Primes as the sum of two squares

Last lecture we investigated the Gaussian integers $\mathbb{Z}[i]$. A main result was that this ring is a UFD, which means that it has a version of the Fundamental Theorem of Arithmetic, where every element in the ring can be factored in an essentially unique way.

In this lecture we continue this discussion. We begin by taking a step back and asking: why do we care about factorization in general rings? One reason is that it helps us understand concrete arithmetic problems in simpler rings like the integers. We illustrate this point with Fermat's "Sum of two squares" theorem, which is a statement about the integers, but whose proof will use factorization in the Gaussian integers.

We ask the following question: which prime numbers p can be written

$$p = a^2 + b^2$$

where a and b are integers? For example, we have $2 = 1^2 + 1^2$. It is not hard to see that 3 cannot be written as $a^2 + b^2$. On the other hand, we have

$$\begin{aligned} 5 &= 1^2 + 2^2 \\ 13 &= 2^2 + 3^2 \\ 17 &= 1^2 + 4^2 \end{aligned}$$

whereas 7 and 11 cannot be written as sums of two squares. After writing out enough examples you may recognize the pattern that the odd primes p that can be written as the sum of two squares satisfy $p \equiv 1 \pmod 4$.

To see why this is indeed true, suppose p is odd and $p = a^2 + b^2$. Then we consider

$$p = a^2 + b^2 \pmod 4$$

Now $a \pmod 4$ and $b \pmod 4$ are elements in $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Noting that in $\mathbb{Z}_4$ we have $0^2 = 0$, $1^2 = 1$, $2^2 = 0$, $3^2 = 1$, we must have $a^2 \pmod 4$ and $b^2 \pmod 4$ each equal to either 0 or 1 in $\mathbb{Z}_4$. Since the sum $p \mod 4$ is odd, we must have one of them equal to 0, say $a^2 \pmod 4$, and the other equal to 1, say $b^2 \pmod 4$. But then

$$p = a^2 + b^2 \equiv 0 + 1 \equiv 1 \pmod 4$$

Thus we have shown that if an odd prime p can be written as $p = a^2 + b^2$ then $p \equiv 1 \pmod 4$. This explains why $3, 7, 11, \ldots$ cannot be written as a sum of two squares. More surprising is that the converse is true, i.e. if $p \equiv 1 \pmod 4$ then p can be written as a sum of two squares.

▸ **(Fermat) A prime number $p \neq 2$ is the sum of two integer squares, i.e. $p = a^2 + b^2$ where $a, b \in \mathbb{Z}$, if and only if $p \equiv 1 \pmod 4$.**

Let us first make the connection to Gaussian integers. For $x = a + bi \in \mathbb{Z}[i]$ recall that we have the norm (or rather norm squared) function defined by

$$N(x) = |x|^2 = a^2 + b^2$$

This function is multiplicative in the sense that for all $x, y \in \mathbb{Z}[i]$ we have

$$N(xy) = N(x)N(y)$$

This is just a recasting of the property that $|xy| = |x||y|$ for the usual complex norm.

For any $x \in \mathbb{Z}[i]$ note that $N(x)$ is a non-negative integer. Furthermore, x is a unit if and only if $N(x) = 1$. Indeed, you have already classified the units in this ring: $\mathbb{Z}[i]^\times = \{1, -1, i, -i\}$.

Next, suppose $p = a^2 + b^2$. In the ring $\mathbb{Z}[i]$ this can be factored as follows:

$$p = a^2 + b^2 = (a + bi)(a - bi)$$

and each factor is not a unit. In other words, if $p = a^2 + b^2$ then p is no longer "prime", i.e. irreducible, in the ring $\mathbb{Z}[i]$. In fact, suppose $p \in \mathbb{Z}$ is not irreducible in $\mathbb{Z}[i]$. Then $p = xy$ for some non-units $x, y \in \mathbb{Z}[i]$, because of factorization in the ring $\mathbb{Z}[i]$. Then

$$p^2 = N(p) = N(xy) = N(x)N(y)$$

Since x, y are non-units, $N(x), N(y) > 1$ and hence $N(x) = N(y) = p$. Writing $x = a + bi$ we then have $p = N(x) = a^2 + b^2$. In other words, we have shown: if a prime $p \in \mathbb{Z}$ is *not irreducible* in the ring $\mathbb{Z}[i]$ then $p = a^2 + b^2$ for some integers $a, b \in \mathbb{Z}$.

Therefore, to complete the proof of Fermat's sum of two squares theorem we must show that primes $p \equiv 1 \pmod 4$ are not irreducible in $\mathbb{Z}[i]$.

To prove this last statement we will use the following lemma.

▸ **If a prime p satisfies $p \equiv 1 \pmod 4$ then there is some $m \in \mathbb{Z}$ such that**

$$m^2 + 1 \equiv 0 \pmod p$$

Proof. Recall that $\mathbb{Z}_p$ for p prime has the structure of a field. Note that $(p-1)! \pmod p$ is the product of all non-zero elements in $\mathbb{Z}_p$. For $x \neq \pm 1$ in $\mathbb{Z}_p^\times$ the multiplicative inverse of x is not equal to x. Thus apart from $\pm 1 \bmod p$, elements in $\mathbb{Z}_p^\times$ pair off as inverses. Thus

$$(p-1)! \equiv 1 \cdot 2 \cdots (p-1) \equiv 1 \cdot (-1) \equiv -1 \pmod p$$

This holds for any prime p. (This is sometimes called Wilson's Theorem.) Next suppose $p \equiv 1 \pmod 4$ and write $p = 1 + 4n$. Then we compute

$$\begin{aligned} -1 \equiv (p-1)! &\equiv (1 \cdot 2 \cdots (2n)) \left((p-1) \cdot (p-2) \cdots (p-2n)\right) \\ &\equiv ((2n)!) \left((-1)^{2n} (2n)!\right) \equiv ((2n)!)^2 \pmod p \end{aligned}$$

Now take $m = (2n)!$. Then $m^2 + 1 \equiv 0 \pmod p$, and we are done. □

Now we return to our main goal and show that if a prime number $p \in \mathbb{Z}$ satisfies $p \equiv 1 \pmod 4$ then p is not irreducible in $\mathbb{Z}[i]$.

From the lemma, we have some integer $m \in \mathbb{Z}$ such that p divides $m^2 + 1$. Then in the ring $\mathbb{Z}[i]$ we have that p divides the product of Gaussian integers

$$m^2 + 1 = (m + i)(m - i)$$

Thus we have an expression of the form

$$p \cdot x = (m + i)(m - i) = p_1 \cdots p_k \cdot q_1 \cdots q_l$$

where we have written out factorizations of $(m + i)$ and $(m - i)$ into irreducibles. Now, if p is irreducible, it is some p_i or some q_i up to multiplication by a unit, by unique factorization in $\mathbb{Z}[i]$. This implies p divides one of $m + i$ or $m - i$. However, $m/p \pm i/p$ are not Gaussian integers, so we have a contradiction. Thus we have proved that p is not irreducible. This completes the proof of Fermat's theorem on which primes are the sum of two squares.

With just a bit more work, the following more precise version of a "Fundamental Theorem of Arithmetic" for the Gaussian integers $\mathbb{Z}[i]$ can be proved.

▶ **The irreducible elements in $\mathbb{Z}[i]$ up to mulitplication by units are the following:**

- $1 + i$
- $a + bi$ **where** $a, b \in \mathbb{Z}$ **with** $a > |b| > 0$ **and** $a^2 + b^2 = p$ **a prime, with** $p \equiv 1 \pmod 4$
- $p \in \mathbb{Z}_{>0}$ **prime with** $p \equiv 3 \pmod 4$

Thus every element in $\mathbb{Z}[i]$ can be written uniquely as a product of the above elements and a unit, which is one of $1, -1, i, -i$.