SUMS OF TWO SQUARES AND LATTICES

KEITH CONRAD

One of the basic results of elementary number theory is Fermat's two-square theorem.

Theorem 1 (Fermat, 1640)**.** *An odd prime p is a sum of two squares if and only if $p \equiv 1 \bmod 4$. Furthermore, a representation of a prime as a^2+b^2 in $\mathbf{Z}$ is unique up to the order and signs of a and b.*

For example, $5 = 1+4 = 1^2+2^2$, and the only way we can write 5 as a^2+b^2 is by letting a and b be 1 and 2 up to order and sign ($5 = (-1)^2+2^2 = (-2)^2+(-1)^2$, *etc.*).

An odd prime that is a sum of two squares has to be $1 \bmod 4$ since the only squares mod 4 are 0 and 1, so they can't sum to $3 \bmod 4$. To prove, conversely, that any prime $p \equiv 1 \bmod 4$ is a sum of two squares, there are several methods available: descent [6, Chap. 26] (this was Fermat's own approach, according to [7, p. 67]), factorization of p in the Gaussian integers [2, p. 120], Jacobi sums [2, p. 95], the pigeonhole principle [1, pp. 264–265], continued fractions [5, pp. 132–133], quadratic forms [3, pp. 163–164], and Minkowski's convex body theorem [3, pp. 454–455]. One of the virtues of the proof using Gaussian integers is that, thanks to unique factorization in $\mathbf{Z}[i]$, along with existence of a sum of two squares representation for p we obtain the uniqueness of this representation (up to order and sign of the terms being squares) from the uniqueness in unique factorization. This uniqueness can also be proved using simple congruence and divisibility arguments [1, pp. 265–266].

The question that motivated the present note is whether or not there is a proof of the uniqueness part of Theorem 1 using lattice methods, in the spirit of Minkowski's proof of the existence part of Theorem 1. We will give such a proof, as suggested by D. Clausen. Let p be an odd prime and assume $p = a^2+b^2$ for some integers a and b. We want to show this is the only representation of p as a sum of two squares up to the order and signs of a and b.

Since $a^2+b^2 \equiv 0 \bmod p$, both a and b are nonzero modulo p, so dividing the congruence by b shows there is a solution to $k^2+1 \equiv 0 \bmod p$. For x and y in $\mathbf{Z}$, $x^2+y^2 \equiv 0 \bmod p$ if and only if $y^2 \equiv -x^2 \equiv (kx)^2 \bmod p$, which is equivalent to $y \equiv \pm kx \bmod p$. Set

$$L = \{(x,y) \in \mathbf{Z}^2 : y \equiv kx \bmod p\} = \mathbf{Z}\begin{pmatrix}1\\k\end{pmatrix} + \mathbf{Z}\begin{pmatrix}0\\p\end{pmatrix},$$

which is a lattice in $\mathbf{R}^2$ with fundamental parallelogram having area $|\det(\begin{smallmatrix}1 & 0\\ k & p\end{smallmatrix})| = p$. (This lattice appears in the existence part of the proof of Theorem 1 when using Minkowski's theorem.) Let C be the circle $\{(x,y) \in \mathbf{R}^2 : x^2+y^2 = p\}$. The uniqueness in Theorem 1 amounts to saying C contains only 8 integral points:

$$(1) \qquad (a,b), \quad (a,-b), \quad (-a,b), \quad (-a,-b), \quad (b,a), \quad (b,-a), \quad (-b,a), \quad (-b,-a).$$

None of these points can be equal, since otherwise $b = \pm a$ and then $a^2+b^2 = 2a^2$ is even, but $a^2+b^2 = p$ is an odd prime.

For each integral point (x, y) on C, one of (x, y) or $(x, -y)$ is in the lattice L since $x^2 + y^2 = p \Rightarrow x^2 + y^2 \equiv 0 \bmod p \Rightarrow y \equiv \pm kx \bmod p$, and the points (x, y) and $(x, -y)$ can't both be in L since then $-y \equiv y \bmod p$, which implies $y \equiv 0 \bmod p$ (because p is odd), but that contradicts $x^2 + y^2 = p$. Therefore the total number of integral solutions to $x^2 + y^2 = p$ is $2|C \cap L|$, so showing C contains exactly 8 integral points is the same as showing $|C \cap L| = 4$.

When $p = a^2 + b^2$ with integers a and b, we have $b \equiv ka \bmod p$ or $b \equiv -ka \bmod p$. Change the sign on b if necessary to make $\boxed{b \equiv ka \bmod p}$, so $(a, b) \in C \cap L$. From $b \equiv ka \bmod p$ we get $a \equiv k(-b) \bmod p$ by multiplying both sides of the congruence by k, so four integral points in $C \cap L$ are (a, b), $(-a, -b)$, $(-b, a)$, and $(b, -a)$. (The other four points on C in (1) are not in L, but are in the alternate lattice $L' = \{(x, y) \in \mathbf{R}^2 : y \equiv -kx \bmod p\} = \mathbf{Z}\binom{1}{-k} + \mathbf{Z}\binom{0}{p}$.) If there is any additional integral point on C then we will get 4 such points by swapping coordinates and signs, and none of these points will be ones we had before (why?), so $|C \cap L|$ is a multiple of 4.

We will now count $|C \cap L|$ in a second way, using areas. Construct the convex polygon having as its vertices the points in $C \cap L$. This polygon is contained in and on the circle C, so its area is less than the area of C, which is πp. The area of the polygon can be described by an exact formula in terms of $|C \cap L|$ using Pick's theorem:

Theorem 2 (G. Pick, 1899)**.** *Let Λ be a lattice in $\mathbf{R}^2$ and let Π be a polygon with vertices on Λ. If Π is convex then the area of Π is $(I + B/2 - 1)\Delta$ where I is the number of points in Λ that are in the interior of Π, B is the number of points in Λ on the boundary of Π, and Δ is the area of a fundamental parallelogram for Λ.*

Pick's theorem is often stated for polygons with vertices on the standard integral lattice $\mathbf{Z}^2$, but the formulation of the theorem for more general lattices will be convenient for us. This more general case can be reduced by linear algebra to the standard lattice case of $\mathbf{Z}^2$. A proof of Pick's theorem is in [4].

The only point of L inside C is the origin since (from the definition of L) every point in L has squared distance from the origin $x^2 + y^2$ equal to a multiple of p, so for the convex polygon with vertex set $C \cap L$ we have $I = 1$. Since $B = |C \cap L|$ and $\Delta = p$, the area of the convex polygon is $(1 + |C \cap L|/2 - 1)p = |C \cap L|p/2$ by Pick's theorem. An upper bound on the area of the polygon is πp, so $|C \cap L|p/2 < \pi p$, and thus $|C \cap L| < 2\pi \approx 6.28$. The number $|C \cap L|$ is a multiple of 4, so $|C \cap L|$ must be 4 and that's what we wanted to show.

References

[1] D. M. Burton, "Elementary Number Theory," 6th ed., McGrawHill, New York, 2007.
[2] K. Ireland and M. Rosen, "A Classical Introduction to Modern Number Theory," 2nd ed., Springer-Verlag, New York, 1990.
[3] J. R. Goldman, "The Queen of Mathematics: A Historically Motivated Guide to Number Theory," A.K. Peters, Natick, MA, 2004.
[4] I. Niven and H. Zuckerman, *Lattice points and polygonal area*, Amer. Math. Monthly **74** (1967), 1195-1200.
[5] C. D. Olds, "Continued Fractions," Math. Assoc. America, Washington, D.C., 1963.
[6] J. H. Silverman, "A Friendly Introduction to Number Theory," 3rd ed., Prentice Hall, Upper Saddle River, NJ, 2006.
[7] A. Weil, "Number Theory: An Approach Through History from Hammurapi to Legendre," Birkhäuser, Boston, 1984.