SUMS OF TWO SQUARES

WILLIAM GAO

1. Introduction

Which prime numbers p may be written as a sum of two squares? How about natural numbers n?

To answer these questions, we consider the quadratic field

$$\mathbb{Q}[\sqrt{D}] = \{a + b\sqrt{D} : a, b \in \mathbb{Q}\},$$

where D is assumed to be square-free. In particular consider its ring of integers

$$\mathcal{O} = \begin{cases} \mathbb{Z}[\sqrt{D}] & D \equiv 2 \text{ or } 3 \bmod 4, \\ \mathbb{Z}[\frac{1+\sqrt{D}}{2}] & D \equiv 1 \bmod 4. \end{cases}$$

We will be especially interested in the case where $\mathcal{O} = \mathbb{Z}[\sqrt{D}]$. We endow $\mathcal{O}$ with the field norm $N : \mathcal{O} \to \mathbb{Z}$ defined by

$$N(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - b^2 D.$$

It is easy to verify this is multiplicative. A few preparatory remarks follow.

Lemma 1.1. *$\alpha \in \mathcal{O}$ has norm ± 1 if and only if α is a unit.*

Proof. ($\Longleftarrow$) Suppose $\alpha \in \mathcal{O}$ is a unit with inverse α^{-1}. Then

$$N(\alpha)N(\alpha^{-1}) = N(\alpha\alpha^{-1}) = N(1) = 1,$$

so $N(\alpha)$ is a unit in $\mathbb{Z}$, or $N(\alpha) = \pm 1$.

($\Longrightarrow$) If $\alpha = a + b\sqrt{D}$ is such that $N(\alpha) = \pm 1$, let $\overline{\alpha} = a - b\sqrt{D}$. Then

$$\alpha\overline{\alpha} = a^2 - b^2 D = N(\alpha) = \pm 1.$$

Thus $\pm\overline{\alpha}$ is the inverse of α, accordingly. $\square$

Lemma 1.2. *Suppose $\pi \in \mathcal{O}$ is such that $N(\pi) = \pm p$ where $p \in \mathbb{Z}$ is prime. Then π is irreducible in $\mathcal{O}$.*

Proof. If $\pi = \alpha\beta$ for some $\alpha, \beta \in \mathcal{O}$ then

$$p = N(\pi) = N(\alpha\beta) = N(\alpha)N(\beta),$$

which implies one of $N(\alpha)$ and $N(\beta)$ is ± 1 and the other is $\pm p$. Since having norm ± 1 implies the element is a unit, it follows that π is irreducible in $\mathcal{O}$. The case where $N(\pi) = -p$ is identical. $\square$

Lemma 1.3. *Suppose $\pi \in \mathcal{O}$ is prime and let (π) be the prime ideal generated by π in $\mathcal{O}$. Then $(\pi) \cap \mathbb{Z}$ is a prime ideal in $\mathbb{Z}$.*

Proof. If $a, b \in \mathbb{Z}$ are such that $ab \in (\pi) \cap \mathbb{Z}$, then $ab \in \mathbb{Z}$ and $ab \in (\pi)$, so $a \in (\pi)$ or $b \in (\pi)$. This implies a or b is in $(\pi) \cap \mathbb{Z}$. $\square$

We know every ideal in $\mathbb{Z}$ is in the form (p) for some prime $p \in \mathbb{Z}$, so we may write $(\pi) \cap \mathbb{Z} = (p)$.

Now $p \in (\pi)$ implies that π divides p in $\mathcal{O}$, so the prime elements of $\mathcal{O}$ can be determined by the primes in $\mathbb{Z}$ which factor in $\mathcal{O}$.

Lemma 1.4. *Suppose $\pi \mid p$ in $\mathcal{O}$. Then p is either irreducible or a product of irreducibles.*

Proof. If $p = \pi\pi'$, then $N(\pi)N(\pi') = N(p) = p^2$. Thus either $N(\pi) = N(\pi') = \pm p$ so both π, π' are irreducible, or one of $N(\pi), N(\pi')$ is equal to 1, so the corresponding factor is a unit, and thus p is irreducible. $\square$

2. The Gaussian Integers

We now focus on the special case $D = -1$, which yields the Gaussian Integers $\mathbb{Z}[i]$. This is a Euclidean Domain, and thus a Principal Ideal Domain and a Unique Factorization Domain. The units are precisely $\pm 1, \pm i$, and by PID the primes and irreducibles coincide.

The norm becomes $N(a + bi) = a^2 + b^2$, and by our previous work, p factors in $\mathbb{Z}[i]$ into precisely two irreducibles if and only if $p = a^2 + b^2$ is the sum of two integer squares (otherwise, p is irreducible in $\mathbb{Z}[i]$).

Clearly $2 = 1^2 + 1^2 = (1 + i)(1 - i)$ is the sum of two squares. Now assume p is an odd prime. Consider $\mathbb{F}_p^\times$, which is an abelian group of order $p - 1$.

Lemma 2.1. *$-1 \in \mathbb{F}_p^\times$ is the unique element of order 2.*

Proof. If $m^2 \equiv 1 \bmod p$ then p divides $m^2 - 1 = (m-1)(m+1)$. Thus p divides $m-1$ or $m + 1$. In the former case, $m \equiv 1 \bmod p$, and in the latter case $m \equiv -1 \bmod p$, so -1 is the unique element of order 2 in $\mathbb{F}_p^\times$. $\square$

Lemma 2.2. *$p \mid (n^2 + 1)$ for some $n \in \mathbb{Z}$ if and only if $p \equiv 1 \bmod 4$.*

Proof. We first notice that

$$\begin{aligned} p \mid (n^2+1) &\iff n^2 + 1 \equiv 0 \bmod p, \\ &\iff n^2 \equiv -1 \bmod p, \\ &\iff -1 \in \mathbb{F}_p^\times \text{ is a square}, \\ &\iff \mathbb{F}_p^\times \text{ has an element of order 4.} \end{aligned}$$

$(\implies)$ If $\mathbb{F}_p^\times$ has an element of order 4, then by Lagrange, $4 \mid p-1$ so $p \equiv 1 \bmod 4$.

$(\impliedby)$ We show that $\mathbb{F}_p^\times$ has an element of order 4. We observe that $\left|\mathbb{F}_p^\times/\{\pm 1\}\right| = \frac{|\mathbb{F}_p^\times|}{|\{\pm 1\}|} = \frac{p-1}{2}$ is even. Thus $\mathbb{F}_p^\times/\{\pm 1\}$ contains an element $\overline{x}$ of order 2. In $\mathbb{F}_p^\times$, x must have order 2 or 4. But $x \neq -1$; otherwise $\overline{x}$ would be the identity in $\mathbb{F}_p^\times/\{\pm 1\}$, so $x \in \mathbb{F}_p^\times$ has order 4. $\square$

Now if p is prime and $p \equiv 1 \bmod 4$ then there exists $n \in \mathbb{Z}$ such that p divides $n^2 + 1$ in $\mathbb{Z}$. Consequently, p divides $(n + i)(n - i)$ in $\mathbb{Z}[i]$. If p were irreducible in $\mathbb{Z}[i]$ then p would divide either $n + i$ or $n - i$ in $\mathbb{Z}[i]$, and thus would divide their difference $2i$, which is absurd. Thus p is not irreducible in $\mathbb{Z}[i]$, and thus is not prime in $\mathbb{Z}[i]$. We have thereby shown:

Theorem 2.3. *A prime p is the sum of two integers squares, $p = a^2 + b^2$ for $a, b \in \mathbb{Z}$, if and only if $p = 2$ or $p \equiv 1 \bmod 4$.*

3. Summary

More generally, we observe that the product of two sums of two squares is again a sum of two squares. So given $n = p_1^{k_1} \cdots p_r^{k_r}$ where each $p_i \equiv 1 \bmod 4$, n is a sum of squares. More powerfully, for any $n \in \mathbb{Z}^+$, write the prime factorization as

$$n = 2^a p_1^{k_1} \cdots p_r^{k_r} q_1^{s_1} \cdots q_\ell^{s_\ell},$$

where each $p_i \equiv 1 \bmod 4$ and each $q_i \equiv 3 \bmod 4$. n is a sum of two squares if and only if every s_i is even.