1.8.4. Suppose that p is a prime in the first factorization of n and q is a prime in the second. Show that pq does not divide n, otherwise there would be a smaller number with two different prime factorizations.

We now let $p < q$ be any two primes dividing the hypothetical least n with two prime factorizations. The following exercises derive a contradiction by showing that pq *does* divide n.

1.8.5. $\frac{n}{p} - \frac{n}{q} = \frac{n}{pq}(q-p)$ is a natural number. (Why?) Deduce that $\frac{n}{q}(q-p)$ is a natural number $< n$, hence with unique prime factorization, and that p divides $\frac{n}{q}(q-p)$.

1.8.6. But p does not divide $q-p$. (Why not?) Deduce from the unique prime factorization of $\frac{n}{q}(q-p)$ that p divides $\frac{n}{q}$, and hence that pq divides n, as required.

1.9* Foundations

The aim of mathematics is to prove things, which is hard, so mathematicians continually search for clearer and more powerful methods of proof. From time to time, this leads to criticism of existing methods as being unclear, or too complicated, or too narrow. Attempts are then made to find methods to replace them, which may lead to some parts of mathematics being rebuilt on different foundations. Historically, most of the rebuilding has been in the foundations of geometry and calculus, which we'll look at later, but in the 19th and 20th centuries it went as far as the foundations of arithmetic. The new foundations of arithmetic did not make the old ones obsolete, because in practice one gets along fine using induction and ring properties of $\mathbb{Z}$. But they were a revelation all the same, because they showed why induction is crucial to arithmetic: the ring properties can be derived from it. Thus arithmetic is *entirely* about the implications of the counting process!

This surprising discovery, which had been missed by all mathematicians from Euclid to Gauss, is due mainly to Hermann Grassmann (1861) and Richard Dedekind (1888).

Grassmann made the breakthrough by noticing that induction can be used not only as a method of proof, but as a method of *defi-*

nition. To define a function f on the natural numbers by induction, one writes down a value of $f(1)$, and a definition of $f(i+1)$ in terms of $f(i)$. It then follows by induction that $f(n)$ is defined for any natural number n. One function can be regarded as given along with the natural numbers themselves—the *successor* function $f(n) = n + 1$. All other standard functions, as Grassmann and Dedekind found, can be defined by induction.

In particular, Grassmann found that $+$ and $\times$ can be defined by induction, as follows. The defining equations for $+$ are

$$m + 1 = m + 1 \qquad \text{for all } m \tag{1}$$

$$m + (i + 1) = (m + i) + 1 \qquad \text{for all } m, i. \tag{i+1}$$

These equations are not as empty as they look! Equation (1) defines $m + n$ for $n = 1$ (and all natural numbers m) as the successor of m. Equation $(i+1)$ defines $m+(i+1)$ as the successor of $m+i$ (again for all natural numbers m). Thus the set of n for which $m + n$ is defined includes 1, and it includes $i + 1$ when it includes i, hence it includes all natural numbers, by induction.

Once $+$ is defined, $\times$ is defined inductively by the equations

$$m \times 1 = m \qquad \text{for all } m \tag{1}$$

$$m \times (i + 1) = m \times i + m \qquad \text{for all } m, i, \tag{i+1}$$

because the value of $m \times (i + 1)$ is defined in terms $m \times i$ and the previously defined function $+$.

The advantage of defining $+$ and $\times$ this way is that their properties can also be *proved* by induction. With suitable definitions of 0 and the negative integers, similar to those in Section 1.3, Grassmann found inductive proofs of all the ring properties of $\mathbb{Z}$ (see the exercises). Thus induction is a complete foundation for arithmetic.

Richard Dedekind (1888) asked himself the question: what are the properties of the successor function that allow it to serve as a basis for the rest of arithmetic? He found the answer to this question in terms of *sets*—a radical idea at the time, but one that has since been accepted as the most reasonable way to provide a foundation for all of mathematics.

The essential properties of the successor function are very simple.

1. It is defined on an infinite set (namely, the set of natural numbers).
2. It is one-to-one (that is, unequal numbers have unequal successors).
3. It is not onto the whole set (in particular, 1 is not a successor).

Dedekind realized that *any* function f with these properties gives rise to a set that "behaves like" the natural numbers. If a is an element that is not a value of f, then $a, f(a), f(f(a)), \ldots$ behave like $1, 2, 3, \ldots$. Thus the *abstract structure* of the natural numbers is completely described by an infinite set and a function that is one-to-one but not onto.

However, it is a deep philosophical question whether infinite sets can actually be proved to exist. Dedekind gave an answer that is very interesting, although it lies outside mathematics. He said that such a set is the set of possible ideas, because for every idea I there is another idea $f(I)$ = the idea of I, which is distinct from I. Indeed the "idea of" function f behaves like a successor function on the set of ideas.

Mathematicians have not accepted Dedekind's set of ideas as a genuine set, and the existence of infinite sets is taken as an axiom, so we will not attempt to prove it. However, the statement of this *axiom of infinity* (as it is called) is remarkably similar to Dedekind's description of the set of ideas. For each set X we define a "successor" of X by taking X as a member of a new set $\{X\}$ (rather like forming the "idea of X"). The actual successor of X is taken to be $X \cup \{X\}$, the set whose members are the members of X *and* X itself, for technical reasons. Then the axiom of infinity says that there is a set Ω rather like the set of ideas: Ω is not empty (in fact, take the empty set to be one of its members), and along with each X in Ω, the successor of X is also in Ω.

Thus when we pursue the natural numbers into the depths of set theory, what we find is nothing but the empty set and its successors. But this is all we need! John von Neumann (1923) suggested that this is the best way to define the natural numbers, or rather, the natural numbers together with zero, because the empty set is surely the best possible set to represent zero. Here is what the first few numbers look like, according to his definition.

$0 = \{\}$ (the empty set)
$1 = \{0\}$ (the set whose only member is 0)
$2 = \{0, 1\}$
$3 = \{0, 1, 2\}$
$\vdots$

In other words, 0 is the empty set, and each natural number is the set of its predecessors. You must admit that this definition can hardly be beaten for economy, because everything is built out of "nothing"—the empty set. It is also quite natural and elegant, because the ordering of natural numbers is captured by membership, the basic concept of set theory: $m < n \Leftrightarrow m$ is a member of n. Last but not least, von Neumann's definition is a very snappy answer if anyone ever forces you to give a definition of the natural numbers!

Exercises

You may have noticed that the second equation in the definition of $+$, namely,

$$m + (i + 1) = (m + i) + 1,$$

is a special case of the associative law for $+$. In fact, this was precisely Grassmann's starting point in his inductive proof of the ring properties of $\mathbb{Z}$. The associative law for $+$ (in $\mathbb{N}$) may be formulated as a statement about n by letting $S_1(n)$ be the statement:

$$l + (m + n) = (l + m) + n \qquad \text{for all natural numbers } l \text{ and } m.$$

1.9.1. Show that $S_1(1)$ is true by definition of $+$.

1.9.2. Prove $S_1(i) \Rightarrow S_1(i + 1)$ with the help of $S_1(1)$.

Grassmann's next goal was to use associativity of $+$ to prove commutativity of $+$ in $\mathbb{N}$, again inductively. However, it is not even clear that $1 + n = n + 1$, so the latter statement, call it $S_2(n)$, must be proved first. $S_2(1)$ is $1 + 1 = 1 + 1$, so $S_2(1)$ is true!

1.9.3. Prove $S_2(i) \Rightarrow S_2(i + 1)$ using associativity of $+$.

Finally, we can let $S_3(n)$ be the full commutativity statement for $\mathbb{N}$:

$$m + n = n + m \qquad \text{for all natural numbers } m.$$

$S_3(1)$ is $1 + n = n + 1$, which has just been proved, so it remains to do the following.

1.9.4. Prove $S_3(i) \Rightarrow S_3(i + 1)$ using associativity of $+$ and $S_3(1)$.

Now let us switch to Dedekind's work. When we said that $a, f(a), f(f(a)), \ldots$ "behave like" $1, 2, 3 \ldots$, you may have wanted to ask: what is the exact meaning of the three dots? This is a fair question, because f could be defined beyond where we intend the sequence $a, f(a), f(f(a)), \ldots$ to go. Here is an example.

1.9.5. Let S be the union of $\mathbb{N}$ with the set $\mathbb{Z} + 1/2 = \{m + 1/2 : m \in \mathbb{Z}\}$, and let $f(x) = x + 1$. Now show

(a) f is defined for all members of S, is one-to-one, but not onto S.

(b) S does *not* behave like $\mathbb{N}$, because infinite descent is possible in S.

In this example, $a = 1$, and the intended meaning of $\{a, f(a), f(f(a)), \ldots\}$ is the set $\mathbb{N}$. But how do we capture the meaning of "..." in other cases without using expressions like "obtainable in a finite number of steps" which assume what we are trying to define? Dedekind also had an answer to this question. He said that $\{a, f(a), f(f(a)), \ldots\}$ consists of the elements that belong to *all* sets that include a, and that include $f(x)$ when they include x.

1.9.6. Ponder Dedekind's definition, and show that it also enables us to define $\mathbb{N}$ from the set Ω asserted to exist by the axiom of infinity.

1.10 Discussion

The Euclidean Algorithm

The Euclidean algorithm is a splendid example of the universality of mathematics. It seems to have been discovered in three different cultures and for several different mathematical purposes. In ancient Greece it was crucial in Euclid's theory of divisibility and primes, as we have seen, and it was also important in the study of irrational

numbers (see Section 8.6*). Euclid's proof of the prime divisor property was actually more complicated than the one given in Section 1.6, because he apparently did not know that $\gcd(a, b) = ma + nb$ for integers m and n.

This linear representation of the gcd was discovered in India and China, perhaps first by Âryabhaṭa and Bhâskara I around 500 AD. The Indian mathematicians were interested in integer solutions of equations $ax + by = c$, and this depends on finding $\gcd(a, b)$ in the form $ma + nb$, as we saw in Section 1.5. Such problems also arose in Chinese mathematics, particularly in the so-called "Chinese remainder" problems we shall study in Section 6.6.

The algorithm became familiar in Europe by the 16th century, but for another 200 years it was considered just a useful tool rather than a revealing property of numbers. Gauss avoided use or mention of the Euclidean algorithm in his *Disquisitiones Arithmeticae*. He avoided using it for the fundamental theorem of arithmetic by giving a direct proof, by descent, of the prime divisor property (the one covered in Exercises 1.6.4 and 1.6.5). He did not even mention it when discussing the gcd and lcm, giving instead the rules for computing them from prime factorizations, and saying only that

> we know from elementary considerations how to solve these problems when the resolution of the numbers A, B, C, etc. into factors are not given (*Disquisitiones*, article 18).

And he hid its role in the solution of $ax + by = 1$ (article 28) by referring only to the so-called "continued fraction" method, which is equivalent.

Dirichlet simplified, and in some ways extended, the *Disquisitiones* in his *Vorlesungen über Zahlentheorie* (lectures on number theory) of 1867. One of his reforms was reinstatement of the Euclidean algorithm. He used it to derive the fundamental theorem and related results much as we have in this chapter, and went so far as to say:

> It is now clear that the whole structure rests on a *single* foundation, namely the algorithm for finding the greatest common divisor of two numbers. All the subsequent theorems, even when they depend on the later concepts of relative and abso-

> lute prime numbers, are still only simple consequences of the result of this initial investigation ... (Dirichlet (1867), §16).

One of the reasons Dirichlet was enthusiastic about the Euclidean algorithm was that it could be used in other situations, a fact that also converted Gauss in the end. In 1831, Gauss found it useful to introduce what are now called *Gaussian integers* —numbers of the form $a+b\sqrt{-1}$, which we shall study in Chapter 7—and found that the key to their arithmetic was the applicability of the Euclidean algorithm. Perhaps it was with this generalization in mind that Dirichlet based his number theory on the Euclidean algorithm from the beginning, because the passage quoted above continues:

> ... so one is entitled to make the following claim: any analogous theory, for which there is a similar algorithm for the greatest common divisor, must also have consequences analogous to those in our theory.

Induction

The ascent form of induction is now considered indispensible in all fields of mathematics that use natural numbers, so one would expect to find it in the earliest mathematical works. Surprisingly, it does not seem to be there. The first clear statement of the "base step, induction step" format first appeared in 1654. How did mathematicians manage for so long without this essential tool?

The answer, I believe, is that until recently mathematicians preferred *descent* to the "base step, induction step" form of induction we called *ascent* in Section 1.8. Descent is not only simpler than ascent because no "base step" is involved; it also seems to occur more naturally at the lower levels of mathematics.

Examples of descent date from ancient times, at least as far back as Euclid's *Elements*, around 300 BC, and conceivably in proofs that $\sqrt{2}$ is irrational. Euclid uses descent in Proposition 31 of Book VII of the *Elements*, to prove that any composite number A has a prime divisor or, as he puts it, that A is "measured" by some prime. He