ON GEOMETRIC PROOFS OF THEOREMS ON SUMS OF SQUARES

STEVE FAN

ABSTRACT. This short note discusses some interesting geometric proofs of the classical theorems on sums of two squares, three squares and four squares.

1. Introduction

An old and interesting problem in number theory concerns the representations of a given positive integers as the sum of a fixed number of integral squares. Precisely speaking, given a fixed positive integer k, we would like to determine for which positive integers n there exist integers $x_1, ..., x_k$ such that

$$n = \sum_{i=1}^{k} x_i^2.$$

The case $k = 1$ is trivial. Fermat considered the case $k = 2$ for primes $n = p$. He showed that a prime congruent to 3 modulo 4 cannot be written as the sum of two squares. He also observed that a prime congruent to 1 modulo 4 is a sum of two squares. The generalization to all positive integers n was already described in 1625 by Girard before Fermat. Neither Girard nor Fermat provided proofs of their observations. The first proof was discovered by Euler in the mid 1700s. The case $k = 4$ was resolved by Lagrange in 1770. He proved that every positive integer can be represented as the sum of four squares. Lagrange's proof is based on Euler's early attempts and is completely elementary. During 1797-1798 Legendre proved that a positive integer n can be written as the sum of three squares precisely when n is not of the form $4^a(8b+7)$ for some non-negative integers a and b. His proof is based on the arithmetical theory of binary and ternary quadratic forms. In 1801 Gauss generalized Legendre's result by determining the number of representations of an integer as the sum of three squares. It is now known that the general case can be explained nicely by the theory of modular forms. Nowadays we have many different proofs of the results of Fermat, Lagrange and Legendre that make use of ideas and tools from various branches of mathematics, such as algebraic number theory, complex analysis (especially, the theory of modular forms), Diophantine approximation, and geometry of numbers. In this note, we shall discuss interesting proofs using ideas from geometry of numbers, a subject invented by Hermann Minkowski.

2. Minkowski's First Theorem

Recall that an n-dimensional lattice $\Lambda \subseteq \mathbb{R}^n$ is an additive subgroup of $\mathbb{R}^n$ which is spanned by an $\mathbb{R}$-basis of $\mathbb{R}^n$ over $\mathbb{Z}$. It is not hard to show that two n-dimensional lattices are the same if and only if one can be obtained from the other by a unimodular transformation (i.e., a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ with $[T] \in M_{n\times n}(\mathbb{Z})$ and $\det[T] = \pm 1$). Suppose that

Λ is spanned by $v_1, ..., v_n$. The fundamental parallelotope spanned by $v_1, ..., v_n$ is defined by

$$\mathcal{P} := \left\{ \sum_{i=1}^{n} a_i v_i \colon a_1, ..., a_n \in [0,1] \right\}.$$

If we think of $v_1, ..., v_n$ as row vectors in $\mathbb{R}^n$, then the volume V of $\mathcal{P}$ is easily seen to be $V = |\det(v_1^T, ..., v_n^T)|$. This quantity is clearly invariant under change of basis (via a unimodular transformation). We denote it by $\mathrm{vol}(\mathbb{R}^n/\Lambda)$. We shall also speak of "bodies" in $\mathbb{R}^n$. A body $B \subseteq \mathbb{R}^n$ in the broad sense is a Lebesgue measurable subset of $\mathbb{R}^n$. Its volume $\mathrm{vol}(B)$ is defined to be the same as its Lebesgue measure. It is called convex if for any two points $x, y \in B$, one has $\lambda x + (1-\lambda)y \in B$ for all $\lambda \in [0,1]$. It is said to be centrally symmetric if $x \in B$ implies $-x \in B$. Our main tool is the following theorem of Minkowski.

Theorem 2.1 (Minkowski's First Theorem). *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $B \subseteq \mathbb{R}^n$ be a convex centrally symmetric body. If $\mathrm{vol}(B) > 2^n \mathrm{vol}(\mathbb{R}^n/\Lambda)$, then B contains a point $x \in \Lambda \setminus \{0\}$.*

Proof. Let $\mathcal{P}$ be a fundamental parallelotope of Λ. Note that $\mathbb{R}^n$ is covered by the non-overlapping family $\{x + \mathcal{P}\}_{x \in \Lambda}$ of copies of $\mathcal{P}$. It follows that $(1/2)B$ is covered by the non-overlapping family $\{(1/2)B \cap (x + \mathcal{P})\}_{x \in \Lambda}$. Since $\mathrm{vol}((1/2)B) = \mathrm{vol}(B)/2^n$, we have

$$\mathrm{vol}(\mathbb{R}^n/\Lambda) < \mathrm{vol}\left(\frac{1}{2}B\right) = \sum_{x \in \Lambda} \mathrm{vol}\left(\frac{1}{2}B \cap (x + \mathcal{P})\right) = \sum_{x \in \Lambda} \mathrm{vol}\left(\left(\frac{1}{2}B - x\right) \cap \mathcal{P}\right),$$

where we have used the fact that the Lebesgue measure is translation-invariant. If the elements of $\{(1/2)B - x\}_{x \in \Lambda}$ were pairwise disjoint, then

$$\sum_{x \in \Lambda} \mathrm{vol}\left(\left(\frac{1}{2}B - x\right) \cap \mathcal{P}\right) = \mathrm{vol}\left(\mathcal{P} \cap \bigcup_{x \in \Lambda}\left(\frac{1}{2}B - x\right)\right) \leq \mathrm{vol}(\mathcal{P}) = \mathrm{vol}(\mathbb{R}^n/\Lambda),$$

a contradiction. Hence there must exist two distinct points $x, y \in \Lambda$ such that $((1/2)B - x) \cap ((1/2)B - y) \neq \emptyset$. Let $z \in ((1/2)B - x) \cap ((1/2)B - y)$. Then we have $2(z+x) \in B$ and $2(z+y) \in B$. Since B is centrally symmetric, we have $-2(z+y) \in B$. By the convexity of B, we find that

$$x - y = \frac{1}{2}[2(z+x) - 2(z+y)] \in B.$$

Since $x - y \in \Lambda \setminus \{0\}$, we conclude that $x - y \in B \cap (\Lambda \setminus \{0\})$. □

The above proof is borrowed from [5]. For alternative proofs, see [4, §3.10], for instance.

3. Sums of Two Squares

We start with Fermat's theorem that every prime $p \equiv 1 \pmod 4$ can be expressed as the sum of two squares.

Theorem 3.1 (Fermat's Theorem). *Every prime $p \equiv 1 \pmod 4$ can be expressed as the sum of two squares.*

Proof. Since $p \equiv 1 \pmod 4$, there exists $a \in \mathbb{Z}$ for which $a^2 + 1 \equiv 0 \pmod p$. Let us consider the lattice

$$\Lambda := \{(x, ax + py) \in \mathbb{Z}^2 \colon x, y \in \mathbb{Z}\} \tag{1}$$

with basis $v_1 = (1, a)$ and $v_2 = (0, p)$. Then $\operatorname{vol}(\mathbb{R}^2/\Lambda) = p$. Let $D \subseteq \mathbb{R}^2$ be the open disk of radius $\sqrt{2p}$ centered at the origin. Then $\operatorname{vol}(D) = 2\pi p > 4\operatorname{vol}(\mathbb{R}^2/\Lambda)$. By Theorem 2.1 we see that D contains a non-zero point $(x, ax + py) \in \Lambda$. Since

$$x^2 + (ax + py)^2 \equiv (1 + a^2)x^2 \equiv 0 \pmod p$$

and $0 < x^2 + (ax + py)^2 < 2p$, we must have $x^2 + (ax + py)^2 = p$. □

Grace [3] provides a geometric proof of Theorem 3.1 without using Minkowski's theorem. Instead, he appeals to the fact from algebra that an additive subgroup of $\mathbb{R}^n$ is a lattice if and only if it is discrete (see [6, Proposition 4.2]). His proof is constructive compared to the one above in that it tells us how to find a pair $(x, y) \in \mathbb{Z}$ for which $p = x^2 + y^2$ for any given prime $p \equiv 1 \pmod 4$. We now present a modified version of Grace's proof of Theorem 3.1 that does not use the fact from algebra mentioned above.

Grace's Proof of Theorem 3.1. Consider the lattice Λ defined as in (1). Let $A = (\xi, \eta) = (x, ax + py) \in \Lambda$ be a non-zero point in Λ such that

$$\|A\| = \sqrt{\xi^2 + \eta^2} = \min_{X \in \Lambda \setminus \{(0,0)\}} \|X\|.$$

Then $A' := (-\eta, \xi)$ is also a point in Λ, since $A' = (x', ax' + py')$ with $x' = -ax - py$ and $y' = ay + (a^2 + 1)x/p$. By construction, we have $\Lambda \cap \triangle OAA' = \{O, A, A'\}$. Let $\mathcal{P}$ be the parallelogram with OA and OA' being its two adjacent sides. Then $\mathcal{P}$ is a fundamental parallelogram of Λ, since $\mathcal{P}$ contains no other points in Λ than its vertices. It is in fact a square with area $p = \xi^2 + \eta^2$. □

Remark 1. We give a concrete example illustrating Grace's method. Let $A = (\xi, \eta) = (x, ax + py) \in \Lambda$ be a non-zero point in Λ and consider

$$\|A\|^2 = x^2 + (ax + py)^2 = (a^2 + 1)x^2 + 2apxy + p^2y^2 = \frac{p}{m}[(mx + ay)^2 + y^2],$$

where $m = (a^2 + 1)/p$. To obtain a representation of p as the sum of two squares, we need only to minimize the quadratic form

$$Q(x, y) := (mx + ay)^2 + y^2$$

for $(x, y) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$. In practice, one can choose $1 \leq a < p/2$ so that

$$1 \leq m \leq \frac{p^2 - 2p + 5}{4p} \leq \frac{p - 1}{4}.$$

For example, let $p = 13$ and take $a = 5$. Then $m = 2$. We need to minimize $Q(x, y) = (2x+5y)^2+y^2$ for $(x, y) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$. Observe that $Q(x, y) \equiv 0 \pmod 2$ for any $(x, y) \in \mathbb{Z}^2$ and that $Q(3, -1) = 2$. Thus the pair $(x, y) = (3, -1)$ generates a representation of 13 as the sum of two squares: $13 = x^2 + (5x + 13y)^2 = 3^2 + 2^2$.

It is now an easy exercise to derive the following result [4, Theorem 366] from Theorem 3.1 by using the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

and the fact that for any prime $p \equiv 3 \pmod 4$, the congruence $x^2 \equiv -1 \pmod p$ is unsolvable.

Theorem 3.2. *A positive integer n can be expressed as the sum of two squares if and only if all primes $p \equiv 3 \pmod 4$ have even exponents in the prime factorization of n.*

4. Sums of Four Squares

Now we prove Lagrange's four-square theorem [4, Theorem 369] using Theorem 2.1. The proof is borrowed from [2]. It is a natural extension of the proof of Theorem 3.1 in the sense that the convex centrally symmetric body we shall consider here is the four dimensional ball. We first prove the following lemma which serves as a substitute for the fact that $(-1/p) = 1$ for all primes $p \equiv 1 \pmod 4$ used in the proof of Theorem 3.1.

Lemma 4.1. *For any odd positive integer n, there exist $a, b \in \mathbb{Z}$ such that $a^2 + b^2 + 1 \equiv 0 \pmod n$.*

Proof. By the Chinese remainder theorem, it suffices to prove our lemma for odd prime powers $n = p^v$. For $v = 1$, consider the subsets

$$\begin{aligned} A &:= \{a^2 \colon 0 \le a < p/2\} \subseteq \mathbb{Z}/p\mathbb{Z}, \\ B &:= \{-b^2 - 1 \colon 0 \le b < p/2\} \subseteq \mathbb{Z}/p\mathbb{Z}. \end{aligned}$$

Then $\#A = \#B = (p+1)/2$. It follows that $\#A + \#B > p$. By the pigeonhole principle there exist $0 \le a, b < p/2$ for which $a^2 \equiv -b^2 - 1 \pmod p$.

Let $v \ge 2$ and suppose that there exist $a_0, b_0 \in \mathbb{Z}$ such that $a_0^2 + b_0^2 + 1 \equiv 0 \pmod{p^{v-1}}$. Without loss of generality, we may assume $p \nmid a_0$. Choose $k \in \mathbb{Z}$ for which

$$2ka_0 \equiv -\frac{1 + a_0^2 + b_0^2}{p^{v-1}} \pmod p.$$

Then we have

$$(a_0 + kp^{v-1})^2 + b_0^2 \equiv a_0^2 + 2ka_0p^{v-1} + b_0^2 \equiv -1 \pmod{p^v}.$$

By induction, for any $v \ge 1$ there exist $a, b \in \mathbb{Z}$ for which $a^2 + b^2 + 1 \equiv 0 \pmod{p^v}$. □

We are now able to prove Lagrange's theorem.

Theorem 4.2 (Lagrange's Four-Square Theorem)**.** *Every positive integer n can be expressed as the sum of four squares.*

Proof. Suppose first that n is odd. By Lemma 4.1 we can find $a, b \in \mathbb{Z}$ such that $a^2 + b^2 + 1 \equiv 0 \pmod n$. Consider the lattice

$$\Lambda := \{(nx_1 + ax_3 + bx_4, nx_2 + bx_3 - ax_4, x_3, x_4) \colon x_1, x_2, x_3, x_4 \in \mathbb{Z}\}$$

with basis $v_1 = (n, 0, 0, 0)$, $v_2 = (0, n, 0, 0)$, $v_3 = (a, b, 1, 0)$ and $v_4 = (b, -a, 0, 1)$. Since

$$\det(v_1^T, v_2^T, v_3^T, v_4^T) = \det \begin{bmatrix} n & 0 & a & b \\ 0 & n & b & -a \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = n^2,$$

we have $\mathrm{vol}(\mathbb{R}^4/\Lambda) = n^2$. Let $B \subseteq \mathbb{R}^4$ be the open ball of radius $\sqrt{2n}$ centered at the origin. We compute by means of the spherical coordinates

$$\mathrm{vol}(B) = (2n)^2 \int_0^\pi \sin^2 \varphi_1 \, d\varphi_1 \int_0^{2\pi} \int_0^\pi \sin \varphi_2 \, d\varphi_2 d\varphi_3 \int_0^1 r^3 \, dr = 2\pi^2 n^2 > 2^4 \, \mathrm{vol}(\mathbb{R}^4/\Lambda).$$

It follows by Theorem 2.1 that B contains a non-zero point $(nx_1 + ax_3 + bx_4, nx_2 + bx_3 - ax_4, x_3, x_4) \in \Lambda$. Since

$$(nx_1 + ax_3 + bx_4)^2 + (nx_2 + bx_3 - ax_4)^2 + x_3^2 + x_4^2 \equiv (a^2 + b^2 + 1)(x_3^2 + x_4^2) \equiv 0 \pmod{n}$$

and

$$0 < (nx_1 + ax_3 + bx_4)^2 + (nx_2 + bx_3 - ax_4)^2 + x_3^2 + x_4^2 < 2n,$$

we conclude that

$$(nx_1 + ax_3 + bx_4)^2 + (nx_2 + bx_3 - ax_4)^2 + x_3^2 + x_4^2 = n.$$

This proves the theorem when n is odd.

Suppose now that n is even. Writing $n = 2^v m$, where m and v are positive integers with $2 \nmid m$, we can find $x_1, x_2, x_3, x_4 \in \mathbb{Z}$ such that $m = x_1^2 + x_2^2 + x_3^2 + x_4^2$. If v is even, then

$$n = (2^{v/2}x_1)^2 + (2^{v/2}x_2)^2 + (2^{v/2}x_3)^2 + (2^{v/2}x_4)^2.$$

If v is odd, then

$$n = [2^{(v-1)/2}(x_1 + x_2)]^2 + [2^{(v-1)/2}(x_1 - x_2)]^2 + [2^{(v-1)/2}(x_3 + x_4)]^2 + [2^{(v-1)/2}(x_3 - x_4)]^2.$$

This completes the proof of the theorem. □

It is worth noting that we need only to appeal to Lemma 4.1 for primes $m = p$ in order to prove Lagrange's theorem if we make use of the following identity of Euler:

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = A^2 + B^2 + C^2 + D^2,$$

where

$$\begin{aligned} A &= x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4, \\ B &= x_1y_2 - x_2y_1 + x_3y_4 + x_4y_3, \\ C &= x_1y_3 - x_3y_1 + x_4y_2 - x_2y_4, \\ D &= x_1y_4 + x_4y_1 + x_2y_3 + x_3y_2. \end{aligned}$$

This identity, which looks uncanny at first glance, can in fact be obtained by considering the the product of the quaternions $X = x_1 + x_2\mathbf{i} + x_3\mathbf{j} + x_4\mathbf{k}$ and $Y = y_1 + y_2\mathbf{i} + y_3\mathbf{j} + y_4\mathbf{k}$ and observing that $N(XY) = N(X)N(Y)$, where

$$N(X) := (x_1 + x_2\mathbf{i} + x_3\mathbf{j} + x_4\mathbf{k})(x_1 - x_2\mathbf{i} - x_3\mathbf{j} - x_4\mathbf{k}) = x_1^2 + x_2^2 + x_3^2 + x_4^2.$$

5. Sums of Three Squares

Finally, we turn to Legendre's three-square theorem. Observe that if $n = 4^a(8b + 7)$ is a positive integer, then n is not representable by a sum of three squares. Indeed, assume that

$$n = x_1^2 + x_2^2 + x_3^2$$

for some $x_1, x_2, x_3 \in \mathbb{Z}$. If $a \geq 1$, then

$$x_1^2 + x_2^2 + x_3^2 \equiv n \equiv 0, 4 \pmod{8}.$$

But $x_i^2 \equiv 0, 1, 4 \pmod{8}$ for each $1 \leq i \leq 3$. It follows that x_1, x_2, x_3 are all even. By induction, we have that x_1, x_2, x_3 are all divisible by 2^a. Thus

$$8b + 7 = (x_1/2^a)^2 + (x_2/2^a)^2 + (x_3/2^a)^2.$$

This is impossible, since the right-hand side cannot be congruent to 7 modulo 8. Hence it is sufficient to show that every positive integer $n \neq 4^a(8b+7)$ can be expressed as the sum of three squares, and it will suffice to prove this for square-free n. To the author's knowledge, there are no easy proofs of this. Here we follow Ankeny [1] to give a geometric proof based on Minkowski's theorem. In the proof below, we shall also make use of the quadratic reciprocity law as well as Dirichlet's theorem on primes in arithmetic progressions which asserts that for any integers k and l with $k \geq 1$ and $\gcd(k,l)=1$, the arithmetic progression $\{kn+l\}_{n=1}^{\infty}$ contains infinitely many primes.

Theorem 5.1 (Legendre's Three-Square Theorem)**.** *Let n be a positive integer which is not of the form $4^a(8b+7)$. Then n can be expressed as the sum of three squares.*

Proof. Throughout the proof, we shall always assume that n is square-free. Then $n \equiv 1,2,3,5,6 \pmod 8$. Suppose that $n = 2^k m$, where $k \in \{0,1\}$ and $2 \nmid m$. By Dirichlet's theorem on primes in arithmetic progressions and the Chinese remainder theorem, there exists a prime $q \equiv 1 \pmod 4$ such that

$$\left(\frac{-2^k}{q}\right) = (-1)^{k(m-1)/2} \tag{2}$$

and

$$\left(\frac{-2^l q}{p}\right) = 1 \tag{3}$$

for all prime factors p of m, where $(\cdot/\cdot)$ is the Jacobi symbol and

$$l = \begin{cases} 1 & \text{if } n \equiv 3 \pmod 8, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\frac{k(m-1)}{2} \equiv \frac{m-1}{2} + l \pmod 2,$$

which implies that

$$\left(\frac{-2^l}{m}\right) = \left(\frac{-1}{m}\right)\left(\frac{2^l}{m}\right) = (-1)^{(m-1)/2+l} = \left(\frac{-2^k}{q}\right).$$

It is easy to see by the law of quadratic reciprocity that

$$\left(\frac{-n}{q}\right) = \left(\frac{-2^k}{q}\right)\prod_{p|m}\left(\frac{p}{q}\right) = \left(\frac{-2^k}{q}\right)\prod_{p|m}\left(\frac{q}{p}\right) = \left(\frac{-2^k}{q}\right)\left(\frac{-2^l}{m}\right)\prod_{p|m}\left(\frac{-2^l q}{p}\right) = 1.$$

Thus there exists $u \in \mathbb{Z}$ for which $u^2 \equiv -n \pmod{4^l q}$. By (3) and the Chinese remainder theorem, there exists $v \in \mathbb{Z}$ for which $v^2 \equiv -\alpha \pmod n$, where $0 < \alpha < n$ is an integer such that $2^l q\alpha \equiv 1 \pmod n$. Put $\beta := \sqrt{2^l q}$ and consider the lattice

$$\Lambda := \{(\beta^2 vx + uvy + nz, \beta x + u\beta^{-1}y, \sqrt{n}\beta^{-1}y) \colon x,y,z \in \mathbb{Z}\}$$

with basis $v_1 = (\beta^2 v, \beta, 0)$, $v_2 = (uv, u\beta^{-1}, \sqrt{n}\beta^{-1})$ and $v_3 = (n,0,0)$. Since

$$\det(v_1^T, v_2^T, v_3^T) = \det\begin{bmatrix} \beta^2 v & uv & n \\ \beta & u\beta^{-1} & 0 \\ 0 & \sqrt{n}\beta^{-1} & 0 \end{bmatrix} = n^{3/2},$$

we have $\operatorname{vol}(\mathbb{R}^4/\Lambda) = n^{3/2}$. Let $B \subseteq \mathbb{R}^3$ be the open ball of radius $\sqrt{2n}$ centered at the origin. Then

$$\operatorname{vol}(B) = \frac{4\pi}{3}(2n)^{3/2} = \frac{8\sqrt{2}\pi}{3}n^{3/2} > 2^3 \operatorname{vol}(\mathbb{R}^4/\Lambda).$$

It follows by Theorem 2.1 that B contains a non-zero point $(X, Y, Z) = (\beta^2 vx + uvy + nz, \beta x + u\beta^{-1}y, \sqrt{n}\beta^{-1}y) \in \Lambda$. Note that $X \in \mathbb{Z}$ and

$$X^2 + Y^2 + Z^2 = (2^l qvx + uvy + nz)^2 + \frac{(2^l qx + uy)^2 + ny^2}{2^l q} \tag{4}$$

with

$$\frac{(2^l qx + uy)^2 + ny^2}{2^l q} = x(2^l qx + 2uy) + \frac{(u^2 + n)y^2}{2^l q} \in \mathbb{Z}.$$

It follows that

$$X^2 + Y^2 + Z^2 \equiv (v^2 + \alpha)(2^l qx + uy)^2 \equiv 0 \pmod{n}.$$

Since $0 < X^2 + Y^2 + Z^2 < 2n$, we conclude that $X^2 + Y^2 + Z^2 = n$.

It is sufficient to show that

$$Y^2 + Z^2 = \frac{(2^l qx + uy)^2 + ny^2}{2^l q} \tag{5}$$

is a sum of two squares. By Theorem 3.2, it suffices to show that all odd prime factors of $Y^2 + Z^2$ with odd exponents are congruent to 1 modulo 4. Let $p > 2$ be a prime factor of $Y^2 + Z^2$ with an odd exponent $t \geq 1$. Suppose first that $p \nmid n$. By $X^2 + Y^2 + Z^2 = n$ we have $(n/p) = 1$. If $p = q$, then $(-n/p) = 1$, since $u^2 \equiv -n \pmod{4^l q}$; if $p \neq q$, then it follows from (5) that $p^t \parallel [(2^l qx + uy)^2 + ny^2]$ and hence $(-n/p) = 1$. In either case, we have $(-1/p) = 1$, which is equivalent to $p \equiv 1 \pmod{4}$. Suppose now that $p \mid n$. Then $p \neq q$, $p \mid (2^l qx + uy)$ and $p \mid X$. It follows from (4) and the equation $X^2 + Y^2 + Z^2 = n$ that

$$(2^l qx + uy)^2 + ny^2 \equiv 2^l qn \pmod{p^2},$$

which implies that $ny^2 \equiv 2^l qn \pmod{p^2}$. Since $p \parallel n$, we obtain $y^2 \equiv 2^l q \pmod{p}$. Thus $(2^l q/p) = 1$. Comparing this with (3) we have $(-1/p) = 1$. Again, this gives $p \equiv 1 \pmod{4}$. This completes the proof. □

Remark 2. Note that the proof breaks down for square-free $n \equiv 7 \pmod{8}$ as expected, because in this case one would have $(-1/n) = (-2/n) = -1$.

References

[1] N. C. Ankeny, *Sums of three squares*, Proc. Am. Math. Soc. **8** (2) (1957), 316–319.

[2] H. Davenport, *The geometry of numbers*, Math. Gaz. **31** (296) (1947), 206–210.

[3] J. H. Grace, *The four square theorem*, J. Lond. Math. Soc. **2** (1927), 3–8.

[4] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th. ed., Oxford Univ. Press, Oxford, 2008. Revised by D. R. Heath-Brown and J. H. Silverman; With a forward by A. J Wiles.

[5] D. A. Marcus, *Number Fields*, 2nd. ed., Universitext, Springer Nature, New York, 2018.

[6] J. Neukirch and N. Schapacher (tran.), *Algebraic Number Theory*, Grundlehren Math. Wiss., vol. 322, Springer-Verlag, Berlin-Heidelberg-New York, 1999.

Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA
Email address: steve.fan.gr@dartmouth.edu