21

Algebraic Number Theory

Preview

Another concept of abstract algebra that emerged from the old algebra of equations was that of *ring*, which arose from attempts to find *integer* solutions of equations. The first steps towards the ring concept were taken by Euler (1770b), who discovered equations whose integer solutions are most easily found with the help of irrational or imaginary numbers.

Gauss realized that these auxiliary numbers work because they *behave like* integers. In particular, they admit a concept of "prime" for which unique prime factorization holds.

In the 1840s and 1850s the idea of "algebraic integers" was pushed further by various mathematicians, and it reached maturity when Dedekind (1871) defined the concept of *algebraic integer* in a *number field of finite degree*. By this time, considerable experience with number fields had been acquired, and Kummer had noticed that such fields do *not* always admit unique prime factorization.

Kummer found a way around this difficulty by introducing new objects that he called *ideal numbers* (in analogy with "ideal" objects in geometry, such as points at infinity). Dedekind replaced Kummer's undefined "ideal numbers" by concrete sets of numbers that he called *ideals*. He was then able to restore unique prime factorization by proving that it holds for ideals.

Ring theory as we know it today is largely the result of building a general setting for Dedekind's theory of ideals. It owes its existence to Emmy Noether, who used to say that "it's already in Dedekind."

J. Stillwell, *Mathematics and Its History*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-1-4419-6053-5_21,

21.1 Algebraic Numbers

The integers are the simplest objects in mathematics but, as history shows, their secrets are deeply hidden. A vast range of mathematical disciplines—such as geometry, algebra, and analysis—has been called upon to clarify the apparently simple concept of integer. In particular, a broader *concept of integer* itself seems to be useful. We have seen in Section 5.4, for example, how integer solutions of the Pell equation $x^2 - Ny^2 = 1$ can be produced with the help of irrational numbers of the form $a + b\sqrt{N}$, and in Section 10.6 how the number $(1 + \sqrt{5})/2$ helps explain the mysterious sequence of Fibonacci numbers. These are examples of the way *algebraic* numbers help elucidate the behavior of integers.

In the 19th century, a powerful theory of algebraic numbers was developed, with the aim of throwing more light on ordinary number theory. It was very successful in this respect, but it also developed a life of its own, and in the 20th century its concepts were appropriated by the abstract theories of rings, fields, and vector spaces. Later in the chapter we sketch how this happened, but our main goal is to explain algebraic number theory itself, the inspiration for this whole development.

First we should state the definition: an *algebraic number* is one that satisfies an equation of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad \text{where } a_0, a_1, \ldots, a_n \in \mathbb{Z}.$$

The symbol $\mathbb{Z}$ for integers comes from the German word "Zahlen," meaning "numbers." We sometimes call these integers the "ordinary," or *rational*, integers, to avoid confusion with the algebraic integers defined in Section 21.3.

The algebraic numbers obviously include $\sqrt{2}$ (a solution of $x^2 - 2 = 0$), $\sqrt[3]{2}$ (a solution of $x^3 - 2 = 0$), and less obviously $\sqrt{2} + \sqrt{3}$ (see Exercise 21.1.1). The first mathematicians to use algebraic numbers systematically in number theory were Lagrange and Euler around 1770. A spectacular example was given by Euler (1770b), when he used the algebraic number $\sqrt{-2}$ to prove the following claim of Fermat: $x = 5$ *and* $y = 3$ *is the only positive integer solution of* $y^3 = x^2 + 2$. (The equation in fact goes back to Diophantus, who mentioned its integer solution in his Book VI, Problem 17.)

Euler's argument is incomplete but essentially correct, and we complete it later by closer study of the set $\mathbb{Z}[\sqrt{-2}]$ of numbers $a + b\sqrt{-2}$, where $a, b \in \mathbb{Z}$. It goes as follows.

Suppose x and y are integers such that $y^3 = x^2 + 2$. Then

$$y^3 = (x + \sqrt{-2})(x - \sqrt{-2}).$$

Assuming that numbers of the form $a + b\sqrt{-2}$ "behave like" ordinary integers, we can conclude that $x + \sqrt{-2}$ and $x - \sqrt{-2}$ are *cubes* (since their product is the cube y^3). That is, there are $a, b \in \mathbb{Z}$ such that

$$\begin{aligned} x + \sqrt{-2} &= (a + b\sqrt{-2})^3 \\ &= a^3 + 3a^2b\sqrt{-2} + 3ab^2(-2) + b^3(-2\sqrt{-2}) \\ &= a^3 - 6ab^2 + (3a^2b - 2b^3)\sqrt{-2}. \end{aligned}$$

Equating real and imaginary parts, we get

$$\begin{aligned} x &= a^3 - 6ab^2, \\ 1 &= 3a^2b - 2b^3 = b(3a^2 - 2b^2) \quad \text{for some } a, b \in \mathbb{Z}. \end{aligned}$$

Now the only integer products equal to 1 are 1×1 and $(-1) \times (-1)$; hence $b = \pm 1$, and therefore $a = \pm 1$, from the second equation. Then the only positive solution for x occurs with $a = -1$, $b = \pm 1$, in which case $x = 5$ and hence $y = 3$. □

This wonderful flight of fancy, that the numbers $a + b\sqrt{-2}$ "behave like" ordinary integers, can actually be justified. It depends on the theory of divisibility in $\mathbb{Z}[\sqrt{-2}]$, which turns out to be similar to divisibility in $\mathbb{Z}$, already studied in Section 3.3.

Exercises

21.1.1 Show that the number $\sqrt{2} + \sqrt{3}$ satisfies the equation $x^4 - 10x^2 + 1 = 0$.

Before starting to investigate divisibility in $\mathbb{Z}[\sqrt{-2}]$, it will be useful to renew our acquaintance with $\mathbb{Z}$, particularly with regard to the behavior of squares, cubes, and their divisors.

21.1.2 Use unique prime factorization to show that a positive integer n is a square if and only if each prime in the prime factorization of n occurs to an even power.

21.1.3 If l and m are positive integers with no common prime divisor, and lm is a square, use Exercise 21.1.2 to show that l and m are both squares.

21.1.4 Show similarly that if l and m are integers with no common prime divisor, and if lm is a cube, then l and m are both cubes.

Thus to prove such results about the numbers $x + \sqrt{-2}$ and $x - \sqrt{-2}$ we need to know, first, that they have no common prime divisor. In the next section we introduce the concept of *norm*, which reduces such divisibility questions to questions about divisibility in the ordinary integers.

21.2 Gaussian Integers

Beyond $\mathbb{Z}$ itself, the simplest set to "behave like" integers is $\mathbb{Z}[i]$, the set of numbers of the form $a + bi$, where $a, b \in \mathbb{Z}$. These are called the *Gaussian integers*, because Gauss (1832c) was the first to study them and prove their basic properties. $\mathbb{Z}[i]$ is like $\mathbb{Z}$ in being closed under the operations $+$, $-$, and $\times$, but also in having primes and unique prime factorization.

An ordinary prime may be defined as an integer of size >1 that is not the product of integers of smaller size. A *Gaussian prime* may be defined in the same way, provided we make a sensible definition of "size." The ordinary absolute value $|a + bi| = \sqrt{a^2 + b^2}$ is a suitable measure, so we say that a Gaussian integer α is a Gaussian prime if $|\alpha| > 1$ but α is not the product of Gaussian integers of smaller absolute value.

An equivalent definition of Gaussian primes is in terms of the *square* of the absolute value, the *norm* of α, $N(\alpha)$. Namely, α is a Gaussian prime if $N(\alpha) > 1$ and α is not the product of Gaussian integers of smaller norm.

The norm has the advantage that $N(a + ib) = a^2 + b^2$ is an ordinary positive integer, so we can exploit the known properties of integers. For example, we can see immediately why *every Gaussian integer has a Gaussian prime factorization*. Namely, if α is not itself a Gaussian prime, then $\alpha = \beta\gamma$, where $N(\beta), N(\gamma) < N(\alpha)$. If β, γ are Gaussian primes, then we have a Gaussian prime factorization of α; if not, at least one of them factorizes into Gaussian integers of smaller norm, and so on. *This process must terminate*, because norms are ordinary nonnegative integers and hence they cannot decrease in size indefinitely. At termination, we have a Gaussian prime factorization of α.

The *uniqueness* of this prime factorization is a deeper result, for which it is convenient to revert to the absolute value measure of size and interpret $|a+ib|$ as the distance of $a+ib$ from O. This gives a surprisingly geometric proof that Gaussian integers have "division with remainder."

Division property of $\mathbb{Z}[i]$. *For any α and $\beta \neq 0$ in $\mathbb{Z}[i]$, there are μ and ρ in $\mathbb{Z}[i]$ such that*

$$\alpha = \mu\beta + \rho \quad \text{with} \quad |\rho| < |\beta|.$$

Proof. The multiples $\mu\beta$ for $\mu \in \mathbb{Z}[i]$ are sums of terms $\pm\beta$ and $\pm i\beta$. It follows, since the lines from O to β and $i\beta$ are perpendicular, that the numbers $\mu\beta$ lie at the corners of a lattice of squares of side $|\beta|$, as in Figure 21.1.

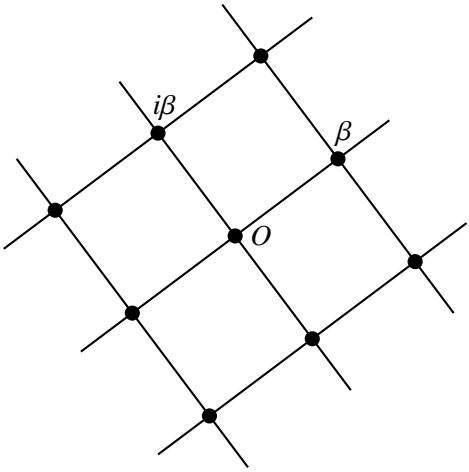

Figure 21.1: Multiples of β in $\mathbb{Z}[i]$

Now α lies in one of these squares, and if we let

$$\rho = \alpha - \text{nearest corner } \mu\beta,$$

it follows that the perpendiculars from α to the nearest sides are of length $\leq |\beta|/2$ (draw a picture). Therefore, since two sides of a triangle have total length greater than the third, we have

$$|\rho| < \frac{|\beta|}{2} + \frac{|\beta|}{2} = |\beta|,$$

as required. □

The division property of $\mathbb{Z}[i]$ has the following consequences, parallel to those for natural numbers described in Section 3.3.

1. There is a *Euclidean algorithm* for $\mathbb{Z}[i]$, which takes any $\alpha, \beta \in \mathbb{Z}[i]$ and repeatedly divides the larger of the pair by the smaller, keeping the smaller number and the remainder. It ends by finding $\gcd(\alpha, \beta)$, a common divisor of α, β that is greatest in norm.

2. There are $\mu, \nu \in \mathbb{Z}[i]$ such that $\gcd(\alpha, \beta) = \mu\alpha + \nu\beta$.

3. If ϖ is a Gaussian prime that divides $\alpha\beta$, then ϖ divides α or β.

4. The *Gaussian prime factorization of a Gaussian integer is unique*, up to the order of factors and factors of norm 1 (that is, factors ± 1, $\pm i$).

Exercises

We know from Section 20.2 that the absolute value is multiplicative, and hence so is the norm: $N(\alpha\beta) = N(\alpha)N(\beta)$. Indeed, this is just a restatement of Diophantus's identity. It follows that *if α divides γ* (that is, if $\gamma = \alpha\beta$ for some β), *then $N(\alpha)$ divides $N(\gamma)$* [because $N(\gamma) = N(\alpha)N(\beta)$].

Thus we have a criterion for divisibility in the Gaussian integers based on divisibility in the ordinary integers. Among other things, this enables us to show that certain Gaussian integers are Gaussian primes.

21.2.1 By considering $N(4 + i)$, show that $4 + i$ is a Gaussian prime.

21.2.2 Show that an ordinary prime of the form $a^2 + b^2$ is *not* a Gaussian prime, and find its Gaussian prime factorization.

Now we modify the above argument for the division property of $\mathbb{Z}[i]$ to show that $\mathbb{Z}[\sqrt{-2}]$ also has it. That is, *if α and $\beta \neq 0$ are in $\mathbb{Z}[\sqrt{-2}]$, then there are μ and ρ in $\mathbb{Z}[\sqrt{-2}]$ such that*

$$\alpha = \mu\beta + \rho \quad \text{with} \quad |\rho| < |\beta|.$$

21.2.3 Show that the multiples $\mu\beta$ of any $\beta \in \mathbb{Z}[\sqrt{-2}]$ lie at the corners of a grid of rectangles, each of which has sides of length $|\beta|$ and $\sqrt{2}|\beta|$.

21.2.4 Deduce from Exercise 21.2.3 and the Pythagorean theorem that any α lies at distance $< |\beta|$ from the nearest multiple $\mu\beta$ of $\beta \neq 0$, and hence that $\mathbb{Z}[\sqrt{-2}]$ has the division property.

As in $\mathbb{Z}[i]$, the division property leads to a Euclidean algorithm for gcd, and eventually to unique prime factorization in $\mathbb{Z}[\sqrt{-2}]$. This enables us to fill in the gaps of Euler's argument in the previous section, as soon as we have checked that $\gcd(x + \sqrt{-2}, x - \sqrt{-2}) = 1$ when $y^3 = x^2 + 2$.

21.2.5 Show that if x and y are ordinary integers with $y^3 = x^2 + 2$, then x is odd.

Finally, we invoke the norm in $\mathbb{Z}[\sqrt{-2}]$,

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

21.2.6 Show that $N(x + \sqrt{-2})$ is odd, whereas $N(2\sqrt{-2}) = 2^3$, and hence that

$$1 = \gcd(x + \sqrt{-2}, 2\sqrt{-2}) = \gcd(x + \sqrt{-2}, x - \sqrt{-2}).$$

Unique prime factorization in $\mathbb{Z}[\sqrt{-2}]$ gives an easy proof of one of the results of Fermat proved in Section 11.4. This was pointed out to me by Lin Tan.

21.2.7 Suppose that $t^2 = u^2 + 2s^2$, for ordinary integers s, t, u. By considering the prime factorization of both sides in $\mathbb{Z}[\sqrt{-2}]$, show that t is also of the form $p^2 + 2q^2$, for ordinary integers p, q.

21.3 Algebraic Integers

The Gaussian integers are an excellent example of algebraic numbers that "behave like" integers, but it is not yet clear what the general concept of "integer" should be. After a period of exploration by Dirichlet, Kummer, Eisenstein, Hermite, and Kronecker in the 1840s and 1850s, the following definition was proposed by Dedekind (1871): an *algebraic integer* is a root of an equation of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad \text{where } a_0, a_1, \ldots, a_{n-1} \in \mathbb{Z}. \qquad (*)$$

Thus the definition of algebraic integer results from the definition of algebraic number (Section 21.1) by restricting the polynomials to those with leading coefficient 1, or *monic* polynomials as they are often called.

One reason that this definition suggested itself was a result proved by Eisenstein (1850) that the numbers satisfying such equations are closed under $+$, $-$, and $\times$. It follows, since algebraic numbers inherit the properties of $+$, $-$, and $\times$ from $\mathbb{C}$, that algebraic integers form a *commutative ring with unit*, as defined in Section 20.3.

Another reason for the restriction to monic polynomials is that the *rational* algebraic integers are precisely the ordinary integers. This property of monic polynomials was pointed out by Gauss (1801), Article 11, and it is quite easy to prove. We suppose that the equation (*) has a rational solution that is not an ordinary integer. Then we may assume that the solution is of the form $x = r/pq$, where p, q, r are ordinary integers and p is a prime not dividing r. Substituting this value for x in (*), and multiplying through by $(pq)^n$, we get

$$r^n = -a_{n-1}r^{n-1}(pq) - \cdots - a_1 r(pq)^{n-1} - a_0(pq)^n.$$

However, this is impossible, because p divides the right-hand side but not the left.

In practice, it is difficult to work in the ring of all algebraic integers, and we prefer to work in smaller rings such as $\mathbb{Z}[i]$ or $\mathbb{Z}[\sqrt{-2}]$. The exercises

in the previous section show that $\mathbb{Z}[\sqrt{-2}]$ is the perfect setting for Euler's proof that $y^3 = x^2 + 2$ has only one positive solution in $\mathbb{Z}$.

The advantage of rings such as $\mathbb{Z}[i]$ or $\mathbb{Z}[\sqrt{-2}]$ is that they have the concept of norm, which allows us to define the concept of prime and to show that each element of the ring has a prime factorization. However, the *uniqueness* of prime factorization is not guaranteed, and in a sense we were lucky to find it in $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$.

A more typical ring of algebraic integers is

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}.$$

In this ring $|a + b\sqrt{-5}| = \sqrt{a^2 + 5b^2}$, and hence the norm is

$$N(a + b\sqrt{-5}) = a^2 + 5b^2.$$

As before, we define a *prime* to be a number of norm >1 that is not the product of numbers of smaller norm, and it follows as in $\mathbb{Z}[i]$ that every member of $\mathbb{Z}[\sqrt{-5}]$ factorizes into primes of $\mathbb{Z}[\sqrt{-5}]$.

It is likewise true that if β divides α in $\mathbb{Z}[\sqrt{-5}]$, then $N(\beta)$ divides $N(\alpha)$ in $\mathbb{Z}$. Hence α is a prime of $\mathbb{Z}[\sqrt{-5}]$ if $N(\alpha)$ is not divisible by any smaller norm $\neq 1$, that is, by any smaller integer of the form $a^2+5b^2 \neq 1$. Examples of primes in $\mathbb{Z}[\sqrt{-5}]$ are

$$\begin{aligned} 2, &\quad \text{because } N(2) = 4, \\ 3, &\quad \text{because } N(3) = 9, \\ 1 + \sqrt{-5}, &\quad \text{because } N(1 + \sqrt{-5}) = 6, \\ 1 - \sqrt{-5}, &\quad \text{because } N(1 - \sqrt{-5}) = 6. \end{aligned}$$

Hence it follows that 6 has *two different prime factorizations* in $\mathbb{Z}[\sqrt{-5}]$:

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

In the 1840s Kummer noticed examples of the failure of unique prime factorization, and he realized that it is a serious problem. He wrote:

> It is greatly to be lamented that this virtue of the real numbers [that is, of the ordinary integers] to be decomposable into prime factors, always the same ones for a given number, does not also belong to the complex numbers [that is, the algebraic integers]; were this the case, the whole theory,

> which is still laboring under such difficulties, could easily be brought to a conclusion. For this reason, the complex numbers we have been considering seem imperfect, and one may well ask whether one ought not to look for another kind which would preserve the analogy with the real numbers with respect to such a fundamental property.
>
> Translation by Weil (1975) from Kummer (1844)

Kummer found "another kind of number" that preserved the property of unique prime factorization, and he called them *ideal numbers*. Today we know them under the name of *ideals*.

Exercises

Although ordinary fractions, such as 1/2, are not algebraic integers, some "algebraic fractions" are.

21.3.1 Show that the golden ratio $(1 + \sqrt{5})/2$ is an algebraic integer.

21.3.2 Find the three algebraic integers that satisfy the equation $x^3 - 1 = 0$.

Eisenstein's theorem that the algebraic integers are closed under $+$, $-$, and $\times$ was given a new proof by Dedekind (1871) using linear algebra.

21.3.3 Suppose that α and β are algebraic integers satisfying the equations

$$\alpha^a + p_{a-1}\alpha^{a-1} + \cdots + p_1\alpha + p_0 = 0,$$
$$\beta^b + q_{b-1}\beta^{b-1} + \cdots + q_1\beta + q_0 = 0.$$

Deduce from these that any power $\alpha^{a'}$ may be written as a linear combination of $1, \alpha, \alpha^2, \ldots, \alpha^{a-1}$ with ordinary integer coefficients, and any power $\beta^{b'}$ as a linear combination of $1, \beta, \beta^2, \ldots, \beta^{b-1}$ with ordinary integer coefficients.

21.3.4 Now let $\omega_1, \omega_2, \ldots, \omega_n$ denote the $n = ab$ products of the form $\alpha^{a'}\beta^{b'}$, where $a' < a$ and $b' < b$. Show that, if ω denotes any one of $\alpha + \beta$, $\alpha - \beta$, or $\alpha\beta$, then we have n equations with ordinary integer coefficients $k_j^{(i)}$:

$$\begin{aligned}
\omega\omega_1 &= k_1'\omega_1 + k_2'\omega_2 + \cdots + k_n'\omega_n,\\
\omega\omega_2 &= k_1''\omega_1 + k_2''\omega_2 + \cdots + k_n''\omega_n,\\
&\vdots\\
\omega\omega_n &= k_1^{(n)}\omega_1 + k_2^{(n)}\omega_2 + \cdots + k_n^{(n)}\omega_n.
\end{aligned}$$

21.3.5 Explain why the equations in Exercise 21.3.4 have a nonzero solution for $\omega_1, \omega_2, \ldots, \omega_n$, and hence that

$$\begin{vmatrix} k_1' - \omega & k_2' & \ldots & k_n' \\ k_1'' & k_2'' - \omega & \ldots & k_n'' \\ \ldots & \ldots & \ldots & \ldots \\ k_1^{(n)} & k_2^{(n)} & \ldots & k_n^{(n)} - \omega \end{vmatrix} = 0.$$

Also explain why this is a monic equation, with ordinary integer coefficients, for $\omega = \alpha + \beta, \alpha - \beta$, or $\alpha\beta$.

21.4 Ideals

Kummer did not explicitly define his "ideal numbers." Rather, he observed that prime algebraic integers sometimes behave *as if* they were nontrivial products, and from their behavior he inferred the behavior of their "ideal factors." Dedekind (1871) showed that "ideal factors" could be realized by sets of actual numbers, and he called these sets *ideals*. In his (1877) work he used the numbers in $\mathbb{Z}[\sqrt{-5}]$ to illustrate his method, showing that 2 and 3 behave as if they were products of primes—$2 = \alpha^2$ and $3 = \beta_1\beta_2$—and then showing how α, β_1, and β_2 may be realized as ideals.

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

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, as we saw in Figure 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 $\mathbb{Z}[\sqrt{-5}]$

The ring $\mathbb{Z}[\sqrt{-5}]$ 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.

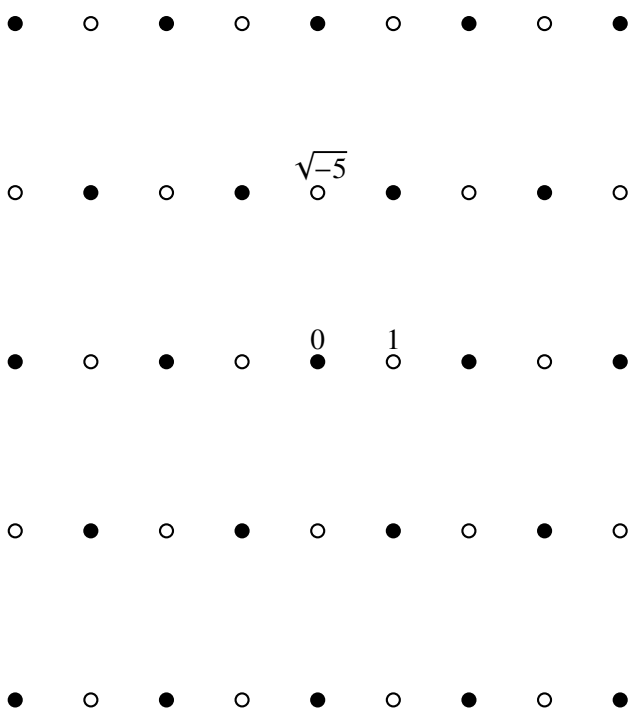

Figure 21.2: The nonprincipal ideal $(2) + (1 + \sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$

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 $\mathbb{Z}[\sqrt{-5}]$.

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 is 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 $\mathbb{Z}[\sqrt{-5}]$. For more on classes of quadratic forms, see Section 21.6.

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 that $2 \in [(2) + (1 + \sqrt{-5})]^2$. 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 $[(2) + (1 + \sqrt{-5})]^2$ is contained in (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 $[(3) + (1 + \sqrt{-5})][(3) + (1 - \sqrt{-5})]$ contains 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.

$$p = x^2 + y^2 \iff p \equiv 1 \pmod 4 \qquad \text{(Fermat (1640b))}$$

$$p = x^2 + 2y^2 \iff p \equiv 1 \text{ or } 3 \pmod 8 \qquad \text{(Fermat (1654))}$$

$$p = x^2 + 3y^2 \iff p \equiv 1 \pmod 3 \qquad \text{(Fermat (1654))}$$

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 so 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 (mod 20), while the primes of the form $2x^2 + 2xy + 3y^2$ are those $\equiv 3$ or 7 (mod 20). And products of the latter primes are $\equiv 1$ or 9 (mod 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 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*.

In Kummer's time it was thought that $\mathbb{Z}[\zeta_n]$ was the key to Fermat's last theorem, because if $a, b, c \in \mathbb{Z}$ are such that $a^n + b^n = c^n$, then the nth power $a^n + b^n$ factorizes into n linear factors in $\mathbb{Z}[\zeta_n]$. In fact, this was the basis of a mistaken "proof " by Lamé (1847). However, Kummer noticed that such arguments break down, precisely because *unique prime factorization fails in* $\mathbb{Z}[\zeta_n]$. Kummer showed that this happens for $n \geq 23$, and he created the theory of ideal numbers in an attempt to repair the damage. In this respect, ideal numbers were only partially successful (not that it matters, now that we have Wiles's proof of Fermat's last theorem), but they proved their worth elsewhere. Dedekind's revision of Kummer's idea gave us the concept of ideal, which is indispensable in algebra today.

For a treatment of primes of the form $x^2 + 5y^2$ using ideals, see Artin (1991) or Stillwell (2003), and for more on the history of $x^2 + ny^2$, see the introduction to Dedekind (1877), and Cox (1989). The latter pursues another remarkable thread in number theory—the modular function. As mentioned in the exercises to Section 16.5, the modular function is a function of lattice shapes, which is why it can reflect ideals of imaginary quadratic integers. For more, see Cox's book, or McKean and Moll (1997).

Exercises

There is an "easy direction" of Fermat's theorems about $x^2 + y^2$, $x^2 + 2y^2$, and $x^2 + 3y^2$ that can be proved with the help of congruences. This direction shows that primes are *not* representable in the given forms if they have the wrong remainders on division by 4, 8, and 3, respectively. (Compare with Exercises 1.5.2 and 3.2.1.)

21.6.1 Show that

1. An odd prime $x^2 + y^2 \not\equiv 3 \pmod 4$.
2. An odd prime $x^2 + 2y^2 \not\equiv 5 \text{ or } 7 \pmod 8$.
3. An odd prime $x^2 + 3y^2 \not\equiv 2 \pmod 3$.

The "hard direction" of Fermat's theorems, finding the x^2 and y^2 to represent primes with the right remainders, involves more than we can cover completely here. However, for $x^2 + y^2$ and $x^2 + 2y^2$ it involves unique prime factorization in $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$, both of which were discussed earlier in this chapter.

For $x^2 + 3y^2$, the proof involves not so much $\mathbb{Z}[\sqrt{-3}]$ as the larger ring

$$\mathbb{Z}\left[\frac{1+\sqrt{-3}}{2}\right] = \left\{m + \frac{1+\sqrt{-3}}{2}n : m, n \in \mathbb{Z}\right\}.$$

21.6.2 Show that $(1 + \sqrt{-3})/2$ is an algebraic integer and that $\mathbb{Z}[(1 + \sqrt{-3})/2]$ contains $\mathbb{Z}[\sqrt{-3}]$.

21.6.3 Show that 2, $1 + \sqrt{-3}$, and $1 - \sqrt{-3}$ are primes of $\mathbb{Z}[\sqrt{-3}]$, and deduce that 4 has two distinct prime factorizations in $\mathbb{Z}[\sqrt{-3}]$.

21.6.4 By a geometric argument like those used for $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$, show that $\mathbb{Z}[(1 + \sqrt{-3})/2]$ has unique prime factorization.

21.7 Rings and Fields

Kronecker is famous for saying "God made the natural numbers, the rest is the work of man." (This is reported, for example, in his obituary by Weber (1892).) Algebraic number theory was very much what he had in mind, because Kronecker, like Dedekind, saw number theory as the source of the most interesting problems, and the inspiration for all mathematical concepts. We can at least agree that number theory was the inspiration for two of the most important *algebraic* concepts: rings and fields.

Perhaps the first step toward abstract algebra was the introduction of negative numbers, creating the ring $\mathbb{Z}$ of integers from the natural numbers. This seems to have been a very difficult step, because mathematicians for many centuries (say, from the time of Diophantus to Descartes) lived in a halfway house where negative numbers were only partially accepted—sometimes being admitted in intermediate calculations, but not allowed as answers. Likewise, it was a long time before the "ratios" of the Greeks became the *field* $\mathbb{Q}$ of rational numbers.

Thus the first level of abstraction, the creation of inverses for addition and multiplication, took place unconsciously over thousands of years. The next level, identifying *axioms* for rings and fields, took place in the 19th century, mainly under the influence of algebraic number theory. The ring axioms are essentially the result of writing down the properties of + and ×

that algebraic integers share with the ordinary integers, and the field axioms are the properties that algebraic numbers share with rational numbers.

The concept of field was implicit in the work of Abel and Galois in the theory of equations, but it became explicit when Dedekind introduced *number fields of finite degree* as the setting for algebraic number theory. He saw that the ring of all algebraic integers is not a convenient ring, because it has no "primes." This is because $\sqrt{\alpha}$ is an algebraic integer if α is, so there is always a nontrivial factorization $\alpha = \sqrt{\alpha}\,\sqrt{\alpha}$ in the ring of all algebraic integers. On the other hand, the algebraic integers in a field generated from a single algebraic number α of degree n,

$$\mathbb{Q}(\alpha) = \{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in \mathbb{Q}\},$$

have better behavior. The algebraic integers β in $\mathbb{Q}(\alpha)$ have a norm $N(\beta)$ that is an ordinary integer, and this guarantees the existence of primes, as we have seen in special cases like $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$, which are the algebraic integers in the fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$ of degree 2.

By drawing attention to the field $\mathbb{Q}(\alpha)$ of degree n, Dedekind also brought to light some *vector space* structure: the *basis* $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ of $\mathbb{Q}[\alpha]$, the *linear independence* of these basis elements over $\mathbb{Q}$, and the *dimension* (equal to the degree) of $\mathbb{Q}[\alpha]$ over $\mathbb{Q}$. Despite the long history of linear algebra, dating back 2000 years in China at least, again it was the greater generality afforded by algebraic number theory that finally brought its fundamental concepts to light.

The next level of abstraction was reached in the 20th century and it was (in a new twist to Kronecker's words) the work of a woman, Emmy Noether. In the 1920s she developed concepts for discussing common properties of different algebraic structures, such as groups and rings. One of the things groups and rings have in common is *homomorphisms*, or structure-preserving maps. A map $\varphi : G \to G'$ is a *homomorphism of groups* if $\varphi(gh) = \varphi(g)\varphi(h)$ for any $g, h \in G$. Similarly, a map $\varphi : R \to R'$ is a homomorphism of rings if $\varphi(r + s) = \varphi(r) + \varphi(s)$ and $\varphi(rs) = \varphi(r)\varphi(s)$ for any $r, s \in R$. From this higher vantage point, normal subgroups (Section 19.2) and ideals can be seen as instances of the same concept. Each is the *kernel* of a homomorphism φ: the set of elements mapped by φ to the identity element (1 for a group, 0 for a ring).

Exercises

It is not clear that $\mathbb{Q}(\alpha)$ (as defined above) is a field for any algebraic number α. The hardest part is to prove that the quotient of any two of its elements is also

an element. Some inkling of the difficulty may be grasped by working out the special case of $\mathbb{Q}(i)$.

21.7.1 Show that, if $a_1, b_1, a_2, b_2 \in \mathbb{Q}$, then $\frac{a_1+ib_1}{a_2+ib_2}$ is of the form $a + ib$, where $a, b \in \mathbb{Q}$.

It is also not obvious that the kernel of a group homomorphism is a normal subgroup, partly because the definition of normal subgroup in Section 19.2 is not the most convenient for this purpose. It is easier to prove that the kernel of a ring homomorphism is an ideal, using the definition of an ideal given in Section 21.4.

21.7.2 Suppose that R is a ring and φ maps R into another ring in such a way that $\varphi(r + s) = \varphi(r) + \varphi(s)$ and $\varphi(rs) = \varphi(s)\varphi(s)$ for any $r, s \in R$. Show that the set

$$\{r : \varphi(r) = 0\}$$

has the two defining properties of an ideal.

The equivalence of kernels and ideals may be illustrated in $\mathbb{Z}$ by the ideal (3) of multiples of 3.

21.7.3 Find a homomorphism of $\mathbb{Z}$ whose kernel is (3).

21.8 Biographical Notes: Dedekind, Hilbert, and Noether

Richard Dedekind (Figure 21.3) was born in 1831 in Brunswick, the home town of Gauss, into an academic family. His father, Julius, was professor of law at the Collegium Carolinum, and his mother, Caroline Emperius, was the daughter of another professor there. Richard was the youngest of four children in a close-knit family. They remained in Brunswick for most of their lives, and Richard lived with his sister Julie (both of them being unmarried) until 1914. Sounds dull, but this seemingly eventless life was the background to revolutionary activity in mathematics, in its way as provocative as the work of Galois.

Dedekind became interested in mathematics in high school, after coming to the conclusion that chemistry and physics were not sufficiently logical. He attended the Collegium Carolinum, the scientific academy that Gauss also attended, before entering Göttingen University in 1850. There he became friends with Riemann and made rapid academic progress, completing a thesis under Gauss's supervision in 1852. After the death of Gauss in 1855, Dirichlet was appointed to Gauss's chair, and he became the third major influence on Dedekind's career. After a brief period at the

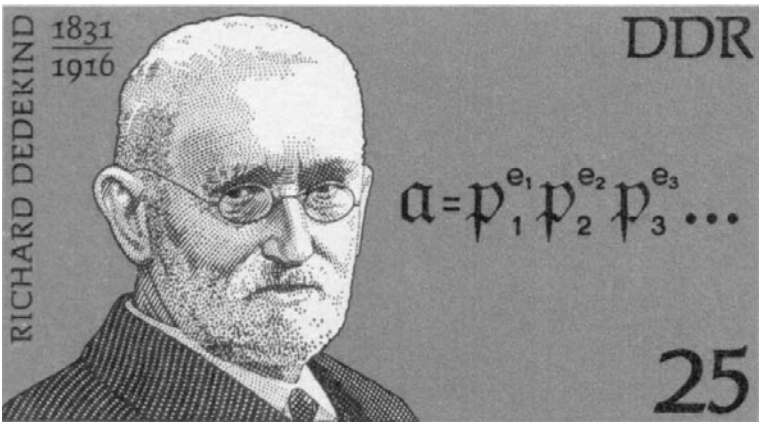

Figure 21.3: Richard Dedekind

Polytechnikum in Zurich (now known as the ETH), a position that he won in competition with Riemann, Dedekind returned to the Polytechnikum in Brunswick, where he remained for the rest of his life. It was not a prestigious position, but the home comforts enabled him to concentrate on mathematics.

Dedekind was the last student of Gauss, and Gauss's number theory was the inspiration for much of Dedekind's work, as it was for many of the great German mathematicians of the 19th century. When Dedekind started, the new generation of Eisenstein, Dirichlet, and Kronecker was beginning to understand Gauss's ideas, and making further progress. Dirichlet in particular made Gauss more approachable with his elegant and readable *Vorlesungen über Zahlentheorie* (Lectures on Number Theory, Dirichlet (1863)), which simplified much of Gauss's difficult theory of quadratic forms and added stunning new results and proofs of his own. The climax of Dirichlet's lectures is a *class number formula*, giving a uniform description of the number of inequivalent quadratic forms with given discriminant. The lectures were edited by Dedekind and first published in 1863, four years after Dirichlet's death. Dedekind took this project very seriously and made it virtually his life's work, bringing out further editions in 1871, 1879, and 1894, each time adding supplementary material, until the supplements amounted to more than Dirichlet's book itself. The theory of ideals made its first appearance in the 1871 edition, and was expanded and deepened in 1879 and 1894, eventually including a lot of Galois theory as well.

However, Dedekind was disappointed in the low enthusiasm for ideals shown by other mathematicians, and in 1877 he attempted a more popular approach. Dedekind (1877) is nearly perfect for the modern reader—clear, concise, and well motivated—but apparently it was still too abstract for his contemporaries. The theory of ideals did not really catch on until it was given a new exposition by Hilbert (1897), as we shall see below.

In the meantime, Dedekind had made several other great contributions to mathematics that were slowly taking root:

- the theory of real numbers as "Dedekind cuts,"
- the theory of Riemann surfaces as algebraic function fields,
- the characterization of natural numbers as an "inductive set."

What these contributions had in common, and what made them hard for Dedekind's contemporaries to grasp, was the idea of treating *infinite sets* as mathematical objects. Dedekind actually started doing this in 1857, when he treated congruence modulo n as the arithmetic of congruence classes

$$\begin{aligned} 0 \bmod n &= \{0, \pm n, \pm 2n, \ldots\}, \\ 1 \bmod n &= \{1, 1 \pm n, 1 \pm 2n, \ldots\}, \\ &\vdots \\ n-1 \bmod n &= \{n-1, n-1 \pm n, n-1 \pm 2n, \ldots\}, \end{aligned}$$

which are added and multiplied according to the rules

$$\begin{aligned} (i \bmod n) + (j \bmod n) &= (i + j) \bmod n, \\ (i \bmod n)(j \bmod n) &= (i \cdot j) \bmod n. \end{aligned}$$

(We mentioned multiplication mod n in Section 19.1, but without mention of congruence classes.)

The idea of adding or multiplying sets by adding or multiplying *representatives* transfers directly to Dedekind cuts and, with some modification, to ideals and Riemann surfaces. Dedekind hoped that this cornucopia of applications would convince his colleagues of the value of the idea that "mathematical objects are sets," but it was a hard idea to sell. At first he was joined only by Cantor, who took up the theory of infinite sets as enthusiastically as Dedekind took up the applications (see Chapter 24).

Dedekind had to wait decades before his ideas entered the mainstream (and in some cases after they had been rediscovered by others—for example, his theory of natural numbers became the "Peano axioms"), but fortunately he lived long enough. He died in 1916 at the age of 84.

David Hilbert (Figure 21.4) was born in 1862 in Königsberg and died in Göttingen in 1943. His father, Otto, was a judge, and David may have inherited his mathematical ability from his mother, about whom we know little except that her maiden name was Erdtmann. Königsberg was in the remote eastern part of Prussia (it is now Kaliningrad, a small, disconnected piece of Russia), but with a strong mathematical tradition dating back to Jacobi. When Hilbert attended university there in the 1880s he became friends with Hermann Minkowski, a former child mathematical prodigy two years his junior, and Adolf Hurwitz, who was three years older and a professor in Königsberg from 1884. The three used to discuss mathematics on long walks, and Hilbert seems to have picked up his basic mathematical education in this way. In later life he made "mathematical walks" an important part of the education of his own students.

Figure 21.4: David Hilbert

Hilbert's first research interest was in the theory of invariants, an algebraic topic then held in high esteem. An elementary example of an invariant is the discriminant $b^2 - 4ac$ of a quadratic form, which Lagrange (1773b) noticed is invariant when the form is transformed into an equivalent form (Section 21.6). By Hilbert's time, invariant theory had become a jungle, with success depending mainly on the ability to hack through formidable calculations. The "king of invariant theory," Paul Gordan of Erlangen, was notorious for papers consisting almost entirely of equations—in fact, the story goes that he had assistants fill in any words that were necessary. In 1888 Hilbert swept all this away by solving the main problem of invariant theory, in a simple and purely conceptual manner: the *Hilbert basis theorem* showed the existence of the invariants above the quadratic level, without needing to calculate them!

Gordan was at first incredulous and exclaimed, "This is not mathematics, it is theology!" but eventually Hilbert's idea was developed further, to calculate the invariants, and Gordan had to concede that it was mathematics after all. Hilbert, for his part, moved on to conquer other worlds. In fact, this became his modus operandi for most of his career: investigate a topic thoroughly for a few years, turn it upside down, then do something completely different.

Hilbert's triumph in invariant theory secured his position in Königsberg, and in 1892 he married Käthe Jerosch, a very capable woman who acted as secretary and research assistant for many of his works. In particular, she compiled the bibliography for his massive *Zahlbericht* ("Number Report") of 1897, the work in which algebraic number theory came of age. Hilbert was commissioned by the German Union of Mathematicians in 1893 to write a report on algebraic number theory, and the report became a 300-page book (Hilbert (1897)), looking back to quadratic forms and Fermat's last theorem, and forward to *class field theory*, a major topic of the 20th century.

The mathematical public, which had not been ready when Dedekind presented algebraic number theory a few years earlier, now saw the point, and Klein invited Hilbert to assume a chair in mathematics at Göttingen, which he held from 1895 until the end of his life.

After the *Zahlbericht*, Hilbert turned to the foundations of geometry, which we have touched on in Sections 1.6, 2.1, 19.6, and 20.7. Again he scored several triumphs—finally filling the gaps in Euclid, discovering the algebraic meaning of the Pappus and Desargues theorems—but also

leaving some unfinished business. Hilbert realized that modeling Euclid's geometry by real-number coordinates is not exactly a proof that geometry is consistent; one still needs to prove that the theory of real numbers is consistent. Hilbert found this far from obvious and made it second on his list of mathematical problems presented in Paris in 1900. Then he dropped the subject in favor of mathematical physics.

However, no one found a consistency proof for the theory of real numbers, and by the 1920s Hilbert felt compelled to return to the subject. *Hilbert's program*, as it became known, called first for a formal language of mathematics, in which the concept of proof itself was mathematically definable, by precise rules for manipulating formulas. This phase of the program was in fact feasible, and was essentially carried out by Whitehead and Russell in their *Principia Mathematica* of 1910. The hard part, however, was proving that the rules of proof could not lead to a contradiction. This is where Hilbert's program stalled, and in 1931 Gödel showed that it could never be completed. His famous *incompleteness theorems* (Chapter 24) showed that such a consistency proof does not exist, and that enlarging the formal language by new axioms only puts the consistency proof further out of reach.

To his credit, Hilbert was among the first to publicize Gödel's work. The first complete proofs of Gödel's theorems are in the book of Hilbert and Bernays (1938). But it was Hilbert's misfortune to end his career, not only with the failure of one of his mathematical dreams, but also with his mathematical community in ruins. The eclipse of Göttingen began in 1933, when the Nazis came to power in Germany and began dismissing Jewish professors. In a few years, most of Germany's mathematical talent had fled, leaving the elderly and frail Hilbert in Göttingen virtually alone. He died on 14 February 1943.

One of the Jewish mathematicians forced to leave Göttingen in 1933 was Emmy Noether (Figure 21.5), who was in many ways a natural successor of Dedekind and Hilbert. Emmy Noether was born in 1882 in Erlangen and died in 1935 in Bryn Mawr, Pennsylvania. She was the oldest of four children of the mathematician Max Noether and of Ida Kaufmann. As a child she loved music, dance, and languages and planned to become a language teacher, qualifying as a teacher of English and French in 1900.

At this time in Germany, women were permitted to study at universities only unofficially, and very few did so, since the permission of the lecturer was also required. However, a few teachers were permitted to attend for

Figure 21.5: Emmy Noether

purposes of "further education," and in 1900 Emmy Noether became one of them, studying mathematics at the University of Erlangen. Here she met the "king of invariants," Paul Gordan, and wrote a thesis under his supervision in 1907. It was on invariant theory, naturally, and Emmy later described it as "crap," but it was not a complete waste of time. Physicists today admire one of her early results, on the invariants of mechanical systems.

In 1910 Gordan retired and there was a reshuffle of positions, leading to the appointment of Ernst Fischer in 1911. Fischer is not well known today, but it seems that Noether's algebraic talent suddenly blossomed through working with him. She dropped the computational approach of Gordan and rapidly mastered the conceptual approach of Dedekind and Hilbert, so much so that Hilbert invited her to Göttingen in 1915. Getting a position was another matter—Hilbert is said to have ridiculed Göttingen's exclusion of women professors by saying "this is a university, not a bathing establishment"—but she was eventually granted an unofficial chair in 1922.

In the 1920s Noether was at the height of her powers, and she found students worthy of her ability. Among them were Emil Artin, who solved two of Hilbert's problems, and B. L. van der Waerden, who brought the ideas of Noether to the world in his *Moderne Algebra* of 1930. Noether herself modestly used to claim that "es steht schon bei Dedekind" ("it's already in Dedekind") and encouraged her students to see for themselves by reading all of Dedekind's supplements. Thus, despite the highly abstract nature of Noether's algebra, her students were made aware of its direct descent from the number theory of Gauss and Dirichlet. In van der Waerden's *Algebra* this connection was unfortunately broken, and many in the next generation of students grew up unaware of it. In recent years there has been a welcome reversal of this trend; in particular, the *Algebra* of Emil Artin's son Michael uses number theory to illustrate the theory of ideals (Artin (1991)).