Elem. Math. 67 (2012) 169 – 181
0013-6018/12/040169-13
DOI 10.4171/EM/207

Elemente der Mathematik

On certain permutation groups and sums of two squares

Pietro Corvaja, Carlo Petronio, and Umberto Zannier

Pietro Corvaja was born in 1967. He graduated in mathematics at Pisa University in 1989. He obtained a PhD in mathematics in Paris in 1995. In Udine since 1995, he is mainly interested in Diophantine equations and Arithmetic Geometry.

Carlo Petronio was born in 1966. He graduated at the University of Pisa in 1989 and obtained his PhD at the Scuola Normale Superiore of Pisa in 1995. He is a Professor of Geometry at the University of Pisa. His research topic is Geometric Topology.

Umberto Zannier graduated in Pisa in 1980. He is now a Professor of Geometry at Scuola Normale Superiore in Pisa. He is mainly interested in Number Theory.

1 Introduction

The present note does not contain new results. It may be considered as an *addendum* to [6], presenting several different viewpoints around one of the results proved there, and some links of this result with classical works. In particular, we provide a new reading of the proofs of the following known facts:

Theorem 1.1 ([8, 6])**.** *Let d be an integer congruent to* 1 *modulo* 4*. Then the following conditions are equivalent to each other:*

(a) *d is the sum of the squares of two integers.*

(b) *There exist permutations $\sigma_0, \sigma_1, \sigma_\infty \in \mathfrak{S}_d$ such that:*

Albert Girard machte 1632 die Beobachtung, dass eine ungerade Primzahl p genau dann Summe zweier Quadrate ist, wenn $p \equiv 1 \mod 4$. Da Fermat am 25. Dezember 1640 in einem Brief an Mersenne berichtete, er habe einen Beweis dafür gefunden, wird der Satz gelegentlich Fermats Weihnachtstheorem genannt. Euler gelang 1749 ein Beweis mit Hilfe der Methode des unendlichen Abstiegs. Neuere Forschungen zeigten, dass eine beliebige natürliche Zahl $d \equiv 1 \mod 4$ Summe zweier Quadrate ist, genau dann wenn drei Permutationen von d Objekten existieren, die gewissen einfachen kombinatorischen Bedingungen genügen. In der vorliegenden Arbeit zeigen die Autoren dieses Resultat aus einem neuen Blickwinkel und beweisen dabei auch den klassischen Satz für Primzahlen neu.

(i) $\sigma_0\sigma_1 = \sigma_\infty$.

(ii) *The cycles in the decompositions of* σ_0, σ_1, *and* σ_∞ *have lengths, respectively,* $(1, 4, \ldots, 4)$, $(1, 4, \ldots, 4)$, *and* $(1, 2, \ldots, 2)$.

(iii) *The subgroup of* $\mathfrak{S}_d$ *generated by* σ_0, σ_1, σ_∞ *acts transitively on the set* $\{1, \ldots, d\}$.

Theorem 1.2 (Fermat-Euler)**.** *If* p *is a prime and* p *is congruent to* 1 *modulo* 4 *then* p *is the sum of the squares of two integers.*

In this work, we focus on the relation between these purely algebraic results and the theory of (ramified) covers of surfaces. The latter, in turn, is connected to algebraic geometry in the following way: every smooth complex algebraic curve can be viewed as a real surface, and every non-constant morphism between complex algebraic curves can be viewed as a branched (or ramified) cover of the corresponding topological surfaces. (We shall give later more precise definitions.) Especially, the algebraic curves coming into play will be the projective line (topologically the sphere) and genus-one curves, or elliptic curves (topologically two-dimensional tori).

It is well-known that the crucial point for the classification of integers which are sums of two squares resides in the above Fermat-Euler theorem, which concerns the representation of primes. Among the many different proofs of this classical theorem, the one provided by Ritt [8] in the 1920's makes use of the construction of three permutations as in Theorem 1.1 and of the corresponding unramified cover of $\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$.

The full Theorem 1.1, for arbitrary d and not only for primes, was established in [6] using the translation of condition (b) into the geometric question of the existence of a branched cover over the sphere matching certain prescribed branching data. This is the real-differential analogue of the algebraic viewpoint of Ritt, with the complex projective line replaced by the real sphere and covers defined in the sense of differential topology.

In the differential context, a branched cover is a surjective map π between surfaces, locally modeled on maps of the form $(\mathbb{C}, 0) \ni z \mapsto z^a \in (\mathbb{C}, 0)$ for some $a \in \mathbb{N}$; moreover 0 in the target $\mathbb{C}$ is a branching point of π if $a > 1$, in which case a is called the local degree of π at 0 in the source $\mathbb{C}$, and the branching data of π consist of the local degrees of π at the preimages of its branching points. Note that if the map is proper and all local degrees are equal to one, then the map is a cover in the classical sense of algebraic topology. We also say it is an unramified cover. A modern treatment of this theory can be found in Fulton's book [2].

Given a branched cover π as above, of total degree d, one gets permutations in $\mathfrak{S}_d$ by considering the monodromy of π around the branching points. More precisely, for any branching point one gets a permutation whose cycle decomposition lengths are the local degrees of π at the preimages of the point.

Equivalence between (a) and (b) was proved in [6] by interpreting branched covers as covers between 2-orbifolds, and by showing, using the geometry of orbifolds, that the degree of an orbifold cover arising from permutations as in (b) must satisfy (a), and conversely. In this note we will reprove Theorem 1.1 using the fact that, by the Riemann Existence Theorem, any topological branched cover over the sphere can be realized as a ramified cover of $\mathbb{P}_1(\mathbb{C})$ by a complex algebraic curve (see [3, 12]). Using this complex-algebraic viewpoint

we will show that the existence of permutations as in condition (b) is equivalent to the existence of a certain endomorphism of the elliptic curve E with Weierstrass equation $y^2 = x^3 - x$, namely of the complex torus $E = \mathbb{C}/\mathbb{Z}[i]$. Equivalence with condition (a) will then follow from the fact that the degree of an endomorphism of this elliptic curve is always the sum of two squares. We mention that the link between branched covers and elliptic curves already appears implicitly in [10], and especially in [3]. The latter paper, among many other things, also contains explicit constructions of algebraic covers with monodromy given by the permutations as in condition (b), and of several other similar ones.

As we mentioned, we shall construct in our proofs some branched covers of the sphere, described algebraically by morphisms $\mathbb{P}_1(\mathbb{C}) \to \mathbb{P}_1(\mathbb{C})$, ramified only above $\{0, 1, \infty\}$. These covers, which are connected with the celebrated Grothendieck theory of *dessins d'enfants* [9], can be described by rational functions of the form $P(t)/Q(t)$, for polynomials $P(t)$ and $Q(t)$ such that the product $P(t)Q(t)(P(t) - Q(t))$ has "a few" distinct zeros. We take the opportunity to explore this connection, which will lead us to mention the celebrated abc-theorem for polynomials.

Our version of the proof of the classical Fermat-Euler theorem follows the mentioned idea from the old work by Ritt [8], and it builds on Theorem 1.1 and on the different viewpoints on branched covers developed to show it. Of course, such a proof has a small interest in itself, since it relies on deep results, whereas many elegant and simple proofs are known. However, the connection seems a striking one to us, and it raises the question whether a direct proof exists in purely combinatorial terms related to permutations as in condition (b).

This paper is organized as follows: in Section 2 we spell out the correspondence between condition (b) and branched covers, introducing the reader to the more general context of the Hurwitz existence problem. Then we exploit the Riemann Existence Theorem to interpret and reprove Theorem 1.1 in terms of the endomorphisms of an elliptic curve. In Section 3 we show how the algebraic and geometric machinery thus developed can be employed to prove the Fermat-Euler theorem. We conclude the paper by describing an alternative approach to the geometric version of Theorem 1.1, based on the notion of universal ramified cover with signature. We note that universal ramified covers with signature appear implicitly also in [6], where they are presented in the language of 2-orbifolds and defined making use also of the metric. Once again, we present the link between different viewpoints about the same objects.

2 Certain branched covers of the Riemann sphere

As announced in the Introduction, we begin by providing a geometric translation of condition (b) in Theorem 1.1. To do so we will need some definitions and notation. Let $\pi : \widetilde{\Sigma} \to \Sigma$ be a branched topological cover between real, closed, and connected surfaces. We say that π has *branching type* $(a_1, \ldots, a_r)$ over a point $P \in \Sigma$, where $a_1, \ldots, a_r$ are positive integers, if $\pi^{-1}(P)$ consists of r distinct points $Q_1, \ldots, Q_r$ such that locally at Q_i the map π may be represented as $z \mapsto z^{a_i}$, on viewing Σ and $\widetilde{\Sigma}$ as locally homeomorphic to the complex plane, with P and Q_i corresponding to 0. Note that π has a well-defined total degree d and that the branching type of any point of Σ is a partition of d. The *branching data* of π is given by the collection of all these partitions.

By considering the monodromy of a branched cover (a representation of the fundamental group of the complement of the branching points into the symmetric group $\mathfrak{S}_d$ on the d letters $\{1, \dots, d\}$, already employed in [4], see also [10, 12]), one gets the following:

Proposition 2.1. *Condition* (b) *in Theorem* 1.1 *is equivalent to:*

(b′) *There exists a branched cover* $\widetilde{\Sigma} \to \mathbb{P}_1(\mathbb{C})$ *of degree* d*, ramified over three points, with branching data* $(1, 4, \dots, 4)$*,* $(1, 4, \dots, 4)$*,* $(1, 2, \dots, 2)$.

Proof. Suppose a cover $\pi : \widetilde{\Sigma} \to \mathbb{P}_1(\mathbb{C})$ as in (b′) exists. Without loss of generality we can assume that the branching points in $\mathbb{P}_1(\mathbb{C})$ are $0, 1, \infty$. Setting $X = \mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$ and $\widetilde{X} = \pi^{-1}(X)$ we see that π induces a genuine degree-d cover $\widetilde{X} \to X$. Moreover, choosing loops around the points removed from $\mathbb{P}_1(\mathbb{C})$ we see that $\pi_1(X)$ has a natural presentation as $\langle \gamma_0, \gamma_1, \gamma_\infty |\ \gamma_0\gamma_1 = \gamma_\infty \rangle$. Upon fixing a basepoint P_0 in X and identifying $\pi^{-1}(P_0)$ with $\{1, \dots, d\}$ we then get the monodromy representation $\rho : \pi_1(X) \to \mathfrak{S}_d$, obtained by lifting to $\widetilde{X}$ the loops in X. We now set $\sigma_z = \rho(\gamma_z)$ for $z = 0, 1, \infty$, and we prove that $\sigma_0, \sigma_1, \sigma_\infty$ satisfy the points (i)–(iii) of condition (b) in Theorem 1.1. Point (i) follows from the fact that ρ is a representation. Moreover σ_z has one cycle for each point in $\pi^{-1}(z)$, and the length of the cycle is the local degree of π at that point, whence (ii). Finally point (iii) follows from the fact that $\widetilde{X}$ is connected.

We have shown that condition (b′) implies condition (b) in Theorem 1.1, and we now prove that (b) implies (b′) by reversing the construction. Given $\sigma_0, \sigma_1, \sigma_\infty \in \mathfrak{S}_d$ as in (b), by point (i), if we set $\rho(\gamma_z) = \sigma_z$ for $z = 0, 1, \infty$, we get a representation $\pi_1(X) \to \mathfrak{S}_d$, whence a cover $\widetilde{X} \to X$, that is connected by (iii). We then get the desired cover $\widetilde{\Sigma} \to \mathbb{P}_1(\mathbb{C})$ by adding to $\widetilde{X}$ one point for each cycle of σ_z for all $z = 0, 1, \infty$. □

Remark 2.2. For a branched cover $\widetilde{\Sigma} \to \mathbb{P}_1(\mathbb{C})$ of degree $d = 4k + 1$ and ramified over three points, the branching data $(1, 4, \dots, 4)$, $(1, 4, \dots, 4)$, and $(1, 2, \dots, 2)$ force $\widetilde{\Sigma}$ to be $\mathbb{P}_1(\mathbb{C})$ too. In fact the partitions have lengths $k+1$, $k+1$, and $2k+1$, and the Riemann-Hurwitz formula shows that if the genus of $\widetilde{\Sigma}$ is g then $2(1 - g) - (k + 1) - (k + 1) - (2k + 1) = (4k + 1) \cdot (2 - 3)$, which implies that $g = 0$.

By this remark, from now on we only deal with the case $\widetilde{\Sigma} = \mathbb{P}_1(\mathbb{C})$.

Remark 2.3. Up to an automorphism of the target $\mathbb{P}_1(\mathbb{C})$ one can suppose without loss of generality that, if a branched cover $\mathbb{P}_1(\mathbb{C}) \to \mathbb{P}_1(\mathbb{C})$ has three branching points, these points are 0, 1, and ∞. For a cover as in Theorem 2.4 we will always assume that the branching types are $(1, 4, \dots, 4)$ over 0 and 1, and $(1, 2, \dots, 2)$ over ∞.

Taking into account Proposition 2.1, the following result established as Theorem 0.4 in [6] is equivalent to Theorem 1.1:

Theorem 2.4. *Suppose* $d = 4k + 1$ *for some* $k \in \mathbb{N}$*. The following conditions are equivalent to each other:*

(b′) *There exists a branched cover* $\widetilde{\Sigma} \to \mathbb{P}_1(\mathbb{C})$ *of degree* d*, ramified over three points, with branching data* $(1, 4, \dots, 4)$*,* $(1, 4, \dots, 4)$*,* $(1, 2, \dots, 2)$.

(a) $d = x^2 + y^2$ *for some* $x, y \in \mathbb{N}$.

The proof given in [6] of this result employs the geometry of 2-obifolds. In particular, it exploits the fact that $S^2(4, 4, 2)$, namely the sphere with three cone points of orders 4, 4, and 2, bears a Euclidean geometric structure that is rigid up to rescaling. We also mention that [6] contains several other results giving partial solutions to the so-called Hurwitz existence problem, which asks whether a set of branching data satisfying the necessary condition given by the Riemann-Hurwitz formula is actually realized by an existent topological branched cover. We address the reader to [7] for an account of the history of this old and still not completely solved problem.

We will now illustrate a proof of Theorem 2.4, and hence of Theorem 1.1, in terms of branched covers of complex algebraic curves. To begin we spell out the following consequence of the Riemann Existence Theorem already anticipated in the Introduction:

Proposition 2.5. *If a topological cover of the sphere onto itself matching certain branching types exists, it can also be realized as a cover of algebraic curves $\pi : \mathbb{P}_1 \to \mathbb{P}_1$, defined over $\mathbb{C}$ or even over the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$.*

Note that a map π as in this proposition will be a rational function with complex coefficients of a complex variable t, and the coefficients may actually be assumed, by specialization or operating with an automorphism of the domain, to lie in $\overline{\mathbb{Q}}$. This last fact implies that the absolute Galois group of $\overline{\mathbb{Q}}$ acts on the set of such covers, leading to Grothendieck's theory of *dessins d'enfants*, for which the interested reader is referred to [9].

Let us now concentrate on branching types as in Theorem 2.4. To discuss the existence of a corresponding map π as in Proposition 2.5 we further normalize the situation noting that, up to composition with an automorphism of the domain $\mathbb{P}_1$, we can assume without loss of generality that 0 (respectively, 1, and ∞) is the unique unramified point above 0 (respectively, 1, and ∞). The branching conditions then imply that the map π, if any, can be expressed as

$$\pi(t) = \frac{t\,P^4(t)}{R^2(t)} = 1 + \frac{(t-1)\,Q^4(t)}{R^2(t)}$$

with $\deg P(t) = \deg Q(t) = k$, and $\deg R(t) = 2k$. More precisely, a cover as in Theorem 2.4 exists if and only if there exist polynomials $P(t), Q(t), R(t) \in \overline{\mathbb{Q}}[t]$ without multiple roots such that $t\,P(t)$, $(t-1)\,Q(t)$, and $R(t)$ are pairwise coprime, $\deg P(t) = \deg Q(t) = k$, $\deg R(t) = 2k$, and

$$t\,P^4(t) = (t-1)\,Q^4(t) + R^2(t). \tag{1}$$

Remark 2.6. It is worth noticing that when three such polynomials exist, they provide an *extremal example* of the so-called abc-theorem for polynomials, due to Stothers and Mason. We recall the statement of this result, given for instance in [5, Theorem 7.1]: *Let $a(t)$, $b(t)$, $c(t)$ be relatively coprime polynomials with complex coefficients satisfying $a(t) + b(t) = c(t)$. Then the number of distinct zeros of $a(t)b(t)c(t)$ is at least*

$$1 + \max\{\deg a(t),\ \deg b(t),\ \deg c(t)\}.$$

In our case, setting

$$a(t) = (t-1)\,Q^4(t), \quad b(t) = R^2(t), \quad c(t) = t\,P^4(t)$$

we get that the product $a(t)b(t)c(t)$ has at most $1 + k + 2k + 1 + k = 4k + 2$ distinct zeros, while the maximum of the degrees is $4k + 1$, so the above bound is attained. The abc-theorem for polynomials, that admits generalizations to function fields of curves of arbitrary genus, is nowadays a well-known tool for studying diophantine equations over function fields. See in this respect [13] and [14]; other extremal examples for the abc-theorem, still coming from geometric constructions, are treated in [13]. Its analogue for number fields is the celebrated abc-conjecture of Masser and Oesterlé, which is discussed, e.g., in [5, Chapter 7]. It is still open, and it is one of the deepest problems in diophantine analysis.

Before starting the proof that (b′) $\Rightarrow$ (a) in Theorem 2.4, we recall some basic facts about elliptic curves (or complex tori). Every smooth cubic curve has genus one, and conversely every smooth genus-one curve is isomorphic to a smooth cubic. In turn, such curves have an algebraic group structure, unique up to the choice of the neutral element, and conversely every irreducible projective algebraic group of dimension one is a genus-one curve, whence isomorphic to a smooth plane cubic. Such one-dimensional algebraic groups are called elliptic curves. From the complex analytic viewpoint, elliptic curves are complex tori, i.e., quotients $\mathbb{C}/\Lambda$, for a lattice $\Lambda \subset \mathbb{C}$. One particular elliptic curve will be relevant for our purposes: it arises as the quotient $E := \mathbb{C}/\mathbb{Z}[i]$ by the lattice of Gaussian integers, where $i = \sqrt{-1}$, and it is characterized by the fact of having an automorphism of order 4 fixing the origin (the neutral element). It can also be defined algebraically as the projective completion of the affine curve of equation $y^2 = x^3 - x$, where as the origin one can take the point at infinity; the automorphism $(x, y) \mapsto (-x, iy)$ sends the curve to itself, has order 4 and fixes the point at infinity. Alternatively, E can be defined as the smooth completion of the affine curve $v^2 = u^4 - 1$. The ring of endomorphisms of E as an elliptic curve (i.e., the group of the endomorphisms fixing the origin) is isomorphic to the ring $\mathbb{Z}[i]$; and actually, each such endomorphism is induced by the multiplication by a complex number λ; since the endomorphism sends $\mathbb{Z}[i]$ to itself, one necessarily has that $\lambda \in \mathbb{Z}[i]$. The degree of such an endomorphism is then easily seen to be the norm $\lambda \cdot \overline{\lambda}$ of the algebraic integer λ, that is a sum of squares.

Given an elliptic curve E defined over a field k, the set $E(k)$ of its rational points has a group structure. When $k = \mathbb{Q}$ is the field of rational numbers, $E(k)$ is a finitely generated group, by a famous theorem of Mordell. When k is a function field, an analogous result holds. Also, a *height function* can be defined in both cases; it is a function $H : E(k) \to \mathbb{R}$, satisfying for every $P \in E(k)$ and $n \in \mathbb{N}$ the identity $H(nP) = n^2 H(P)$, where nP denotes the sum $P + \ldots + P$, with P taken n-times, with respect to the mentioned group structure. This function is a semi-definite quadratic form, vanishing on the torsion points, and it is called the Néron-Tate height.

Proof of (b′) $\Rightarrow$ (a) *in Theorem* 2.4. Suppose the relevant branched cover exists, so there are polynomials $P(t)$, $Q(t)$, and $R(t)$ satisfying equation (1) and the other conditions. Dividing by $Q^4(t)$ in equation (1) we obtain the $\overline{\mathbb{Q}}(t)$-point

$$\left(\frac{P(t)}{Q(t)}, \frac{R(t)}{Q^2(t)}\right)$$

on the genus-1 curve over $\mathbb{Q}(t)$ with affine equation $y^2 = tx^4 - (t-1)$. Its points over $\mathbb{Q}(t)$ form a finitely generated group, and the involved degrees correspond to the values of a Néron-Tate height.

In this particular case the elliptic curve turns out to have constant j-invariant (equal to 1728), so the curve can in fact be defined over the constant field $\mathbb{Q}$, that allows one to analyze the situation in a much simpler way than in more general circumstances.

Indeed, consider the curve E which is the normalization of the closure in $\mathbb{P}_2$ of the affine curve (still denoted by the same symbol)

$$E: \quad v^2 = u^4 - 1.$$

Since it has genus 1, it becomes an elliptic curve E if we choose, e.g., the point $O := (0, i)$ as origin. As we said, E is isomorphic, as a complex torus, to the quotient $\mathbb{C}/\mathbb{Z}[i]$, namely it admits an automorphism of order 4 fixing the origin, given by $(u, v) \mapsto (iu, v)$. We shall prove the following result:

Proposition 2.7. *Let $P(t)$, $Q(t)$, and $R(t)$ be three polynomials satisfying* (1)*, with* $\deg P(t) = \deg Q(t) = k$, $\deg R(t) = 2k$*, such that $t \cdot (t-1) \cdot P(t) \cdot Q(t) \cdot R(t)$ has no multiple roots. Then, up to replacing the polynomial $R(t)$ by $-R(t)$, the map $(u, v) \mapsto (x, y)$, where*

$$x = u\frac{P(t)}{Q(t)}, \quad y = v\frac{R(t)}{Q^2(t)}, \quad t = \frac{u^4}{v^2}$$

induces an endomorphism of E (as an elliptic curve) of degree $d = 4k + 1$.

Proof. From $v^2 = u^4 - 1$ and $t = u^4/v^2$ it immediately follows that $u^4 = \frac{t}{t-1}$ and $v^2 = \frac{1}{t-1}$. Substituting in the expression for x and y one obtains $x^4 = \frac{t}{t-1}\frac{P^4(t)}{Q^4(t)}$ and $y^2 = \frac{1}{t-1}\frac{R^2(t)}{Q^4(t)}$, which shows that the equality $x^4 - 1 = y^2$ is equivalent to (1). This proves that the map $(u, v) \mapsto (x, y)$ indeed sends E to itself. Since x vanishes at O, the morphism sends O either to itself or to the point $(0, -i)$, in which case we replace $R(t)$ by $-R(t)$, and then get that the morphism fixes the origin, so it is also an endomorphism of E in the sense of elliptic curves. Its degree is easily seen to be $d = 4k + 1$. □

Now, condition (a) of the statement follows since, as explained above, the ring $\mathrm{End}(E)$ of the endomorphisms of E (as an elliptic curve) is isomorphic to $\mathbb{Z}[i]$, with the degree given by the square of the absolute value. More precisely, the endomorphism in Proposition 2.7 corresponds to the multiplication by a Gaussian integer $\lambda = a + ib$ and we have $d = a^2 + b^2$, concluding the proof. □

Proof of (a) $\Rightarrow$ (b′) *in Theorem* 2.4. Since for every Gaussian integer $a + ib$ there exists an endomorphism of E of degree $d = a^2 + b^2$, we must show that every endomorphism φ of E of odd degree d can be obtained as above for some polynomials $P(t)$, $Q(t)$, and $R(t)$. (Similar considerations are valid for even degrees, leading to slightly different analogue conclusions.) Let $\varphi \in \mathrm{End}(E)$ be an endomorphism of degree $d = 4k + 1$. We have an

expression of the form $\varphi(u,v) = (x(u,v), y(u,v))$ for suitable rational functions $x, y \in \mathbb{C}(E)$. Now, the degree-8 map $t : E \to \mathbb{P}_1$ given by $t = u^4/v^2$ is clearly invariant under the action of the subgroup G of the automorphisms of E (as algebraic curves) generated by

$$\alpha : (u,v) \mapsto (iu, v), \qquad \beta : (u,v) \mapsto (u, -v).$$

Note that α generates the isotropy group of O, whereas β may be also described as the map $p \mapsto \delta - p$ where $\delta := (0,-i)$; of course, α has order 4 and β has order 2; note that β is central in G, and $\alpha^2 \circ \beta : \ p \mapsto p + \delta$ is the (only) central translation in the automorphism group of E. Since G has order 8, t generates the field of invariants for G. On the other hand, it is easy to check that $t \circ \varphi = x^4/y^2$ is invariant under the action of G, therefore it is a rational function $Z(t)$ of t.

The function t has a divisor on E of the shape $4((O)+(\delta)) - 2((Q_1)+(Q_i)+(Q_{-1})+(Q_{-i}))$, where $Q_l = (l,0)$. Also, the divisor of $u - i^s$ (for $s = 1,\dots,4$) is $2(Q_{i^s}) - (\infty_+) - (\infty_-)$, that of $v + u^2$ is $2((\infty_+) - (\infty_-))$ for some labeling of the poles of u, v, and finally that of $v - i - u^2$ is $2(\delta) - 2(\infty_+)$. It easily follows that δ, $\infty_\pm$ have order 2 on E whereas the Q_l's have order 4.

With this information, considering zeros and multiplicities, we see that $Z(t) = x^4/y^2 = tP^4(t)/R^2(t)$ for suitable polynomials $P(t)$ and $R(t)$, where $\deg(tP^4(t)) > \deg R^2(t)$ – here we use the fact that d is odd, so φ fixes δ and sends the set of poles of t to itself. Similarly, we have $x^4/y^2 - 1 = 1/y^2$, that we can rewrite as $(t-1)Q^4(t)/R^2(t)$. Finally, the equation for E shows that $P(t)$, $Q(t)$, and $R(t)$ satisfy (1) and thus lead to a cover as in part (b′) of the statement. □

Remark 2.8. As a byproduct of our argument we have obtained a correspondence between permutations as in Proposition 2.1 and polynomials satisfying relation (1). Our proof actually also produces a relevant field of definition for the coefficients, as in [3].

We recall in passing that a Weierstrass model of the curve E employed above is obtained by the inverse transformations $\eta := \frac{u}{v-i}$, $\xi := \frac{u^2}{v-i} = u\eta$ and $u = \frac{\xi}{\eta}$, $v = i + \frac{\xi}{\eta^2}$, that lead to the equation $\eta^2 = \frac{1}{2i}(\xi^3 - \xi)$.

Galois structure. We conclude this paragraph with some extra considerations on the constructions we encountered so far. First of all we prove the following:

Proposition 2.9. *With the above notation (in particular $G = \langle \alpha, \beta \rangle$ is the group defined in the previous proof), the map $t \circ \varphi = Z(t) =: z$ defines a Galois cover $E \to \mathbb{P}_1$, whose Galois group Γ (of order $8d$) is*

$$\Gamma = \big\{ p \mapsto gp + \kappa : \ g \in G, \ \kappa \in \operatorname{Ker} \varphi \big\}. \tag{2}$$

Before starting the proof, let us recall some classical facts about endomorphisms of elliptic curves. We already said that the endomorphisms fixing the origin (called isogenies) correspond to scalar multiplications on $\mathbb{C}$: namely, if $E = \mathbb{C}/\Lambda$ is given as the quotient of the complex plane modulo a lattice, then every complex number λ with $\lambda \cdot \Lambda \subset \Lambda$ defines an isogeny induced by the map $z \mapsto \lambda z$. Clearly, every translation is also a morphism

$E \to E$ as an algebraic curve. Every endomorphism $\psi : E \to E$ is the product of an isogeny and a translation, namely is of the form $\psi : P \mapsto \varphi(P) + A$, where φ is an isogeny and $A \in E$ is fixed. The "linear" part φ of ψ can be obtained as the "differential" of ψ; more precisely, ψ acts on the one-dimensional space of holomorphic 1-forms (which are automatically invariant by translations) and its action only depends on φ. If φ is induced by the multiplication by λ on the complex plane, then ψ also acts on 1-forms by multiplication by λ.

We begin with a preliminary result:

Lemma 2.10. *Let $G = \langle \alpha, \beta \rangle$ be the group defined above. Let $\varphi : E \to E$ be an isogeny of odd degree. Then for every $g \in G$ one has $g \circ \varphi = \varphi \circ g$.*

Proof. Clearly φ commutes with α, since both are isogenies. To prove that φ commutes with β, we shall prove the equivalent fact that φ commutes with $\alpha^2 \circ \beta$, which is the translation by δ. To this end, it is useful to think in terms of the actions of φ and $\alpha^2 \circ \beta$ on the complex plane $\mathbb{C}$. The latter corresponds to the multiplication by $\frac{1+i}{2}$, while φ corresponds to the multiplication by a Gaussian integer $a + ib$ with $a^2 + b^2 \equiv 1 \pmod 2$. Now, we have to prove that the functions $\mathbb{C} \ni z \mapsto (a+ib)z + \frac{1+i}{2}$ and $\mathbb{C} \ni z \mapsto (a+ib)(z + \frac{1+i}{2})$ coincide modulo $\mathbb{Z}[i]$. This is equivalent to the fact that $(a + ib)\frac{1+i}{2} - \frac{1+i}{2} \in \mathbb{Z}[i]$, which easily follows from the hypothesis that $a^2 + b^2$ is odd. □

Proof of Proposition 2.9. The map $z = t \circ \varphi : E \to \mathbb{P}_1$ induces a field extension $\overline{\mathbb{Q}}(E)/\overline{\mathbb{Q}}(z)$ of degree $8d$. Let us prove that it is invariant under the action of the group Γ defined above. Let $p \in E$, $g \in G$, and $\kappa \in \operatorname{Ker}\varphi$. Clearly, $\varphi(gp + \kappa) = \varphi(gp)$; now, by Lemma 2.10, we have $\varphi(gp) = g(\varphi(p))$ and since t is invariant by G we have $z(p) = (t \circ \varphi)(p) = z(gp + \kappa)$ as wanted. It remains to prove that Γ has order $8d$; this is due to the fact that φ has odd degree, so $\operatorname{Ker}\varphi$ has odd order. Then the subgroup of translations in Γ has order $2 \deg \varphi$; more precisely Γ is also given as the extension

$$\{0\} \to \langle \delta \rangle \oplus \operatorname{Ker}\varphi \to \Gamma \to \{1, i, -1, -i\} \to \{1\},$$

where the map $\Gamma \to \{1, i, -1, -i\}$ denotes the action on the invariant differentials. From this representation, it is clear that its order is $8d$. Hence Γ is the Galois group of the cover $z = t \circ \varphi : E \to \mathbb{P}_1$. □

Proposition 2.9 implies in particular that the Galois closure of the equation in t over $\overline{\mathbb{Q}}(z)$ given by $Z(t) = z$ is contained in the above extension $\overline{\mathbb{Q}}(E)/\overline{\mathbb{Q}}(z)$. However, the Galois closure, whose Galois group is generated by the permutations $\sigma_0, \sigma_1, \sigma_\infty$ corresponding to our cover as in Proposition 2.1, is actually smaller: in fact, as already noticed, the element $\gamma := \alpha^2 \circ \beta$ acts on (u, v) as $\gamma(u, v) = (-u, -v)$, namely as a translation by δ, and hence fixes the field $\overline{\mathbb{Q}}(t)$. We have also already remarked that γ is in the center of Γ. Therefore the said Galois closure is contained in the fixed field of γ, and is in fact equal to it, because no subgroup of G larger then $\langle \gamma \rangle$ is normal in Γ. This fixed field of γ is easily seen to be $\overline{\mathbb{Q}}(u^2, v^2, uv)$. If we set $\sigma := v/u^3$ and $\tau := -1/u^2$ we find that σ and τ generate this field and $\sigma^2 = \tau^3 - \tau$. This is a Weierstrass equation for an elliptic curve E^* (again isomorphic to E) which is the quotient of E by the order-2 group of automorphisms generated by γ.

Remark 2.11. Equation (2) yields an explicit representation of the group generated by our three permutations $\sigma_0, \sigma_1, \sigma_\infty$, which is isomorphic to $\Gamma/\langle\gamma\rangle$, and has order $4d$.

Remark 2.12. Alternative proofs based on techniques similar to those employed here are possible also for Theorems 0.5 and 0.6 in [6].

3 Sums of two squares

In this section, as announced in the Introduction, we follow an idea of Ritt [8], that we present in a modern language, leading to a proof of the Fermat-Euler Theorem 1.2. According to the implication (b) $\Rightarrow$ (a) in Theorem 1.1, to show that a prime number d congruent to 1 modulo 4 is a sum of two squares, it is sufficient to show that there exist permutations $\sigma_0, \sigma_1, \sigma_\infty \in \mathfrak{S}_d$ satisfying condition (b) of the theorem. We also note that if such $\sigma_0, \sigma_1, \sigma_\infty$ exist then they necessarily generate a group of order $4d$. Our explicit construction of the permutations for prime d is essentially the same as Ritt's one:

Proposition 3.1. *Let p be a prime congruent to* 1 *modulo* 4. *Then the group $\mathbb{F}_p^*$ has an element ℓ of order* 4. *Consider the affine automorphisms L and T of the line $\mathbb{A}^1$ over $\mathbb{F}_p$ defined by $L(x) = \ell x$ and $T(x) = x + 1$, and the permutations*

$$\sigma_0 := L, \qquad \sigma_1 = T^{-1}LT, \qquad \sigma_\infty := LT^{-1}LT$$

of the set $\mathbb{F}_p = \mathbb{A}^1(\mathbb{F}_p)$. Then σ_0, σ_1, σ_∞ satisfy condition (b) *of Theorem* 1.1 *with $d = p$.*

Proof. Existence of $\ell \in \mathbb{F}_p^*$ of order 4 readily follows from the assumption $p \equiv 1 \pmod 4$. Let us proceed and prove that the permutations $\sigma_0, \sigma_1, \sigma_\infty$ defined in the statement satisfy the items (i)–(iii) of condition (b) in Theorem 1.1; (i) asserts that $\sigma_\infty = \sigma_0\sigma_1$, which is indeed true by definition.

The cycle type of σ_0 is clearly $(1, 4, \ldots, 4)$, because $\ell^2 = -1$, whence L and L^2 have 0 as a fixed point and act injectively on $\mathbb{F}_p^*$. Since σ_1 is conjugate to σ_0, it also has such a cycle type. Turning to σ_∞, and using again the fact that $\ell^2 = -1$, we see that σ_∞ takes the form $\sigma_\infty(x) = -x + c$, for a suitable $c \in \mathbb{F}_p$ (actually $c = -\ell - 1$). Therefore it is an (affine) involution of the line $\mathbb{A}^1$, and since $p \neq 2$ its cycle type is of the form $(1, 2, \ldots, 2)$, which completes the proof of condition (ii).

To establish (iii) we note that the commutator $[\sigma_0, \sigma_1]$ is a nontrivial translation, so it acts transitively on $\mathbb{F}_p$. □

Combining Theorem 1.1 and Proposition 3.1 one readily deduces the well-known Fermat-Euler theorem, stated in the Introduction as Theorem 1.2. As already mentioned, the resulting proof of this classical result is extraordinarily demanding: a closer look shows that, in addition to the construction of the permutations in Proposition 3.1, it depends also on:

(A) The topological construction of a finite cover of $\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$ such that lifting three simple disjoint loops based at a point P_0 and encircling $0, 1, \infty$ one obtains the given permutations $\sigma_0, \sigma_1, \sigma_\infty$ on the fiber over P_0. This construction appears, e.g., in [12, Theorems 4.27 and 4.31]; it may be proved by patching local covers or taking a suitable quotient of the universal cover of $\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$.

(B) The Riemann Existence Theorem, used to realize the said topological cover as the unramified part of a ramified cover of complex algebraic curves. This step is delicate and requires fairly hard analysis, based either on the Dirichlet principle or on the vanishing of suitable cohomology of holomorphic sheaves on Riemann surfaces. (See again [12, Theorem 4.27].)

(C) The structure of the endomorphism ring of the elliptic curve E of Section 2, namely its identification with $\mathbb{Z}[i]$. Recall that E may be identified with the torus $\mathbb{C}/\mathbb{Z}[i]$ and that it is characterized by the property of having a vanishing Weierstrass invariant g_3 (an easy direct computation which amounts to showing that $\sum_{\omega\in\mathbb{Z}[i]\setminus\{0\}} \omega^{-6}$ is 0 – see [11] for careful proofs of all of these facts).

In conclusion, all of these steps involve some nontrivial mathematics, and (B) is particularly delicate. Combining the (self-contained) proof of Proposition 3.1 with the arguments used in [6] to establish Theorem 2.4, one gets instead a proof of Theorem 1.2 based on item (A) above and on the existence and rigidity (up to scaling) of a Euclidean structure on the orbifold $S^2(4, 4, 2)$.

On the other hand there exist many few-lines self-contained proofs of the Fermat-Euler result. Nevertheless, one cannot say that the proof given above *contains*, from the logical viewpoint, any classical proof, as for instance the argument based on the unique factorization of $\mathbb{Z}[i]$. In fact, although this ring plays an implicit role in item (C) above, its factorization properties are not employed, neither explicitly nor implicitly.

Ramified covers with signature. An alternative approach to Theorem 2.4, to which the argument in [6] is closer and that does not require items (B) and (C) above, is described in a sketchy but complete fashion in [10, pp. 60–63]. This avoids the viewpoint of complex algebraic curves altogether, being based on the notion of *ramified cover with signature* which, roughly speaking, consists of a topological cover of a space deprived of finitely many points, together with a ramified structure above the remaining points, of the same type as a map of the shape $z \mapsto z^n$. It corresponds to what is called *geometric universal cover* in the language of 2-orbifolds used in [6].

In our case we have a *universal covering with signature* $(4, 4, 2)$, meaning a space Y that is obtained by suitably completing the quotient of the universal cover of $\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$ by the normal subgroup N of $\pi_1(\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\})$ generated by c_0^4, c_1^4, c_∞^2, where c_0, c_1, c_∞ are the simple disjoint loops already mentioned above. As stated in [10, p. 63], one realizes Y as the Euclidean plane $\mathbb{C}$, with covering group Γ given by the rigid motions of the plane preserving the orientation and the lattice $\mathbb{Z}[i]$; it is then easy to check that $\Gamma \cong \pi_1(\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\})/N$. This covering has the following universal property: *Let* $\Sigma \to \mathbb{P}_1(\mathbb{C})$ *be a ramified cover, of degree* d*, unramified over* $\mathbb{P}_1(\mathbb{C}) \setminus \{0, 1, \infty\}$ *and with ramifications of order dividing* 4 *over* 0 *and over* 1*, and dividing* 2 *over* ∞*. Then there exists a subgroup* $H < \Gamma$ *of index* d *such that* $\Sigma \simeq \mathbb{C}/H$*; in addition, the cover* $\mathbb{C} \to \mathbb{C}/\Gamma \simeq \mathbb{P}_1(\mathbb{C})$ *factors as* $\mathbb{C} \to \mathbb{C}/H = \Sigma \to \mathbb{C}/\Gamma \simeq \mathbb{P}_1(\mathbb{C})$.

An explicit description of the elements of Γ as affine transformations of the complex line is as follows:

$$\Gamma = \{\mathbb{C} \ni z \mapsto uz + \lambda \,:\, u \in \{1, i, -1, -i\},\ \lambda \in \mathbb{Z}[i]\}. \tag{3}$$

There are three orbits of points in $\mathbb{C}$ having non trivial stabilizers: the first one is $\mathbb{Z}[i]$, where each point has a stabilizer of order 4; the second one is $\frac{1+i}{2} + \mathbb{Z}[i]$, also having a stabilizer of order 4; the third one is $\left(\frac{1}{2} + \mathbb{Z}[i]\right) \cup \left(\frac{i}{2} + \mathbb{Z}[i]\right)$, having a stabilizer of order 2. They correspond to the pre-images of $0, 1, \infty$.

Theorem 3.2. *The group* Γ *defined by equation* (3) *is the universal group of type* (4, 4, 2).

Proof. We confine ourselves to a sketch. Let G be a group of type $(4, 4, 2)$, so G is generated by two elements α, β with $\alpha^4 = \beta^4 = (\alpha\beta)^2 = 1$. It is immediate from this presentation that the congruence class modulo 4 of the length of a word representing an element of G only depends on that element. Also, it is easily checked that the words of length divisible by 4 commute and can be generated by $u := \alpha^3\beta$ and $v := \alpha\beta^3$. Hence we always have a group homomorphism $G \to \mathbb{Z}/4\mathbb{Z}$ whose kernel is an Abelian (normal) subgroup generated by two elements. So we have an exact sequence $\{0\} \to \langle u, v\rangle \to G \to \mathbb{Z}/4\mathbb{Z}$, where $\langle u, v\rangle \cong (\mathbb{Z}/a\mathbb{Z}) \times (\mathbb{Z}/b\mathbb{Z})$ for integers a, b (possibly zero or one) and the last morphism (which need not be surjective) is the reduction modulo 4 of the word length. We note that the action (by conjugation) of G on $\langle u, v\rangle$ is uniquely determined by the initial relations $\alpha^4 = \beta^4 = (\alpha\beta)^2 = 1$. Coming back to our group Γ, let us consider the three elements c_0, c_1, c_∞ of Γ defined by $c_0(z) = iz$, $c_1(z) = 1 + iz$, $c_\infty(z) = -z + i$. Then $c_0^4 = c_1^4 = c_\infty^2 = 1$ and $c_0c_1 = c_\infty$. Clearly, $\tilde{u} := c_0^3c_1$ and $\tilde{v} := c_0c_1^3$ act as $\tilde{u}(z) = z + i$ and $\tilde{v}(z) = z - 1$, so they generate the subgroup of translations, isomorphic to $\mathbb{Z}[i] \cong \mathbb{Z}^2$. Hence we have the exact sequence

$$\{0\} \to \mathbb{Z}^2 \to \Gamma \to \mathbb{Z}/4\mathbb{Z} \to \{0\}.$$

Note that the action of Γ on $\mathbb{Z}[i]$ is compatible with the action of G on $\langle u, v\rangle$, in the natural sense: for instance we have in G the relation $\alpha u \alpha^{-1} = v^{-1}$, which corresponds in Γ to the relation $c_0\tilde{u}c_0^{-1} = \tilde{v}^{-1}$. From this fact it easily follows that G is a quotient of Γ. □

Let us get back to the setting of Theorem 2.4, and let us use the universal covering with signature $(4, 4, 2)$, denoted by Y as above. A cover X of the Riemann sphere with the relevant branching types exists if and only if it can be realized as the quotient of Y by a subgroup Δ of Γ, of *odd* index d in Γ. The permutations σ_0, σ_1, σ_∞ then correspond to the images of c_0, c_1, c_∞ in the permutation representation of Γ on the right cosets Γ/Δ. One easily sees that if Δ exists then it must contain an element σ of order 4, which must be a rotation of $\pi/2$ around some point. The orbit of the origin by Δ is a lattice stable under σ, which corresponds to an ideal in $\mathbb{Z}[i]$. This ideal is principal, and we find again the conclusion that d is a sum of two squares. As a matter of fact, to conclude one can also avoid invoking the principal-ideal ring structure, by observing that the said lattice must have a basis of type v, σv, whence its index is necessarily a sum of two squares.

Remark 3.3. When the degree is an odd prime p, this approach also allows one to elucidate the structure of the group G generated by the permutations σ_0, σ_1, σ_∞. In fact, as stated in the proof of Theorem 3.2, the group generated by the words of length 4 in σ_0, σ_1 is commutative. Hence G has an Abelian subgroup G_0 of index at most 4. Since G is a transitive subgroup of $\mathfrak{S}_p$ it contains a p-cycle g, which must lie in G_0. Then G_0 must be the group generated by g.

References

[1] Edmonds, A.L.; Kulkarni, R.S.; Stong, R.E.: Realizability of branched coverings of surfaces. *Trans. Amer. Math. Soc.* 282 (1984), 773–790.

[2] Fulton, W.: *Algebraic Topology. A first course*. Grad. Texts in Math. 153, 1995.

[3] Guralnick, R.M.; Müller, P.; Saxl, J.: The rational function analogue of a question of Schur and exceptionality of permutation representations. *Mem. Amer. Math. Soc.* 162 (2003), viii+79.

[4] Hurwitz, A.: Riemann'sche Flächen mit gegebenen Verzweigungspunkten. *Math. Ann.* 39 (1891), 1–61.

[5] Lang, S.: *Algebra*. Revised Third Edition, Grad. Texts in Math. 211, 2002.

[6] Pascali, M.A.; Petronio, C.: Surface branched covers and geometric 2-orbifolds. *Trans. Amer. Math. Soc.* 361 (2009), 5885–5920.

[7] Petronio, C.; Pervova, E.: Realizability and exceptionality of candidate surface branched covers: methods and results. Seminari di Geometria 2005–2009, Università degli Studi di Bologna, Dipartimento di Matematica, 2010, 205–210.

[8] Ritt, J.F.: New proofs of two well known theorems on quadratic forms. *Ann. of Math.* 26 (1925), 202–204.

[9] Schneps, L. (ed.): *The Grothendieck Theory of Dessins d'Enfants*. London Math. Soc., Lecture Note Series, Vol. 200, Cambridge Univ. Press, Cambridge 1994.

[10] Serre, J.-P.: *Topics in Galois Theory*. Research Notes in Mathematics, Vol. 1, Jones and Bartlett, Boston MA 1992.

[11] Silverman, J.: *The Arithmetic of Elliptic Curves*. Grad. Texts in Math. 106, 1985.

[12] Völklein, H.: *Groups as Galois Groups, an Introduction*. Cambridge Studies in Advanced Mathematics, Vol. 53, Cambridge Univ. Press, Cambridge 1996.

[13] Zannier, U.: A note on the S-unit equation over function fields. *Acta Arith.* 64 (1993), 87–98.

[14] Zannier, U.: On Davenport's lower bound and Riemann Existence Theorem. *Acta Arith.* 71 (1995), 107–137.

Pietro Corvaja
Dipartimento di Matematica e Informatica
Università di Udine
Via delle Scienze, 206
I-33100 Udine, Italy
e-mail: pietro.corvaja@dimi.uniud.it

Carlo Petronio
Dipartimento di Matematica Applicata
Università di Pisa
Via Filippo Buonarroti, 1C
I-56127 Pisa, Italy
e-mail: petronio@dm.unipi.it

Umberto Zannier
Scuola Normale Superiore
Piazza dei Cavalieri, 7
I-56126 Pisa, Italy
e-mail: u.zannier@sns.it