and Chevalley-Weil we may conclude that the equation

$$u + v = 1$$

has only finitely many solutions in $\mathcal{O}_{L,S}^{\times}$ where L is an arbitrary number field and S is any finite set of places of L. Historically, Siegel was the first to prove this result. Of course, he proved it without appealing to Faltings' theorem.

For the rest of this section we would like to discuss in more details the variant of the Chevalley-Weil theorem that was used in the proof of Theorem 1.1. Let $(X_K, \pi, \mathbb{P}^1_K)$ be the algebraic Galois branched covering of degree d, with Galois group G and signature (p,q,r) constructed in Section 2. Let us fix an embedding of K into $\mathbb{C}$. Since π is defined over K we have a natural action of $Gal(\mathbb{C}/K)$ on all the fibers of π above points $t \in \mathbb{P}^1(K)$. Moreover, for every $t \in \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\}$, we have a simply transitive action of G on $\pi^{-1}(t)$ since π is Galois. We thus get two group homomorphisms:

$$\rho_1 : Gal(\mathbb{C}/K) \to Sym(\pi^{-1}(t)) \quad \text{and} \quad \rho_2 : G \to Sym(\pi^{-1}(t)).$$

It is important to know how ρ_1 and ρ_2 are related. Let us choose a complex embedding $\varphi : X(\mathbb{C}) \hookrightarrow \mathbb{P}^N(\mathbb{C})$. For every $P \in X(\mathbb{C})$ let us denote the image of P by φ by $\varphi(P) = [\varphi_0(P), \varphi_1(P), \ldots, \varphi_N(P)] \in \mathbb{P}^N(\mathbb{C})$. For $t \in \mathbb{P}^1(K)\backslash\{0,1,\infty\}$ define the

number field L_t to be the field generated over K by all the coordinates of $\varphi(P)$ for all $P \in \pi^{-1}(t)$. Let $\pi^{-1}(t) = \{P_1, \ldots, P_d\}$. The first thing to notice is that the number field $L' := K(\varphi(P_1))$ is equal to L_t. For every $i \in \{1, \ldots, d\}$ there exists an element $g \in G$ such that $g(\varphi(P_1)) = \varphi(P_i)$. Therefore the coordinates of $\varphi(P_i)$ can be expressed algebraically in terms of the coordinates of $\varphi(P_1)$ so $L' = L_t$. It thus follows that the action of an element $\sigma \in Gal(\overline{K}/K)$ on L_t is completely determined by its action on the coordinates of $\varphi(P_1)$. Since $\sigma(\varphi(P_1)) = \varphi(P_i)$ for a unique i we readily see that every automorphism of L_t/K can be realized "algebraically"by the action of a unique element $g \in G$ (G acts simply transitively on the fibers). We have the following identification

$$Gal(L_t/K) = \{g \in G : \exists \sigma \in Gal(\overline{K}/K) \text{ such that } \sigma(\varphi(P_1)) = g\varphi(P_1)\} \subseteq G.$$

We would like now to understand the ramification of L_t over K when $t \in \mathbb{P}^1(K)$. The morphism $\pi : X_K \to \mathbb{P}^1_K$ induces an inclusion of fields $K(\mathbb{P}^1) \simeq K(x) \hookrightarrow K(X)$ where x is a variable. Note that $K(X)/K(x)$ is Galois. Let $t \in K$. We define the specialization of π at t to be the K-algebra map

$$K \simeq K[x]/(x-t) \hookrightarrow K[X]/(x-t)$$

where $K[X]$ corresponds to the integral closure of $K[x]$ in $K(X)$. Let $t \in \mathbb{P}^1(K)\backslash\{0, 1, \infty\}$. Since π is unramified at all the points above t we have $(x-t)K[X] = \mathfrak{p}_1 \ldots \mathfrak{p}_r$ where the $\mathfrak{p}_i$'s are distinct prime ideals of $K[X]$. We thus find that $K[X]/(x-t) \simeq L_1 \oplus \ldots \oplus L_r$ where $L_i = K[X]/\mathfrak{p}_i$. Note that all L_i's are Galois over K with Galois group $D(\mathfrak{p}_i/(x-t)) = \{g \in G : g(\mathfrak{p}_i) = \mathfrak{p}_i\}$ so that all the L_i's collapse to the same number field in a fixed algebraic closure of K.

Exercise 5 Show that $L_i/K \simeq L_t/K$.

In order to understand the ramification of L_i over K we need to define the arithmetic intersection between two points $a, b \in \mathbb{P}^1(K)$ at a prime ideal $\wp$ of K.

Definition 3.1 *Let $\wp$ be a prime ideal of K and $a, b \in K \cup \{\infty\}$. We define*

$$I_\wp(a,b) := \begin{cases} ord_\wp(a-b) & \text{if } ord_\wp(a) \geq 0, ord_\wp(b) \geq 0 \\ ord_\wp(\frac{1}{a} - \frac{1}{b}) & \text{if } ord_\wp(\frac{1}{a}) \geq 0, ord_\wp(\frac{1}{b}) \geq 0 \\ 0 & otherwise \end{cases}$$

where $ord_\wp(0) = \infty$ and $ord_\wp(\infty) = -\infty$.

Before stating the Chevalley-Weil theorem for branched covering we need to make one more definition.

Definition 3.2 *Let $X \xrightarrow{G} \mathbb{P}^1$ be a Galois branched covering over $\mathbb{C}$. A Galois branched covering $x : X_K \xrightarrow{G} \mathbb{P}^1_K$ is called a good model for $X \xrightarrow{G} \mathbb{P}^1$ over K if*

the primes of $\mathcal{O}_K$ (when viewed as primes in $\mathcal{O}_K[x]$) that ramify in $\mathcal{O}_K[X_K]$ are contained in S_{bad}. The ring $\mathcal{O}_K[X_K]$ stands for the integral closure of $\mathcal{O}_K[x]$ in $K(X_K)$ and the set S_{bad} is the union of the set of primes that divide the order of G and the set of primes at which two branch points meet.

We can now state in more details a result due to Beckmann which implies the "ramification control"of $L_1/K \simeq L_t/K$ (by Exercise 5) where

$$K[X]/(x-t) \simeq L_1 \oplus \ldots \oplus L_r \tag{7}$$

and L_t for $t \in \Sigma_{p,q,r}$ is defined as in Step 2 of the proof of Theorem 1.1. We have the following theorem which is a special case of Theorem 1.2 of [1].

Theorem 3.2 *(Chevalley-Weil for branched coverings) Assume that $X_K \stackrel{G}{\to} \mathbb{P}^1_K$ is a good model where $G = \langle \overline{\gamma}_0, \overline{\gamma}_1, \overline{\gamma}_\infty \rangle$ and let $L = L_1$ be as in (7). Then L is ramified only at the places $S = S_{bad} \cup S_t$ where*

$$S_{bad} := \{\wp \text{ is a finite prime of } K : \wp | \#G\}$$

and

$$S_t = \{\wp \text{ is a finite prime of } K : I_\wp(t,j) > 0 \text{ for some } j \in \{0,1,\infty\}\}.$$

Moreover, if t meets $j \in \{0,1,\infty\}$ at $\wp$, i.e., $I_\wp(t,j) > 0$ (note that j can at most meet one of those values) then

$$I(\mathfrak{p}/\wp) = \langle \overline{\gamma}_j^{I_\wp(t,j)} \rangle$$

up to conjugation in G where $\mathfrak{p}$ is some prime ideal of L above $\wp$.

The last part of the theorem says basically that the geometric ramification "controls"the arithmetic ramification.

Remark 3.3 In general one cannot always guarantee the existence of a good model but nevertheless, Theorem 3.2 remains valid if we add to the set S_{bad} the finite set of primes that prevent the model to be good.

Using the previous theorem one deduces the following proposition.

Proposition 3.1 *Let $t \in \Sigma_{p,q,r}$ and $\wp \nmid S_{bad}$ then L_t/K is unramified at $\wp$.*

Proof Since $t \in \Sigma_{p,q,r} \subseteq \mathbb{P}^1(\mathbb{C})\backslash\{0,1,\infty\}$ we have $t = \frac{a^p}{c^r}$ for coprime integers a, c. Moreover $t - 1 = -\frac{b^q}{c^r}$ for b and c coprime. By Lemma 1.1 we have

$$\begin{aligned} I_\wp(t,0) &\equiv 0 \pmod{p}, \\ I_\wp(t,1) &\equiv 0 \pmod{q}, \\ I_\wp(t,\infty) &\equiv 0 \pmod{r}. \end{aligned}$$

Using the last part of Theorem 3.2 we deduce that $I(\mathfrak{p}/\wp) = 1$. Therefore L_t is unramified at $\wp$. $\square$

References

[1] S. Beckmann, *On extensions of number fields obtained by specializing branched coverings*, J. Reine Angew. Math., **419**, 27-53, 1991.

[2] G.V. Belyi. *Galois extensions of a maximal cyclotomic field* Izv. Akad. Nauk SSSR Ser., **43**, 267-276, 1979.

[3] F. Beukers. *The generalized Fermat equation*, Duke Math. Journal, **91**, 61-88, 1998.

[4] A. Clebsch. *Über diejenige ebenen Curven, deren Coordinaten rationale Functionen eines Parameters sind*, J. Reine Angew. Math. **64** (1865), p. 43-65.

[5] H. Darmon *Rational points on curves*, in this volume.

[6] H. Darmon. *Rigid local systems, Hilbert modlar forms, and Fermat's last theorem*, Duke. Math. J. (3), **102**, 413-449, 2000.

[7] H. Darmon. *A fourteenth lecture on Fermat's last theorem*, CRM Proc. Lecture Notes 36 Amer. Math. Soc., **36**, 103-115, 2004.

[8] H. Darmon and A. Granville. *On the equations* $x^p + y^q = z^r$ *and* $z^m = f(x, y)$, Bulletin of the London Mathematical Society, **129**, 513-544, 1995.

[9] H. Darmon and L. Merel. *Winding quotients and some variants of Fermat's Last Theorem*, J. Reine Angew. Math., **490** (1997), p. 81-100.

[10] G. Gonzàlez-Diez, *Variations on Belyi's theorem*, Quart. J. Math., **57**, 355-366, 2006.

[11] J. Jost. *Compact Riemann surfaces, third edition*, Springer-Verlag, 2006.

[12] B. Köck. *Belyi's theorem revisited*, Beiträge Algebra Geometrie, **45**, 253-265, 2004.

[13] R. Miranda. *Algebraic curves and Riemann surfaces*, American Mathematical Society, 1991.

[14] J. R. Munkres. *Topology, Second edition*, Prentice Hall, 2000.

[15] H. Poincaré. *Sur les fonctions fuchsiennes*, C.R. Acad. Sci. Paris, 92 (1882), p. 1038-1040.

[16] B. Riemann. *Theorie der Abel'schen Functionen*, J. Reine und Angew. Math., **54** (1857), p. 115-155.

[17] G. Roch. *Über die Anzahl der willkürlichen Constanten in algebraischen Functionen*, J. Reine Angew. Math., **64** (1865), p. 372-376.

[18] J.-P. Serre. *Topics in Galois theory*, Jones and Bartlett Publishers, 1992.

[19] J.-P. Serre. *Géométrie algébrique et géométrie analytique*, Ann. Inst. Fourier, **6** (1956), p. 1-42.

[20] R. Taylor and A. Wiles. *Ring theoretic properties of certain Hecke algebras*, Ann. Math. (2), **141**, 553-572, 1995.

[21] A. Weil. *The field of definition of a variety*, Amer. J. Math., **78**, 509-524, 1956.

[22] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. Math. (2), **141**, 443-551, 1995.