Two squares and four squares: the simplest proof of all?

(Published in *Math. Gazette* **94** (2010), 119–123, with minor variations)

Introduction

Two of the classic theorems of number theory are:

Fermat's two-squares theorem: If p is prime and $p \equiv 1 \bmod 4$, then p is the sum of two squares.

Lagrange's four-squares theorem: Every positive integer is the sum of (at most) four squares.

Fermat stated the two-squares theorem in a letter to his friend Mersenne in 1640. He claimed to have a proof, but did not leave one for posterity. The first known proof was given by Euler in 1754. The four-squares theorem was proved by Lagrange in 1770, and his proof was promptly simplified by Euler (see [4]).

Numerous proofs are now known for both theorems. Three of the best known methods for the two-squares theorem, all completely different, can be seen in [3], chapter 10: they use, respectively, (a) the "method of descent", (b) the Gaussian integers, (c) the geometry of lattices in the plane. Methods similar to (a) and (c) are then described for the four-squares theorem: (a) is essentially Euler's method, and it is the one found in most textbooks.

However, there is a proof of Fermat's theorem that is decidedly simpler and more elegant than any of these. It is by no means new: in fact, the method is attributed to the Norwegian mathematician Axel Thue, in the early 20th century! Despite this, not many books present it: one that does is [1] (Theorem 12–2). A partly similar method appears in [2, p. 300], following Landau, but it differs by requiring a previous result on approximation by rationals. Since the proof is very short and apparently less widely known than it should be, we repeat it here. We then show how the method can be extended, with a little more work, to the four-squares theorem. The result is a proof that is arguably at least a little simpler than those found in any book of which I am aware.

Two squares

Like nearly all methods for the two-squares theorem, our proof uses the following fact:

Lemma 1: If p is prime and congruent to 1 mod 4, then there exists u such that $u^2 \equiv -1 \bmod p$.

This is a standard result of elementary number theory. Again, many proofs are known. My preferred one, since it depends on a minimum of other results, is (briefly) as follows: in the group of residue classes mod p, exactly $\frac{1}{2}(p-1)$ elements are squares, and the squares consist of distinct pairs $(\hat{r}, \hat{r}^{-1})$, together with $\hat{1}$, and $(-\hat{1})$ if it is a square. If $p \equiv 1 \bmod 4$, then $\frac{1}{2}(p-1)$ is even, so $-\hat{1}$ has to be a square.

Proof of the two-squares theorem. Let u be as in Lemma 1. Let $k = [p^{1/2}]$, so that $k < p^{1/2} < k+1$. Consider the numbers $a - bu$ corresponding to pairs (a, b) with $0 \leq a \leq k$, $0 \leq b \leq k$. There are $(k+1)^2$ such pairs. Since $(k+1)^2 > p$, there must be two distinct pairs (a_1, b_1) and (a_2, b_2) whose corresponding numbers lie in the same congruence class mod p, so that

$$a_1 - b_1 u \equiv a_2 - b_2 u \mod p.$$

Write $a = a_1 - a_2$ and $b = b_1 - b_2$. Then a and b are not both zero (since the pairs are distinct), $|a| \leq k$, $|b| \leq k$ and $a \equiv bu \bmod p$. Hence $a^2 \equiv b^2 u^2 \equiv -b^2 \bmod p$, so p divides $a^2 + b^2$. But $0 < a^2 + b^2 \leq 2k^2 < 2p$, so in fact $a^2 + b^2 = p$, as required.

The reasoning did not really use the fact that p is prime, but only the existence of u.

Four squares

The proof of the four-squares theorem consists of two main steps:

Step 1: every *prime* number is expressible as the sum of at most four squares,

Step 2: if m and n are expressible as sums of four squares, then so is mn.

It then follows that every positive integer is so expressible. Step 2 is a heavy, but elementary, piece of algebra, somewhat simplified if quaternions are used ([3], Lemma 10.5). We do not offer anything different for this step. Our aim is to simplify Step 1.

Again we use a lemma common to nearly all proofs of the theorem. It can be proved quite simply, in a similar way to the method indicated for Lemma 1 (e.g. [3, p. 203], [1, section 12.3].

Lemma 2: If p is prime, then there exist u and v such that $u^2 + v^2 \equiv -1 \bmod p$.

Consider the set $\mathbb{Z}(i)$ of Gaussian integers, that is, complex numbers $\alpha = a + ib$ for $a, b \in \mathbb{Z}$. We do not need any of the factorisation theory of $\mathbb{Z}(i)$. We simply say that a positive integer p divides $a+ib$ if it divides a and b, and that $\alpha \equiv \beta \bmod p$ if p divides $\alpha - \beta$. There are clearly p^2 congruence classes, represented by $a+ib$, where a and b are chosen from $\{0, 1, \ldots, p-1\}$.

Suppose that $\alpha = a + ib$, $\beta = c + id$ and $\alpha \equiv \beta \bmod p$, so that p divides $a - c$ and $b - d$. Then $\alpha\bar{\alpha} \equiv \beta\bar{\beta} \bmod p$, since

$$\alpha\bar{\alpha} - \beta\bar{\beta} = (a^2 + b^2) - (c^2 + d^2) = (a + c)(a - c) + (b + d)(b - d).$$

Proof of Step 1. The cases $p = 2$ and $p = 3$ are easy, so we assume that $p > 3$. Let u, v be as in Lemma 2, and let $k = [p^{1/2}]$. Consider the numbers $(a + ib) - (c + id)(u + iv)$ corresponding to quadruples (a, b, c, d) with a, b, c, d chosen from $\{0, 1, \ldots, k\}$. There are $(k + 1)^4$ such quadruples, and $(k + 1)^4 > p^2$, so two of the corresponding numbers must be in the same congruence class mod p: call them $(a_j + ib_j) - (c_j + id_j)(u + iv)$ $(j = 1, 2)$. Let $a = a_1 - a_2$ etc. Then $|a|$, $|b|$, $|c|$, $|d| \leq k$, not all of a, b, c, d are 0 and

$$a + ib \equiv (c + id)(u + iv) \mod p.$$

By the remark preceding the proof,

$$a^2 + b^2 \equiv (c^2 + d^2)(u^2 + v^2) \equiv -c^2 - d^2 \mod p,$$

so p divides S, where $S = a^2 + b^2 + c^2 + d^2$. Now $S \leq 4k^2 < 4p$, so $S = rp$, where r is 1, 2 or 3. If $r = 1$, we have our result. We must deal with the cases $r = 2$ and $r = 3$.

Case $r = 2$. Out of a, b, c, d, either (i) all are even, (ii) all are odd, or (iii) exactly two are even, say a, b. In each case, $a \pm b$ and $c \pm d$ are even, and we have

$$p = [\tfrac{1}{2}(a + b)]^2 + [\tfrac{1}{2}(a - b)]^2 + [\tfrac{1}{2}(c + d)]^2 + [\tfrac{1}{2}(c - d)]^2.$$

Case $r = 3$. We now have $a^2 + b^2 + c^2 + d^2 = 3p$. Squares are congruent to 0 or 1 mod 3. If a^2, b^2, c^2, d^2 are all multiples of 3, then they are multiples of 9. But then p is a multiple of 3, contradicting the fact that it is prime and greater than 3. So three of the four squares, say a^2, b^2, c^2 must be congruent to 1 mod 3, while $d^2 \equiv 0$ mod 3. Then $d \equiv 0$ mod 3 and $a \equiv \pm 1$ mod 3. Replacing a by $-a$ if necessary, we may assume that $a \equiv 1$ mod 3. Similarly for b, c. Let

$$n_1 = a + b + c, \quad n_2 = a - b + d, \quad n_3 = -a + c + d, \quad n_4 = b - c + d.$$

Then each n_j is congruent to 0 mod 3, and it is easily checked that

$$n_1^2 + n_2^2 + n_3^2 + n_4^2 = 3(a^2 + b^2 + c^2 + d^2) = 9p,$$

so that

$$p = \sum_{j=1}^{4} \left(\frac{n_j}{3}\right)^2.$$

This proof, admittedly, has elements in common with the well-known proof by the method of descent (e.g. [1], Theorem 12–6), but at least some simplification has resulted from the restriction of r to the values 2 and 3.

Variant using four-dimensional volume. Could we have avoided the cases $r = 2$ and $r = 3$, as in the proof for two squares? The harder case $r = 3$ can indeed be avoided, at the cost of an appeal to four-dimensional volumes. The volume of a four-dimensional sphere of radius R is $(\pi^2/2)R^4$. This leads to the following four-dimensional analogue of Gauss's "circle theorem" [2, p. 270].

Lemma 3: The number of integer lattice points (a, b, c, d) with $a^2 + b^2 + c^2 + d^2 \leq R^2$ is between $(\pi^2/2)(R-1)^4$ and $(\pi^2/2)(R+1)^4$.

Proof: Denote this set of points by S, and its number of members by $|S|$. Let $E(a, b, c, d)$ be the four-dimensional cube with side 1 and centre (a, b, c, d), and let $E(S)$ be the union of these cubes for all (a, b, c, d) in S. Each cube has volume 1, so $|S|$ is clearly equal to the volume of $E(S)$. All points of a unit cube are within distance δ from its centre, where $\delta^2 \leq 4 \times \frac{1}{4} = 1$, hence $\delta \leq 1$. Now take any point (x_1, x_2, x_3, x_4) of the sphere with centre the origin and radius $R - 1$. This point lies in some cube $E(a, b, c, d)$, and by the triangle inequality and the preceding remark, $a^2 + b^2 + c^2 + d^2 \leq R^2$, so (a, b, c, d) is in S. In other words, the sphere of radius $R-1$ is contained in $E(S)$. The lower inequality follows. Similar reasoning gives the upper inequality (but we don't need it.)

Alternative proof of Step 1. With u, v as in Lemma 2, consider the numbers $(a+ib) - (c+id)(u+iv)$ corresponding to quadruples (a, b, c, d) with $a^2+b^2+c^2+d^2 \leq R^2$, where R is to be chosen. We need the number of quadruples to be greater than p^2, hence $\pi^2(R-1)^4 > 2p^2$, or

$$R - 1 > \frac{2^{1/4}}{\sqrt{\pi}}\sqrt{p}. \tag{1}$$

As before, we choose two points in the same congruence class, and let $a = a_1 - a_2$, etc. Then $a^2 \leq 2(a_1^2 + a_2^2)$, so $a^2 + b^2 + c^2 + d^2 \leq 4R^2$. To exclude the case $r = 3$, we require $4R^2 < 3p$. For there to exist R satisfying both this condition and (1), we need

$$\left(\frac{\sqrt{3}}{2} - \frac{2^{1/4}}{\sqrt{\pi}}\right)\sqrt{p} \approx 0.1951\sqrt{p} > 1,$$

which is satified for all $p > 27$. We have to finish by checking directly that primes smaller than 27 are sums of at most four squares, which of course is easy.

References

1. D.M. Burton, *Elementary Number Theory* (3rd edition), William C. Brown (1994).

2. G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford (1938) (revised edition 1979).

3. G.A. Jones and J.M. Jones, *Elementary Number Theory*, Springer (1998).

4. P. Shiu, Euler's contribution to number theory, *Math. Gazette* **91** (2007), 453–461.

G.J.O. JAMESON

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF

e-mail: g.jameson@lancaster.ac.uk