Yet another Proof of an old Hat

Roland Bacher

Yet another Proof of Fermat's Two Squares Theorem for Prime Numbers

Roland Bacher

February 8, 2022

Abstract

Every odd prime number p can be written in exactly $(p+1)/2$ ways as a sum $ab+cd$ of two ordered products ab and cd such that $\min(a,b) > \max(c,d)$. This gives an elegant proof of Fermat's Theorem expressing primes in $1+4\mathbb{N}$ as sums of two squares [1].

Theorem 0.1. *For every odd prime number p there exist exactly $(p+1)/2$ sequences (a,b,c,d) of four elements in the set $\mathbb{N} = \{0,1,2,\dots\}$ of non-negative integers such that $p = ab+cd$ and $\min(a,b) > \max(c,d)$.*

As a consequence we obtain a new proof of an old result discovered by Fermat. (The old rascal did not want to spoil his margins and left the proof to Euler who had no such qualms.)

Corollary 0.2. *Every prime number in $1+4\mathbb{N}$ is a sum of two squares.*

Proof of Corollary 0.2. If p is a prime number congruent to 1 (mod 4), the number $(p+1)/2$ of solutions (a,b,c,d) defined by Theorem 0.1 is odd. The involution $(a,b,c,d) \longmapsto (b,a,d,c)$ has thus a fixed point (a,a,c,c) expressing p as a sum of two squares. □

Nowadays a venerable old hat, Corollary 0.2 has of course already quite a few proofs. Some are described in the entry "Fermat's theorem on sums of two squares" of [7]. The author enjoyed the presentation of a few 'elementary' proofs given in [3]. Several of these proofs are based on the existence of a fixed point:

Zagier (based on previous work of Liouville and Heath-Brown) published a one sentence proof in [8].

Elsholtz, see Section 1.6 of [3], combined two approaches based on Euclidean lattices by Grace and Lucas resulting in a proof somewhat reminiscent of the present proof.

[1] Keywords: Primes, sum of two squares, lattice. Math. class: Primary: 11A41. Secondary: 11H06.

Christopher, see [2], gave a proof based on the existence of a fixed point of an involution acting on suitable partitions with parts of exactly two different sizes (amounting essentially to solutions of $p = ab+cd$ without requirements of inequalities).

The set $\mathcal{S}_p$ of solutions defined by Theorem 0.1 is invariant under the action of Klein's Vierergruppe with non-trivial elements acting by

$$(a, b, c, d) \longmapsto (b, a, c, d), (a, b, d, c), (b, a, d, c)$$

(i.e. by exchanging either the first two elements, or the last two elements, or the first two and the last two elements). The following tables list all elements (a, b, c, d) with a, b, c, d decreasing together with the size $\sharp(\mathcal{O})$ of the corresponding orbit under Klein's Vierergruppe for the sets $\mathcal{S}_{29}, \mathcal{S}_{31}, \mathcal{S}_{37}$:

a	b	c	d	$\sharp(\mathcal{O})$
29	1	0	0	2
14	2	1	1	2
7	4	1	1	2
9	3	2	1	4
5	5	4	1	2
5	5	2	2	1
5	4	3	3	2
				15

a	b	c	d	$\sharp(\mathcal{O})$
31	1	0	0	2
15	2	1	1	2
10	3	1	1	2
6	5	1	1	2
7	4	3	1	4
9	3	2	2	2
5	5	3	2	2
				16

a	b	c	d	$\sharp(\mathcal{O})$
37	1	0	0	2
18	2	1	1	2
12	3	1	1	2
9	4	1	1	2
6	6	1	1	1
7	5	2	1	4
11	3	2	2	2
7	4	3	3	2
5	5	4	3	2
				19

Establishing complete lists $\mathcal{S}_p$ of solutions for small primes is a rather pleasant pastime and rates among the author's more confessable procrastinations.

A rough sketch for proving Theorem 0.1 goes along the following lines: Every solution $p = ab+cd$ in $\mathcal{S}_p$ can be encoded by two vectors $u = (a, c), v = (-d, b)$ generating a sublattice $\mathbb{Z}u + \mathbb{Z}v$ of index p in $\mathbb{Z}^2$. It is thus enough to understand the number of solutions encoded by every sublattice of index p in $\mathbb{Z}^2$. Such sublattices are in one-to-one correspondence with all $p + 1$ elements of the the projective line $\mathbb{F}_p \cup \{\infty\}$ over the finite field $\mathbb{F}_p$. An element μ encoding the slope $\mu = \frac{b}{a}$ of $[a : b]$ (using the obvious convention for $\mu = \infty$) defines the sublattice $\Lambda_\mu(p) = \{(x, y) \in \mathbb{Z}^2 \mid ax + by \equiv 0 \pmod p\}$ of index p in $\mathbb{Z}^2$. The two lattices $\Lambda_0(p) = \mathbb{Z}(p, 0) + \mathbb{Z}(0, 1)$ and $\Lambda_\infty(p) = \mathbb{Z}(1, 0) + \mathbb{Z}(0, p)$ with singular slopes $0, \infty \notin \mathbb{F}_p^*$ give rise to the two exceptional solutions $p \cdot 1 + 0 \cdot 0$ and $1 \cdot p + 0 \cdot 0$. The two lattices $\Lambda_1(p)$ and $\Lambda_{-1}(p)$ with self-inverse slopes 1 and -1 do not correspond to a solution. Every pair of distinct lattices $\Lambda_\mu(p), \Lambda_{\mu^{-1}}(p)$ with mutually inverse slopes $\mu, \mu^{-1} \in \mathbb{F}_p^* \setminus \{1, -1\}$ gives rise to exactly one solution in $\mathcal{S}_p$. Theorem 0.1 follows now easily.

Nailing down all the pesky details is tedious but elementary and is the content of Section 1.

A very sketchy last Section discusses algorithmic and statistical aspects and ends with Theorem 2.3 giving a formula for the number of solutions of $n = ab - cd$ with (a, b, c, d) in $\mathbb{N}^4$ such that $\min(a, b) > \max(c, d)$.

1 Proof of Theorem 0.1

Henceforth, a *lattice* denotes always a discrete subgroup of a Euclidean vector space. We will mainly work with sublattices of the lattice $\mathbb{Z}^2$ of all integral points of the Euclidean plane $\mathbb{R}^2$ (endowed with its standard orthogonal basis). A *basis* of a lattice Λ of rank (or dimension) 2 is a set u, v of two elements in Λ such that $\Lambda = \mathbb{Z}u + \mathbb{Z}v$.

The following result is a special case of Pick's Theorem[2]:

Lemma 1.1. *Two linearly independent elements u, v of a 2-dimensional lattice Λ form a basis of the lattice Λ if and only if the triangle with vertices $(0, 0), u, v$ contains no other elements of Λ.*

Proof. The parallelogram $\mathcal{P}$ with vertices $(0, 0), u, v, u + v$ is a fundamental domain of the sublattice $\mathbb{Z}u+\mathbb{Z}v$ of Λ spanned by u and v. The two elements u, v generate thus Λ if and only if only vertices of $\mathcal{P}$ belong to Λ.

Since $\mathcal{P}$ and Λ are invariant under the affine involution $x \longmapsto -x+u+v$ exchanging the two triangles with vertices $(0, 0), u, v$ and $u, v, u + v$, the parallelogram $\mathcal{P}$ intersects Λ exactly in its vertices if and only if the triangle defined by $(0, 0), u, v$ intersects Λ exactly in its vertices. □

Two bases f_1, f_2 and g_1, g_2 of $\mathbb{R}^2$ are *interlaced* if $\mathbb{R}f_1 \cup \mathbb{R}f_2 \cup \mathbb{R}g_1 \cup \mathbb{R}g_2$ intersects the unit circle $\mathbb{S}^1$ in eight distinct points such that points of $(\mathbb{R}f_1 \cup \mathbb{R}f_2) \cap \mathbb{S}^1$ and points of $(\mathbb{R}g_1 \cup \mathbb{R}g_2) \cap \mathbb{S}^1$ alternate on $\mathbb{S}^1$.

Lemma 1.2. *Two bases f_1, f_2 and g_1, g_2 of a 2-dimensional lattice $\Lambda = \mathbb{Z}f_1 + \mathbb{Z}f_2 = \mathbb{Z}g_1 + \mathbb{Z}g_2$ are never interlaced.*

Proof. Up to sign-changes and up to exchanging the roles of f_1 and f_2 we have otherwise $f_1 = \alpha g_1 + \beta g_2$ and $f_2 = \gamma g_1 - \delta g_2$ where $\alpha, \beta, \gamma, \delta$ are strictly positive integers. This implies that g_1 belongs to the segment joining $\frac{1}{\alpha}f_1$ to $\frac{1}{\gamma}f_2$ contained in the convex hull of $(0, 0), f_1, f_2$. Lemma 1.1 implies thus $g_1 = f_1$ or $g_1 = f_2$ in contradiction with strict positivity of $\alpha, \beta, \gamma, \delta$. □

Remark 1.3. *Lemma 1.2 follows also from the inequality* $\det \begin{pmatrix} \alpha & \beta \\ \gamma & -\delta \end{pmatrix} = -\alpha\delta - \beta\gamma < -1$.

[2]Pick's theorem gives the area $\frac{1}{2}b + i - 1$ of a closed lattice polygon P (with vertices in $\mathbb{Z}^2$) containing b lattice points $\partial P \cap \mathbb{Z}^2$ in its boundary and i lattice points in its interior.

We consider the eight open cones of $\mathbb{R}^2$ forming the complement of the four lines defined by $xy(x^2 - y^2) = 0$. We call these eight open cones *windmill-cones* (in the hope of turning the content of this paper into a piece of loftier mathematics) and we colour them alternately black and white, starting with a black E-NE windmill-cone $\{(x, y) \mid 0 < y < x\}$ (using the conventions of wind-roses).

A basis e, f of $\mathbb{R}^2$ is a *black windmill basis* if e and f are contained in the open upper half-plane, one element in $\{e, f\}$ belongs to the open black E-NE windmill cone and the other element in $\{e, f\}$ belongs to the open black N-NW windmill cone. A *white windmill* basis is defined similarly by exchanging the role of the two black E-NE and N-NW windmill cones with the two white N-NE and W-NW windmill cones of the upper half-plane.

A 2-dimensional lattice Λ has a black (respectively white) windmill basis if $\Lambda = \mathbb{Z}e + \mathbb{Z}f$ is generated by a black (respectively white) windmill basis e, f.

Lemma 1.4. *All windmill bases of a lattice have the same colour.*

Proof. Otherwise we get a contradiction with Lemma 1.2 since windmill bases of different colours are obviously interlaced. □

An odd prime-number p and an element μ of $\mathbb{F}_p$ (henceforth identified with $\{0, \ldots, p-1\}$) define a sub-lattice

$$\Lambda_\mu(p) = \{(x, y) \in \mathbb{Z}, \mid x + \mu y \equiv 0 \pmod{p}\} \tag{1}$$

of index p in $\mathbb{Z}^2$. We set $\Lambda_\infty(p) = \{(x, y) \in \mathbb{Z}, \mid y \equiv 0 \pmod{p}\}$. All $p+1$ lattices $\Lambda_0(p), \ldots, \Lambda_{p-1}(p), \Lambda_\infty(p)$ are distinct and $\mathbb{Z}^2$ contains no other sublattices of prime index p.

Proposition 1.5. *The four lattices $\Lambda_0(p), \Lambda_\infty(p), \Lambda_1(p), \Lambda_{p-1}(p)$ have no windmill basis.*

Proof. Each of these four lattices is invariant under an orthogonal reflection σ_L with respect to a suitable line L defined by $xy(x^2 - y^2) = 0$. Up to sign changes, such an orthogonal reflection σ_L exchanges white and black windmill bases. Lemma 1.4 shows now non-existence of (black or white) windmill bases for such lattices. □

The next result shows optimality of Proposition 1.5:

Proposition 1.6. *Every lattice $\Lambda_\mu(p)$ with $2 \leq \mu \leq p-2$ has a windmill basis.*

Proof. $\Lambda_\mu(p)$ contains obviously no elements of the form $(\pm m, 0)$ or $(\pm m, \pm p)$ with m in $\{1, 2, \ldots, p-1\}$. Since p is prime, $\Lambda_\mu(p)$ contains no elements of the form $(0, \pm m), (\pm p, \pm m)$ with m in $\{1, \ldots, p-1\}$. Moreover, for μ in

$\{2, \ldots, p-2\}$ considered as a subset of the finite field $\mathbb{Z}/p\mathbb{Z}$, the elements $1+\mu$ and $1-\mu$ are invertible in $\mathbb{Z}/p\mathbb{Z}$. This implies that $\Lambda_\mu(p)$ contains also no elements of the form $\pm(m,m), \pm(m,-m)$ for m in $\{1, \ldots, p-1\}$. The intersection of a (black or white) windmill-cone with $[-p,p]^2$ defines thus a triangle of area $p^2/2 > p/2$ whose boundary contains no lattice-points of $\Lambda_\mu(p)$ except for its three vertices. Lemma 1.1 implies now that every open (black or white) windmill-cone contains a non-zero element (x,y) of $\Lambda_\mu(p)$ with coordinates x, y in $\{\pm 1, \pm 2, \ldots, \pm(p-1)\}$.

There exists thus a parallelogram $\mathcal{P}$ of minimal area with vertices $\pm e, \pm f$ in $\Lambda_\mu(p) \cap \{-p+1, \ldots, p-1\}^2$ such that $\{\pm e, \pm f\}$ intersects either all four open black windmill-cones or all four open white windmill-cones.

Suppose for simplicity that all elements of $\{\pm e, \pm f\}$ are black (i.e. belong to open black windmill cones). (The case where $\pm e$ and $\pm f$ are all white is completely analogous.)

Since $\Lambda_\mu(p)$ intersects the diagonal $\mathbb{R}(1,1)$ and the antidiagonal $\mathbb{R}(1,-1)$ in $\mathbb{Z}(p,p)$ and in $\mathbb{Z}(p,-p)$, and since $\Lambda_\mu(p)$ contains obviously no elements of the form $(\pm a, 0), (0, \pm a)$ with a in $\{1, \ldots, p-1\}$, all non-zero elements of $\mathcal{P} \cap \Lambda_\mu(p)$ belong to open windmill-cones. Suppose that $\mathcal{P} \setminus \{\pm e, \pm f\}$ contains a non-zero element g of $\Lambda_\mu(p)$. Area-minimality of $\mathcal{P}$ and the absence of non-zero elements in $\Lambda_\mu(p) \cap (\mathbb{Z}(1,1) \cup \mathbb{Z}(1,-1)) \cap \{-p+1, \ldots, p-1\}^2$ shows that g is contained in a white windmill cone (under the assumption that e and f are black). Up to a sign change, the element g belongs either to the triangle with vertices $(0,0), e, f$ or to the triangle with vertices $(0,0), e, -f$. Lemma 1.1 applied to the two sets e, f and $e, -f$ generating the same sublattice $\mathbb{Z}e + \mathbb{Z}f$ of $\Lambda_\mu(p)$ implies thus the existence of a non-zero element h in $\mathcal{P} \cap \Lambda_\mu(p)$ such that $\{\pm g, \pm h\}$ intersects all four open white windmill-cones. The parallelogram with vertices $\pm g, \pm h$ in all four open white windmill-cones is thus strictly included in $\mathcal{P}$ in contradiction with area-minimality of $\mathcal{P}$.

Lemma 1.1 shows now that e, f is a windmill basis (perhaps up to sign changes) of $\Lambda_\mu(p)$. □

Proposition 1.7. *Two windmill bases e, f and e', f' of a lattice $\Lambda = \mathbb{Z}e + \mathbb{Z}f = \mathbb{Z}e' + \mathbb{Z}f'$ share a common element.*

C. Leuridan communicated to the author the following elegant proof of Proposition 1.7:

Proof. Up to replacing Λ by its reflection $\sigma(\Lambda)$ with respect to the vertical coordinate axis (given by $\sigma(x,y) = (-x,y)$) and up to permuting e with f and e' with f', Lemma 1.4 shows that we can take e, e' in the open black E-NE windmill cone and f, f' in the open black N-NW windmill cone.

There is nothing to do if $e = e'$ or if $f = f'$.

Otherwise we write

$$\begin{array}{ll} e = (x_e, y_e), & f = (-x_f, y_f), \\ e' = (x_{e'}, y_{e'}), & f' = (-x_{f'}, y_{f'}) \end{array}$$

where x_*, y_* are all strictly positive.

Up to exchanging the basis e, f with the basis e', f', Lemma 1.2 and the assumptions $e \neq e', f \neq f'$ imply that e', f' belong to the open cone spanned by e and f.

Since e, f and e', f' are generators of a common lattice Λ, there exist strictly positive integers $\alpha, \beta, \gamma, \delta$ such that $\alpha\delta - \beta\gamma = 1$ and

$$\begin{aligned} e' &= \alpha e + \beta f = (\alpha x_e - \beta x_f, \alpha y_e + \beta y_f), \\ f' &= \gamma e + \delta f = (\gamma x_e - \delta x_f, \gamma y_e + \delta y_f). \end{aligned}$$

The inclusion of e' in the open E-NE windmill cone amounts to the inequality $x_{e'} = \alpha x_e - \beta x_f > y_{e'} = \alpha y_e + \beta y_f$. We have thus

$$\alpha x_e > \alpha(x_e - y_e) > \beta(x_f + y_f) \ .$$

Since $f = (-x_f, y_f)$ belongs to open N-NW windmill cone we have $y_f > x_f > 0$ implying $x_f + y_f > 2x_f$ and leading to the inequality

$$\alpha x_e > 2\beta x_f \tag{2}$$

involving only strictly positive integers.

The inclusion of $f' = (-x_{f'}, y_{f'}) = (\gamma x_e - \delta x_f, \gamma y_e + \delta y_f)$ in the open N-NW windmill cone shows

$$\delta x_f > \gamma x_e \ . \tag{3}$$

Multiplying corresponding sides of the two inequalities (2) and (3) we get $\alpha\delta > 2\beta\gamma$ after simplification by the strictly positive integer $x_e x_f$. This leads to a contradiction by considering

$$1 = \alpha\delta - \beta\gamma > \beta\gamma \geq 1 \ .$$

□

Proposition 1.8. *A lattice Λ of rank 2 in $\mathbb{R}^2$ has only a finite number of windmill bases. If Λ has $k \geq 1$ windmill bases, then there exists a unique windmill basis e, f such that every windmill basis of Λ is of the form $e, f + se$ with s in $\{0, \ldots, k-1\}$.*

Proof. Suppose first that a lattice Λ has three distinct windmill bases e_i, f_i for $i = 1, 2, 3$ such that $\bigcap_{i=1}^{3} \{e_i, f_i\} = \emptyset$. Up to exchanging e_i and f_i for some indices i, Proposition 1.7 shows that we can assume $e_1 = e_2$. Proposition 1.7 implies now $\{e_3, f_3\} = \{f_1, f_2\}$. The underlying windmill basis e_3, f_3 is thus

contained in the same open windmill cone. This contradicts the definition of a windmill basis.

All windmill bases of Λ start thus with a common element e. Their second element is of the form $f_i = f + s_i e$ for suitable integers s_i where f is the second basis element of a fixed windmill basis e, f for Λ. Since e and f belong to different open windmill cones, the affine line $f + \mathbb{R}e$ intersects the open windmill cone containing f in an open interval of bounded length. The set of all possible integers s_i giving rise to windmill bases $e, f + s_i e$ is thus a finite set of consecutive integers. Replacing f with $f + \min(s_1, \ldots, s_k)e$ we get $\{s_1, \ldots, s_k\} = \{0, 1, \ldots, k-1\}$. □

We call a black windmill basis u, v of a lattice $\Lambda_\mu(p)$ (with μ in $\{2, \ldots, p-2\}$) *standard* if $u = (a, c), v = (-d, b)$ with $a, b, c, d \in \mathbb{N}$ such that $\min(a, b) > \max(c, d)$.

Proposition 1.9. *Given an odd prime number p, a lattice $\Lambda_\mu(p)$ with μ in $\{2, \ldots, p-2\}$ has either only white windmill bases or it has a unique standard black windmill basis.*

Proof. Proposition 1.6 shows that such a lattice has windmill bases. They are all of the same colour by Lemma 1.4 or Proposition 1.8.

We can thus assume that $\Lambda_\mu(p)$ has $k \geq 1$ black windmill bases. Proposition 1.8 shows that all these windmill bases are given by $e, f + se$ with s in $\{0, \ldots, k-1\}$ for a uniquely defined black windmill basis e, f. We set $u = (a, c) = e$, $v = (-d, b) = f + (k-1)e$ if e belongs to the open E-NE windmill cone and $u = (a, c) = f$, $v = (-d, b) = e$ otherwise (i.e. if e belongs to the open N-NW windmill cone).

We claim that u, v is a standard black windmill basis of $\Lambda_\mu(p)$: We have $a > c$ since $u = (a, c)$ belongs to the open E-NE windmill cone and $b > d$ since $v = (-d, b)$ belongs to the open N-NW windmill cone.

Since $u - v = (a + d, c - b), v = (-d, b)$ is not a windmill basis we have $b \geq c$. If $b = c$, the vectors $u - v = (a + d, 0), v = (-d, b)$ are a basis of $\Lambda_\mu(p)$. This implies $a + d = p$ and $b = 1$ contradicting the inequalities $1 \leq d < b$.

Since $u = (a, c), v + u = (a - d, b + c)$ is not a windmill basis, we have $a \geq d$. If $a = d$, the vectors $u = (a, c), v + u = (0, b + c)$ are a basis of $\Lambda_\mu(p)$. This implies $b + c = p$ and $a = 1$ contradicting the inequalities $1 \leq c < a$.

Unicity follows easily from Proposition 1.8 describing the complete set of windmill bases for $\Lambda_\mu(p)$. □

Proof of Theorem 0.1. Given an odd prime number p, we denote by $\mathcal{S}_p$ the set of all solutions (a, b, c, d) as defined by Theorem 0.1.

We associate to a solution (a, b, c, d) in $\mathcal{S}_p$ the two vectors $u = (a, c)$, $v = (-d, b)$ and we consider the sub-lattice $\Lambda = \mathbb{Z}u + \mathbb{Z}v$ of index $p = ab - c(-d)$ in $\mathbb{Z}^2$ generated by u and v. Since p is prime, there are exactly two solutions

with $cd = 0$, given by $(p, 1, 0, 0)$ and $(1, p, 0, 0)$ corresponding to the lattices $\mathbb{Z}(p, 0) + \mathbb{Z}(0, 1)$ and $\mathbb{Z}(1, 0) + \mathbb{Z}(0, p)$.

We suppose henceforth $cd > 0$. The vectors u and v are then contained respectively in the black E-NE and the black N-NW windmill-cone and form a standard black windmill basis of the lattice Λ.

Sub-lattices of prime-index p in $\mathbb{Z}^2$ are in bijection with the set of all $p+1$ points on the projective line $\mathbb{P}^1\mathbb{F}_p$ over the finite field $\mathbb{F}_p$. More precisely, a point $[a : b]$ of the projective line defines the lattice

$$\Lambda_{[a:b]} = \{(x, y) \in \mathbb{Z}^2 \mid ax + by \equiv 0 \pmod{p}\}$$

corresponding to the lattice $\Lambda_\mu(p)$ defined by (1) where $\mu \equiv b/a \pmod{p}$ using obvious conventions. We have already considered lattices associated to the two solutions with $cd = 0$. By Proposition 1.5, the lattices corresponding to $\mu \equiv \pm 1 \pmod{p}$ have no windmill basis and yield thus no solutions. All $(p - 3)$ lattices $\Lambda_\mu(p)$ with $\mu \in \{2, \ldots, p - 2\}$ have windmill bases by Proposition 1.6.

Since $\Lambda_\mu(p)$ and $\Lambda_{p-\mu}(p)$ (respectively $\Lambda_{\mu^{-1} \pmod{p}}(p)$) differ by a horizontal (respectively diagonal) reflection, they have windmill bases of different colours. Proposition 1.9 shows that exactly $(p - 3)/2$ values of μ in $\{2, \ldots, p-2\}$ correspond to lattices $\Lambda_\mu(p)$ with unique standard black windmill bases. These $(p - 3)/2$ lattice are thus in one-to-one correspondence with solutions in (a, b, c, d) in $\mathcal{S}_p$ such that $cd > 0$. Taking into account the two degenerate solutions $p \cdot 1 + 0 \cdot 0$ and $1 \cdot p + 0 \cdot 0$, the number of elements in the set $\mathcal{S}_p$ equals thus $(p - 3)/2 + 2 = (p + 1)/2$. □

Remark 1.10. *The lattice $\Lambda = \mathbb{Z}(a, c) + \mathbb{Z}(-d, b)$ associated to a solution (a, b, c, d) in $\mathcal{S}_p$ has a fundamental domain given by the union of the rectangle of size $a \times b$ with vertices $(0, 0), (a, 0), (a, b), (0, b)$ and of the rectangle of size $d \times c$ with vertices $(a, 0), (a + d, 0), (a + d, c), (a, c)$.*

2 Complements

2.1 Constructing the solution associated to $\{\mu, p-\mu\} \subset \{2, \ldots, p-2\}$

Given an odd prime number p, every pair $\{\mu, p - \mu\}$ with μ in $\{2, \ldots, p - 2\}$ defines exactly one solution in $\mathcal{S}_p$ and all solutions except $(p, 1, 0, 0)$ and $(1, p, 0, 0)$ are of this form. The associated solution can be constructed as follows: Gaußian lattice-reduction applied to

$$\Lambda_\mu(p) = \mathbb{Z}(p, 0) + \mathbb{Z}(-\mu, 1) = \{(x, y) \mid x + \mu y \equiv 0 \pmod{p}\}$$

yields a basis containing a shortest vector w in $\Lambda_\mu(p)$. Proposition 2.1 gives an efficient construction of a black windmill basis either for $\Lambda_\mu(p)$ or for

$\Lambda_{p-\mu}(p)$. Propositions 1.8 and 1.9 show how to construct the standard black windmill basis $(a,c),(-d,b)$ from an arbitrary black windmill basis for $\Lambda_\mu(p)$ or for $\Lambda_{p-\mu}(p)$.

Given a prime $p \equiv 1 \pmod 4$, the fixed point (a,a,c,c) corresponds to $u=(a,c), v=(-c,a)$. We have thus $a+\mu c \equiv -c+\mu a \equiv 0 \pmod p$ showing $\mu^2 \equiv -1 \pmod p$. The associated pair $\{\mu, p-\mu\}$ in $\{2,\dots,p-2\}$ defines thus both square roots of -1 modulo p. The corresponding construction of the associated solution $a^2+c^2=p$ boils now down to Grace's proof of Fermat's Theorem as described for example in [3].

Proposition 2.1. *Given an odd prime number p and μ in $\{2,\dots,p-2\}$, let w be a shortest non-zero element of $\Lambda_\mu(p)$. There exists a windmill basis of $\Lambda_\mu(p)$ which contains either w or a shortest element of $\Lambda_\mu(p) \setminus \mathbb{Z}w$.*

A solution (a,b,c,d) in $\mathcal{S}_p$ is of *hexagonal type* if $(a+d)^2+(c-b)^2 < \min(a^2+c^2, b^2+d^2)$. Solutions of hexagonal type correspond to standard black windmill bases u,v not containing a shortest non-zero lattice element (i.e. with $u-v$ shorter than u and v) of $\Lambda = \mathbb{Z}u+\mathbb{Z}v$. The associated lattice Λ is close to a regular hexagonal lattice.

Sketch of proof for Proposition 2.1. After a rotation by a suitable angle $k\pi/2$ and perhaps a horizontal reflection, we end up with a lattice Λ having a shortest non-zero element w in the open black E-NE windmill-cone. Let L_+ be the closest affine line above $\mathbb{R}w$ which is parallel to $\mathbb{R}w$ and intersects $\Lambda \setminus \mathbb{Z}w$. If the intersection of L_+ with the open black N-NW windmill-cone contains an element r of Λ, we get a black windmill basis by considering w,r.

Otherwise a geometric argument (using minimality of w and the definition of L_+) shows that L_+ intersects Λ in a rightmost point v of the open white W-NW windmill-cone and in a leftmost point u of the open white N-NE windmill-cone and we get a white windmill basis by considering u,v. Since u,v are separated by the black N-NW windmill-cone containing the orthogonal line to $\mathbb{R}w$, either u or v is a shortest element of $\Lambda \setminus \mathbb{Z}w$. □

2.2 Statistical properties

The lattice $\mathbb{Z}(a,c)+\mathbb{Z}(-d,b)$ associated to a solution (a,b,c,d) in $\mathcal{S}_p$ defines (up to homothety) a point $\frac{1}{\sqrt{p}}\begin{pmatrix} a & -d \\ c & b \end{pmatrix} \Big/ \mathrm{SL}(2,\mathbb{Z})$ on $\mathcal{U} = \mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$. Except on a set of measure zero contained in the set of all lattices having more than one pair of minimal vectors (non-zero vectors with smallest norm), the set U can be identified with a quotient of the unit tangent bundle of the modular curve $\mathcal{M}$ classifying lattices of $\mathbb{C}$ up to orientation-preserving similarity as follows: Given a basis u,v of an arbitrary lattice $\Lambda = \mathbb{Z}u+\mathbb{Z}v$ in $\mathbb{C}$ consisting of a vector u of minimal length and of v in the open halfspace

$iu(\mathbb{R} + i\mathbb{R}_{>0})$, we can consider the geodesic with slope $\Im(u)/\Re(u)$ at the marked point v/u of the hyperbolic plane $\mathbb{H}$. The projection of this marked geodesic onto the modular curve is well defined, except if $v/u + \mathbb{Z}$ intersects the unit-circle.

The set of lattices without windmill bases or with (at least) two pairs of minimal vectors defines a set of measure 0 on $\mathcal{U}$ which partitions its complement in $\mathcal{U}$ into two open subsets $\mathcal{U}_b$ and $\mathcal{U}_w$ of equal measure corresponding to all lattices with black, respectively white, windmill bases.

The distribution of all lattices of fixed large (not necessarily prime) index tends to the natural probability measure on $\mathcal{U}_b$, see for example Theorem 1.2 in [4]. This implies the existence of asymptotic limit-laws for all suitably rescaled smooth quantities related to solutions in $\mathcal{S}_p$. Such limit-laws are thus given by integrals over $\mathcal{U}_b$ of suitable integrable functions.

We are going to describe a few features without (the obvious) proofs:

For every strictly positive ϵ there exists a constant $B = B(\epsilon)$ such that only at most $\epsilon\frac{p+1}{2}$ elements in $\mathcal{S}_p$ satisfy the inequality $\max(a, b) > B\sqrt{p}$ or $B\min(c, d) \leq \sqrt{p}$, i.e. most solutions involve integers a, b, c, d of size $O(\sqrt{p})$. (Suitable neighbourhoods of the cusp of $\mathcal{U}$ can have arbitrarily small measures.)

The angular distribution $\arctan(c/a)$ of the vectors (a, c) is not uniform in $[0, \pi/4]$. This is due to the contribution of lattices with a black standard basis involving a small minimal vector $(-d, b)$ in the black N-NW windmill cone (such lattices give rise to vectors (a, c) with c much smaller than a) and to the existence of hexagonal solutions (giving rise to lattices with $\arctan(c/a)$ in $[\pi/6, \pi/4]$). The non-uniformity of this angular distribution is at first sight slightly surprising in comparision with uniformity of the angular distribution $\arctan(c/a)$ defined by $a^2 + c^2 = p$, $0 < c < a$ for primes $p \equiv 1 \pmod 4$, see [6].

The proportion of solutions (a, b, c, d) of hexagonal type (i.e. satisfying $(a + d)^2 + (c - b)^2 < \min(a^2 + c^2, b^2 + d^2)$) in $\mathcal{S}_p$ is fairly small (there are for example only 4370 solutions of hexagonal type among all 500002 solutions in $\mathcal{S}_{1000003}$) and tends to a limit corresponding to suitably oriented lattices in a neighbourhood of the hexagonal lattice.

Remark 2.2. *All reasonable asymptotical statistical properties continue to hold for sets of solutions $\mathcal{S}_n$ with n not necessarily prime.*

There is however perhaps no nice formula for the number of elements in sets $\mathcal{S}_n$ with n composite. (By the way, there are two possible definitions for $\mathcal{S}_n$ if n is composite: If $n = ab$ is a non-trivial factorisation of n, one can either accept or reject solutions of the form $ab + 0 \cdot c = ab + c \cdot 0 = n$ with $1 \leq c < \min(a, b)$.)

2.3 A variation

The equation $n = ab - cd$ with $\min(a, b) > \max(c, d)$ is also interesting:

Theorem 2.3. *The number of elements (a,b,c,d) in $\mathbb{N}^4$ such that $n = ab+cd$ and $\min(a,b) > \max(c,d)$ is given by*

$$\sum_{d|n,\ d^2\geq n} \left(d+1-\frac{n}{d}\right) .$$

If p is a prime number, we get p such solutions. They correspond to all p sublattices of index p in $\mathbb{Z}^2$ which do not contain the vector $(1,1)$.

Rough sketch of proof. As for the case of $p = ab+cd$, the set $\mathcal{S}_n^-$ of solutions (to $n = ab-cd$ with a,b,c,d in $\mathbb{N}$ such that $\min(a,b) > \max(c,d)$) is in bijection with a subset of lattices of index n in $\mathbb{Z}^2$: A solution (a,b,c,d) corresponds to the sublattice $\mathbb{Z}(a,c)+\mathbb{Z}(d,b)$ of index n in $\mathbb{Z}^2$. Conversely, given a sublattice Λ of index n in $\mathbb{Z}^2$, we denote by $\mathcal{C} = \mathcal{C}(\Lambda)$ the convex hull of all non-zero lattice points in $\mathbb{N}^2 \cap \Lambda \setminus \{(0,0)\}$. The lattice Λ corresponds to a solution if and only if $\Lambda \cap \mathbb{N}(1,1)$ does not intersect the boundary $\partial\mathcal{C}$ of the closed (but non-compact) convex set $\mathcal{C}$. Given such a lattice Λ, let $u = (a,c)$ and $v = (d,b)$ be the two lattice points on $\Lambda \cap \partial\mathcal{C}(\Lambda)$ which are closest to the diagonal $\mathbb{R}(1,1)$. They are always separated by the diagonal $\mathbb{R}(1,1)$ and we can assume $a > c$ and $d < b$ (up to exchanging u and v). Lemma 1.1 shows that $\Lambda = \mathbb{Z}u + \mathbb{Z}v$ and it is not difficult to check that (a,b,c,d) is in $\mathcal{S}_n^-$. Every solution (a,b,c,d) corresponds to such a basis u,v of a suitable lattice. In order to count the number of elements in $\mathcal{S}_n^-$, it is thus enough to subtract the number of 'bad sublattices'(with $\partial\mathcal{C}$ intersecting $\Lambda \cap \mathbb{Z}(1,1)$) from the total number of lattices of index n in $\mathbb{Z}^2$ (formulae giving the number of sublattices of index n in $\mathbb{Z}^d$ are for example given in [5] and in [9]). □

The sets $\mathcal{S}_n^-$ give again rise to limit statistics defined by integrals of suitable integrable functions on the unit tangent bundle of the modular curve. Asymptotically, typical solutions in $\mathcal{S}_n^-$ involve integers a,b,c,d of size roughly $\sqrt{n}$.

Remark 2.4. *The convex hull $\mathcal{C}$ occuring in the sketched proof is closely related to continued fractions: Given a real number θ, vertices of the convex hull of $e^{-i\arctan(\theta)}(\mathbb{Z}+i\mathbb{Z}) \cap [0,\infty] + i[0,\infty] \setminus \{0\}$ correspond essentially to convergents of θ, see for example [1].*

Acknowledgements: I thank M. Decauwert, P. Dehornoy, P. De la Harpe, C. Elsholtz, A. Guilloux, C. Leuridan, C. MacLean, E. Peyre, L. Spice for comments, remarks or questions. Special thanks to C. Leuridan whose numerous suggestions improved the text hugely over a previous version and to A. Guilloux who explained me the equirepartition stuff of Section 2.2.

References

[1] V. I. Arnold, Arnold, *Higher dimensional continued fractions.* Regul. Chaotic Dyn. **3**, No. 3, 10–17 (1998).

[2] D.A. Christopher, *A partition-theoretic proof of Fermat's two squares theorem.* Discrete Math. **339**, No. 4, 1410–1411 (2016).

[3] C. Elsholtz, *A combinatorial approach to sums of two squares and related problems* in Additive number theory. Festschrift in honor of the sixtieth birthday of Melvyn B. Nathanson (edited by D. Chudnovsky, David et al.), Springer, 115–140 (2010).

[4] A. Eskin, H. Oh, *Ergodic theoretic proof of equidistribution of Hecke points.* Ergodic Theory Dyn. Syst. **26**, No. 1, 163–167 (2006).

[5] B. Gruber, Alternative formulae for the number of sublattices, Acta Cryst. **A53**, 807–808 (1997).

[6] E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen. II.* Math. Z. **6**, 11–51 (1920).

[7] Wikipedia, *Fermat's theorem on sums of two squares*, November 2021, https://en.wikipedia.org/wiki/Fermat's_theorem_on_sums_of_two_squares

[8] D. Zagier, *A one-sentence proof that every prime* $p \equiv 1 \pmod 4$ *is a sum of two squares.* Am. Math. Mon. **97**, No. 2, 144 (1990).

[9] Y.M. Zhou, Gaussian binomials and the number of sublattices, Acta Cryst. **A62**, 409–410 (2006).

Roland BACHER,
Univ. Grenoble Alpes, Institut Fourier,
F-38000 Grenoble, France.

e-mail: Roland.Bacher@univ-grenoble-alpes.fr