Primes which are sums of two squares

November 27, 2022

A lot of the material in this handout comes from Keith Conrad's notes which can be found here:

`https://kconrad.math.uconn.edu/blurbs/ugradnumthy/Zinotes.pdf`

Those notes are also much more in-depth and cover more applications. So if you find this stuff interesting, check those out for more!

(1) The goal of this handout is to prove the following theorem of Girard (often attributed to Fermat):

Theorem 1.1. *A prime number p can be written as a sum of two squares if and only if $p = 2$ or $p \equiv 1 \bmod 4$.*

We will prove this theorem by studying a new number system called the *Gaussian integers*. But before we do, let's get acquainted with the above theorem.

Exercise 1.2. Write each of the primes

$$2, 5, 13, 17, 29$$

as a sum of two squares.

Exercise 1.3. Prove directly that 11 cannot be written as a sum of two squares.

Exercise 1.4. In this exercise you will prove one of the directions of the main theorem.

(a) List the possible values of x^2 modulo 4.

(b) List the possible values of $x^2 + y^2$ modulo 4.

(c) Prove that if p is a prime and $p = x^2 + y^2$ for some integers x and y, then $p \equiv 1 \bmod 4$ or $p = 2$.

(2) In order to prove the theorem on primes which are the sum of two squares, we will need a preliminary, mod p version of the result.

Lemma 2.5 (Lagrange)**.** *If $p \equiv 1 \bmod 4$, then there is some integer m such that $m^2 + 1$ is divisible by p.*

Before proving the lemma, let's play with it a bit.

Example 2.6. Let's try it with $p = 5$. Then $2^2 + 1 = 5$ so $m = 2$ works. It's not the only solution, we could also use $m = 3$, since $3^2 + 1 = 10$ is divisible by 5.

Exercise 2.7. For each of the primes $p = 13, 17$, and 29, find an integer m so that $m^2 + 1 \equiv 0 \bmod p$.

Now let's prove the lemma.

Proof of Lagrange's lemma. We are interested in finding solutions to the equation

$$x^2 + 1 \equiv 0 \bmod p$$

By Fermat's little theorem, we already know that every number satisfies

$$x^{p-1} - 1 \equiv 0 \bmod p$$

If p is odd (and most primes are!), then we can factor the left hand side as:

$$(x^{(p-1)/2} - 1)(x^{(p-1/2)} + 1)$$

So every number x either makes the first term vanish or the second, modulo p. A polynomial can only have as many roots as its degree[1], so not every one of the $p - 1$ congruence classes can be a root of the first factor above.

Therefore there is some integer n so that

$$n^{(p-1)/2} + 1 \equiv 0 \bmod p$$

If $p = 4k + 1$, then let $m = n^k$. Then

$$m^2 + 1 = n^{2k} + 1 = n^{(p-1)/2} + 1 \equiv 0 \bmod p$$

This completes the proof. □

Problem 2.8. In the course of the proof, we used the following fact: If

$$q(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0$$

is a monic[2] degree d polynomial with integer coefficients, and $d > 0$, then there are at most d congruence classes of integers r such that

$$q(r) \equiv 0 \bmod p.$$

In this exercise you will prove that this is true using induction on d.

[1]This is a somewhat subtle point- how do you prove this for polynomial equations *modulo* p?

[2]That just means the coefficient of the highest power of x is 1.

(a) First prove the base case, when $d = 1$: if $q(x) = x + b$, prove that there is exactly one congruence class of solutions to $q(x) \equiv 0 \bmod p$. (Hint: This is not hard.)

(b) Before moving on to the inductive step, you will have to prove the following. Suppose $q(x)$ is a monic, degree d polynomial and r is an integer such that

$$q(r) \equiv 0 \bmod p$$

Find a monic, degree $(d-1)$ polynomial $f(x)$ such that

$$q(x) \equiv (x - r)f(x) \bmod p,$$

by which we mean that the coefficients are equivalent modulo p. This step is the key one; it might be worth trying some examples to get a feel for what's going on. Also, remember: when working modulo p, you can *divide* by any number not divisible by p.

(c) Show that if

$$q(x) \equiv f(x)g(x) \bmod p$$

then any root of $q(x)$ modulo p must be a root of $f(x)$ or a root of $g(x)$ modulo p. (Do you see why it's important that p is prime here?)

(d) Use the previous two parts to complete the inductive step.

(3) Okay, now it's time to introduce the star of the show.

Definition 3.9. A **Gaussian integer** is a number of the form

$$a + bi$$

where $a, b \in \mathbb{Z}$ and i is a fixed square root of -1, that is:

$$i^2 = -1$$

The number a is called the **real part** of $a + bi$ and the number b is called the **imaginary part**. The set of Gaussian integers is denoted $\mathbb{Z}[i]$.

Exercise 3.10. Prove that $a + bi = c + di$ if and only if $a = c$ and $b = d$. Hint: One way to do this is to solve for i and get a contradiction.

We can add Gaussian integers by adding the real and imaginary parts separately:

$$(a + bi) + (u + vi) = (a + u) + (b + v)i$$

We can multiply by distributing and using the rule that $i^2 = -1$. So:

$$\begin{aligned}(a + bi)(u + vi) &= au + avi + bui + (bi)(vi) \\ &= au + bvi^2 + (av + bu)i \\ &= (au - bv) + (av + bu)i\end{aligned}$$

Exercise 3.11. For each pair of Gaussian integers below, compute their product and their sum.

(a) $1 + 2i$ and $3 - 4i$

(b) $1 + i$ and $1 - i$

(c) $a + bi$ and $a - bi$

(4) Our proof of the main theorem will rest on understanding the arithmetic of Gaussian integers in a similar way to how we understood the arithmetic of the ordinary integers. In particular, we will be interested in understanding primes and prime factorization. The foundation to our understanding of primes in the usual integers is the Euclidean algorithm, which repeatedly used the fact that we can always 'divide with remainder'. That is, if $m \neq 0$ and n is another integer, then we can write

$$n = mq + r$$

for some $0 \leqslant r < |m|$.

In trying to carry this out for the Gaussian integers, we run into an issue: how do we express that some Gaussian integer is 'smaller' than another? For example, which of these two is 'smaller'?

$$2 + 3i, 1 + 4i$$

A first guess might be to look at real or imaginary parts, or maybe their sum. These are all different measurements of relative size, but it turns out the one that's the most convenient is the following:

Definition 4.12. If $a + bi$ is a Gaussian integer then the **norm** of $a + bi$ is defined to be

$$N(a + bi) = a^2 + b^2$$

One reason this is convenient is because of the following properties:

Exercise 4.13. Show that if α and β are Gaussian integers, then $N(\alpha\beta) = N(\alpha)N(\beta)$.

Exercise 4.14. Show that $N(\alpha) \geqslant 0$ always and that $N(\alpha) = 0$ if and only if $\alpha = 0$.

Exercise 4.15. Show that if z has a multiplicative inverse, i.e. if there is a Gaussian integer w with $wz = 1$, then $N(z) = 1$. Prove that the only Gaussian integers of norm 1 are $1, -1, i$, and $-i$.

It will be convenient to have another way to write the norm.

Definition 4.16. If $\alpha = a + bi$ is a Gaussian integer, then the **conjugate** of α is defined to be $a - bi$ and is denoted $\overline{\alpha}$.

Example 4.17. The conjugate of i is $-i$. The conjugate of any ordinary integer n is just n back again. The conjugate of $1+i$ is $1-i$.

Exercise 4.18. Prove that $N(\alpha) = \alpha\overline{\alpha}$.

Exercise 4.19. Prove that the real part of α is $\frac{\alpha+\overline{\alpha}}{2}$ and the imaginary part is $\frac{\alpha-\overline{\alpha}}{2i}$.

(5) Now we'll explain how to perform division with remainder with Gaussian integers. Before we do, we'll need to modify the usual division of integers slightly.

Lemma 5.20 (Modified division algorithm)**.** *If m and n are integers and $m \neq 0$, then there are integers q and r so that*

$$n = mq + r$$

and $|r| \leqslant \frac{1}{2}|m|$.

Proof. First perform the usual division algorithm to write

$$n = mq' + r'$$

where $0 \leqslant r' < |m|$. If it happens to be the case that $r' \leqslant \frac{1}{2}|m|$, then we can stop here.

Otherwise, suppose $r' > \frac{1}{2}|m|$. Then take $q = q' \pm 1$ (depending on whether m is positive or negative) and $r = n - mq$. □

Example 5.21. The usual division algorithm for dividing 11 by 4 tells us that

$$11 = 4 \cdot 2 + 3$$

But 3 is not less than or equal to 2. The modified division algorithm say that

$$11 = 4 \cdot 3 + (-1) = 4 \cdot 3 - 1$$

Notice that $|-1| = 1 \leqslant 2$.

Warning 5.22. There can be more than one quotient-remainder pair for the modified division algorithm. For example, when dividing 6 by 4 we could write

$$6 = 4 \cdot 1 + 2$$

or

$$6 = 4 \cdot 2 - 2$$

So 'the quotient' and 'the remainder' are not uniquely determined any more.

Proposition 5.23. *Let α and β be Gaussian integers with $\beta \neq 0$. Then there exist Gaussian integers γ and ρ such that*

$$\alpha = \beta\gamma + \rho$$

where $N(\rho) \leqslant \frac{1}{2}N(\beta)$.

Proof. We'll give the algorithm and leave it to you (or Keith Conrad's notes) for a proof.

Step 1. Write $\alpha\overline{\beta}$ as $m + ni$.

Step 2. Perform *modified* division to find q_j and r_j with

$$m = N(\beta)q_1 + r_1, \quad n = N(\beta)q_2 + r_2$$

and $|r_j| \leqslant \frac{1}{2}N(\beta)$.

Step 3. Then take $\gamma = q_1 + q_2 i$ and $\rho = \alpha - \beta\gamma$. □

Exercise 5.24. Find quotients and remainders for dividing α by β in each of the following, and check that the norm of the remainder is bounded above by half the norm of β.

(a) $\alpha = 11 + 10i$, $\beta = 4 + i$.

(b) $\alpha = 41 + 24i$, $\beta = 11 - 2i$.

(c) $\alpha = 37 + 2i$, $\beta = 11 + 2i$.

(d) $\alpha = 1 + 8i$, $\beta = 2 - 4i$. (In this case the algorithm from the proof gives two different possible answers, can you find both?)

(6) Now that we have the power of division with remainder, we can develop arithmetic in exactly the same way as with ordinary integers. The one caveat is that, whereas in $\mathbb{Z}$ the only numbers with multiplicative inverses are ± 1, in $\mathbb{Z}[i]$ we have $\pm i$ as well. The numbers ± 1 and $\pm i$ are called the **units** of the Gaussian integers. You'll see them crop up in our definitions:

Definition 6.25. We say that a nonzero Gaussian integer δ is a **divisor** of α if there is a Gaussian integer γ so that

$$\alpha = \gamma\delta,$$

and we write $\delta|\alpha$. We say that δ is a **common divisor** of α and β if

$$\delta|\alpha \text{ and } \delta|\beta.$$

We say that a common divisor δ is **a greatest common divisor** of α and β if, whenever ε is another common divisor of α and β, then

$$N(\delta) \geqslant N(\varepsilon).$$

We say that α and β are **relatively prime** if the only common divisors are units. We say that α is **prime** if the only divisors of α are *unit multiples* of 1 or α.

Warning 6.26. Greatest common divisors are *not unique*! We will see shortly that they are almost unique: any two gcd's differ by multiplication by ± 1 or $\pm i$.

Exercise 6.27. Prove that if α is a Gaussian integer with $N(\alpha)$ prime, then α is prime. Use this to show that

$$1+i, 2+i, 3-4i, 4+i, 2+5i$$

are all prime.

Exercise 6.28. Prove that 2 is not prime in $\mathbb{Z}[i]$ by exhibiting a factorization into two non-unit factors.

Exercise 6.29. Prove that if $p \in \mathbb{Z}$ is an ordinary prime with $p \equiv 3 \bmod 4$, then p remains prime in $\mathbb{Z}[i]$. Hint: Suppose that $p = \alpha\beta$ with $N(\alpha), N(\beta) > 1$. By taking norms of both sides, conclude that p can be written as a sum of two squares and then apply an exercise from the first section.

Here are the main results:

Theorem 6.30 (Euclidean algorithm)**.** *Let $\alpha, \beta \in \mathbb{Z}[i]$ be non-zero. If we recursively apply the division algorithm:*

$$\begin{aligned} \alpha &= \beta\gamma_1 + \rho_1 & N(\rho_1) &< N(\beta) \\ \beta &= \rho_1\gamma_2 + \rho_2 & N(\rho_2) &< N(\rho_1) \\ \rho_1 &= \rho_2\gamma_3 + \rho_3 & N(\rho_3) &< N(\rho_1) \\ &\vdots \end{aligned}$$

then it eventually terminates with a zero-remainder, and the last non-zero remainder δ is a greatest common divisor of α and β. Moreover, any common divisor of α and β is a divisor of δ.

Example 6.31. Compute a GCD for $\alpha = 32 + 9i$ and $\beta = 4 + 11i$.

Exercise 6.32. Compute a GCD for $\alpha = 11 + 3i$ and $\beta = 1 + 8i$.

The following results are all proven in exactly the same way as their counterparts over $\mathbb{Z}$. To check your understanding, see if you can prove all of these results.

Corollary 6.33 (Uniqueness of GCD up to units)**.** *If δ and δ' are greatest common divisors of nonzero Gaussian integers α and β then there is a unit $u \in \{1, -1, i, -i\}$ with $\delta = u\delta'$.*

Corollary 6.34 (Bezout's theorem)**.** *If δ is a greatest common divisor of two non-zero Gaussian integers α and β, then there are some $x, y \in \mathbb{Z}[i]$ with*

$$\alpha x + \beta y = \delta$$

Example 6.35. Find x and y as in Bezout's theorem for $\alpha = 32 + 9i$ and $\beta = 4 + 11i$.

Exercise 6.36. Find x and y as in Bezout's theorem for $\alpha = 11 + 3i$ and $\beta = 1 + 8i$.

Corollary 6.37. *Two non-zero Gaussian integers α and β are relatively prime if and only if we can find $x, y \in \mathbb{Z}[i]$ with*

$$\alpha x + \beta y = 1$$

Corollary 6.38. *Suppose $\alpha|\beta\gamma$ and α and β are relatively prime. Then $\alpha|\gamma$.*

Theorem 6.39 (Fundamental Theorem of Gaussian Arithmetic)**.** *Any $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) > 1$ has a unique factorization into (Gaussian) primes in the sense that, if*

$$\alpha = \pi_1 \cdots \pi_r = \pi_1' \cdots \pi_s'$$

where π_j and π_k' are primes, then $r = s$ and, after reordering, each π_j is a unit multiple of each π_j'.

(7) Now we are ready to prove the main theorem!

Theorem 7.40. *A prime number p can be written as a sum of two squares if and only if $p = 2$ or $p \equiv 1 \bmod 4$.*

Proof. We have already shown that if p can be written as a sum of two squares then $p = 2$ or $p \equiv 1 \bmod 4$. So we need to prove the converse.

If $p = 2$ then certainly $2 = 1^2 + 1^2$. So we need to prove that if p is a prime with $p \equiv 1 \bmod 4$ then p can be written as a sum of two squares. By Lagrange's lemma (Lemma 2.5), there is some integer m such that p divides $m^2 + 1$. Thus:

$$p|(m + i)(m - i).$$

I claim that p is composite in $\mathbb{Z}[i]$. Suppose, for the sake of contradiction, that p is a Gaussian prime. Then p must divide $m + i$ or $m - i$. In other words, there is some $a + bi$ with

$$pa + pbi = m \pm i$$

But then we would have $pb = \pm 1$ by Exercise 3.10, and that can't happen. Therefore we learn that $p = \alpha\beta$ for some non-unit Gaussian integers α and β. Then

$$p^2 = N(\alpha)N(\beta)$$

Since $N(\alpha), N(\beta) > 1$, we conclude from the usual Fundamental Theorem of Arithmetic that $p = N(\alpha) = N(\beta)$. But, if we write $\alpha = x + yi$, then we conclude

$$p = N(x + yi) = x^2 + y^2$$

and we're done! □