A few different proofs of the Fermat theorem on sum of two squares

*Maryna SPEKTROVA**

*Student, University of Cambridge

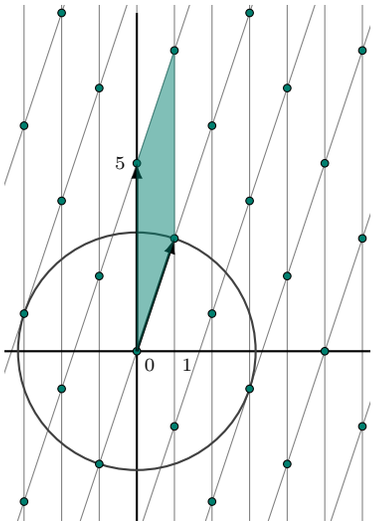
Let us begin by presenting the well-known theorem – the main subject of this article.

Theorem (Fermat's theorem on the sum of two squares). *An odd prime number p can be expressed as a sum of two squares if and only if $p \equiv 1 \pmod 4$.*

The notation $n \equiv l \pmod k$ means that $n - l$ is divisible by k. In other words, n and l leave the same remainder when divided by k.

One of the most intriguing aspects of this theorem is the vast number of different proofs, connected to various fields of mathematics. In particular, it can be proved using Gaussian integers, finite and infinite continued fractions, Thue's lemma, the method of infinite descent, and more. In this article, we present three proofs: two geometric interpretations and one proof based on approximation theory.

Remark. Since for any integer x, it holds that $x^2 \equiv 0, 1 \pmod 4$, no number congruent to 3 modulo 4 can be represented as the sum of two squares. Therefore, we will only prove one implication.



An example of the grid given by vectors $(0, 5)$ and $(1, 3)$.

First Proof

This proof will be based on a lemma concerning parallelogram grids. We do not define this concept formally but instead illustrate an example in the margin.

Lemma (Minkowski's Lemma). *Consider a parallelogram grid and a convex figure Φ, which is centrally symmetric with respect to the origin. Suppose that $S(\Phi) > 4S_0$, where S_0 is the area of the fundamental parallelogram, and $S(\Phi)$ is the area of Φ. Then, Φ contains a point with integer coordinates in this grid, different from the origin.*

Consider a parallelogram grid defined by the vectors $\vec{u} = (1, m)$ and $\vec{v} = (0, p)$, where m is an integer such that $p \mid m^2 + 1$. The existence of such an m for $p = 4k + 1$ follows from Wilson's theorem, as shown in the margin.

Now let's apply Minkowski's lemma to the circle with the center in the origin and the radius $\sqrt{2p}$. As $S(\Phi) = 2\pi p > 4p = 4S_0$, we obtain a lattice point A inside this circle, which in the basis $(\vec{u}, \vec{v})$ has integer coordinates (x_0, y_0). Thus, the Cartesian coordinates of A are given by

$$x_0 \cdot \vec{u} + y_0 \cdot \vec{v} = (x_0, mx_0 + py_0).$$

The squared distance of A from the origin is

$$x_0^2 + (mx_0 + py_0)^2 = x_0^2 + m^2x_0^2 + 2mx_0py_0 + p^2y_0^2 = x_0^2(m^2 + 1) + p(2mx_0y_0 + py_0^2),$$

which is divisible by p. Since A lies inside the circle of radius $\sqrt{2p}$, this value is in the open interval $(0, 2p)$, so it must be equal to p. Therefore, the two squares we are looking for are x_0^2 and $(mx_0 + py_0)^2$. □

We will identify a natural number m satisfying $p \mid m^2 + 1$. We use Wilson's theorem, which states that $(p-1)! \equiv -1 \pmod p$ for any prime p. From this theorem, it follows that

$$\begin{aligned} -1 &\equiv (p-1)! \\ &= 1 \cdot 2 \cdot \ldots \left(\frac{p-1}{2}\right) \cdot \left(\frac{p+1}{2}\right) \cdot \ldots (p-1) \\ &\equiv 1 \cdot 2 \cdot \ldots \left(\frac{p-1}{2}\right) \cdot \left(-\frac{p-1}{2}\right) \cdot \ldots (-1) \\ &= \left(\left(\frac{p-1}{2}\right)!\right)^2 \cdot (-1)^{\frac{p-1}{2}} \\ &= \left(\left(\frac{p-1}{2}\right)!\right)^2 \pmod p. \end{aligned}$$

The last step holds because $p \equiv 1 \pmod 4$. Thus, $p \mid m^2 + 1$ for $m = (\frac{p-1}{2})!$.

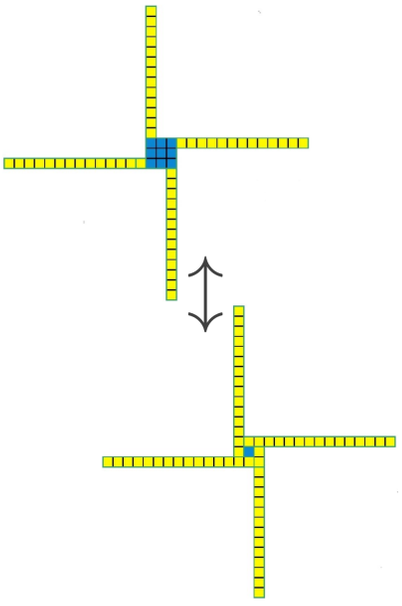
Second proof

This proof is often referred to as the "one-sentence proof by Don Zagier", and indeed, when first published in 1990, it contained only one sentence. For better readability, we present it here in a slightly more detailed form.

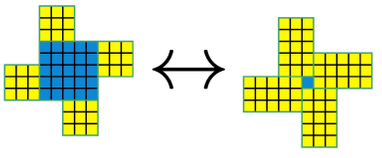
Wojciech Czerwiński wrote about Zagier's proof in Δ^7_{17}.

Consider the set S containing all triples (x,y,z) of natural numbers that satisfy $x^2 + 4yz = p$. Clearly, for each prime p, the set S is finite. Let us define the following function $f : \mathbb{N}^3 \to \mathbb{N}^3$:

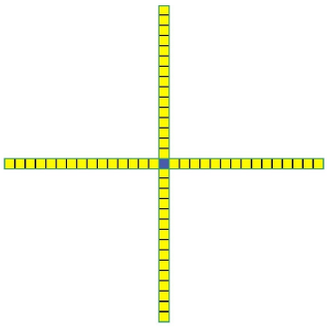
$$f(x, y, z) = \begin{cases} (x + 2z, z, y - x - z), & \text{when } x < y - z, \\ (2y - x, y, x - y + z), & \text{when } y - z < x < 2y, \\ (x - 2y, x - y + z, y), & \text{when } 2y < x. \end{cases}$$



Example of the involution for $p = 61$ ($(3, 1, 13) \leftrightarrow (1, 15, 1)$)



Example of the involution for $p = 61$ ($(5, 3, 3) \leftrightarrow (1, 3, 5)$)



The only fixed point of f for $p = 61$ is $(1, 1, 15)$. Observe how this configuration differs from $(1, 15, 1)$

Wojciech Czerwiński wrote about Dirichlet's approximation theorem in *Delta* 3/21.

We will refer to the triples appearing in the respective cases of f as first, second, and third type triples.

By performing straightforward calculations, we can easily verify that:

- $f(S) \subseteq S$;
- f is a bijection and, in fact, an involution, meaning that $f(f(x, y, z)) = (x, y, z)$;
- Triples of the first, second, and third type map respectively to triples of the third, second, and first type.

Now, let us investigate which triples $t \in S$ are fixed points, i.e., satisfy $f(t) = t$. A triple can be a fixed point only if it is of the second type, which leads to the following system of equations:

$$\begin{cases} 2y - x = x, \\ y = y, \\ x - y + z = z, \end{cases}$$

which simplifies to $x = y$. We know that $x^2 + 4yz = p$, so $x^2 + 4xz = p$. Since p is a prime number, we easily obtain $x = y = 1, z = \frac{p-1}{4}$, meaning that there is exactly one fixed point in S. This implies that S has an odd number of elements.

Observe that we can pair the elements of S such that each triple (x, y, z) is paired with (x, z, y). Since S has an odd number of elements, at least one triple must be paired with itself. For this triple (x_0, y_0, z_0), we have $y_0 = z_0$, so $p = x_0^2 + 4y_0^2 = x_0^2 + (2y_0)^2$. □

An intriguing aspect of this proof is the geometric interpretation of the function f, surprisingly published only in 2007. For each triple in S, we can consider a square with side x and attach four rectangles with sides y and z to it, as shown in the diagram. Each such figure (a "mill" or "square with wings") can be obtained by two different cuts: one forming a smaller square and one forming a larger one. Thus, we obtain an involution between all triples in S (which in fact is just f), with crosses as fixed points. Since for a cross-shaped figure, p must be divisible by the side of the main square, this side must be equal to 1. Thus, there is exactly one fixed point. Now, the same reasoning as above completes the proof.

Third proof

This proof is based on the following theorem, which intuitively describes how well we can approximate a given real number by rational numbers with denominators not exceeding a given integer N.

Theorem (Dirichlet's approximation theorem). *For every $\alpha \in \mathbb{R}$ and $N \in \mathbb{N}$, there exist $r, q \in \mathbb{Z}$ such that:*

- $1 \leqslant q \leqslant N$, and
- $\left|\alpha - \frac{r}{q}\right| \leqslant \frac{1}{Nq}$.

Now consider such an m that $p \mid m^2 + 1$, and apply Dirichlet's approximation theorem for $\alpha = \frac{m}{p}$ and $N = \lceil \sqrt{p} \rceil$. Then, there exist $q \in \mathbb{N}, r \in \mathbb{Z}$ such that $1 \leqslant q \leqslant N$ and $\left|\frac{m}{p} - \frac{r}{q}\right| \leqslant \frac{1}{Nq}$, which is equivalent to $|mq - rp| \leqslant \frac{p}{N} < \sqrt{p}$.

Define M as $(mq - rp)^2 + q^2$. It is easy to check that $M \equiv m^2q^2 + q^2 \equiv 0 \pmod{p}$.

- **Case 1:** $q \neq N$
  Since both $(mq - rp)^2$ and q^2 are less than p, we have $0 < M < 2p$. But M is divisible by p, so $M = p$, which gives $p = (mq - rp)^2 + q^2$.

- **Case 2:** $q = N$
  It is easy to show that $M < 3p$, so $M \in \{p, 2p\}$. If $M = p$, the theorem is already proved. Now consider the case $M = 2p$. Then
  $$(mq - rp)^2 > 2p - (\sqrt{p} + 1)^2 = (\sqrt{p} - 1)^2 - 2,$$
  so
  $$|mq - rp| \in \{\lfloor \sqrt{p} - 1 \rfloor, \lfloor \sqrt{p} \rfloor\} = \{q - 2, q - 1\}.$$
  Moreover, in this case, $|mq - rp|$ and q have the same parity, so $|mq - rp| = q - 2$. Thus, $(q - 2)^2 + q^2 = 2p$, which simplifies to $(q - 1)^2 + 1 = p$, meaning that we have represented p as the sum of two squares. □

When we compare the first and third proofs, we can notice several similarities. Is this merely a coincidence? It turns out that Dirichlet's theorem can actually be derived from Minkowski's theorem.

To see this, consider the set
$$S = \left\{(x, y) \in \mathbb{R}^2 : |x| < N + \frac{1}{2}, |y - \alpha x| \leqslant \frac{1}{N}\right\}.$$
The set S is a parallelogram centered at the origin with area equal to $2(N + \frac{1}{2}) \cdot \frac{2}{N} > 4$, so we can apply Minkowski's theorem to S on the standard Cartesian lattice. This yields an integer point (q, r) in S (by symmetry, we can take q positive), and these q, r satisfy the condition from Dirichlet's theorem.

Theorem (Jacobi's two-square theorem)

Not only does Fermat's theorem have several surprising proofs, but the same is true for one of its generalizations. The following theorem can be proven using quadratic forms, elliptic functions, or Gaussian integers, but here we present only a proof using generating functions.

There are actually several directions in which Fermat's theorem may be generalized or expanded: one can look on the primes of the form $x^2 + ny^2$, obtain a criterion to all natural numbers to be representable as a sum of two squares (Sum of two squares theorem), consider larger amount of squares (Legendre's three-square theorem, Lagrange's four-square theorem), and so on.

Theorem (Jacobi's two-square theorem). *Let $r_2(n)$ denote the number of ways to represent a natural number as a sum of squares of two integers. Moreover, let $d_k(n)$ denote the number of natural divisors of n that leave a remainder k modulo 4. Then, for any natural number n, the following holds:*
$$r_2(n) = 4(d_1(n) - d_3(n)).$$

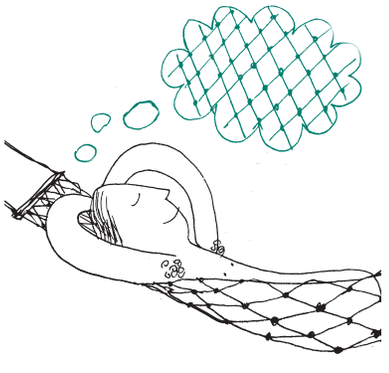
Remark. *Since for every prime number of the form $p = 4k + 1$ we have $d_1(p) = 2$ and $d_3(p) = 0$, Fermat's theorem follows directly from this one.*

Our proof of this theorem (actually, just a very brief sketch) will be based on the following theorem:

Theorem (Jacobi triple product). *For any complex numbers q and z satisfying $|q| < 1$ and $z \neq 0$, the following identity holds:*
$$\prod_{m=1}^{\infty} \left(1 - q^{2m}\right)\left(1 + q^{2m-1}z^2\right)\left(1 + q^{2m-1}z^{-2}\right) = \sum_{n=-\infty}^{\infty} q^{n^2} z^{2n}.$$

Using this theorem multiple times, the following identity can be proven:

$$k^{-1} \prod_{m=1}^{\infty} (1 - x^{4m})(1 - x^{4m-3}k^2)(1 - x^{4m-1}k^{-2}) + \prod_{m=1}^{\infty} (1 - x^{4m})(1 - x^{4m-3}k^{-2})(1 - x^{4m-1}k^2) = \prod_{m=1}^{\infty} (1 - (-x)^m)(1 + (-x)^m k)(1 + (-x)^{m-1}k^{-1}).$$



Differentiating both sides of this equation with respect to k and substituting $k = -1$, we obtain the following identity:
$$\left(\sum_{n=-\infty}^{\infty} x^{n^2}\right)^2 = 1 + 4\sum_{n=1}^{\infty} \left(\frac{x^{4n-3}}{1 - x^{4n-3}} - \frac{x^{4n-1}}{1 - x^{4n-1}}\right).$$

In fact, this equation is equivalent to Jacobi's theorem. To see this, let us first rewrite the right-hand side. The standard formula for the sum of a geometric series implies:

$$\sum_{n=1}^{\infty}\left(\frac{x^{4n-3}}{1-x^{4n-3}}-\frac{x^{4n-1}}{1-x^{4n-1}}\right)$$
$$=\sum_{n=1}^{\infty}\Big(\sum_{k=1}^{\infty}x^{(4n-3)k}-\sum_{k=1}^{\infty}x^{(4n-1)k}\Big)$$
$$=\sum_{n\in\mathbb{N}\setminus\{0\},k\in\mathbb{N}}(x^{(4n-3)k}-x^{(4n-1)k})=\sum_{k=1}^{\infty}(d_1(k)-d_3(k))x^k.$$

Now let us rewrite the left-hand side:

$$\Big(\sum_{n=-\infty}^{\infty}x^{n^2}\Big)^2=\Big(\ldots+x^{(-1)^2}+x^{0^2}+x^{1^2}+\cdots\Big)^2=1+\sum_{k=1}^{\infty}r_2(k)x^k.$$

Thus, we have proved that

$$1+\sum_{k=1}^{\infty}r_2(k)x^k=1+\sum_{k=1}^{\infty}(4d_1(k)-4d_3(k))x^k,$$

which completes the proof of Jacobi's theorem. □

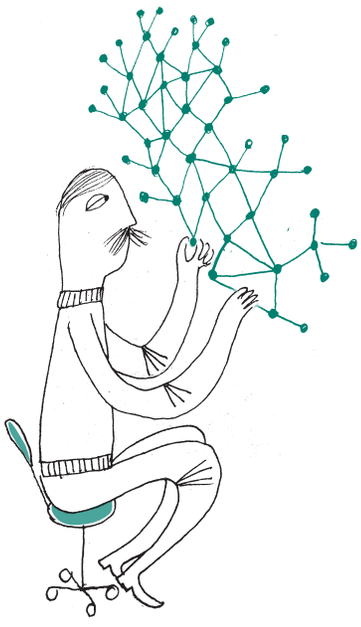
Paul Erdős, one of the greatest mathematicians of the 20th century, often referred to The Book, where God keeps the perfect proofs for mathematical theorems. He famously said: "You don't have to believe in God, but, as a mathematician, you should believe in The Book". If such a book truly exists, I believe that these proofs, which are an excellent example of the deep connections between different areas of mathematics, would certainly deserve a special place in it.

Can you see a graph here?

*Sylwia SAPKOWSKA**

* Student, Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw

Graph search algorithms are everywhere. Without them, your favorite web search engine wouldn't work, and you probably optimize your own "graph paths" subconsciously – for example, when planning your day. The connections between you, your friends, and your family also form a graph, which we feverishly search through during social gatherings.



Since graph search algorithms have many applications in everyday life, it's no surprise that numerous math olympiad problems can also be solved using this idea.

Depth First Search

Before moving forward, let's discuss the simplest graph search algorithm – **Depth First Search**, or DFS for short.

We can think of it as a "walk" through a graph. While performing the algorithm, we keep track of whether a given vertex has already been visited. We start by selecting an initial vertex. When we reach some vertex v during the execution of the algorithm, we follow these steps:

1. Mark vertex v as visited.
2. For each adjacent vertex u of the current vertex v, if u has **not** been visited yet – recursively perform this procedure for u.
3. After exploring all adjacent vertices of v, backtrack to the previous vertex – the one from which we arrived at v.

DFS can be thought of as an analogy to sightseeing. Imagine you're a tourist in a bustling city, holding a map and eager to explore every hidden gem the city