$$\Rightarrow 0 + b = -a \qquad \text{by the additive inverse property}$$
$$\Rightarrow b = -a \qquad \text{by the identity property of } 0.$$

There is a similar set of rules describing the behavior of $\times$:

$$a \times (b \times c) = (a \times b) \times c \qquad \text{(associative law)}$$
$$a \times b = b \times a \qquad \text{(commutative law)}$$
$$a \times 1 = a \qquad \text{(identity property of 1)}$$
$$a \times 0 = 0 \qquad \text{(property of 0)}$$

and finally, a rule for the interaction of $+$ and $\times$:

$$a \times (b + c) = a \times b + a \times c \qquad \text{(distributive law)}$$

From these we deduce that $a \times (-1) = -a$ for any integer a because

$$\begin{aligned} a + a \times (-1) &= a \times 1 + a \times (-1) && \text{by the identity property of 1} \\ &= a \times (1 + (-1)) && \text{by the distributive law} \\ &= a \times 0 && \text{by the additive inverse property} \\ &= 0 && \text{by the property of 0} \\ \Rightarrow \quad a \times (-1) &= -a && \text{by the uniqueness of additive inverse.} \end{aligned}$$

It follows in particular that $(-1) \times (-1) = 1$, because $-(-1) = 1$.

We extend the set $\mathbb{Z}$ of integers to the set $\mathbb{Q}$ of *rational numbers*,[4] or simply *rationals*, by adjoining a *multiplicative inverse* a^{-1} of each nonzero integer a. The multiplicative inverse of a^{-1} is defined to be a, and these inverses have the following property:

$$a \times a^{-1} = 1 \qquad \text{(multiplicative inverse property)}$$

These properties of $\mathbb{Q}$ are what we use unconsciously in doing ordinary arithmetic with $+$, $-$, $\times$, and $\div$. The quotient $a \div b$ or a/b is the same as $a \times b^{-1}$. As mentioned earlier, questions about the arithmetic of $\mathbb{Q}$ are really equivalent to questions about $\mathbb{Z}$, or even $\mathbb{N}$, but the extra properties of $\mathbb{Q}$ sometimes make life easier. This is particularly the case in geometry, where the rational numbers pave the way for interpreting *points* as numbers.

---

[4]The symbol $\mathbb{Q}$ stands for "quotients." We do not use the initial letter of "rational" because the same letter is later needed for the real numbers.

Exercises

The rules governing the behavior of $+$, $-$, and $\times$ are called the *ring properties* of $\mathbb{Z}$, and in general any set with functions $+$ and $\times$ satisfying these rules is called a *ring*. As we have already said, the ring properties of $\mathbb{Z}$ are so familiar that we normally use them unconsciously. Becoming conscious of them helps us to understand arithmetic, not only in $\mathbb{Z}$, but also in any other system that satisfies the same rules. We call such a system a *commutative ring with unit*, the "unit" in this case being the number 1. Later we shall find it helpful to use many other rings, even to study $\mathbb{Z}$ itself.

The following exercises help to explain why the ring properties are fundamental to arithmetic, by showing how they determine the values of expressions written using natural numbers, $+$, $-$, and $\times$, and some standard algebraic identities.

1.4.1. Show, using the properties of $+$ and $-$, that $(-1) - (-4) = 3$.

1.4.2. More generally, show that $(-a) - (-b) = b - a$.

1.4.3. Now, using properties of $\times$ and the distributive law, show $(-a)(-b) = ab$.

1.4.4. Also use the ring properties to prove that

$$(a+b)^2 = a^2 + 2ab + b^2 \qquad \text{and} \qquad (a-b)(a+b) = a^2 - b^2,$$

where, as usual, xy stands for $x \times y$.

Incidentally, there is a good mathematical reason for abbreviating $x \times y$ to xy. The distributive law is better written as

$$a(b+c) = ab + ac,$$

because the products on the right-hand side have precedence over the sum—they have to be evaluated first.

A ring with the multiplicative inverse property, such as $\mathbb{Q}$, is called a *field*. In fact, the way we extended $\mathbb{Z}$ to $\mathbb{Q}$ is an instance of a common construction with rings, called "forming the field of fractions." For any a and $b \neq 0$ we form the fraction $a/b = ab^{-1}$, and we add and multiply fractions according to the rules you learned around fifth grade:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \qquad \text{and} \qquad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$$

These rules also arise from the principle of "keeping things natural"—they are needed to make + and × behave the same for fractions as they do for natural numbers.

1.5 Linear Equations

The humble linear equation $ax + by = c$ takes on a new interest when we seek *integer* solutions x and y for given integers a, b, and c. It can very easily fail to have an integer solution, so the problem is first to decide *whether* there is an integer solution, and if so, how to find it.

Take the example $15x + 12y = 4$. For any integers x and y, 3 divides $15x + 12y$, because 3 divides 15 and 12. But 3 does not divide 4, hence there is no integer solution of $15x + 12y = 4$. In general, we can see that $ax + by = c$ has no integer solution if a and b have a divisor that does not divide c.

But what if the divisors of a and b divide c? It is not at all clear there are integers x and y with $ax + by = c$, though they seem to exist in every case we try. For example, if we consider $17x + 12y = 1$, the only divisors of both 17 and 12 are ± 1, which certainly divide the right-hand side. And with some difficulty (say, by searching down lists of the multiples of 17 and 12) we indeed find a solution, $x = 5$ and $y = -7$.

This presumably depends on some connection between divisors of a and b and numbers of the form $ax + by$. We can already see part of it: Any divisor of a and b is also a divisor of $ax + by$, for any integers x and y. The less obvious part comes from thinking about the *greatest* common divisor of a and b, which we call $\gcd(a, b)$, and seeing that it has the form $ax + by$.

There is a famous algorithm for finding $\gcd(a, b)$, for natural numbers a and b. It is called the *Euclidean algorithm*, and it was described by Euclid as "repeatedly subtracting the smaller number from the larger." To be precise, we produce pairs of natural numbers $(a_1, b_1), (a_2, b_2), (a_3, b_3), \ldots$ as follows. The first pair (a_1, b_1) is (a, b) itself, and each new pair (a_{i+1}, b_{i+1}) comes from (a_i, b_i) by

$$a_{i+1} = \max(a_i, b_i) - \min(a_i, b_i) \qquad \text{(taking the smaller from the larger)},$$

$$b_{i+1} = \min(a_i, b_i) \qquad \text{(and keeping the smaller)},$$

until $a_k = b_k$, in which case the algorithm halts. Then $\gcd(a, b) = a_k = b_k$.

It is clear that the algorithm does reach a pair of equal numbers a_k and b_k, because the natural numbers $a_1, a_2, a_3, \ldots$ cannot decrease indefinitely. But why does it produce the gcd?

Correctness of the Euclidean algorithm *All pairs produced by the Euclidean algorithm have the same common divisors, hence* $a_k = b_k = \gcd(a, b)$.

Proof Each divisor of a_i and b_i is also a divisor of a_{i+1} (because any divisor of two numbers also divides their difference) and of b_{i+1}. Conversely, any divisor of a_{i+1} and b_{i+1} also divides a_i and b_i, because any divisor of two numbers also divides their sum. Thus each pair (a_{i+1}, b_{i+1}) has the same divisors as all previous pairs, and hence the same gcd. But then

$$\gcd(a, b) = \gcd(a_1, b_1) = \gcd(a_2, b_2) = \cdots = \gcd(a_k, b_k) = a_k = b_k,$$

because $a_k = b_k$. □

Not only does the Euclidean algorithm give $\gcd(a, b)$, it gives it in the form $ax + by$.

Linear representation of the gcd *All the numbers* a_i, b_i *produced by the Euclidean algorithm are of the form* $ax + by$, *for some integers* x *and* y, *hence this is also the form of* $\gcd(a, b)$ *itself.*

Proof The first pair a, b are certainly each of the required form. This is also true of all subsequent numbers a_{i+1}, b_{i+1}, because each is either a previous number or the difference of two of them. In particular, $\gcd(a, b) = a_k = ax + by$ for some integers x and y. □

We illustrate the Euclidean algorithm on $a = 17$ and $b = 12$ in the first two columns of the following table. The third column keeps track of what happens to a and b, eventually giving x and y with $17x + 12y = \gcd(17, 12) = 1$.

$$
\begin{aligned}
(a_1, b_1) = (17, 12) &= (a, b) \\
(a_2, b_2) = (5, 12) &= (a - b, b) \\
(a_3, b_3) = (7, 5) &= (b - (a - b), a - b) \\
&= (2b - a, a - b) \\
(a_4, b_4) = (2, 5) &= ((2b - a) - (a - b), a - b) \\
&= (3b - 2a, a - b) \\
(a_5, b_5) = (3, 2) &= ((a - b) - (3b - 2a), 3b - 2a) \\
&= (3a - 4b, 3b - 2a) \\
(a_6, b_6) = (1, 2) &= ((3a - 4b) - (3b - 2a), 3b - 2a) \\
&= (5a - 7b, 3b - 2a)
\end{aligned}
$$

The last line shows the gcd, 1, expressed as $5a - 7b$. Thus we have the solution $x = 5$, $y = -7$ to $17x + 12y = 1$, just as we found by trial before. The Euclidean algorithm does not have any computational advantage in a small example such as this, but it does in large examples. If a and b are integers with many digits, the Euclidean algorithm can be completed in roughly as many steps as there are digits (see the exercises), whereas listing the multiples of a and b takes an exponentially larger number of steps.

In addition to being computationally powerful, the Euclidean algorithm gives us remarkable theoretical insight. For a start, we have confirmed our guess about integer solutions $ax + by = c$.

Test for integer solvability of $ax+by = c$ *The equation* $ax+by = c$ *has an integer solution if and only if* $\gcd(a, b)$ *divides* c.

Proof We have already seen that if $\gcd(a, b)$ does *not* divide c, then the equation $ax + by = c$ has no integer solution.

Conversely, if $\gcd(a, b)$ divides c, suppose $c = \gcd(a, b) \times d$. We now know that there are integers x' and y' such that $\gcd(a, b) = ax' + by'$. Therefore, $c = (ax' + by')d = a(x'd) + b(y'd)$, and hence we have the solution $x = x'd$, $y = y'd$ of $ax + by = c$. □

Exercises

In practice, we usually speed up the Euclidean algorithm by dividing the larger number by the smaller and keeping the remainder, instead of subtracting the smaller number from the larger. (Halting then occurs

when the remainder is 0.) Because division of a by b is really subtraction of b from a until the difference is less than b, the division form of the algorithm produces the same result—it simply skips any steps where the same number is subtracted more than once. This saves many steps when a is much larger than b.

1.5.1. Show that the remainder, when a is divided by a smaller number b, is less than $a/2$.

1.5.2. Deduce from Exercise 1.5.1 that the number of steps to find $\gcd(a, b)$, by the division form of the Euclidean algorithm, is at most twice the number of binary digits in a.

An interesting showcase for the Euclidean algorithm is the *Fibonacci sequence*, $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$, in which each number is the sum of the previous two.

1.5.3. Use the Euclidean algorithm to verify that $\gcd(55, 34) = 1$.

We use the notation F_n for the nth term of the Fibonacci sequence, starting with $F_0 = 0$. The whole sequence can then be defined by the equations

$$F_0 = 0, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1}.$$

1.5.4. Show that one step of the Euclidean algorithm on the pair (F_{n+2}, F_{n+1}) produces the pair (F_{n+1}, F_n), hence

$$1 = \gcd(F_2, F_1) = \gcd(F_3, F_2) = \gcd(F_4, F_3) = \cdots .$$

You probably also noticed that "division of F_{n+2} by F_{n+1}" is really subtraction, so in the case of consecutive Fibonacci numbers, the subtraction form of the Euclidean algorithm cannot be sped up. In fact, the Euclidean algorithm performs at its slowest on consecutive Fibonacci numbers, though it would take us too far afield to explain what this means. The full story may be found in Shallit (1994).

Because $\gcd(F_{n+1}, F_n) = 1$ by Exercise 1.5.4, it follows by the corollary to the correctness of the Euclidean algorithm that there are integers x and y such that $F_{n+1}x + F_n y = 1$.

1.5.5. Find integers x_1 and y_1 such that $F_2x_1 + F_1y_1 = 1$, and integers x_2 and y_2 such that $F_3x_2 + F_2y_2 = 1$.

1.5.6. Show that $F_{n+2}F_n - F_{n+1}F_{n+1} = -F_{n+1}F_{n-1} + F_nF_n$ and hence that

$$1 = -F_2F_0 + F_1F_1 = F_3F_1 - F_2F_2 = -F_4F_2 + F_3F_3 = \cdots .$$

It is worth mentioning that when Euclid proved that there are infinitely many primes, he did not argue exactly as we did in Section 1.3. Instead of using division with remainder to prove that each prime p_j fails to divide $p_1p_1 \cdots p_k+1$, he used the obvious fact that $\gcd(p_1p_1 \cdots p_k+1, p_1p_1 \cdots p_k) = 1$.

1.5.7. Why is this obvious? Use the similar fact $\gcd(p_1p_1 \cdots p_k - 1, p_1p_1 \cdots p_k) = 1$ to give another proof that there are infinitely many primes.

1.6 Unique Prime Factorization

The discovery that the greatest common divisor of a and b is of the form $ax+by$, for some integers x and y, has important repercussions for prime divisors.

Prime divisor property *If a prime p divides the product of integers a and b, then p divides either a or b.*

Proof Suppose that p divides ab and p does *not* divide a. Then we have to show that p divides b. Because p does not divide a, and p is prime, 1 is the only divisor of p that divides a. We therefore have

$$1 = \gcd(a, p) = ax + py \qquad \text{for some integers } x \text{ and } y.$$

It follows, multiplying both sides by b, that

$$b = abx + pby.$$

But p divides each term on the right-hand side of this equation—it divides ab by assumption and pby obviously—hence p divides b. □

This important property was known to Euclid, as were many of its important consequences, which we shall see later. However, he did not state the following consequence, which today is considered the definitive statement about prime divisors.

Unique prime factorization *Each natural number is expressible in only one way as a product of primes, apart from the order of factors.*

Proof By repeatedly finding prime divisors (Section 1.3), we can factorize any natural number into primes. Now suppose, contrary to the theorem, that there is a natural number with two different prime factorizations:

$$p_1p_2p_3\cdots p_s = q_1q_2q_3\cdots q_t.$$

We may assume that any factor common to both sides has already been canceled, hence no factor on the left is on the right.

But p_1 divides the left-hand side, and therefore it divides the right, which is a product of q_1 and $q_2q_3\cdots q_t$. Thus it follows from the prime divisor property that p_1 divides q_1 (in which case $p_1 = q_1$, because q_1 is prime) or else p_1 divides $q_2q_3\cdots q_t$. In the latter case we similarly find either $p_1 = q_2$ or p_1 divides $q_3\cdots q_t$. Continuing in this way, we eventually find that

$$p_1 = q_1 \quad \text{or} \quad p_2 = q_2 \quad \text{or} \quad \cdots \quad \text{or} \quad p_1 = q_t.$$

But this contradicts our assumption that p_1 is not a prime on the right side. Thus there is no natural number with two different prime factorizations. □

A variation on the preceding proof, which some people prefer, starts with $p_1p_2p_3\cdots p_s = q_1q_2q_3\cdots q_t$ but does *not* assume that the factorizations are different. One again finds $p_1 = q_1$ or $p_2 = q_2$ or $\cdots$ or $p_1 = q_t$, but now this simply means that there is a common factor p_1 on both sides. Cancel it, and repeat until no primes remain, at which stage it is clear that the original factorizations were the same.

Exercises

Unique prime factorization is a powerful way to prove results like the irrationality of $\sqrt{2}$, which we first did in Section 1.1 using properties of even and odd numbers, that is, by using special properties of the number 2. We saw that to extend the method to $\sqrt{3}$ requires a new (and longer) argument about the number 3, and presumably it gets worse for $\sqrt{5}$, $\sqrt{6}$, and so on. With unique prime factorization, the argument depends only on the presence of primes, not which particular ones they are.