Every prime of the form $4k+1$ is the sum of two perfect squares

Geoff Smith ©1998

This proof is shamelessly culled from the excellent book, The Theory of Numbers by G. H. Hardy and E. M. Wright. My copy is battered 4th edition, published by Oxford University Press. The relevant theorems are 82, 86 and the third proof of Theorem 366 in section 20.4

We will assume familiarity with the integers modulo $p$ where $p$ is a prime number, as it will be throughout the discussion. Moreover, $p$ will leave remainder 1 on division by 4.

Lemma 1 Let $p$ be prime of the form $4k+1$. It follows that there is an integer $x$ with $x^2 \equiv -1 \mod p$.

Proof. Wilson’s theorem asserts that for any prime $p$ we have $(p-1)! \equiv -1 \mod p$. To prove this you need to know a little about the integers modulo $p$, and we invite you to chase up what you do not know. The important point is that if $1 \leq a \leq p-1$, then there is a unique $b$ with $1 \leq b \leq p-1$ such that $ab \equiv 1 \mod p$. In the product $(p-1)!$ the $p-1$ cancel in pairs (forming 1’s) except for 1 and $p-1$ which are their own multiplicative inverses $(p-1)^2 \equiv 1^1 \equiv 1 \mod p$. To see this, observe that if $1 \leq c \leq p-1$ and $c^2 - 1 \equiv 0$, then $p$ divides $(c-1)(c+1)$. Since $p$ is prime, this forces $c$ to be 1 or $p-1$.

We conclude that $(p-1)! \equiv -1 \mod p$. Now if $(p+1)/2 \leq x \leq p-1$ if and only if $-(p-1)/2 \leq x - p \leq -1$. Thus

$$(p-1)! \equiv (-1)((p-1)/2!2^2 \cdot ((p-1)/2)^2 \equiv ((p-1)/2)!^2 \equiv -1 \mod p.$$ 

Here we have used the fact that $(p-1)/2$ is an even number. The proof of the lemma is complete.

We conclude that $-1$ is a square modulo $p$.

Thus there is an integer $m$ such that $mp = x^2 + 1 = x^2 + y^2$ so $mp$ is the sum of two squares. In fact, of course, $m$ will have to be a natural number, and since we may take $x < p$ we have $0 < m < p$. Now we assume that $m_0$ is the least value of $m$ so that $m_0p$ is the sum of two squares. We want to show $m_0 = 1$. 

1
Theorem 1 It follows that $m_0 = 1$.

Proof. If not, then $1 < m_0 < p$ and $m_0p = x^2 + y^2$ for some integers $x$ and $y$. If $p$ divided both $x$ and $y$ it would follow that $m_0p$ was divisible by $p^2$ which is not possible by the restrictions on $m_0$. Thus we can choose integers $c, d$ so that $x_1 = x - cm_0, y_1 = y - dm_0$, $|x_1| \leq m_0/2, |y_1| \leq m_0/2$ and $x_1^2 + y_1^2 > 0$.

What we have done is to choose $cm_0$ to be a nearest multiple of $m_0$ to $x$, and $dm_0$ to be a nearest multiple of $m_0$ to $y$. Notice that

$$0 < x_1^2 + y_1^2 \leq 2(m_0/2)^2 < m_0^2$$

and

$$x_1^2 + y_1^2 \equiv x^2 + y^2 \equiv 0 \mod m_0.$$

Thus $x_1^2 + y_1^2 = m_1m_0$ where $0 < m_1 < m_0$. Multiplying we obtain

$$m_0^2m_1p = (x^2 + y^2)(x_1^2 + y_1^2) = (xx_1 + yy_1)^2 + (xy_1 - x_1y)^2.$$

Now $xx_1 + yy_1 = x(x - cm_0) + y(y - dm_0) = m_0X$ and $xy_1 - x_1y = x(y - dm_0) - y(x - cm_0) = m_0Y$ where $X = px - cy - dx$ and $Y = cy - dx$. Thus $m_1p = X^2 + Y^2$ and $0 < m_1 < m_0$. The method of descent forces $m_0$ to be 1.