QUADRATIC INTEGERS

KEITH CONRAD

1. Introduction

Does uniqueness of prime factorization in $\mathbf{Z}$ really need a proof? To show why it does, we will meet some number systems generalizing $\mathbf{Z}$ where prime factorization is *not* unique.

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. Let $d = -1$, so $\sqrt{d} = i$. The set of quadratic integers in this case is

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

These are complex numbers with real and imaginary parts in $\mathbf{Z}$, like $4 + i$ and $7 - 8i$.

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

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

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

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 and Divisibility 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 $\mathbf{Z}[i]$, $\mathrm{N}(a + bi) = (a + bi)(a - bi) = a^2 + b^2$. When d is ± 2 and ± 3,

$$\mathrm{N}(a + b\sqrt{2}) = (a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2,$$
$$\mathrm{N}(a + b\sqrt{-2}) = (a + b\sqrt{-2})(a - b\sqrt{-2}) = a^2 + 2b^2,$$
$$\mathrm{N}(a + b\sqrt{3}) = (a + b\sqrt{3})(a - b\sqrt{3}) = a^2 - 3b^2,$$
$$\mathrm{N}(a + b\sqrt{-3}) = (a + b\sqrt{-3})(a - b\sqrt{-3}) = a^2 + 3b^2.$$

For all $m \in \mathbf{Z}$, $\mathrm{N}(m) = m^2$. In particular, $\mathrm{N}(1) = 1$.

While quadratic integers may be irrational or not even real, their norms are always plain integers, *e.g.*, $\mathrm{N}(7+4\sqrt{2}) = 49 - 2\cdot 16 = 17$ and $\mathrm{N}(1+2\sqrt{5}) = 1 - 5\cdot 4 = -19$.

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 $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$. □

Remark 2.4. Everything we have done up to this point can be carried out with coefficients in $\mathbf{Q}$ rather than in $\mathbf{Z}$: let $\mathbf{Q}[\sqrt{d}] = \{r + s\sqrt{d} : r, s \in \mathbf{Q}\}$ and set $\mathrm{N}(r + s\sqrt{d}) = r^2 - ds^2$. We have $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$ for all α and β in $\mathbf{Q}[\sqrt{d}]$ by exactly the same calculations as in the proof of Theorem 2.3. We will avoid using $\mathbf{Q}[\sqrt{d}]$ except in Example 2.9.

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

Definition 2.5. For $\alpha = a + b\sqrt{d} \in \mathbf{Z}[\sqrt{d}]$, its *absolute norm* is $|\,\mathrm{N}(\alpha)| = |a^2 - db^2|$.

Example 2.6. 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.

The norm of $a + b\sqrt{d}$ is *always* $a^2 - db^2$, but in concrete cases students sometimes make sign errors in the second term, such as saying $a + b\sqrt{2}$ has norm $a^2 + 2b^2$ instead of the correct $a^2 - 2b^2$. *Find a way to remember the correct norm formula.* For example, know two special cases with d of opposite signs, such as $\mathrm{N}(a + bi)$ and $\mathrm{N}(a + b\sqrt{2})$, which are templates for the cases $d < 0$ and $d > 0$. Or know that when $d > 0$ the norm on $\mathbf{Z}[\sqrt{d}]$ is both positive and negative (and 0), and when $d < 0$ the norm on $\mathbf{Z}[\sqrt{d}]$ is never negative.

Remark 2.7. 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 $2 = \mathrm{N}(1 + i)$, nothing in $\mathbf{Z}[i]$ has norm 3. There are also no numbers in $\mathbf{Z}[i]$ with norm 6, 7, or 11.

We define divisibility in $\mathbf{Z}[\sqrt{d}]$ just like in $\mathbf{Z}$:

Definition 2.8. For α and β in $\mathbf{Z}[\sqrt{d}]$, we say β *divides* α, and we write $\beta \mid \alpha$, when $\alpha = \beta\gamma$ for some $\gamma \in \mathbf{Z}[\sqrt{d}]$.

Example 2.9. Does $4+5i$ divide $14+3i$? We can do the division by taking a ratio and rationalizing the denominator:

$$\frac{14+3i}{4+5i} = \frac{(14+3i)(4-5i)}{(4+5i)(4-5i)} = \frac{71-58i}{41} = \frac{71}{41} - \frac{58}{41}i \in \mathbf{Q}[i].$$

This ratio is not in $\mathbf{Z}[i]$: its real and imaginary parts are $71/41$ and $-58/41$, which are not in $\mathbf{Z}$. Thus $(4+5i) \nmid (14+3i)$.

Example 2.10. While $(4+5i) \mid (14+3i)$ by the previous example, $(4-5i) \mid (14+3i)$ in $\mathbf{Z}[i]$ since $14+3i = (4-5i)(1+2i)$.

Theorem 2.11. *In $\mathbf{Z}[\sqrt{d}]$, $\alpha = a+b\sqrt{d}$ is divisible by an ordinary integer c if and only if $c \mid a$ and $c \mid b$ in $\mathbf{Z}$.*

Proof. To say $c \mid (a+b\sqrt{d})$ in $\mathbf{Z}[\sqrt{d}]$ is the same as $a+b\sqrt{d} = c(m+n\sqrt{d}) = cm+cn\sqrt{d}$ for some $m, n \in \mathbf{Z}$, and that is equivalent to $a = cm$ and $b = cn$, or $c \mid a$ and $c \mid b$. □

Example 2.12. In $\mathbf{Z}[i]$, $2 \mid (2+2i)$ since $2+2i = 2(1+i)$ or since the real and imaginary parts of $2+2i$ are both even.

However in $\mathbf{Z}[2i] = \{a+b\cdot 2i : a, b \in \mathbf{Z}\}$ we have $2 \nmid (2+2i)$ because we can't write $2+2i = 2(a+b\cdot 2i) = 2a+4bi$ with $a, b \in \mathbf{Z}$ or because $(2+2i)/2 = 1+i \notin \mathbf{Z}[2i]$. Since $(2+2i)^2 = 8i = 4(2i) = 2^2(2i)$, we have $2^2 \mid (2+2i)^2$ in $\mathbf{Z}[2i]$. Marvel at that: when $\alpha = 2$ and $\beta = 2+2i$, we have $\alpha \nmid \beta$ while $\alpha^2 \mid \beta^2$ in $\mathbf{Z}[2i]$. That never happens in $\mathbf{Z}$, where $m \mid n$ and $m^2 \mid n^2$ are equivalent properties.

Taking $b = 0$ in Theorem 2.11 tells us divisibility between ordinary integers does not change when working in $\mathbf{Z}[\sqrt{d}]$: for $a, c \in \mathbf{Z}$, $c \mid a$ in $\mathbf{Z}[\sqrt{d}]$ if and only if $c \mid a$ in $\mathbf{Z}$.

The multiplicativity of the norm turns divisibility relations in $\mathbf{Z}[\sqrt{d}]$ into divisibility relations in the more familiar setting of $\mathbf{Z}$, as follows.

Theorem 2.13. *For α, β in $\mathbf{Z}[\sqrt{d}]$, if $\beta \mid \alpha$ in $\mathbf{Z}[\sqrt{d}]$ then $\mathrm{N}(\beta) \mid \mathrm{N}(\alpha)$ in $\mathbf{Z}$.*

Proof. Write $\alpha = \beta\gamma$ for $\gamma \in \mathbf{Z}[\sqrt{d}]$. Taking the norm of both sides, $\mathrm{N}(\alpha) = \mathrm{N}(\beta)\,\mathrm{N}(\gamma)$ by Theorem 2.3. This equation is in $\mathbf{Z}$, so $\mathrm{N}(\beta) \mid \mathrm{N}(\alpha)$ in $\mathbf{Z}$. □

Theorem 2.13 gives us a quick way to show one element of $\mathbf{Z}[\sqrt{d}]$ does *not* divide another: check the corresponding norm divisibility fails. For example, if $(3+7i) \mid (10+3i)$ in $\mathbf{Z}[i]$, then (taking norms), $58 \mid 109$ in $\mathbf{Z}$, but that isn't true. Therefore $(3+7i) \nmid (10+3i)$ in $\mathbf{Z}[i]$. Turning a divisibility problem in $\mathbf{Z}[\sqrt{d}]$ into one in $\mathbf{Z}$ has an obvious appeal, since we are more comfortable with divisibility in $\mathbf{Z}$.

However, Theorem 2.13 only says norm-divisibility in $\mathbf{Z}$ follows from divisibility in $\mathbf{Z}[\sqrt{d}]$. The converse is usually *false*. Consider $\alpha = 14+3i$ and $\beta = 4+5i$. While $\mathrm{N}(\beta) = 41$ and $\mathrm{N}(\alpha) = 205 = 41\cdot 5$, so $\mathrm{N}(\beta) \mid \mathrm{N}(\alpha)$ in $\mathbf{Z}$, we saw in Example 2.9 that $(4+5i) \nmid (14+3i)$.

A foolproof method of checking divisibility in $\mathbf{Z}[\sqrt{d}]$ is testing if the ratio is in $\mathbf{Z}[\sqrt{d}]$ after rationalizing the denominator, as we did in Example 2.9 with $d = -1$.

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

To define primes in $\mathbf{Z}[\sqrt{d}]$, which should have only "trivial factors," we want to define trivial factors. They are analogous to the trivial factors of n in $\mathbf{Z}$ being ± 1 and $\pm n$.

One source of trivial factors are the invertible numbers in $\mathbf{Z}[\sqrt{d}]$: 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 u is a factor of α. Also $\alpha = (u\alpha)v$, so $u\alpha$ is a factor of α.

Example 3.1. In $\mathbf{Z}[\sqrt{3}]$, $2+\sqrt{3}$ is invertible 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. Let $\alpha \in \mathbf{Z}[\sqrt{d}]$ be nonzero.

Call α a *unit* when α has a multiplicative inverse: $\alpha\beta = 1$ for some $\beta \in \mathbf{Z}[\sqrt{d}]$.

Call α *prime* if α is not a unit and its only factors are units and unit multiples of α.

Call α *composite* if it is not a unit and not prime: α has a factor that is not a unit or a unit multiple of α.

Theorem 3.3. *For nonzero α in $\mathbf{Z}[\sqrt{d}]$,*

(1) *α is a unit if and only if $|\,\mathrm{N}(\alpha)| = 1$,*

(2) *α is composite if and only if there is a factorization $\alpha = \beta\gamma$ where $|\,\mathrm{N}(\beta)| < |\,\mathrm{N}(\alpha)|$ and $|\,\mathrm{N}(\gamma)| < |\,\mathrm{N}(\alpha)|$.*

The first property is saying units are exactly the nonzero elements of $\mathbf{Z}[\sqrt{d}]$ with 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 does.

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

(1) First suppose $\mathrm{N}(\alpha) = \pm 1$. 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 $-(a-b\sqrt{d})$. We showed that when $a^2 - db^2 = \pm 1$, $a+b\sqrt{d}$ has inverse $\pm(a-b\sqrt{d})$

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 $\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = 1$. This is an equation in $\mathbf{Z}$, so $\mathrm{N}(\alpha) = \pm 1$.

(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, $|\,\mathrm{N}(\beta)| > 1$. 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 $|\,\mathrm{N}(\gamma)| > 1$. From $|\,\mathrm{N}(\alpha)| = |\,\mathrm{N}(\beta)\,\mathrm{N}(\gamma)| = |\,\mathrm{N}(\beta)||\,\mathrm{N}(\gamma)|$ with both $|\,\mathrm{N}(\beta)|$ and $|\,\mathrm{N}(\gamma)|$ greater than 1, each is also less than $|\,\mathrm{N}(\alpha)|$.

Conversely, suppose $\alpha = \beta\gamma$ in $\mathbf{Z}[\sqrt{d}]$ where $|\,\mathrm{N}(\beta)| < |\,\mathrm{N}(\alpha)|$ and $|\,\mathrm{N}(\gamma)| < |\,\mathrm{N}(\alpha)|$. We have $|\,\mathrm{N}(\alpha)| = |\,\mathrm{N}(\beta)\,\mathrm{N}(\gamma)| = |\,\mathrm{N}(\beta)||\,\mathrm{N}(\gamma)|$, so if β were a unit then $|\,\mathrm{N}(\alpha)| = |\,\mathrm{N}(\gamma)|$, which is not true. Thus β is not a unit. If β were a unit multiple of α, say $\beta = u\alpha$, then $|\,\mathrm{N}(\beta)| = |\,\mathrm{N}(u\alpha)| = |\,\mathrm{N}(u)||\,\mathrm{N}(\alpha)| = |\,\mathrm{N}(\alpha)|$, 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. In $\mathbf{Z}[\sqrt{2}]$, $1+\sqrt{2}$ and $3+2\sqrt{2}$ are units with inverses $-1+\sqrt{2}$ and $3-2\sqrt{2}$. In $\mathbf{Z}[i]$, 5 is composite since $5 = (1+2i)(1-2i)$ and $1 \pm 2i$ are not units since their norms are bigger than 1. That's interesting: 5 is prime in $\mathbf{Z}$ but it is composite in $\mathbf{Z}[i]$.

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

Proof. Set $p = |\,\mathrm{N}(\alpha)|$. Since this is not 1, α is not a unit by Theorem 3.3(1). We will show α is not composite either, and thus α is prime.

Suppose α is composite, so by Theorem 3.3(2) $\alpha = \beta\gamma$ in $\mathbf{Z}[\sqrt{d}]$ where $|\,\mathrm{N}(\beta)| < |\,\mathrm{N}(\alpha)|$ and $|\,\mathrm{N}(\gamma)| < |\,\mathrm{N}(\alpha)|$. Taking absolute norms of both sides of $\alpha = \beta\gamma$, we have $p =$

$|\,\mathrm{N}(\beta)||\,\mathrm{N}(\gamma)|$. This is an equation in the positive integers, and p is a prime number, so either $|\,\mathrm{N}(\beta)|$ or $|\,\mathrm{N}(\gamma)|$ is p. That contradicts $|\,\mathrm{N}(\beta)| < p$ and $|\,\mathrm{N}(\gamma)| < p$. □

Example 3.6. In $\mathbf{Z}[i]$, $1+i$ is prime since its norm is 2. Similarly, $1+2i$, $2+3i$, and $1+4i$ are all prime since their norms are 5, 13, and 17.

Example 3.7. We saw in Example 2.6 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 their norms are -17, -7, 7, 23, and 17, which in absolute value are all prime numbers.

WARNING. The converse of Theorem 3.5 is *false*: a quadratic integer can be prime without having a prime absolute norm. For instance, it can be shown that 3 is prime in $\mathbf{Z}[i]$ although its norm is 9 and $2+\sqrt{10}$ is prime in $\mathbf{Z}[\sqrt{10}]$ although its absolute norm is 6.

Theorem 3.8. *Every $\alpha \in \mathbf{Z}[\sqrt{d}]$ that is not 0 or a unit, meaning $|\,\mathrm{N}(\alpha)| > 1$, 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 (Remark 2.7); 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 in general requires techniques we have not developed here.

That prime factorization exists in $\mathbf{Z}[\sqrt{d}]$ does not mean its elements always have essentially just one prime factorization. Sometimes there can be more than one! What do we mean by two prime factorizations in $\mathbf{Z}[\sqrt{d}]$ being essentially the same or not?

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 $\mathbf{Z}[\sqrt{d}]$.

This is saying that changing the order of the terms in a prime factorization and multiplying the terms in a prime factorization by units are considered to be keeping the prime factorization "essentially the same", and $\mathbf{Z}[\sqrt{d}]$ has unique factorization when all the prime factorizations of an element are essentially the same. This is analogous to the way two irreducible factorizations of a polynomial in $\mathbf{Q}[T]$ are the same up to the order of multiplication and up to multiplication by nonzero constants (the units in $\mathbf{Q}[T]$).

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) shows 4 has nonunique prime factorization in $\mathbf{Z}[\sqrt{-3}]$.

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

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

All factors have absolute norm 4, so if any are composite they have a factor of absolute norm 2, say $x + y\sqrt{5}$ where $x, y \in \mathbf{Z}$. Then $|x^2 - 5y^2| = 2$, so $x^2 - 5y^2 = \pm 2$, and that is impossible since it reduces modulo 5 to $x^2 \equiv \pm 2 \bmod 5$, which has no solution! Thus 2, $\sqrt{5}+1$, and $\sqrt{5}-1$ are all prime in $\mathbf{Z}[\sqrt{5}]$.

The number 2 is not a unit multiple of $\sqrt{5} \pm 1$ in $\mathbf{Z}[\sqrt{5}]$ since the ratios $(\sqrt{5} \pm 1)/2$ are not in $\mathbf{Z}[\sqrt{5}]$. Thus (3.2) shows 4 has nonunique prime factorization in $\mathbf{Z}[\sqrt{5}]$.

Example 3.12. Here two more examples of nonunique factorization:

$$2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}) \quad \text{in } \mathbf{Z}[\sqrt{-5}], \tag{3.3}$$

$$2 \cdot 3 = (\sqrt{10}+2)(\sqrt{10}-2) \quad \text{in } \mathbf{Z}[\sqrt{10}]. \tag{3.4}$$

Why are the numbers in these equations prime? Their absolute norms are 4, 9, and 6, so none are units. If they are composite then they are a product $\beta\gamma$ with factors having smaller absolute norm, so β or γ has absolute norm 2 or 3. In $\mathbf{Z}[\sqrt{-5}]$, where norms are not negative, we'd need $x^2 + 5y^2 = 2$ or $x^2 + 5y^2 = 3$, but both are impossible since they are impossible mod 5. Thus $2, 3$, and $1 \pm \sqrt{-5}$ are all prime in $\mathbf{Z}[\sqrt{-5}]$. In $\mathbf{Z}[\sqrt{10}]$ we'd need $x^2 - 10y^2 = \pm 2$ or $x^2 - 10y^2 = \pm 3$, but these are impossible since they are impossible mod 5 (this is the same argument as in the previous example showing $\mathbf{Z}[\sqrt{5}]$ has no element with norm 2 or -2). Thus $2, 3$, and $\sqrt{10} \pm 2$ are prime in $\mathbf{Z}[\sqrt{10}]$.

Now that we know both sides of (3.3) and (3.4) are prime factorizations, we want to show the left and right sides are not essentially the same prime factorization. In (3.3) that means neither factor on the right side is a unit multiple of 2 or 3, and this is true since $(1 \pm \sqrt{-5})/2$ and $(1 \pm \sqrt{-5})/3$ are not in $\mathbf{Z}[\sqrt{-5}]$. In (3.4) neither factor on the right side is a unit multiple of 2 or 3 since $(\sqrt{10} \pm 2)/2$ and $(\sqrt{10} \pm 2)/3$ are not in $\mathbf{Z}[\sqrt{10}]$.

Example 3.13. Here examples where the number of prime factors changes:

$$3 \cdot 3 \cdot 3 = (2+\sqrt{-23})(2-\sqrt{-23}) \quad \text{in } \mathbf{Z}[\sqrt{-23}], \tag{3.5}$$

$$3 \cdot 3 \cdot 3 = (2\sqrt{79}+17)(2\sqrt{79}-17) \quad \text{in } \mathbf{Z}[\sqrt{79}]. \tag{3.6}$$

Norms in $\mathbf{Z}[\sqrt{-23}]$ are not negative and $\mathrm{N}(3) = 9$, so if 3 were composite in $\mathbf{Z}[\sqrt{-23}]$ then it has a factor of norm 3, but $x^2 + 23y^2 = 3$ has no solution in $\mathbf{Z}$. Since $\mathrm{N}(2 \pm \sqrt{-23}) = 27$, if $2 \pm \sqrt{-23}$ were composite in $\mathbf{Z}[\sqrt{-23}]$ then it is $\beta\gamma$ where $\mathrm{N}(\beta) = 3$ and $\mathrm{N}(\gamma) = 9$, but nothing in $\mathbf{Z}[\sqrt{-23}]$ has norm 3. Thus all numbers in (3.5) are prime in $\mathbf{Z}[\sqrt{-23}]$. Since $\mathrm{N}(2\sqrt{79} \pm 17) = -27$ to show the numbers in (3.6) are prime it suffices to show nothing in $\mathbf{Z}[\sqrt{79}]$ has norm ± 3. If $x^2 - 79y^2 = 3$ in $\mathbf{Z}$ then $x^2 \equiv 3 \bmod 79$, which has no solution. That $x^2 - 79y^2 = -3$ has no solution in $\mathbf{Z}$ is more subtle, since $x^2 - 79y^2 \equiv -3 \bmod m$ is always solvable; we omit details. The two sides of (3.5) and (3.6) are prime factorizations with different numbers of prime factors, so $\mathbf{Z}[\sqrt{-23}]$ and $\mathbf{Z}[\sqrt{79}]$ do not have unique factorization.

We've seen examples of $\mathbf{Z}[\sqrt{d}]$ not having unique factorization. Some $\mathbf{Z}[\sqrt{d}]$ that have unique factorization include $\mathbf{Z}[i]$, $\mathbf{Z}[\sqrt{2}]$, $\mathbf{Z}[\sqrt{-2}]$, and $\mathbf{Z}[\sqrt{3}]$ (but not $\mathbf{Z}[\sqrt{-3}]$: see Example 3.10). One way to prove some $\mathbf{Z}[\sqrt{d}]$ has unique factorization is the way unique factorization is proved in $\mathbf{Z}$ and $F[T]$ where F is a field: establish a division algorithm in $\mathbf{Z}[\sqrt{d}]$, which as in $\mathbf{Z}$ and $F[T]$ leads by a chain of reasoning to the unique factorization.

We saw at the start of this section that in $\mathbf{Z}[\sqrt{d}]$ units divide everything: if $uv = 1$ then $\alpha = u(v\alpha)$, so $u \mid \alpha$. Thus any two numbers in $\mathbf{Z}[\sqrt{d}]$ have all units as common factors. Two numbers in $\mathbf{Z}[\sqrt{d}]$ are called *relatively prime* when their only common factors are units (the automatic common factors). For example, primes in $\mathbf{Z}[\sqrt{d}]$ that are not unit multiples of each other are relatively prime in $\mathbf{Z}[\sqrt{d}]$. To appreciate what can happen in $\mathbf{Z}[\sqrt{d}]$ when it does not have unique factorization, consider the following three properties of relatively prime numbers in $\mathbf{Z}$:

- if $(a, b) = 1$ then $ax + by = 1$ for some x and y in $\mathbf{Z}$,
- if $a \mid bc$ and $(a, b) = 1$ then $a \mid c$,
- if $a \mid c$, $b \mid c$, and $(a, b) = 1$, then $ab \mid c$.

When $\mathbf{Z}[\sqrt{d}]$ does not have unique factorization, the analogues of these three properties in $\mathbf{Z}[\sqrt{d}]$ can have counterexamples!

Example 3.14. The numbers 2 and $1+\sqrt{5}$ are primes in $\mathbf{Z}[\sqrt{5}]$ that are not unit multiples of each other, so they are relatively prime. In $\mathbf{Z}[\sqrt{5}]$ we have

- $(2, 1+\sqrt{5}) = 1$ but we can't write $2x+(1+\sqrt{5})y = 1$ for some x and y in $\mathbf{Z}[\sqrt{5}]$: if $2(a+b\sqrt{5})+(1+\sqrt{5})(m+n\sqrt{5}) = 1$ where a, b, m, and n are in $\mathbf{Z}$, then $2a+m+5n = 1$ and $2b + m + n = 0$, which is impossible since the first equation implies $m + n$ is odd while the second equation implies $m + n$ is even,
- $2 \mid (1+\sqrt{5})(1-\sqrt{5}) = -4$ and $(2, 1+\sqrt{5}) = 1$, but $2 \nmid (1-\sqrt{5})$ since $(1-\sqrt{5})/2 \notin \mathbf{Z}[\sqrt{5}]$,
- $2 \mid 4$, $(1+\sqrt{5}) \mid 4$, and $(2, 1+\sqrt{5}) = 1$, but $2(1+\sqrt{5}) \nmid 4$ since $4/(2(1+\sqrt{5})) = (\sqrt{5}-1)/2 \notin \mathbf{Z}[\sqrt{5}]$.

Example 3.15. The numbers 2 and $1+\sqrt{-5}$ are primes in $\mathbf{Z}[\sqrt{-5}]$ that are not unit multiples of each other, so they are relatively prime. In $\mathbf{Z}[\sqrt{-5}]$ we have

- $(2, 1+\sqrt{-5}) = 1$ but we can't write $2x + (1+\sqrt{-5})y = 1$ for some x and y in $\mathbf{Z}[\sqrt{-5}]$ (similar argument to what is done in the previous example),
- $2 \mid (1+\sqrt{-5})(1-\sqrt{-5}) = 6$ and $(2, 1+\sqrt{-5}) = 1$, but $2 \nmid (1-\sqrt{-5})$ since $(1-\sqrt{-5})/2 \notin \mathbf{Z}[\sqrt{-5}]$,
- $2 \mid 6$, $(1+\sqrt{-5}) \mid 6$, and $(2, 1+\sqrt{-5}) = 1$, but $2(1+\sqrt{-5}) \nmid 6$ since $6/(2(1+\sqrt{-5})) = (1-\sqrt{-5})/2 \notin \mathbf{Z}[\sqrt{-5}]$.