Here we shall take a slightly different route to the same goal; using ideals first to rewrite the theory of divisibility and gcd in $\mathbb{Z}$ and $\mathbb{Z}[i]$, then using them to *introduce* the gcd in $\mathbb{Z}[\sqrt{-5}]$. The ideals realizing α, β_1, and β_2 turn out to be gcds of algebraic integers.

Ideals in $\mathbb{Z}$

In $\mathbb{Z}$ we have the commonplace facts that

$$2 \text{ divides } 6, \quad 3 \text{ divides } 6, \quad \gcd(2,3) = 1.$$

These facts can be rewritten in terms of the sets

$$(2) = \{\text{multiples of } 2\}, \quad (3) = \{\text{multiples of } 3\}, \quad (6) = \{\text{multiples of } 6\},$$

which are examples of ideals. The equivalents of the first two facts are

$$(2) \text{ contains } (6), \quad (3) \text{ contains } (6),$$

which may be summed up by the slogan *to divide is to contain.* To express the third fact we consider another ideal, the *sum* of (2) and (3):

$$(2)+(3) = \{a+b : a \in (2), b \in (3)\}.$$

It is clear that $\gcd(2,3)$ divides any member of the set $(2)+(3)$, and in fact it is not hard to show that

$$(2)+(3) = \{\text{multiples of } 1\} = (1) = (\gcd(2,3)).$$

In general, we call a subset I of a ring R an *ideal* if

- $a \in I$ and $b \in I \Longleftrightarrow a+b \in I$,
- $a \in I$ and $m \in R \Longleftrightarrow am \in I$.

Then, for any $a \in \mathbb{Z}$, the set $(a) = \{\text{multiples of } a\}$ is obviously an ideal, called the *principal ideal* generated by a. It is not hard to prove (see the subsection below and the exercises) that

- every ideal in $\mathbb{Z}$ is (a) for some a,
- a divides $b \Longleftrightarrow (a)$ contains (b),
- $(a)+(b) = (\gcd(a,b))$.

Since ideals in $\mathbb{Z}$ correspond to numbers in $\mathbb{Z}$, the language of ideals tells us nothing we do not already know. However, the *concept* of ideal generalizes to other rings where it might conceivably give us new insight.

Ideals in $\mathbb{Z}[i]$

We know from Section 21.2 that $\mathbb{Z}[i]$ has many similarities to $\mathbb{Z}$, because they both have the division property. These similarities extend to properties of ideals in $\mathbb{Z}[i]$, and the division property explains why. In particular, it explains why every ideal in $\mathbb{Z}[i]$ is of the form $(\beta) = \{\text{multiples of } \beta\}$.

Suppose that I is an ideal of $\mathbb{Z}[i]$, and consider a nonzero element $\beta \in I$ of minimal norm. Then I contains the set (β) of multiples of β, since an ideal contains all multiples of any element. Also, I cannot contain any $\alpha \notin (\beta)$ by the division property: if such an α exists, there is a multiple $\mu\beta$ with $0 < |\alpha - \mu\beta| < |\beta|$. But $-\mu\beta \in I$ and hence $\alpha - \mu\beta \in I$ also, which contradicts the choice of β as a nonzero element of I of minimal norm.

Thus any ideal of $\mathbb{Z}[i]$ consists of all the multiples of some $\beta \in \mathbb{Z}[i]$, which we saw in Section 21.1 is a set with the same shape as $\mathbb{Z}[i]$. The same is true for principal ideals in any $\mathbb{Z}[\sqrt{-n}]$: *they all have the same (rectangular) shape*. In fact the set (β) of multiples of β consists of sums of the elements β and $\beta\sqrt{-n}$, which define a rectangle of the same shape as the rectangle defined by the generating elements 1 and $\sqrt{-n}$ of $\mathbb{Z}[\sqrt{-n}]$.

Ideals in $\mathbb{Z}[\sqrt{-5}]$

$\mathbb{Z}[\sqrt{-5}]$ contains an ideal that is *not* the same shape as $\mathbb{Z}[\sqrt{-5}]$ itself. We expect this, since unique prime factorization fails in $\mathbb{Z}[\sqrt{-5}]$, and so the division property fails too; however, it is satisfying to make this failure visible.

One such ideal is the sum I of the principal ideals (2) and $(1+\sqrt{-5})$,

$$(2) + (1+\sqrt{-5}) = \{2m + (1+\sqrt{-5})n : m, n \in \mathbb{Z}\},$$

part of which is shown in Figure 21.2.

It is clear from the figure that I (consisting of the black dots) is *not* rectangular in shape like $\mathbb{Z}[\sqrt{-5}]$ (consisting of the black and white dots)—the black neighbors of any black dot do not include any two in perpendicular directions.

Thus the members of I are not the multiples of any one $\beta \in \mathbb{Z}[\sqrt{-5}]$. They are, if you like, the multiples of an "ideal number"—a number that is outside $\mathbb{Z}[\sqrt{-5}]$.

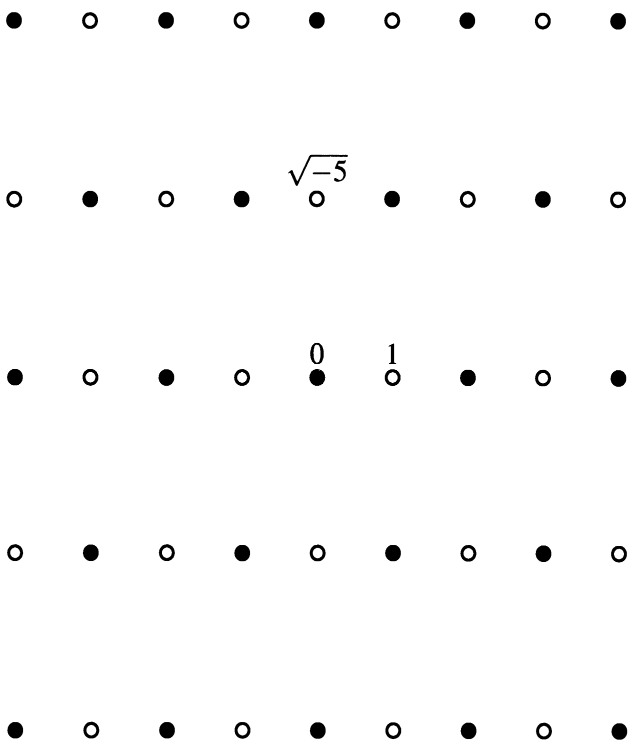

Figure 21.2: The nonprincipal ideal $(2)+(1+\sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$

EXERCISES

Implicit in the discussion above is the following *definition* of the sum of ideals: if A and B are ideals, then

$$A+B=\{a+b : a\in A, b\in B\}.$$

It should also be checked that $A+B$ thus defined is itself an ideal.

21.4.1 Check that $A+B$ has the two defining properties of an ideal.

In $\mathbb{Z}$, we know that $\gcd(a,b)=ma+nb$ for some m and n. This makes it easy to describe the sum of principal ideals $(a)+(b)$ in terms of the gcd.

21.4.2 Show that $(a)+(b)=(\gcd(a,b))$ in $\mathbb{Z}$.

We take up this idea in the next section to find the gcd of any ideals. For the moment, we continue to explore nonprincipal ideals in $\mathbb{Z}[\sqrt{-5}]$, arising as sums of principal ideals.

21.4.3 Show that the vectors from O to 2 and $1+\sqrt{-5}$ define a parallelogram of the same shape as the vectors from O to 3 and $1-\sqrt{-5}$. *Hint*: Consider quotients of complex numbers and what they say about the ratio of side lengths, and the angle between the sides. (The same idea occurs in the exercises for Section 16.5.)

21.4.4 Deduce from Exercise 21.4.3 that the ideal $(3)+(1-\sqrt{-5})$ has the same shape as the ideal $(2)+(1+\sqrt{-5})$.

21.4.5 Show also that the ideal $(3)+(1-\sqrt{-5})$ has the same shape as the ideal $(3)+(1+\sqrt{-5})$.

Thus we have found so far only two different shapes of ideals in $\mathbb{Z}[\sqrt{-5}]$: the shape of $\mathbb{Z}[\sqrt{-5}]$ itself, which is the shape of all principal ideals, and the shape of the nonprincipal ideal $(2)+(1+\sqrt{-5})$.

It can be shown that any ideal in $\mathbb{Z}[\sqrt{-5}]$ has one of these two shapes, which represent what Dedekind called the ideal *classes* of $\mathbb{Z}[\sqrt{-5}]$. This term goes back to the older theory of quadratic forms, where forms $ax^2+bxy+cy^2$ with the same discriminant b^2-4ac were divided into a number of equivalence classes, the number of which was called the *class number*. Lagrange (1773a) showed that any form with discriminant -20 was equivalent to either x^2+5y^2 (the norm of $x+y\sqrt{-5}$) or $2x^2+2xy+3y^2$. These two forms correspond to the two ideal classes of $\mathbb{Z}[\sqrt{-5}]$.

21.5 Ideal Factorization

In $\mathbb{Z}$ we saw that "to divide is to contain," because

$$a \text{ divides } b \iff (a) \text{ contains } (b).$$

In $\mathbb{Z}[\sqrt{-5}]$, we can then say that the nonprincipal ideal $(2)+(1+\sqrt{-5})$ behaves like a common divisor of 2 and $1+\sqrt{-5}$, because

$$(2)+(1+\sqrt{-5}) \text{ contains } (2), \quad (2)+(1+\sqrt{-5}) \text{ contains } (1+\sqrt{-5}).$$

Indeed, we can expect that $(2)+(1+\sqrt{-5})$ is the *greatest common divisor* of 2 and $1+\sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$, since in $\mathbb{Z}$ it is always true that $(a)+(b)=(\gcd(a,b))$.

Not only that, we can expect that $(2)+(1+\sqrt{-5})$ is *prime*. In $\mathbb{Z}$ we notice that p is prime if and only the ideal (p) is *maximal*; that is, the only ideal properly containing (p) is $\mathbb{Z}$ itself. This is because any $a \notin (p)$ is relatively prime to p, hence $ma+np=1$ for some m and n, so 1 is in any ideal containing both a and p.

To prove that $(2)+(1+\sqrt{-5})$ is maximal is even easier. We suppose that $a=m+n\sqrt{-5}\notin(2)+(1+\sqrt{-5})$, which means that m is even. But then $a-1\in(2)+(1+\sqrt{-5})$, hence 1 is in any ideal containing both a and $(2)+(1+\sqrt{-5})$. Such an ideal is therefore $\mathbb{Z}[\sqrt{-5}]$ itself.

To sum up: if ideals in $\mathbb{Z}[\sqrt{-5}]$ have divisibility properties like those in $\mathbb{Z}$, then $(2)+(1+\sqrt{-5})$ is the gcd of 2 and $1+\sqrt{-5}$, and it is prime. Dedekind (1871) defined the product of ideals so that divisibility behaves as expected.

Definition. If A and B are ideals, then

$$AB=\{a_1b_1+a_2b_2+\cdots+a_kb_k : a_1,a_2,\ldots,a_k\in A,\ b_1,b_2,\ldots,b_k\in B\}.$$

It is easily checked that AB is an ideal and (with greater difficulty) that the containment concept of divisibility agrees with the usual concept: B divides A if there is an ideal C such that $A=BC$. However, what is really delightful is that *the product of ideals explains the nonunique prime factorization of 6 in* $\mathbb{Z}[\sqrt{-5}]$,

$$6=2\cdot 3=(1+\sqrt{-5})(1-\sqrt{-5}),$$

by resolving both sides into the same product of prime ideals. In fact, we have

- (2) is the square of the prime ideal $(2)+(1+\sqrt{-5})$,
- (3) is the product of ideals $(3)+(1+\sqrt{-5})$ and $(3)+(1-\sqrt{-5})$, which are prime,
- $(1+\sqrt{-5})$ is the product of $(2)+(1+\sqrt{-5})$ and $(3)+(1+\sqrt{-5})$,
- $(1-\sqrt{-5})$ is the product of $(2)+(1+\sqrt{-5})$ and $(3)+(1-\sqrt{-5})$.

As an example, we prove the first of these claims.

The ideal factorization of 2: $(2)=[(2)+(1+\sqrt{-5})]^2$.

It follows from the definition of product of ideals that

$$\begin{aligned}4=2\times 2&\in[(2)+(1+\sqrt{-5})]^2\\ 2+2\sqrt{-5}=2\times(1+\sqrt{-5})&\in[(2)+(1+\sqrt{-5})]^2\\ -4+2\sqrt{-5}=(1+\sqrt{-5})^2&\in[(2)+(1+\sqrt{-5})]^2.\end{aligned}$$

Adding the elements 4, $2+2\sqrt{-5}$, and $-4+2\sqrt{-5}$ of $[(2)+(1+\sqrt{-5})]^2$, we find $2 \in [(2)+(1+\sqrt{-5})]^2$. It follows that all multiples of 2 are in $[(2)+(1+\sqrt{-5})]^2$, that is, $[(2)+(1+\sqrt{-5})]^2$ contains (2).

Conversely, any element of $[(2)+(1+\sqrt{-5})]^2$ is a sum of products of terms $2m$ and $(1+\sqrt{-5})n$. Any product involving $2m$ is a multiple of 2, and so is any product involving $(1+\sqrt{-5})^2 = -4+2\sqrt{-5}$. Thus any element of $[(2)+(1+\sqrt{-5})]^2$ is a multiple of 2, hence $[(2)+(1+\sqrt{-5})]^2$ contains (2), as required. □

EXERCISES

The other ideal factorizations claimed above, and proofs that the factors are maximal ideals, go along the same lines as the examples just worked out.

21.5.1 Show in turn that 9, 6, and hence 3 belong to the product of ideals

$$[(3)+(1+\sqrt{-5})][(3)+(1-\sqrt{-5})],$$

so $[(3)+(1+\sqrt{-5})][(3)+(1-\sqrt{-5})]$ is contained in the ideal (3).

21.5.2 Show that an element of $(3)+(1+\sqrt{-5})$ times one of $(3)+(1-\sqrt{-5})$ is a multiple of 3, so that (3) contains $[(3)+(1+\sqrt{-5})][(3)+(1-\sqrt{-5})]$.

21.5.3 Consider an ideal A containing $(3)+(1+\sqrt{-5})$ and an element a outside $(3)+(1+\sqrt{-5})$. Show that A contains either 1 or 2, and in the latter case A also contains 1.

21.5.4 Deduce from Exercise 21.5.3 that $(3)+(1+\sqrt{-5})$ is a maximal ideal in $\mathbb{Z}[\sqrt{-5}]$, and show that $(3)+(1-\sqrt{-5})$ is maximal similarly.

21.6 Sums of Squares Revisited

Algebraic number theory has a very long pedigree, which can plausibly be traced back to the Babylonian discovery of Pythagorean triples around 1800 BCE. It is still mysterious how the Babylonians were able to generate triples, seemingly at will, but a method of generation can be clearly recognized in the work of Diophantus. It lies in the Diophantus two-square identity from Section 20.2:

$$(a_1^2+b_1^2)(a_2^2+b_2^2) = (a_1a_2-b_1b_2)^2+(a_1b_2+b_1a_2)^2.$$

This identity allows us to "compose" two Pythagorean triples, (a_1,b_1,c_1) and (a_2,b_2,c_2), to obtain a third triple, $(a_1a_2-b_1b_2, a_1b_2+b_1a_2, c_1c_2)$.

But with Diophantus the focus shifts from the triples (a,b,c) to the pairs (a,b), and particularly to the sums a^2+b^2. As Diophantus said (Section 20.2), 65 is the sum of two squares *because* $65 = 5 \times 13$, and because 5 and 13 are also sums of two squares. To understand which numbers are sums of two squares, we evidently need to look at their factors, and hence the problem boils down to knowing which *primes* are sums of two squares. Apparently Fermat was the first to see that this was the ultimate question about sums of two squares. At any rate, Fermat (1640b) was the first to answer it: *an odd prime p is the sum of two squares if and only if p is of the form* $4n+1$.

Fermat, in his usual manner, stated this theorem without proof. The first published proof was given by Euler (1749), and a series of increasingly elegant proofs was given by illustrious mathematicians, usually when they had new methods to show off: for example, Lagrange (1773b), (theory of quadratic forms), Gauss (1832c) (Gaussian integers), and Dedekind (1877) (ideal theory).

Lagrange's theory of quadratic forms was in fact a precursor of algebraic number theory, stimulated by a trio of theorems stated by Fermat, and by a problem that Fermat was unable to solve. The three theorems are about odd primes p of the forms x^2+y^2 (the one inspired by Diophantus), x^2+2y^2, and x^2+3y^2, and they may be stated as follows.

$$\begin{aligned} p=x^2+y^2 &\iff p \equiv 1 \pmod 4), && \text{[Fermat (1640b)]}\\ p=x^2+2y^2 &\iff p \equiv 1 \text{ or } 3 \pmod 8), && \text{[Fermat (1654)]}\\ p=x^2+3y^2 &\iff p \equiv 1 \pmod 3). && \text{[Fermat (1654)]} \end{aligned}$$

The problem Fermat was unable to solve was to characterize odd primes of the form x^2+5y^2. Here there was a puzzling new phenomenon: primes *not* of the form x^2+5y^2, such as 3 and 7, whose product *is* of the form x^2+5y^2.

Lagrange (1773b) was able to prove Fermat's three theorems, and to explain the anomalous behavior of x^2+5y^2, by his theory of *equivalence of quadratic forms*. If we are interested in the numbers represented by a form $ax^2+bxy+cy^2$, then we also need to survey the forms $a'x'^2+b'x'y'+c'y'^2$ obtainable from $ax^2+bxy+cy^2$ by a change of variables

$$x' = px+qy, \quad y' = rx+sy, \quad \text{where } p,q,r,s \in \mathbb{Z} \text{ and } ps-qr = \pm 1,$$

because such a change of variables $(x,y) \mapsto (x'y')$ is a one-to-one map of

$\mathbb{Z} \times \mathbb{Z}$, and hence the new form represents exactly the same numbers as the old.

Lagrange called such forms *equivalent* and observed that they have the same *discriminant*: $b^2 - 4ac = b'^2 - 4a'c'$. Moreover, he found that

all forms with discriminant -4 are equivalent to $x^2 + y^2$,
all forms with discriminant -8 are equivalent to $x^2 + 2y^2$,
all forms with discriminant -12 are equivalent to $x^2 + 3y^2$,

but *there are two inequivalent forms with discriminant* -20: namely, the forms $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$. By exposing the "invisible companion" $2x^2 + 2xy + 3y^2$ of $x^2 + 5y^2$, Lagrange explained the behavior of numbers of the form $x^2 + 5y^2$. They cannot be understood in isolation, but only as a class that interacts with numbers of the form $2x^2 + 2xy + 3y^2$. In fact, the primes of the form $x^2 + 5y^2$ are those $\equiv 1$ or $9 \pmod{20}$, while the primes of the form $2x^2 + 2xy + 3y^2$ are those $\equiv 3$ or $7 \pmod{20}$. And products of the latter primes are $\equiv 1$ or $9 \pmod{20}$ and of the form $x^2 + 5y^2$.

It appears that Gauss was aware that the theory of quadratic forms could be replaced, at least up to a point, by a theory of "quadratic integers." His theory of $\mathbb{Z}[i]$ is indeed a replacement for Lagrange's theory of the quadratic form $x^2 + y^2$. But Gauss was also aware that in some cases the corresponding quadratic integers failed to have unique prime factorization (which is perhaps why he was the first to recognize the importance of unique prime factorization elsewhere). He was unable to see a way around this obstacle, so Kummer's creation of ideal numbers can be regarded as the solution to a problem that had baffled even the great Gauss.

We do not know how far Kummer developed the theory of ideal numbers in rings of quadratic integers such as $\mathbb{Z}[\sqrt{-5}]$, because he was actually interested in algebraic integers of higher degree, the so-called *cyclotomic integers*. As their name suggests, these arise from the theory of circle division (Sections 2.3 and 14.5), where the solutions $1, \zeta_n, \zeta_n^2, \ldots, \zeta_n^{n-1}$ of the equation

$$x^n - 1 = 0$$

represent n equally spaced points on the unit circle. The numbers

$$a_0 + a_1\zeta_1 + a_2\zeta_n^2 + \cdots + a_{n-1}\zeta_n^{n-1}, \quad \text{where } a_0, a_1, \ldots, a_{n-1} \in \mathbb{Z}$$

form a ring $\mathbb{Z}[\zeta_n]$ of *cyclotomic integers*.