8. Sums of squares

8.1 Primes as a sum of two squares

Recall that the following notation has been used earlier in the study of binomial coefficients.

Definition 8.1 Let p be a prime. Given any nonzero integer n, we denote by $\nu_p(n)$ the unique nonnegative integer k such that $p^k \mid n$ and $p^{k+1} \nmid n$, namely, $\nu_p(n)$ is the power of p in the canonical form of n.

Theorem 8.1 Let x and y be integers, not both zero. For any prime p with $p \equiv 3 \pmod 4$, we have that $\nu_p(x^2+y^2)$ is even.

Proof. Let $n = x^2+y^2$. Note that $n > 0$. Let $d = (x,y)$ and write $x = x_0 d$ and $y = y_0 d$ so $(x_0, y_0) = 1$. Hence, $n = d^2(x_0^2+y_0^2)$.

We first show that $p \nmid (x_0^2+y_0^2)$. If not, then $x_0^2+y_0^2 \equiv 0 \pmod p$, or $x_0^2 \equiv -y_0^2 \pmod p$. Since $(x_0,y_0) = 1$, at least one of x_0 and y_0 is coprime to p. Without loss of generality, we assume that $(y_0,p) = 1$, meaning that y_0 has a modular inverse y_0^{-1} modulo p. Hence, $(x_0 y_0^{-1})^2 \equiv -1 \pmod p$, indicating that -1 is a quadratic residue modulo p. However, this violates Theorem 6.8, saying that -1 is a quadratic non-residue as $p \equiv 3 \pmod 4$.

Thus, $\nu_p(n) = \nu_p(d^2) = 2\nu_p(d)$, which is even. ∎

Theorem 8.2 Any prime p with $p \equiv 1 \pmod 4$ can be written as a sum of two squares.

We will present two proofs of this result: one is based on an important method called "infinite descent" developed by Fermat, and the other relies on a magical involution due to Don Zagier.

Before moving forward, we record a simple but useful formula.

Theorem 8.3 Let $x_1, y_1, x_2, y_2 \in \mathbb{R}$. Then

$$(x_1^2+y_1^2)(x_2^2+y_2^2) = (x_1x_2+y_1y_2)^2 + (x_1y_2-x_2y_1)^2. \tag{8.1}$$

Proof. This formula can be examined by a direct calculation. ∎

R We may also understand (8.1) with recourse to complex numbers. Recall that a *complex number* z is of the form $z = x + yi$ with $x, y \in \mathbb{R}$ where $i = \sqrt{-1}$ is the *imaginary unit*. The *modulus* of z is define by $|z| = \sqrt{x^2 + y^2}$. Let $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$. Note that the left hand side of (8.1) is $|z_1|^2 |z_2|^2$ and the right hand side is $|z_1 z_2|^2$. So, $|z_1|^2 |z_2|^2 = |z_1 z_2|^2$.

8.2 The method of infinite descent

Among different variants of the method of *infinite descent*, we will make use of the following version.

The Method of Infinite Descent Let P be a property that at least one positive integer possesses. Assume that whenever $m > 1$ possesses P, we may find another positive integer m_0 with $m_0 < m$ such that m_0 also possesses P. Then 1 possesses P.

First Proof of Theorem 8.2. Recall from Theorem 6.9 that for primes p with $p \equiv 1 \pmod 4$, there exists an integer x such that $x^2 + 1 = mp$ with $0 < m < p$. In other words, there exists an integer m with $0 < m < p$ such that the equation

$$x^2 + y^2 = mp$$

has an integer solution (x, y).

Assume that $m > 1$. Note that for any integer n, we may always find an integer n_0 with $|n_0| \leq \frac{m}{2}$ such that $n \equiv n_0 \pmod m$. This is because there are at least m consecutive integers in the interval $[-\frac{m}{2}, \frac{m}{2}]$, thereby covering a complete system modulo m.

Now, we find $x \equiv x_0 \pmod m$ with $|x_0| \leq \frac{m}{2}$ and $y \equiv y_0 \pmod m$ with $|y_0| \leq \frac{m}{2}$. Note that we cannot simultaneously have $m \mid x$ and $m \mid y$ for if this is the case, then $m^2 \mid (x^2 + y^2)$ but $m^2 \nmid mp$ since $0 < m < p$ (and hence $(m, p) = 1$), thereby leading to a contradiction. Hence, x_0 and y_0 are not simutaneously 0, and we then have $x_0^2 + y_0^2 > 0$. On the other hand, $x_0^2 + y_0^2 \leq 2 \cdot (\frac{m}{2})^2 < m^2$. Noting that $x_0^2 + y_0^2 \equiv x^2 + y^2 = mp \equiv 0 \pmod m$, we may write $x_0^2 + y_0^2 = m_0 m$ with $0 < m_0 < m$. By Theorem 8.3, we have

$$(xx_0 + yy_0)^2 + (xy_0 - x_0 y)^2 = (x^2 + y^2)(x_0^2 + y_0^2) = (mp) \cdot (m_0 m) = m^2 m_0 p.$$

Meanwhile, we have $xx_0 + yy_0 \equiv x^2 + y^2 \equiv 0 \pmod m$ and $xy_0 - x_0 y \equiv xy - xy = 0 \pmod m$. Hence, $\frac{xx_0 + yy_0}{m}$ and $\frac{xy_0 - x_0 y}{m}$ are integers. It follows that

$$m_0 p = \left(\frac{xx_0 + yy_0}{m} \right)^2 + \left(\frac{xy_0 - x_0 y}{m} \right)^2,$$

a sum of two squares.

Finally, noting that m_0 is a positive integer with $m_0 < m$, we deduce that $x^2 + y^2 = p$ has an integer solution (x, y) with recourse to the method of infinite descent. ■

8.3 Zagier's magical involution

Definition 8.2 Let S be a set. We say that $f : S \to S$ is an *involution* on S if for any $x \in S$, there holds true that $f(f(x)) = x$.

R In fact, every involution f is a bijective map on S. The surjectivity follows by the fact every $x \in S$ in the image of $f(x)$ under f, and the injectivity follows by the fact that if $f(x) = f(y)$, then $x = f(f(x)) = f(f(y)) = y$.

Definition 8.3 Let S be a set and $f : S \to S$ be a map on S. We say that $x \in S$ is a *fixed point* under f if $f(x) = x$.

Theorem 8.4 Let S be a finite set and assume that there is an involution f on S.
(i) If f has no fixed points, then the size $|S|$ of S is even.
(ii) If f has exactly one fixed point, then $|S|$ is odd.

Proof. Since f is an involution on S, we may pair elements of S according to $(x, f(x))$ and treat $(f(x), x)$ as the same pair. Assume that there are s such pairs.

(i). Since f has no fixed points, we have $x \neq f(x)$ in each pair. Thus, every $x \in S$ belongs to exactly one of the pairs. It follows that $|S| = 2s$, which is even.

(ii). Assume that the only fixed point of f is x_0. Every $x \in S$ is either x_0, or belongs to exactly one of the pairs, excluding $(x_0, f(x_0)) = (x_0, x_0)$. Thus, $|S| = 1 + 2(s-1) = 2s-1$, which is odd. ■

Theorem 8.5 Let p be a prime with $p \equiv 1 \pmod 4$. Consider the finite set $S = \{(x,y,z) \in \mathbb{Z}^3_{>0} : x^2 + 4yz = p\}$. Then the following map f on S,

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

is an involution, and it has exactly one fixed point. In particular, $|S|$ is odd.

Proof. We first show that $x \neq y-z$ and $x \neq 2y$. If $x = y-z$, then $p = (y-z)^2 + 4yz = (y+z)^2$ which is impossible since p is prime. If $x = 2y$, then $p = (2y)^2 + 4yz = 4y(y+z)$ which is also impossible. Thus, we may separate S into three disjoint subsets S_1, S_2 and S_3 according to **(1).** $x < y-z$, **(2).** $y-z < x < 2y$, **(3).** $x > 2y$.

A direct calculation reveals that for any $(x,y,z) \in S$, $f(f(x,y,z)) = (x,y,z)$, and hence, f is an involution. Also, if $(x,y,z) \in S_1$, then $f(x,y,z) \in S_3$; if $(x,y,z) \in S_2$, then $f(x,y,z) \in S_2$; and if $(x,y,z) \in S_3$, then $f(x,y,z) \in S_1$. Hence, fixed points (x,y,z) are only in S_2, with

$$x = 2y - x, \qquad y = y, \qquad z = x - y + z,$$

or $x = y$. But in this case, $p = x^2 + 4xz = x(x+4z)$ implies that the only possible x is $x = 1$, and so $y = x = 1$. Finally, since $p \equiv 1 \pmod 4$, that is, $p = 4k+1$ with $k > 0$, we have the unique fixed point $(x,y,z) = (1,1,k)$. We conclude from Theorem 8.4 that $|S|$ is odd. ■

Second Proof of Theorem 8.2. The set S in Theorem 8.5 also has a trivial involution g given by $g(x,y,z) = (x,z,y)$. But g must have a fixed point; otherwise, $|S|$ is even by Theorem 8.4, thereby contradicting to Theorem 8.5. But the fixed point of g means that $z = y$. Hence, we may find positive integers x and y such that $p = x^2 + 4y^2 = x^2 + (2y)^2$. ■

R Don Zagier's proof was published in (*Amer. Math. Monthly* **97** (1990), no. 2, 144). In fact, his involution is a refinement of an equally beautiful argument attributed to Roger Heath-Brown (*Invariant* (1984), 2–5). Heath-Brown's proof, dating back to 1971, was motivated by his study of J. V. Uspensky and M. A. Heaslet's book "*Elementary Number Theory*" (McGraw-Hill Book Co., Inc., New York, 1939), which accounts Liouville's papers on identities for parity functions.

8.4 Fermat's two-square theorem

Now, we are in a position to characterize which integers can be written as a sum of two squares.

Theorem 8.6 (Fermat's Two-Square Theorem). A positive integer n can be written as a sum of two squares if and only if all prime factors p of n with $p \equiv 3 \pmod 4$ have even exponents in the canonical form of n.

Proof. The "only if" part has been shown by Theorem 8.1. For the "if" part, we write in the canonical form

$$n = 2^{\alpha} \prod_{p \equiv 1 \bmod 4} p^{\beta} \prod_{q \equiv 3 \bmod 4} q^{2\gamma}.$$

Here, p runs over all distinct prime factors of n that are congruent to 1 modulo 4, and q runs over all distinct prime factors of n that are congruent to 3 modulo 4. In particular, the exponent of each q is even as assumed. Now, note that $2 = 1^2 + 1^2$, that for each q, we have $q^2 = 0^2 + q^2$, and that for each p, we have $p = x^2 + y^2$ for certain integers x and y by Theorem 8.2. A repeated application of Theorem 8.3 gives the desired result. ∎

8.5 Lagrange's four-square theorem

Concerning sums of four squares, we first require an analog of Theorem 8.3.

Theorem 8.7 Let $x_1, y_1, z_1, w_1, x_2, y_2, z_2, w_2 \in \mathbb{R}$. Then

$$\begin{aligned}(x_1^2 + y_1^2 + z_1^2 + w_1^2)(x_2^2 + y_2^2 + z_2^2 + w_2^2) \\ = (x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2)^2 + (x_1y_2 - y_1x_2 + z_1w_2 - w_1z_2)^2 \\ + (x_1z_2 - y_1w_2 - z_1x_2 + w_1y_2)^2 + (x_1w_2 + y_1z_2 - z_1y_2 - w_1x_2)^2. \end{aligned} \tag{8.2}$$

Proof. This formula can also be examined by a direct calculation. ∎

Theorem 8.8 (Lagrange's Four-Square Theorem). Every positive integer can be written as a sum of four squares.

Proof. Note that $1 = 0^2 + 0^2 + 0^2 + 1^2$ and $2 = 0^2 + 0^2 + 1^2 + 1^2$. In view of Theorem 8.7, it suffices to show that every odd prime can be written as a sum of four squares.

Recall from Theorem 6.10 that for odd primes p, there exists integer x and y such that $x^2 + y^2 + 1 = mp$ with $0 < m < p$. In other words, there exists an integer m with $0 < m < p$ such that the equation

$$x^2 + y^2 + z^2 + w^2 = mp$$

has an integer solution (x, y, z, w).

Assume that $m > 1$. We have two cases.

(i). If m is even, then two of the integers x, y, z and w have the same parity, and the remaining two also have the same parity. Without loss of generality, we assume that x and y have the same parity, and z and w have the same parity. Thus, the four integers $x+y$, $x-y$, $z+w$, $z-w$ are even. Note that if $m_0 = \frac{m}{2}$, then $0 < m_0 < m$. Also,

$$\begin{aligned} m_0 p &= \frac{1}{2}(x^2 + y^2 + z^2 + w^2) \\ &= \left(\frac{x+y}{2}\right)^2 + \left(\frac{x-y}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2, \end{aligned}$$

a sum of four squares.

(ii). If m is odd, then similar to the first proof of Theorem 8.2, we find $x \equiv x_0 \pmod{m}$ with $|x_0| < \frac{m}{2}$, $y \equiv y_0 \pmod{m}$ with $|y_0| < \frac{m}{2}$, $z \equiv z_0 \pmod{m}$ with $|z_0| < \frac{m}{2}$ and $w \equiv w_0 \pmod{m}$ with $|w_0| < \frac{m}{2}$. Here, we use strict "$<$" since m is odd. Therefore, $x_0^2+y_0^2+z_0^2+w_0^2 < 4 \cdot (\frac{m}{2})^2 = m^2$. Also, we cannot simultaneously have $m \mid x$, $m \mid y$, $m \mid z$ and $m \mid w$, and hence, $x_0^2+y_0^2+z_0^2+w_0^2 > 0$. Noting that $x_0^2+y_0^2+z_0^2+w_0^2 \equiv x^2+y^2+z^2+w^2 = mp \equiv 0 \pmod{m}$, we may write $x_0^2+y_0^2+z_0^2+w_0^2 = m_0 m$ with $0 < m_0 < m$. By Theorem 8.7, we have

$$\begin{aligned} m^2 m_0 p &= (mp) \cdot (m_0 m) = (x^2+y^2+z^2+w^2)(x_0^2+y_0^2+z_0^2+w_0^2) \\ &= (xx_0+yy_0+zz_0+ww_0)^2 + (xy_0-yx_0+zw_0-wz_0)^2 = \\ &\quad + (xz_0-yw_0-zx_0+wy_0)^2 + (xw_0+yz_0-zy_0-wx_0)^2 \\ &=: \tilde{x}^2+\tilde{y}^2+\tilde{z}^2+\tilde{w}^2. \end{aligned}$$

Since $x \equiv x_0 \pmod{m}$, $y \equiv y_0 \pmod{m}$, $z \equiv z_0 \pmod{m}$, $w \equiv w_0 \pmod{m}$ and $x^2+y^2+z^2+w^2 \equiv 0 \pmod{m}$, we find that $\tilde{x}$, $\tilde{y}$, $\tilde{z}$ and $\tilde{w}$ are all multiples of m. Hence,

$$m_0 p = \left(\frac{\tilde{x}}{m}\right)^2 + \left(\frac{\tilde{y}}{m}\right)^2 + \left(\frac{\tilde{z}}{m}\right)^2 + \left(\frac{\tilde{w}}{m}\right)^2,$$

a sum of two squares.

Finally, noting that in both cases of the above, m_0 is a positive integer with $m_0 < m$, we deduce that $x^2+y^2+z^2+w^2 = p$ has an integer solution (x,y,z,w) with recourse to the method of infinite descent. ■