QUADRATIC INTEGERS

KEITH CONRAD

1. Introduction

Does uniqueness of prime factorization in $\mathbf{Z}$ really need a proof? Isn't it just obvious? To show why this should not be accepted without proof, we will describe here number systems generalizing $\mathbf{Z}$ where prime factorization is *not* unique. The prime factorization exists but some numbers can have essentially more than one prime factorization!

Definition 1.1. Let d be an integer that is not a perfect square. We set

$$\mathbf{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbf{Z}\}$$

and call such a set of numbers, for a specified choice of d, a set of *quadratic integers.*

Example 1.2. When $d = -1$, so $\sqrt{d} = i$, these quadratic integers are

$$\mathbf{Z}[i] = \{a + bi : a, b \in \mathbf{Z}\}.$$

These are complex numbers whose real and imaginary parts are integers. Examples include $4 - i$ and $7 + 8i$.

Example 1.3. When $d = 2$, $\mathbf{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbf{Z}\}$. Examples include $3 + \sqrt{2}$ and $1 - 4\sqrt{2}$.

We can add, subtract, and multiply in $\mathbf{Z}[\sqrt{d}]$, and the results are again in $\mathbf{Z}[\sqrt{d}]$:

$$\begin{aligned}
(a + b\sqrt{d}) + (a' + b'\sqrt{d}) &= (a + a') + (b + b')\sqrt{d},\\
(a + b\sqrt{d}) - (a' + b'\sqrt{d}) &= (a - a') + (b - b')\sqrt{d},\\
(a + b\sqrt{d})(a' + b'\sqrt{d}) &= (aa' + dbb') + (ab' + ba')\sqrt{d}.
\end{aligned}$$

For example, in $\mathbf{Z}[\sqrt{5}]$, $(2 + 3\sqrt{5})(4 - \sqrt{5}) = 8 - 2\sqrt{5} + 12\sqrt{5} - 15 = -7 + 10\sqrt{5}$.

2. The Norm on $\mathbf{Z}[\sqrt{d}]$

Before we define primes in $\mathbf{Z}[\sqrt{d}]$ we will explain how to measure the size of a number in $\mathbf{Z}[\sqrt{d}]$. In $\mathbf{Z}$, size is measured by the absolute value. For polynomials in $\mathbf{Q}[T]$ or $\mathbf{R}[T]$, size is measured by the degree regardless of how big or small the coefficients are. In $\mathbf{Z}[\sqrt{d}]$, size will be measured by the absolute value of the norm. What's the norm?

Definition 2.1. For $\alpha = a + b\sqrt{d} \in \mathbf{Z}[\sqrt{d}]$, its *norm* is the product

$$\mathrm{N}(\alpha) = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - db^2.$$

Example 2.2. In $\mathbf{Z}[i]$, $\mathrm{N}(a + bi) = a^2 + b^2$. In $\mathbf{Z}[\sqrt{2}]$, $\mathrm{N}(a + b\sqrt{2}) = a^2 - 2b^2$. In $\mathbf{Z}[\sqrt{-2}]$, $\mathrm{N}(a+b\sqrt{-2}) = a^2+2b^2$. In $\mathbf{Z}[\sqrt{3}]$, $\mathrm{N}(a+b\sqrt{3}) = a^2-3b^2$. In $\mathbf{Z}[\sqrt{-3}]$, $\mathrm{N}(a+b\sqrt{-3}) = a^2+3b^2$.

Quadratic integers may be irrational or not even real, but their norm is always a plain integer, *e.g.*, $\mathrm{N}(7+4\sqrt{2}) = 49 - 2\cdot 16 = 17$ and $\mathrm{N}(1+2\sqrt{5}) = 1 - 5\cdot 4 = -19$. For $m \in \mathbf{Z}$, $\mathrm{N}(m) = m^2$. In particular, $\mathrm{N}(1) = 1$.

Here is the key algebraic property of norms.

Theorem 2.3. *The norm is multiplicative: for α and β in $\mathbf{Z}[\sqrt{d}]$, $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$.*

Proof. Write $\alpha = a + b\sqrt{d}$ and $\beta = a' + b'\sqrt{d}$. Then $\alpha\beta = (aa' + dbb') + (ab' + ba')\sqrt{d}$. We now compute $\mathrm{N}(\alpha)\,\mathrm{N}(\beta)$ and $\mathrm{N}(\alpha\beta)$:

$$\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = (a^2 - db^2)(a'^2 - db'^2) = (aa')^2 - d(ab')^2 - d(ba')^2 + d^2(bb')^2$$

and

$$\begin{aligned}
\mathrm{N}(\alpha\beta) &= (aa' + dbb')^2 - d(ab' + ba')^2 \\
&= (aa')^2 + 2aa'bb'd + (dbb')^2 - d(ab')^2 - 2aa'bb'd - d(ba')^2 \\
&= (aa')^2 + (dbb')^2 - d(ab')^2 - d(ba')^2 \\
&= (aa')^2 + d^2(bb')^2 - d(ab')^2 - d(ba')^2.
\end{aligned}$$

The two results agree, so $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$. □

When $d > 0$, $\mathrm{N}(a + b\sqrt{d}) = a^2 - db^2$ might be negative (*e.g.*, $\mathrm{N}(\sqrt{2}) = -2 < 0$). When $d < 0$, so $-d > 0$, $\mathrm{N}(a+b\sqrt{d}) = a^2 - db^2$ is never negative (*e.g.*, $\mathrm{N}(a+b\sqrt{-2}) = a^2+2b^2 \geq 0$). Since a notion of size should be be ≥ 0 and norms might be negative (if $d > 0$), we will use $|\,\mathrm{N}(\alpha)|$ rather than $\mathrm{N}(\alpha)$ as the measure of how "big" a quadratic integer $\alpha \in \mathbf{Z}[\sqrt{d}]$ is.

Example 2.4. In $\mathbf{Z}[\sqrt{2}]$, check $\mathrm{N}(7 + 6\sqrt{2}) = -23$ and $\mathrm{N}(11 + 7\sqrt{2}) = 23$, so $7 + 6\sqrt{2}$ and $11 + 7\sqrt{2}$ both have absolute norm 23. This is analogous to two different polynomials having the same degree.

Remark 2.5. Unlike polynomials, for which there are examples of degree n for all $n \geq 1$, not every positive integer is the absolute norm of a quadratic integer in $\mathbf{Z}[\sqrt{d}]$. For example, in $\mathbf{Z}[i]$ we have $\mathrm{N}(a + bi) = a^2 + b^2$, so while $1 = \mathrm{N}(1)$ and $2 = \mathrm{N}(1 + i)$, there is nothing in $\mathbf{Z}[i]$ with norm 3. There are also no numbers in $\mathbf{Z}[i]$ with norm 6, 7, or 11.

3. Primes and prime factorization in $\mathbf{Z}[\sqrt{d}]$

To define prime elements in $\mathbf{Z}[\sqrt{d}]$, which should have only "trivial factors," we want to define what the trivial factors of a quadratic integer are. This would be analogous to the trivial factors of an integer n being ± 1 and $\pm n$.

One source of trivial factors are the invertible numbers in $\mathbf{Z}[\sqrt{d}]$, also called the *units* of $\mathbf{Z}[\sqrt{d}]$: if $uv = 1$ in $\mathbf{Z}[\sqrt{d}]$, so u and v are inverses of each other, then for every $\alpha \in \mathbf{Z}[\sqrt{d}]$ we have $\alpha = u(v\alpha)$, so every unit in $\mathbf{Z}[\sqrt{d}]$ is a factor of α. Also $\alpha = (u\alpha)v$, so every unit multiple of α is a factor of α.

Example 3.1. In $\mathbf{Z}[\sqrt{3}]$, $2 + \sqrt{3}$ is a unit since $(2 + \sqrt{3})(2 - \sqrt{3}) = 1$, so for every α in $\mathbf{Z}[\sqrt{3}]$ we have $\alpha = (2 + \sqrt{3})((2 - \sqrt{3})\alpha)$: all numbers in $\mathbf{Z}[\sqrt{3}]$ are divisible by $2 + \sqrt{3}$.

Definition 3.2. For nonzero $\alpha \in \mathbf{Z}[\sqrt{d}]$, we call α *prime* if α is not a unit and its only factors are units and unit multiples of α.

We call α *composite* if it is not a unit and not prime: α has a factor other than a unit or a unit multiple of α.

Theorem 3.3. *Let α be nonzero in $\mathbf{Z}[\sqrt{d}]$.*
(1) *α is a unit if and only if $|\operatorname{N}(\alpha)| = 1$.*
(2) *α is composite if and only if there is a factorization $\alpha = \beta\gamma$ where $|\operatorname{N}(\beta)| < |\operatorname{N}(\alpha)|$ and $|\operatorname{N}(\gamma)| < |\operatorname{N}(\alpha)|$.*

The first property is saying units are the nonzero elements of smallest possible absolute norm. The second property is saying that, in terms of size (the absolute norm), a number in $\mathbf{Z}[\sqrt{d}]$ is composite precisely when it has a factorization into two parts that both have smaller size than the original number.

Proof. Set $\alpha = a + b\sqrt{d}$, where a and b are in $\mathbf{Z}$. Then $|\operatorname{N}(\alpha)| = 1 \iff \operatorname{N}(\alpha) = \pm 1$.

(1) First suppose $\operatorname{N}(\alpha) = \pm 1$. Then $(a+b\sqrt{d})(a-b\sqrt{d}) = \pm 1$. If $(a+b\sqrt{d})(a-b\sqrt{d}) = 1$ then $a + b\sqrt{d}$ has inverse $a - b\sqrt{d}$. If $(a+b\sqrt{d})(a-b\sqrt{d}) = -1$ then $a+b\sqrt{d}$ has inverse $-(a - b\sqrt{d})$.

For the converse direction, suppose $\alpha \in \mathbf{Z}[\sqrt{d}]$ is invertible, say $\alpha\beta = 1$ for some β in $\mathbf{Z}[\sqrt{d}]$. Taking the norm of both sides of the equation $\alpha\beta = 1$, we find $\operatorname{N}(\alpha)\operatorname{N}(\beta) = 1$. This is an equation in $\mathbf{Z}$, so $\operatorname{N}(\alpha) = \pm 1$.

(2) Suppose α is composite, so there is a factor β of α that is not a unit or a unit multiple of α. Let γ be the complementary factor of β in α, so $\alpha = \beta\gamma$. Since β is not a unit, $|\operatorname{N}(\beta)| > 1$. If γ were a unit then $\beta = \alpha\gamma^{-1}$, so β would be a unit multiple of α, and that's a contradiction. Thus γ is not a unit in $\mathbf{Z}[\sqrt{d}]$, so $|\operatorname{N}(\gamma)| > 1$. From $|\operatorname{N}(\alpha)| = |\operatorname{N}(\beta)\operatorname{N}(\gamma)| = |\operatorname{N}(\beta)||\operatorname{N}(\gamma)|$ with both $|\operatorname{N}(\beta)|$ and $|\operatorname{N}(\gamma)|$ greater than 1, each is also less than $|\operatorname{N}(\alpha)|$.

Conversely, suppose $\alpha = \beta\gamma$ in $\mathbf{Z}[\sqrt{d}]$ where $|\operatorname{N}(\beta)| < |\operatorname{N}(\alpha)|$ and $|\operatorname{N}(\gamma)| < |\operatorname{N}(\alpha)|$. We have $|\operatorname{N}(\alpha)| = |\operatorname{N}(\beta)||\operatorname{N}(\gamma)|$, so if β were a unit we'd have $|\operatorname{N}(\alpha)| = |\operatorname{N}(\gamma)|$, which is not true. Thus β is not a unit. If β were a unit multiple of α, say $\beta = u\alpha$, then $|\operatorname{N}(\beta)| = |\operatorname{N}(u)||\operatorname{N}(\alpha)| = |\operatorname{N}(\alpha)|$, which is not true either. Thus β is a factor of α that is not a unit or a unit multiple of α, so α is composite in $\mathbf{Z}[\sqrt{d}]$. □

Example 3.4. Since $2 + \sqrt{3}$ is a unit in $\mathbf{Z}[\sqrt{3}]$, with inverse $2 - \sqrt{3}$, a trivial factorization of $5 + 2\sqrt{3}$ is

$$5 + 2\sqrt{3} = (2 + \sqrt{3})(4 - \sqrt{3})$$

since the first factor is a unit.

Example 3.5. A non-trivial factorization of 11 in $\mathbf{Z}[\sqrt{3}]$ is $(2\sqrt{3}+1)(2\sqrt{3}-1)$ since both factors have norm -11. How interesting: 11 is prime in $\mathbf{Z}$ but it is composite in $\mathbf{Z}[\sqrt{3}]$.

The following test for primality in $\mathbf{Z}[\sqrt{d}]$, using the norm, provides a way to generate many primes in $\mathbf{Z}[\sqrt{d}]$ if we can recognize primes in $\mathbf{Z}$.

Theorem 3.6. *For $\alpha \in \mathbf{Z}[\sqrt{d}]$, if $|\operatorname{N}(\alpha)|$ is a prime number then α is prime in $\mathbf{Z}[\sqrt{d}]$.*

Proof. Set $p = |\operatorname{N}(\alpha)|$. Since this is not 1, α is not a unit. We will show α is not composite either, and thus α is prime.

Suppose α is composite, so $\alpha = \beta\gamma$ in $\mathbf{Z}[\sqrt{d}]$ where $|\operatorname{N}(\beta)| < |\operatorname{N}(\alpha)|$ and $|\operatorname{N}(\gamma)| < |\operatorname{N}(\alpha)|$. Taking absolute norms of both sides of $\alpha = \beta\gamma$, we have $p = |\operatorname{N}(\beta)||\operatorname{N}(\gamma)|$. This is an equation in the positive integers, and p is a prime number, so either $|\operatorname{N}(\beta)|$ or $|\operatorname{N}(\gamma)|$ is p. That contradicts $|\operatorname{N}(\beta)| < p$ and $|\operatorname{N}(\gamma)| < p$. □

Example 3.7. We saw in Example 2.4 that $7+6\sqrt{2}$ and $11+7\sqrt{2}$ both have absolute norm 23, so they are each prime in $\mathbf{Z}[\sqrt{2}]$. More prime elements of $\mathbf{Z}[\sqrt{2}]$ are $1+3\sqrt{2}$, $1-2\sqrt{2}$, $3+\sqrt{2}$, $-5+\sqrt{2}$, and $5+2\sqrt{2}$ since each of their absolute norms is a prime number.

WARNING. The converse of Theorem 3.6 is *false*: a quadratic integer can be prime without having a prime norm. For instance, it can be shown that 3 is prime in $\mathbf{Z}[i]$ even though its norm is 9 and $3+\sqrt{5}$ is prime in $\mathbf{Z}[\sqrt{5}]$ even though its norm is 4.

Theorem 3.8. *Every* $\alpha \in \mathbf{Z}[\sqrt{d}]$ *with* $|\,\mathrm{N}(\alpha)| > 1$ *is a product of primes in* $\mathbf{Z}[\sqrt{d}]$.

Proof. Use strong induction on $|\,\mathrm{N}(\alpha)|$. This is analogous to the proof by strong induction on the degree that every nonconstant polynomial in $\mathbf{Q}[T]$ or $\mathbf{R}[T]$ is a product of irreducibles. Details are left to the reader. A new phenomenon in $\mathbf{Z}[\sqrt{d}]$ is that not all positive integers are absolute norms; skip over them in the induction. □

Proving a prime factorization exists in $\mathbf{Z}[\sqrt{d}]$ is completely different from actually finding it. For example, in $\mathbf{Z}[\sqrt{5}]$ what is a prime factorization of $7+\sqrt{5}$? It's not clear at all how to find it! We know it exists thanks to Theorem 3.8, but explicitly finding a prime factorization requires more techniques than we have developed here.

Definition 3.9. We say $\mathbf{Z}[\sqrt{d}]$ has *unique factorization* if whenever

$$p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

for prime quadratic integers p_i and q_j in $\mathbf{Z}[\sqrt{d}]$, we have $r = s$ and, after rearranging terms, $p_i = u_i q_i$ for all i, where u_i is a unit of $\mathbf{Z}[\sqrt{d}]$.

Having "uniqueness" of prime factorization in $\mathbf{Z}[\sqrt{d}]$ be about matching different primes up to unit multiples is analogous to matching irreducibles in $\mathbf{Q}[T]$ up to constant multiples.

Example 3.10. The following equation shows $\mathbf{Z}[\sqrt{-3}]$ does *not* have unique factorization:

$$2 \cdot 2 = (1+\sqrt{-3})(1-\sqrt{-3}). \tag{3.1}$$

We will show 2, $1+\sqrt{-3}$, and $1-\sqrt{-3}$ are all prime in $\mathbf{Z}[\sqrt{-3}]$. The numbers 2, $1+\sqrt{-3}$, and $1-\sqrt{-3}$ all have norm 4. If a number in $\mathbf{Z}[\sqrt{-3}]$ with norm 4 is composite, it has a factor with norm 2 (not -2; why?). That means we can solve $x^2+3y^2 = 2$ in integers x and y, which we plainly can't. So every number in $\mathbf{Z}[\sqrt{-3}]$ with norm 4 is prime in $\mathbf{Z}[\sqrt{-3}]$.

The number 2 is not a unit multiple of $1 \pm \sqrt{-3}$ since $(1 \pm \sqrt{-3})/2$ is not in $\mathbf{Z}[\sqrt{-3}]$. Thus (3.1) is an example of nonunique factorization.

Example 3.11. The following equation shows $\mathbf{Z}[\sqrt{5}]$ does *not* have unique factorization:

$$2 \cdot 2 = (\sqrt{5}+1)(\sqrt{5}-1). \tag{3.2}$$

The factors here have absolute norm 4, so if any are composite they have a factor of absolute norm 2. Then we can solve $x^2 - 5y^2 = \pm 2$ for some $x, y \in \mathbf{Z}$, but this is impossible because it reduces modulo 5 to $x^2 \equiv \pm 2 \bmod 5$, which has no solution!

Could 2 be a unit multiple of $\sqrt{5} \pm 1$? No, since the ratio $(\sqrt{5} \pm 1)/2$ is not in $\mathbf{Z}[\sqrt{5}]$. Thus (3.2) is an example of nonunique factorization.

Here are more examples of nonunique prime factorization among quadratic integers:

$$\begin{aligned} 2 \cdot 3 &= (1+\sqrt{-5})(1-\sqrt{-5}) \quad \text{in } \mathbf{Z}[\sqrt{-5}], \\ 3 \cdot 3 \cdot 3 \cdot 3 &= (5+2\sqrt{-14})(5-2\sqrt{-14}) \quad \text{in } \mathbf{Z}[\sqrt{-14}]. \end{aligned}$$