3. Infinite Descent

Theorem 3.1 (Fermat)**.** *An odd prime p is the sum of two squares if and only if $p \equiv 1 \mod 4$.*

Note that one direction is very easy, since $u^2 \equiv 0$ or $1 \mod 4$, so that the sum of two squares is never congruent to 3 modulo 4.

First we present Euler's original argument and then a more modern proof due to Zagier.

Lemma 3.2. *If n is a sum of two squares and $n = pm$, and the prime p is a sum of two squares then m is a sum of two squares.*

Proof. Indeed suppose that $n = a^2 + b^2$ and $p = u^2 + v^2$. Then p divides

$$\begin{aligned}(ub - va)(ub + va) &= u^2b^2 - v^2a^2 \\ &= u^2(a^2 + b^2) - a^2(u^2 + v^2) \\ &= u^2n - a^2p.\end{aligned}$$

As p is prime, it divides one of the factors. By symmetry we may suppose that it divides $ub - va$.

As

$$(a^2 + b^2)(u^2 + v^2) = (au + bv)^2 + (av - bu)^2$$

and the LHS is $np = mp^2$, it follows that p divides $au + bv$. As both terms on the right are divisible by p, both terms on the RHS are divisible by p^2. Now divide through by p^2. □

Lemma 3.3. *If $n = n_1n_2$ is a sum of squares and n_1 is not a sum of squares then some factor of n_2 is not a sum of squares.*

Proof. Suppose that $n_2 = p_1p_2 \dots p_k$ is the prime factorisation of n_2. If every $p_1, p_2, \dots, p_k$ is a sum of squares then n_1 is a sum of squares by (3.2) and induction on k. □

Proposition 3.4. *If n has a primitive representation then every factor of n is a sum of squares.*

Proof. Suppose that $n = a^2 + b^2$, where $(a, b) = 1$.

Suppose that $n_1|n$. We may write

$$a = cn_1 + r \qquad \text{and} \qquad b = dn_1 + s,$$

where $2|r|$ and $2|s| \leq n_1$. It follows that

$$\begin{aligned}n &= a^2 + b^2 \\ &= (cn_1 + r)^2 + (dn_1 + s)^2 \\ &= c^2n_1^2 + 2crn_1 + r^2 + d^2n_1^2 + 2dsn_1 + s^2 \\ &= An_1 + r^2 + s^2.\end{aligned}$$

It follows that $r^2 + s^2$ is divisible by n_1,

$$r^2 + s^2 = n_1 m_1.$$

Suppose that $d = (r, s)$. Then d is coprime to n_1 as a and b are coprime. Dividing through by d^2, we may assume that $(r, s) = 1$. Note that $m_1 \leq n_1/2$ as

$$\begin{aligned} r^2 + s^2 &\leq \left(\frac{n_1}{2}\right)^2 + \left(\frac{n_1}{2}\right)^2 \\ &= \frac{n_1^2}{2}. \end{aligned}$$

If n_1 is not a sum of squares then (3.3) implies that some factor n_2 of m_1 is not a sum of squares. Note that n_2 divides $n_1 m_1$ which has a primitive representation as a sum of squares. As $n_2 \leq m_1 < n_1$ we can argue by descent that this is not possible. Thus n_1 is a sum of squares. □

Here is Euler's proof

Proof of (3.1). Suppose that $p = 4n + 1$. Then each of the numbers

$$1^{4n} \qquad 2^{4n} \qquad \dots \qquad \text{and} \qquad (4n)^{4n}$$

is congruent to one, modulo p. Therefore all of the differences

$$2^{4n} - 1^{4n} \qquad 3^{4n} - 2^{4n} \qquad \dots \qquad \text{and} \qquad (4n)^{4n} - (4n-1)^{4n}$$

are divisible by p. Each of these differences factors as

$$a^{4n} - b^{4n} = (a^{2n} + b^{2n})(a^{2n} - b^{2n}).$$

If p divides the first factor then (3.4) implies that p is a sum of squares (note that a and b are coprime as their difference is one).

The only remaining possibility is that it always divides the second factor, that is, p divides $2^{2n} - 1^{2n}$, $3^{2n} - 2^{2n}$, $\dots$, $(4n)^{2n} - (4n-1)^{2n}$. Taking second differences, then third differences and so on, we see that the $(2n)$th difference is also divisible by p. But the $(2n)$th differences of any $2n$ successive $(2n)$th powers is $(2n)!$, which is not divisible by p, a contradiction. □

Here is Zagier's proof.

Proof of (3.1). Consider the set

$$S = \{ (x, y, z) \in \mathbb{N}^3 \,|\, x^2 + 4yz = p \}.$$

Note that S is clearly finite, as x, y and $z \leq p$.

Suppose that $(x, y, z) \in \mathbb{N}^3$. It is clear that if $(x, y, z) \in S$ then x is not even, as p is not even.

Note that if $x = y - z$ then

$$
\begin{aligned}
x^2 + 4yz &= (y - z)^2 + 4yz \\
&= y^2 + 2yz + z^2 \\
&= (y + z)^2 \\
&\neq p
\end{aligned}
$$

and so $(x, y, z) \notin S$.

Let

$$\tau\colon S \longrightarrow S$$

be the function

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

By what we have just proved the recipe for τ gives a well-defined function to $\mathbb{N}^3$. We check that the image lies in S. Let $(a, b, c) = \tau(x, y, z)$. It is not hard to see that all three coordinates a, b and c are natural numbers. We have to also check that (a, b, c) is a solution to the equation. There are three cases:

$$
\begin{aligned}
a^2 + 4bc &= (x + 2z)^2 + 4z(y - x - z) \\
&= x^2 + 4xz + 4z^2 + 4yz + -4zx - 4z^2 \\
&= x^2 + 4yz \\
&= p,
\end{aligned}
$$

so that $(a, b, c) \in S$. The second case is almost the same as the first; just switch y and z and flip the sign of x. For the third case, note that a^2 and $4bc$ are the same as for the second case. Thus $\tau(x, y, z) \in S$ and so τ is a well-defined map.

We check that τ is an involution, that is, it is its own inverse, that is, τ^2 is the identity. There are three cases. If $x < y - z$ then $a > 2b$ and so

$$
\begin{aligned}
\tau^2(x, y, z) &= \tau(a, b, c) \\
&= (a - 2b, a - b + c, b) \\
&= (x + 2z - 2z, x + 2z - z + (y - x - z), z) \\
&= (x, y, z).
\end{aligned}
$$

If $y - z < x < 2y$ then $b - c < a < 2b$ and so

$$\begin{aligned}
\tau^2(x, y, z) &= \tau(a, b, c) \\
&= (2b - a, b, a - b + c) \\
&= (2y - (2y - x), y, (2y - x) - y + (x - y + z)) \\
&= (x, y, z).
\end{aligned}$$

Finally, if $x > 2y$ then $a < b - c$ and so

$$\begin{aligned}
\tau^2(x, y, z) &= \tau(a, b, c) \\
&= (a + 2c, c, b - a - c) \\
&= (x - 2y + 2y, y, x - y + z - (x - 2y) - y) \\
&= (x, y, z).
\end{aligned}$$

We look for fixed points, points such that $(a, b, c) = (x, y, z)$. By the above, we must have $y - z < x < 2y$, in which case

$$x = 2y - x \qquad y = y \qquad \text{and} \qquad z = x - y + z.$$

Thus $x = y$. We then have

$$p = x^2 + 4xz,$$

so that $x = 1$ and this determines z. On the other hand, as $p = 4n + 1$, $(1, 1, n)$ is a fixed point, so that it is the unique fixed point.

It follows that $|S|$ is odd, since every point is matched with another point, except for the fixed point.

Now consider the function

$$\sigma \colon S \longrightarrow S$$

given by

$$\sigma(x, y, z) = (x, z, y).$$

σ is clearly an involution of S. As $|S|$ is odd it follows that σ has at least one fixed point. In this case $y = z$ so that

$$p = x^2 + 4y^2,$$

is a sum of squares. $\square$