FERMAT'S TWO SQUARES THEOREM

Pierre de Fermat (1601–1665) was a French lawyer. He has been described as the greatest amateur mathematician of all time, for his contributions to optics, probability, and, most notably, number theory. Perhaps he is best known for "Fermat's Last Theorem" — the (still unproved) assertion that $x^n + y^n = z^n$ has no solution in positive integers x, y, z for any $n \geq 3$. First year undergraduates encounter another (genuine!) theorem of Fermat's, that $x^p \equiv x (\text{mod } p)$ for any integer x and any prime p, as a consequence of Lagrange's Theorem for finite groups.

Fermat's Two Squares Theorem is the following:

If $p \equiv 1(\text{mod } 4)$ *is prime,*
then p *is a sum of two squares*

This result is remarkable in that it relates primes — objects whose definition only involves multiplication and division — to the *additive* structure of the integers. As examples of the theorem we have $5 = 1 + 4, 13 = 4 + 9, 17 = 1 + 16$, etc. Exercise: show that if $p \equiv 3(\text{mod } 4)$ then p cannot be the sum of two squares (consider the remainder when a square is divide by 4).

More than 50 different proofs of the theorem have been published. Undergraduates may encounter two proofs themselves: one using the unique factorization property of $\mathbb{Z}[\sqrt{-1}]$ and another, in the Elementary Number Theory course, using Dirichlet's Approximation Theorem. The vast majority of published proofs, and indeed the two proofs just mentioned, have much in common. In particular, they depend on the following fact:

If $p \equiv 1(\text{mod } 4)$ is prime, there exists an integer x for which $x^2 + 1 \equiv 0 (\text{mod } p)$ — for example $x = (\frac{(p-1)}{2})!$.

I shall describe a new and completely different proof, using group actions on sets. Let

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 2 & -1 \end{pmatrix}$$

and

$$M = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

One easily checks that $A^2 = B^2 = C^2 = I$, and that $A^T MA = B^T MB = C^T MC = M$. In particular A^{-1}, B^{-1}, and C^{-1} exist, whence the linear mappings produced by A, B, C are one-to-one.

Define

$$S = \{\mathbf{v} = (x, y, z) \in \mathbb{Z}^3 : \mathbf{v}^T M \mathbf{v} = p \text{ and } x, y > 0\}$$

(we think of $\mathbf{v}$ as a column vector) and let

$$T = \{(x, y, z) \in S : z > 0\}, \quad U = \{(x, y, z) \in S : x + z > y\}.$$

Note that $\mathbf{v}^T M \mathbf{v} = p$ merely means that $4xy + z^2 = p$. It follows that S is a finite set, since $x, y > 0$. We shall need to know that A maps S to itself, B maps T to itself, and C maps U to itself. We'll only look at the last case, the others being, if anything, easier. If $\mathbf{v} = (x, y, z) \in S$ with $x + z > y$, then $C\mathbf{v} = (x - y + z, y, 2y - z) = (x', y', z')$, say. Hence $x' > 0$, because $x + z > y$; $y' > 0$, because $y > 0$; and $x' + z' > y'$, because $x > 0$. Moreover

$$(C\mathbf{v})^T M (C\mathbf{v}) = \mathbf{v}^T (C^T M C) \mathbf{v} = \mathbf{v}^T M \mathbf{v} = p,$$

so C maps U to itself.

Next we show that S is the disjoint union of T and AT, and also of U and AU. Again, we shall just check the latter assertion. If $(x, y, z) \in S$ then either $x + z > y$ (whence $(x, y, z) \in U$) or $x + z = y$, or $x + z < y$. The case $x + z = y$ cannot arise, since $\mathbf{v}^T M \mathbf{v} = p$ implies $p = 4xy + z^2 = 4x(x + z) + z^2 = (2x + z)^2$, contradicting the primality of p. We shall show that if

$$U' = \{(x, y, z) \in S : x + z < y\}$$

then $U' = AU$; this gives the required result.

If $\mathbf{v} = (x, y, z) \in U$ then $\mathbf{v} \in S$, whence $A\mathbf{v} \in AS = S$. Moreover $A\mathbf{v} = (y, x, -z) = (x', y', z')$, say, with $x' + z' = y - z < x = y'$ whence $A\mathbf{v} \in U'$. So $AU \subseteq U'$. Similarly $AU' \subseteq U$, whence $U' = A^2 U \subseteq AU$. Thus $U' = AU$.

We have now reached the kernel of the proof. Since A is 1-1 we have $\#T = \#AT, \#U = \#AU$. Moreover, as S is the disjoint union of T and AT we have $\#S = \#T + \#AT = 2\#T$. Similarly $\#S = 2\#U$, whence $\#T = \#U$.

Since $C^2 = I$ the action of C on U produces orbits of length 1 or 2. If (x, y, z) is a fixed point of C then $x - y + z = x, y = y, 2y - z = z$, whence $y = z$, and since $4xy + z^2 = p$ one has $p = 4xy + y^2 = y(4x + y)$. Using the facts that p is prime and that $p \equiv 1 (\text{mod } 4)$ we see that this happens if and only if $y = 1$ and $x = (p - 1)/4$. Consequently C has exactly one fixed point in its action on U. Since all the other orbits have length 2 we deduce that $\#U$ is odd.

We now argue similarly with the action of B on T. Since $\#T (= \#U)$ is odd it follows that B must have an odd number of fixed points on U. So

there is at least one fixed point. However, a fixed point of B must have $x = y$, and so $p = 4xy + z^2$ will have a solution in which $x = y$. It follows that $p = (2x)^2 + z^2$ as required.

D. R. Heath-Brown

Appendix — January 2008

Invariant was an occasional publication of the Invariant Society (the Oxford University undergraduate mathematics society). Since the original was not typeset electronically, I have now now re-set the article with LaTeX, correcting a few misprints.

The history of the argument given here is perhaps of interest. I was led to it from a study of the account of Liouville's papers on identities for parity functions, presented in the book by Uspensky and Heaslet [1]. My original notes date from 1971. I gave a splinter-group talk on the argument at the British Mathematical Colloquium in 1980 (or 1979?), after which it seems to have spread by word of mouth. It subsequently became an exercise for trainee teachers in France (Varouchas [2]). Further interest was generated by Zagier's single sentence version of the proof [3].

References

[1] J.V. Uspensky and M.A. Heaslet, *Elementary Number Theory*, (McGraw-Hill Book Company, Inc., New York, 1939).

[2] I. Varouchas, Une démonstration élementaire du théorème des deux carrés, *I.R.E.M., Bull.,* No. 6 (1984), 31–39.

[3] D. Zagier, A one-sentence proof that every prime $p \equiv 1 (\mathrm{mod}\ 4)$ is a sum of two squares, *Amer. Math. Monthly*, 97 (1990), 144.

Mathematical Institute,
24–29, St. Giles',
Oxford
OX1 3LB
UK

rhb@maths.ox.ac.uk